A Retrieval-Augmented Generation (RAG)-Based Framework to Automate Report Writing for Systematic Literature Reviews (SLRs) in Evidence Synthesis

MSR9

Gagandeep Kaur[1], Ankita Sood[1], Vedant Soni[1], Rajdeep Kaur[1], Shubhram Pandey[1], Barinder Singh[2]

[1]Pharmacoevidence, Mohali, India; [2]Pharmacoevidence, London, UK

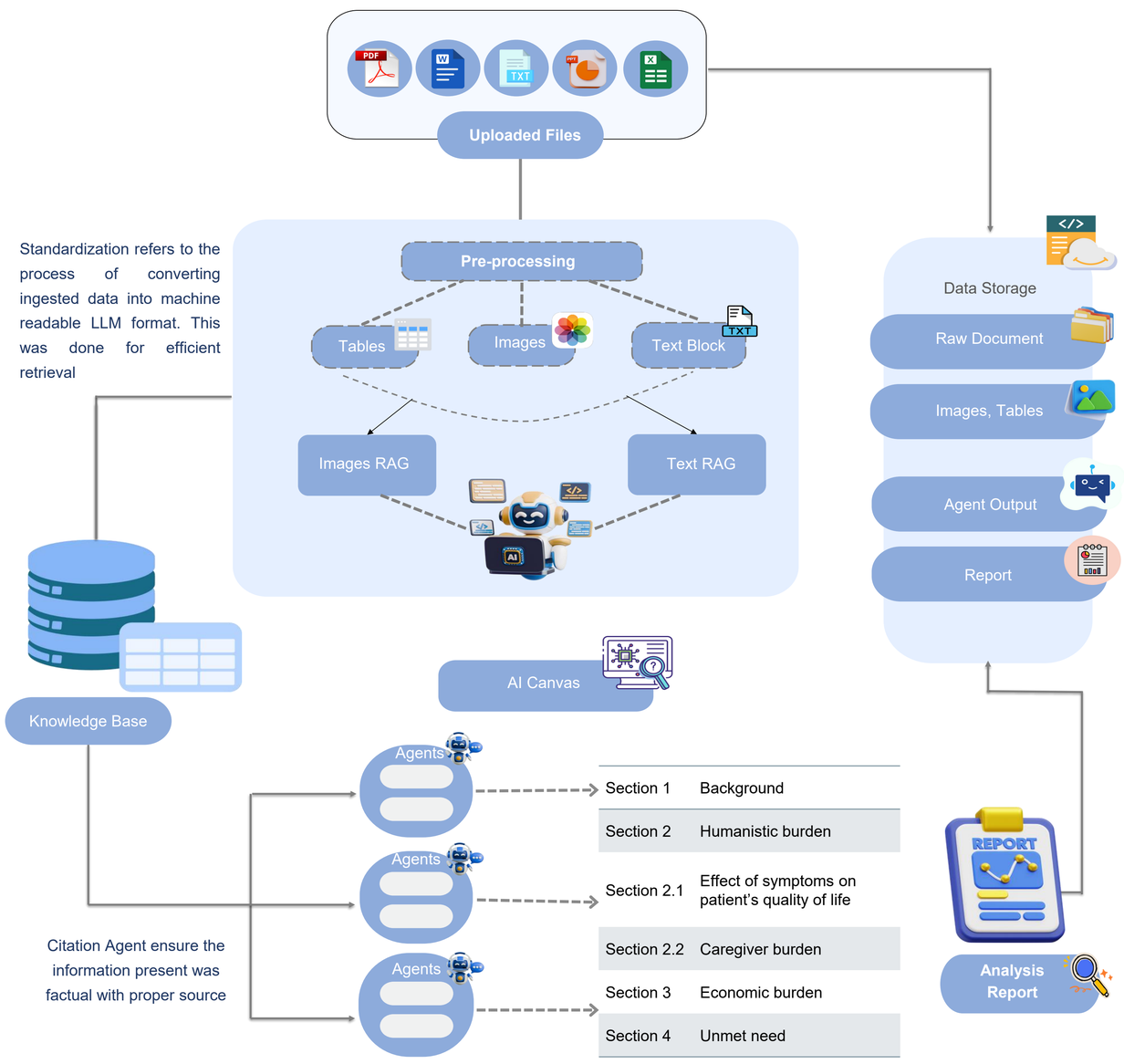
Science-led HEOR & Tech-driven AI

GET POSTER PDF

INTRODUCTION

- Systematic literature reviews (SLRs) are essential for informing clinical guidelines and shaping health policy, characterized by structured expression of ideas, data-driven arguments, and logical reasoning. However, it poses challenges such as handling vast amounts of information and complex ideas[1,2]
- Artificial intelligence (AI) is revolutionizing medical writing by managing complex ideas and extensive information, and enhancing idea generation, content structuring, literature synthesis, data management, editing, and ethical compliance[3]
- The regulatory agencies, including the National Institute for Health and Care Excellence (NICE)[4] and Canada's Drug Agency (CDA)[5], acknowledge the potential of AI to enhance evidence generation in SLRs, while underscoring the need for continued human oversight
