Leveraging Generative Artificial Intelligence for the Creation of Global Value Dossiers Through a RAG Pipeline and Multi-Agent Integration

MSR 136

Sven L. Klijn, MSc[1], Alison Johnson, PhD[1], Sahana Joish, MBA[1], Barinder Singh, RPh[2], Shubhram Pandey, MSc[2], Nicola Waddell, HNC[3], Rajdeep Kaur[2]
[1]Bristol Myers Squibb, Princeton, NJ, USA, [2]Pharmacoevidence Pvt. Ltd., Mohali, India, [3]Pharmacoevidence Pvt. Ltd., London, United Kingdom

Value of GVD automation

- Global Value Dossiers (GVDs) support market access, reimbursement, and pricing decisions by providing comprehensive evidence on the disease and treatment value
- The development of GVDs requires integration of diverse evidence sources, including clinical trial data, real-world evidence, health economics models, and patient-reported outcomes
- Manual GVD creation can be time-consuming and error-prone due to changing market dynamics and competitive settings
- Generative Artificial Intelligence (GenAI) accelerates GVD development by automating evidence synthesis, content generation, and enhancing speed, consistency, and accuracy, while empowering experts to focus on high-value strategic decision-making

Objective

- This study aimed to evaluate the feasibility of using generative artificial intelligence (GenAI) to automate the creation of key sections within a GVD, integrating Retrieval-Augmented Generation (RAG) pipelines, and a multi-agent approach to produce accurate, traceable outputs with human oversight

Methods

- The platform integrates a RAG framework, enhancing language models with external evidence, together with a multi-agent system to process and synthesize clinical, economic, and access evidence needed for comprehensive GVD development
- It is built on a layered architecture using Python 3.11, FastAPI microservices, and PostgreSQL databases, with a React-based interface connected to Amazon Web Services (AWS) Bedrock-hosted Large Language Models (LLMs) (claude sonnet 3.7) to automate and streamline dossier creation and updates

Phase-1

Data Migration & Pre-processing

- In Phase 1, Input data were uploaded and migrated into the RAG data pipeline for processing and retrieval
- All external materials used in the RAG pipeline were sourced under appropriate licenses, ensuring compliance with intellectual property rights
- Multiple input data formats (PDF, Word, PPT, and TXT) were processed through RAG data pipeline
- Each file format has a different data structure and required customized preprocessing workflows to extract content appropriately
- Scanned or image-based content was converted into machine-readable text using Optical Character Recognition (OCR)
- Text and tables converted into Markdown, a lightweight and structured text format that preserves headings, lists, and basic formatting
- Images extracted separately for multimodal processing
- Markdown documents broken down into meaningful, context-aware chunks using Node-based Markdown Chunking
- AI agents analyzed extracted images to generate descriptive captions
- Text chunks and image captions were converted into vector embeddings to enable semantic search and retrieval

Data Storage for Access and Traceability

- For centralized, scalable, and secure access, preprocessed text and images were stored in an S3 repository. This further ensured traceability and expedited retrieval

AI-Enabled Autoclassification

- The evidence was automatically categorized using a specialized AI agent classifier according to its document type (e.g., Clinical Study Report (CSR), Systematic Literature Review (SLR), etc.

