MSR106

Breaking Linguistic Barriers: Cross-Language HEOR & Evidence Integration through GenAI

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

CONCLUSION

The Generative AI approaches like Large Language models (LLM) with Retrieval-Augmented Generation (RAG) demonstrated high efficiency and accuracy in translating articles from different languages into English and extracting relevant data. These findings highlight the capabilities of GenAI solutions in overcoming language barriers in clinical research, enabling the inclusion of non-English articles in SLRs for more precise evidence analysis.

Introduction

- Systematic literature reviews (SLRs) often encounter non-English studies that are excluded or translated using conventional methods, which can be error-prone and time-consuming
- Large language models (LLMs) offer real-time, context-aware translation, improving both efficiency and accuracy
- This study presents a scalable AI framework integrating LLMs with retrieval-augmented generation (RAG) to enhance semantic alignment during translation and data extraction
- Applied to breast cancer literature, the framework enables reliable extraction from non-English sources, improving the inclusivity and scalability of evidence synthesis

Objective

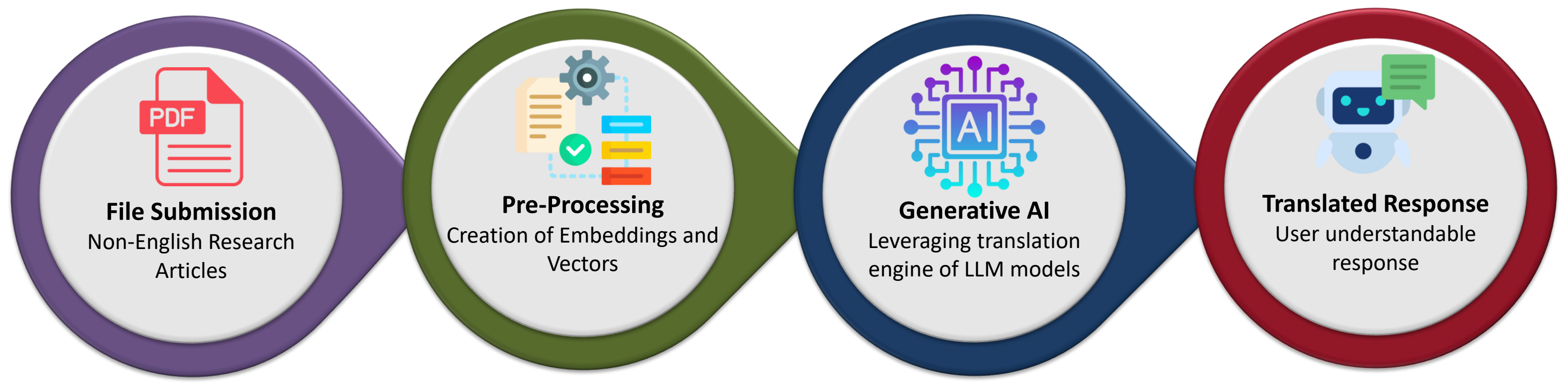
- The objective of this study was to develop and validate a scalable AI-driven framework that leverages large language models (LLMs) and retrieval-augmented generation (RAG) to enable efficient translation, contextual understanding, and data extraction from non-English literature within the SLR process

Methodology

Document processing pipeline

- An AI-powered interface was developed with the Claude model and RAG pipeline to support cross-language translations
- This data pipeline allows the upload and processing of research articles in multiple languages, like Chinese, German, Japanese, French, etc.
- The dynamic RAG pipeline divided these articles into small chunks and created the embeddings to facilitate the efficient retrieval
- These embeddings are useful while calculating the similarity score between one or more chunks during retrieval
- Further, these embeddings are stored in the vector databases
- LLM model operated as the translation engine, utilizing context from the embedded chunks to ensure semantic accuracy during translation
- The architecture was optimized for scalability and user interactivity through a web-based dashboard where researchers could upload, view, and validate translated content
- The translated content is then further used as input for evidence synthesis and data extractions