- This study aimed to evaluate the use of a Retrieval-Augmented Generation (RAG) framework to generate evidence-linked reports for SLRs in alignment with regulatory guidelines

Figure 1: Schematic diagram of the AI-assisted SLR writing process



METHODS

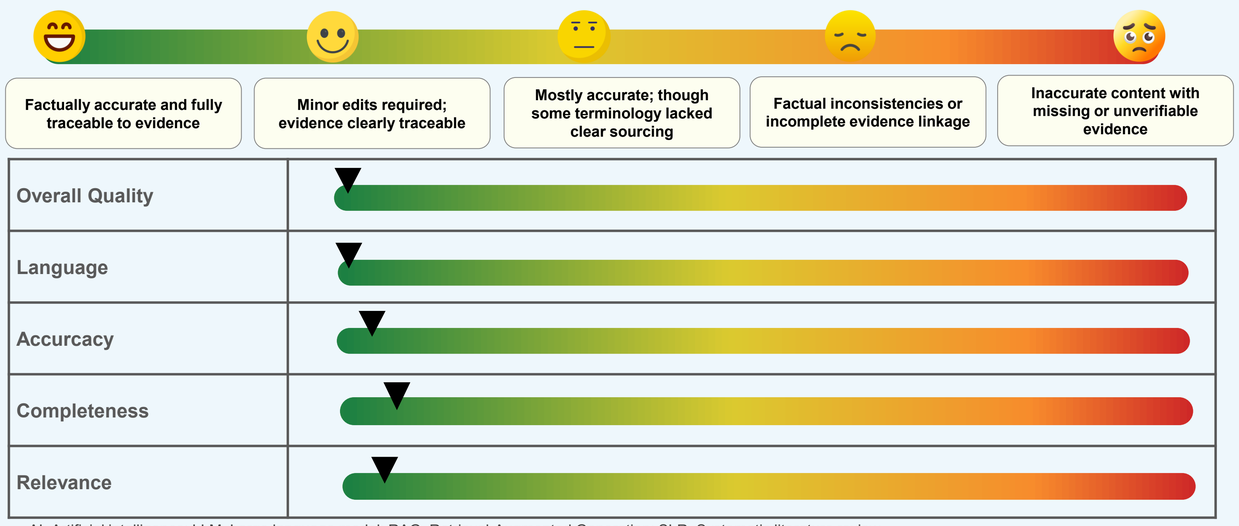
- Subject Matter Experts (SMEs) in SLRs utilised an integrated framework combining the RAG processing pipeline with the multi-agentic approach to generate content related to humanistic and economic burden in Huntington's disease (HD) **(Figure 1)**
- Vast format data ingestion including PDF, PPT, Word, Excel along with context aware RAG pipeline was developed, followed by storage of embedding in vector database. RAG referred these embeddings to generate factual and logical output, eliminating hallucinations
- The prompts directed the large language models (LLMs) to retrieve relevant information from publications related to the symptoms and impact of HD on patients' quality of life, followed by the caregiver burden, economic burden (direct and indirect costs), and unmet needs associated with the disease
- Multi-agentic approach built on top of RAG pipeline to automate the report generation process, and LLM performance at generating the required output using RAG pipeline was evaluated by SMEs across several dimensions **(Figure 2)**, using a five-point Likert Scale

