



INTRODUCTION & OBJECTIVES

- Major Health Technology Assessment (HTA) bodies, including National Institute for Health and Care Excellence (NICE) UK¹ and Canadian Drug Agency (CDA) (Canada)², recognize AI's potential to strengthen methodological rigor, accelerate synthesis, and enhance reproducibility
- This study aimed to evaluate a Retrieval-Augmented Generation (RAG) based generative AI framework designed to ensure transparent, traceable, and reproducible evidence synthesis for HTA submissions. The framework was intended to minimize manual workload thereby supporting responsible Artificial Intelligence (AI) adoption in health evidence generation and regulatory documentation

METHODS

- A proof-of-concept study was conducted to evaluate a RAG based framework designed to enhance transparency, traceability, and reproducibility in evidence generation for HTA submissions
- The system was developed in Python and integrated with Claude Sonnet 3.7
- 30 source documents related to neurodegenerative disorders, including clinical trials, economic models, and regulatory summaries, were uploaded into the RAG pipeline
- All files were pre-processed for AI readiness (cloud migration, OCR, text extraction, metadata like document IDs, section labels, table headers, paragraph markers) (Figure 1)
- To ensure source traceability, metadata such as document identifiers and section references were captured alongside embedded content
- The tool was tested using 40 scenarios developed by subject matter experts (SMEs) across clinical, economic, and regulatory domains
- Each output was independently evaluated by two SMEs for factual accuracy, completeness, and traceability to the original source
- Discrepancies were documented, categorized, and reviewed for potential improvements

RESULTS

- The RAG-based system effectively retrieved and synthesized content from the embedded source documents
- Across 40 SME-designed scenarios, 36 (90%) outputs were rated as factually accurate, complete, and fully traceable and included correct sources
- In 3 scenarios (7.5%), the factual content was accurate; however, minor issues such as missing citations required SME intervention to restore explicit source traceability
- In one scenario (2.5%), the output included only partial paraphrasing of the supporting data source
- SMEs highlighted that the RAG framework-maintained adherence to HTA reporting standards and these reports could be reproduced
- Overall, the system achieved a time-saving of approximately 70-75% compared to manual evidence review, without compromising quality
- Reviewers emphasized that the generated content was scientifically appropriate, logically coherent, followed by balanced interpretation of evidence

CONCLUSIONS

- The study demonstrates that using a RAG-based Gen-AI framework enhances transparency, traceability, and reproducibility of evidence generation
 - This approach reduces human workload, while preserving factual accuracy, and minimizing hallucinations through integrated retrieval
 - This framework supports responsible AI integration in HTA submissions, incorporating automated quality checks to advance evidence-based evaluation with methodological rigor and transparency

Figure 1: Schematic diagram of transparent, traceable and reproducible evidence generation using gen AI and RAG

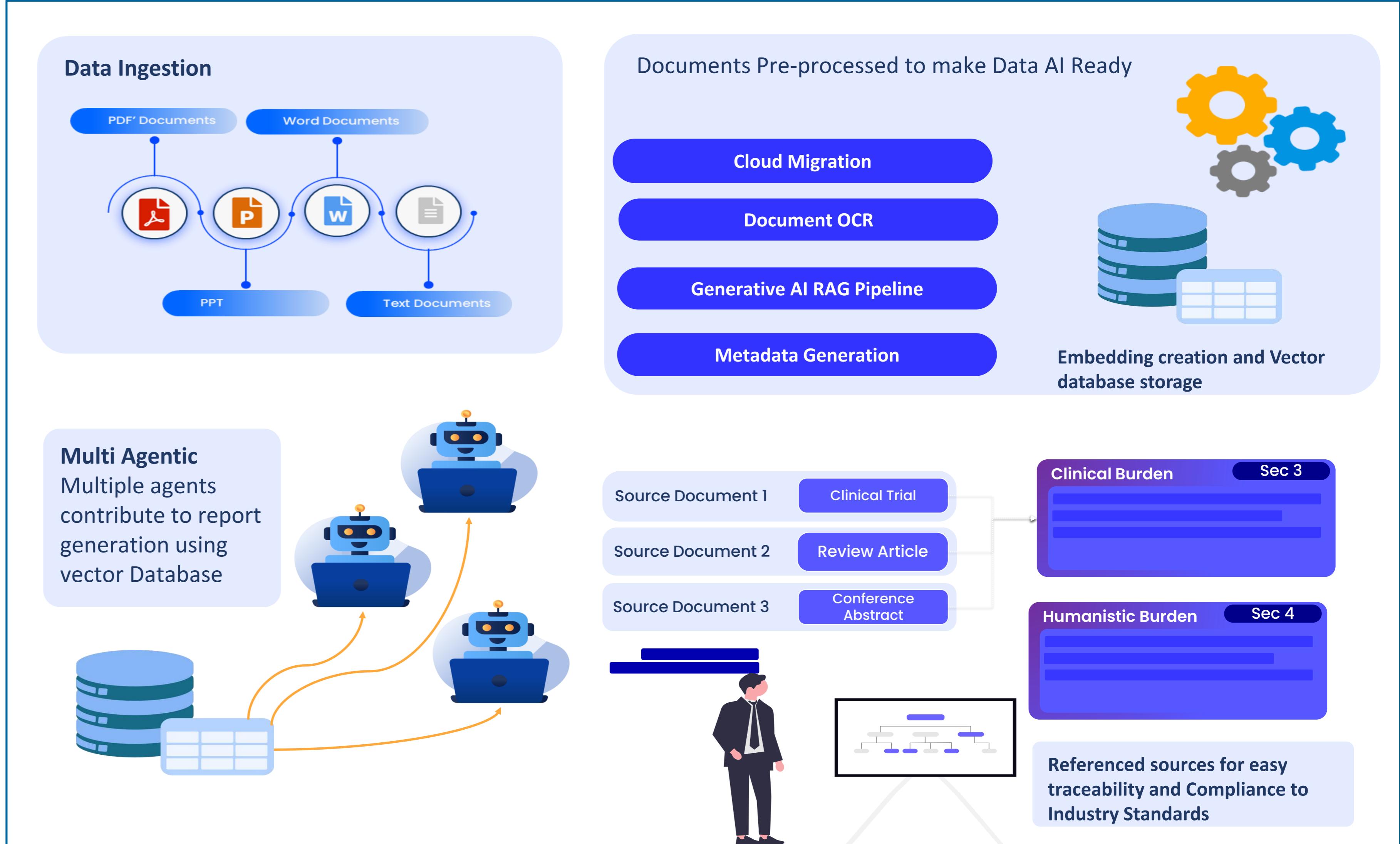


Figure 2: Comparison between direct report generation and RAG based report generation with traceable references