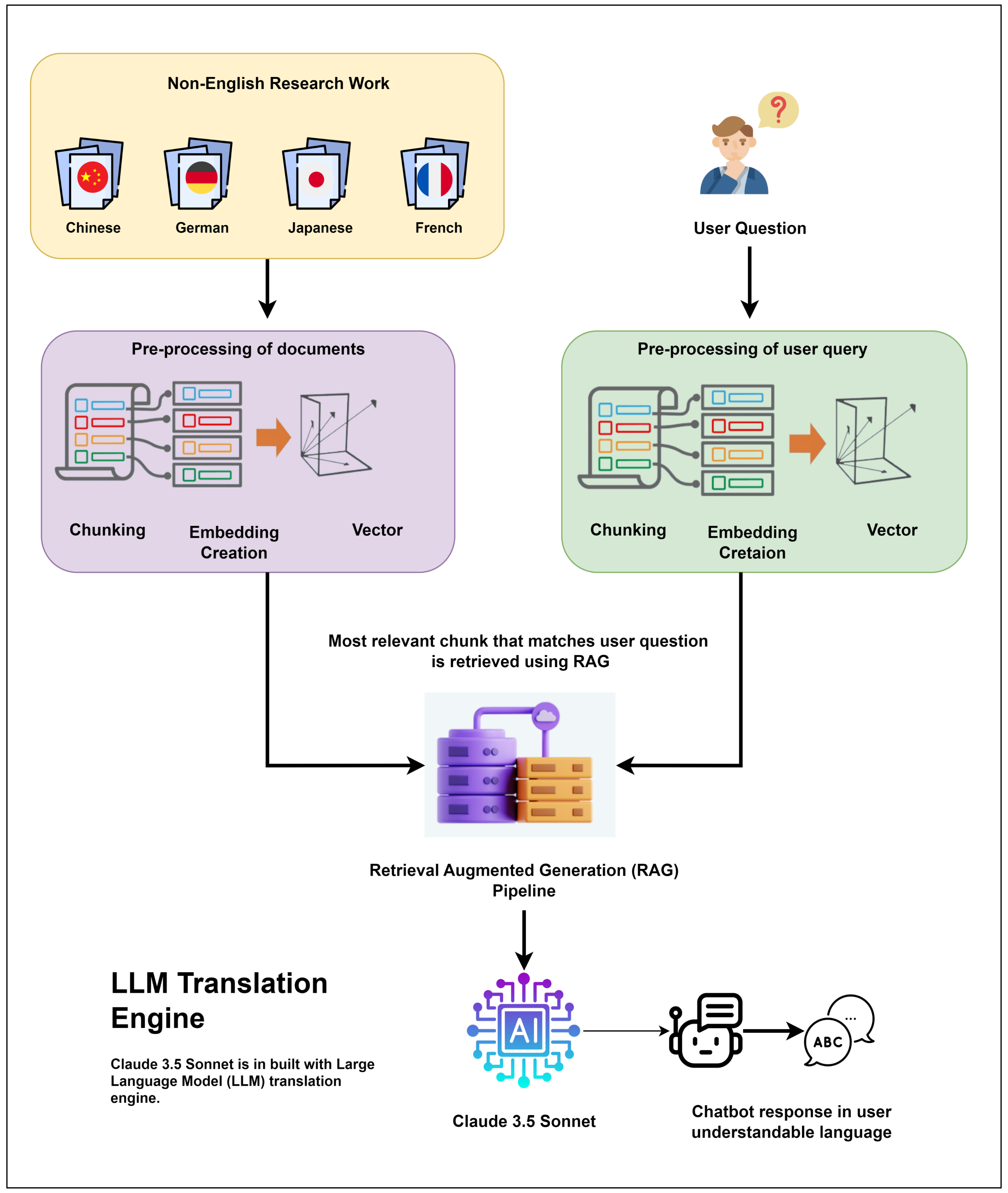
Figure 1. Brief overview of data flow


File Submission
Non-English Research Articles
Pre-Processing
Creation of Embeddings and Vectors
Generative AI
Leveraging translation engine of LLM models
Translated Response
User understandable response

Data Extractions and Expert Validation

- The translated documents were indexed within the interface, allowing researchers to use prompts or predefined queries to extract structured insights (e.g., baseline data, efficacy, safety outcomes, etc.).
- To validate the interface, research articles focusing on the safety and efficacy of Breast Cancer (BC) in different languages were uploaded to the interface
- Five articles in each of the languages, including Chinese, German, Japanese, and French, were uploaded and translated into English
- Specialized prompts were designed by Subject Matter Experts (SMEs) to extract key clinical data elements such as trial endpoints, sample sizes, adverse events, and efficacy rates from translated articles
- SMEs manually reviewed both the extracted data and the corresponding source translation to ensure interpretive accuracy