Figure 2: Evaluation parameters of AI-generated SLR Report

RESULTS

- SMEs strongly agreed that all generated content was relevant and accurate to the topic under consideration
- The responses were largely complete **(Figure 3)** but could be further improved by LLM prompt editing and RAG improvements
- SMEs agreed that the responses were well-written, demonstrating appropriate use of language, clarity of expression, and a professional tone throughout
- All outputs related to humanistic burden were rated as "Strongly agree"
- For economic burden and unmet needs, most content received "strongly agree" or "agree", with only two sentences requiring minor terminology edits
- All generated tables and figures were accurate, except for one instance where a calculated value lacked traceability to the original source, resulting in a "disagree" rating

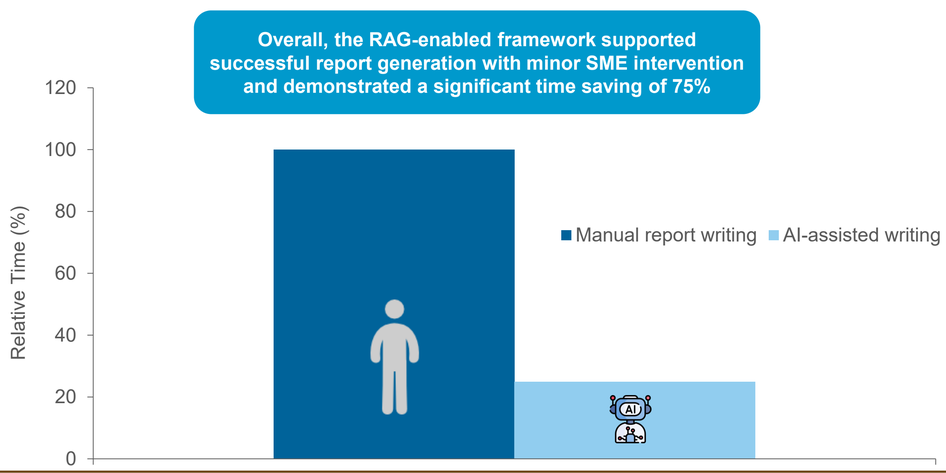
Figure 3: SME evaluation of AI-generated SLR report



AI, Artificial intelligence; LLM, Large language model; RAG, Retrieval-Augmented Generation; SLR, Systematic literature review

- ✓ AI-generated content demonstrated a high degree of accuracy, was well-written, and captured most relevant details with minimal hallucinations or copying errors
- ✓ The outputs closely matched expectations with overall quality of ~90%, though human review remained essential in 5% of cases to ensure completeness and accuracy
- ✓ The overall quality was rated 9/10 for a well-developed draft, with individual components scoring as follows: language (9/10), accuracy (9.5/10), completeness (8/10), and relevance (9/10)

Overall, the RAG-enabled framework supported successful report generation with minor SME intervention and demonstrated a significant time saving of 75%



CONCLUSIONS

- This study demonstrates that a RAG-based generative AI framework can effectively support SLR writing
- Through structured SME validation and minimal human oversight, the approach enables efficient, compliant content generation, supporting responsible AI integration in health economics and outcomes research
- Further testing and refinement of prompts, along with an evaluation of LLM performance in generating other types of literature reports, such as clinical and safety reports, should be undertaken

References

1. Khalifa et al. Comput. Methods Programs Biomed. 2024; 5, 100145;
2. 2. Butler S. Nurse Res. 2025;33(2):27-33.
3. Qudimat et al. Health Sci Rep. 2025;8(2):e70200;
4. NICE position statement. Use of AI in evidence generation. Accessed 29 May 2025
5. CDA-AMC. New Position Statement. Accessed 29 May 2025

Correspondence: Barinder Singh; barinder.singh@pharmacoevidence.com

Disclosure: GK, AS, VS, RK, SP, and BS, the authors declare that they have no conflict of interest

ISPOR 2025; November 09-12; Glasgow, Scotland, UK