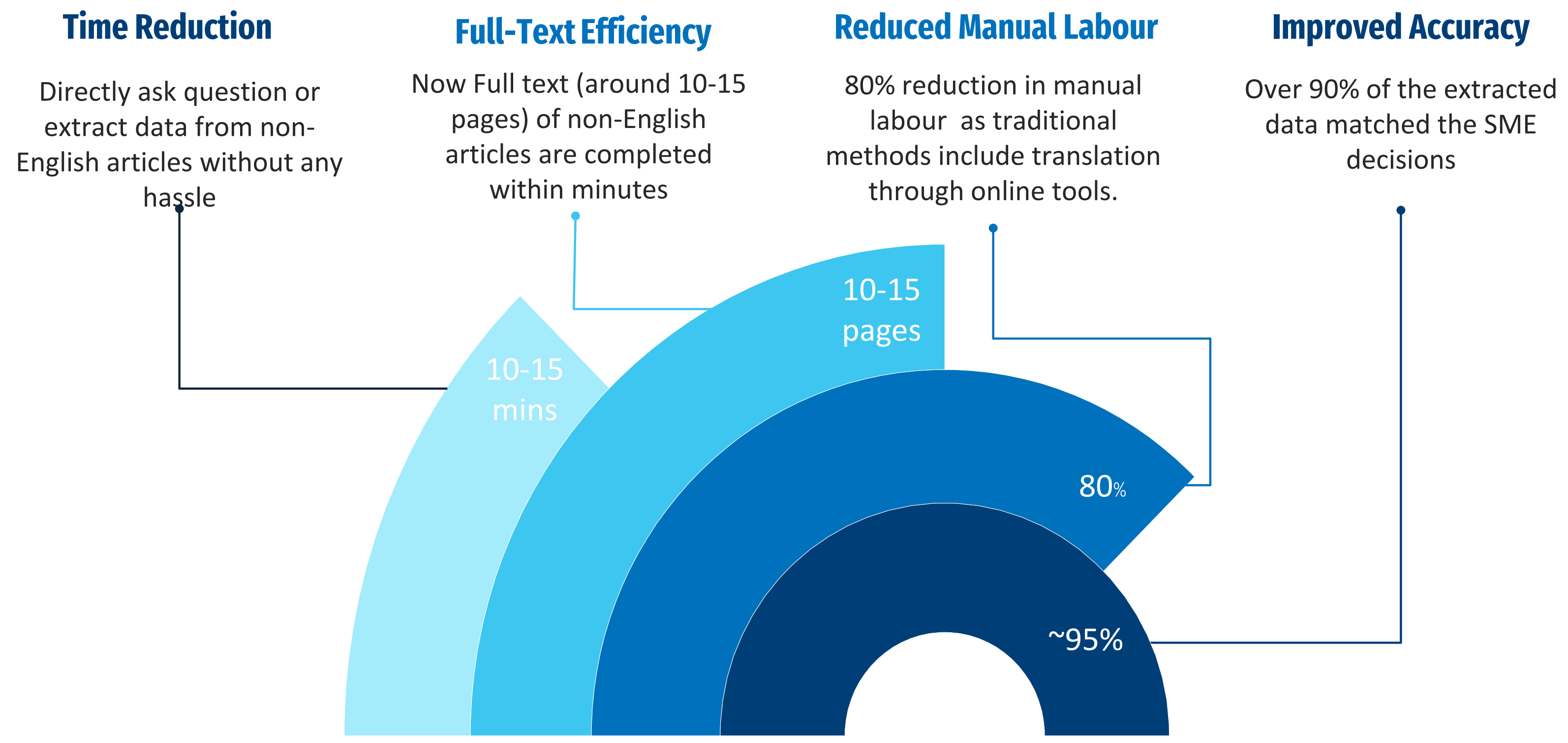
Figure 2. Systematic overview of GenAI based RAG framework for utilizing translation engine of LLM's for analysis of non-English research articles


Non-English Research Work
Chinese
German
Japanese
French
User Question
Pre-processing of documents
Chunking
Embedding Creation
Vector
Pre-processing of user query
Chunking
Embedding Cretaion
Vector
Most relevant chunk that matches user question is retrieved using RAG
Retrieval Augmented Generation (RAG) Pipeline
LLM Translation Engine
Claude 3.5 Sonnet is in built with Large Language Model (LLM) translation engine.
Claude 3.5 Sonnet
ABC
Chatbot response in user understandable language

Results

- Human subject matter experts (SMEs) evaluated the translated content and extracted data, validating it as both contextually coherent and accurate
- LLM with RAG outperformed traditional tools in the following areas :
  - **Fluency & Readability**: LLM-based translation approach yields human-like translations that are easier to read
  - **Medical Context:** Traditional online translation tools sometimes miss or misinterpret clinical terminology. The LLM with enhanced RAG model preserves precise medical meaning for research accuracy
  - **Handling Complex Text:** Bullet points, tables, and numbered lists often get jumbled in online translations. The LLM method preserves the structure, formatting, and provides clear and usable translations, which facilitates accurate evidence synthesis
  - **Image Translation:** Most online translators skip over image-based text completely. With RAG and OCR, even embedded images with text are translated automatically
  - **Text Spacing & Layout**: Translated content from online tools can be complex, with issues like spacing or broken alignment. LLM output keeps spacing and formatting much closer to the original

Figure 3. Visual representation of results



"The RAG-based interface successfully processed and translated articles from Chinese, German, Japanese, and French into English, achieving a 100% completion rate

Across all 20 test articles (five per language), domain experts unanimously validated the translations as accurate, with particular emphasis on the preservation of medical terminology and contextual integrity

As shown in Figure 3, the framework delivered an estimated 80% reduction in manual processing effort compared to traditional translation workflows"

QR code to be added

ISPOR US 2025; May 13-16, 2025; Montreal, Quebec, Canada

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