Phase-2

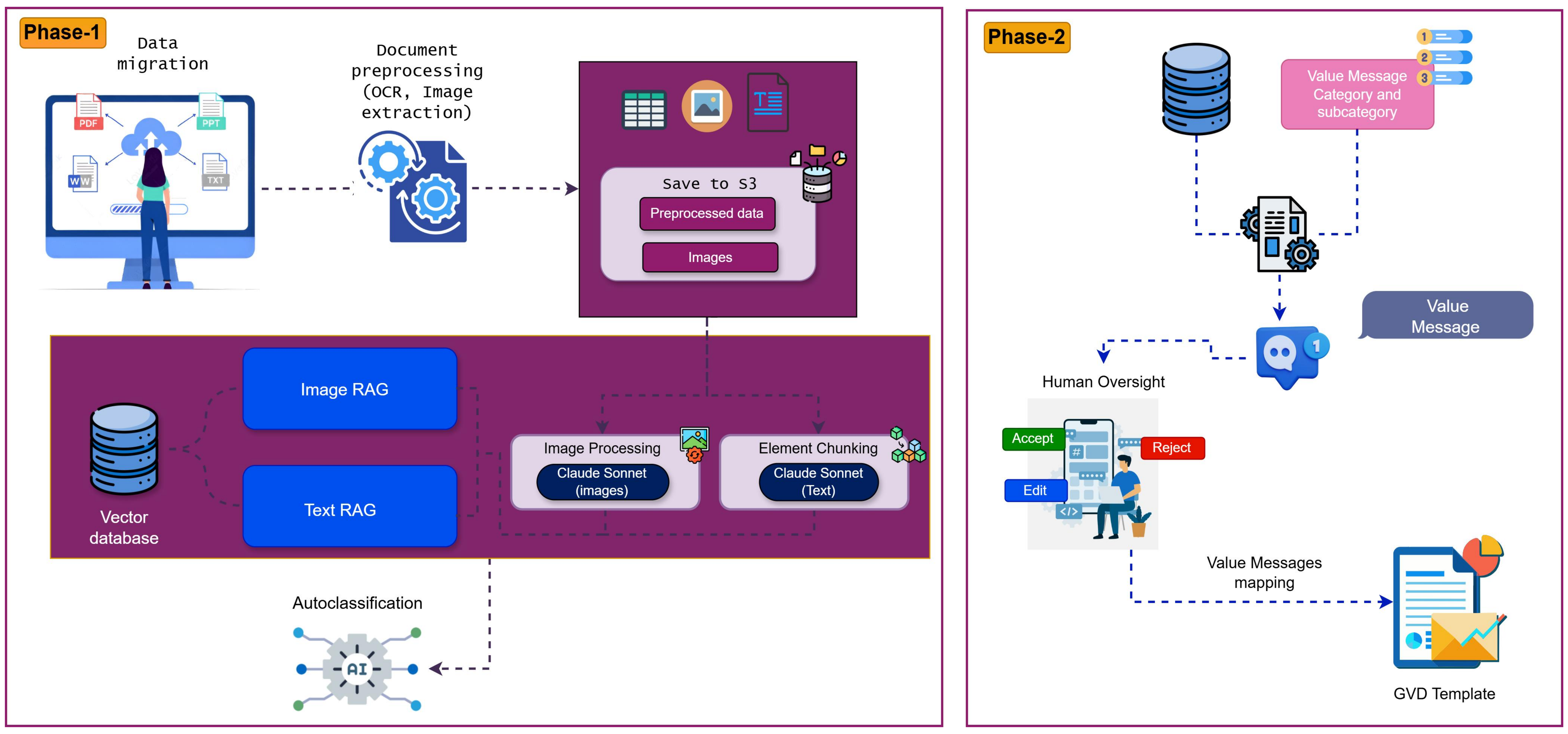
- This phase focused on generating evidence-based value messages that capture the clinical, economic, humanistic, and societal value of a treatment for the GVD
- The dossier value story was generated based on predefined categories (Comparative, Disease, Economic, Humanistic, Societal) (Figure 1)
- Contextual value messages produced by agents using data were stored in the RAG pipeline
- Subject Matter Experts (SMEs) reviewed all the value messages, with the ability to approve, reject, or edit them
- The approved messages were mapped into the predefined GVD template following a validation process, assuring traceable and organized evidence integration

Phase-3

Section-Specific Agents

- Dedicated agents were configured with defined agent name, prompt, and output type (table, graph, or plot) for different GVD sections such as Disease Background, Disease Management, and Unmet Need
- Each agent was initialized with a user-defined prompt specifying the section context, which was then dynamically refined and augmented by an integrated backend agent
- Each agent generated content as per the provided context and in accordance with value messages after retrieving the suitable evidence chunks from the RAG database

Figure 1: Workflow for Data Pre-processing and Value Message Creation in GVD Development



Multi-Format Content Generation

- The outputs were generated in three formats, including text, tables, and graphs, to ensure that both narrative evidence and quantitative data were represented

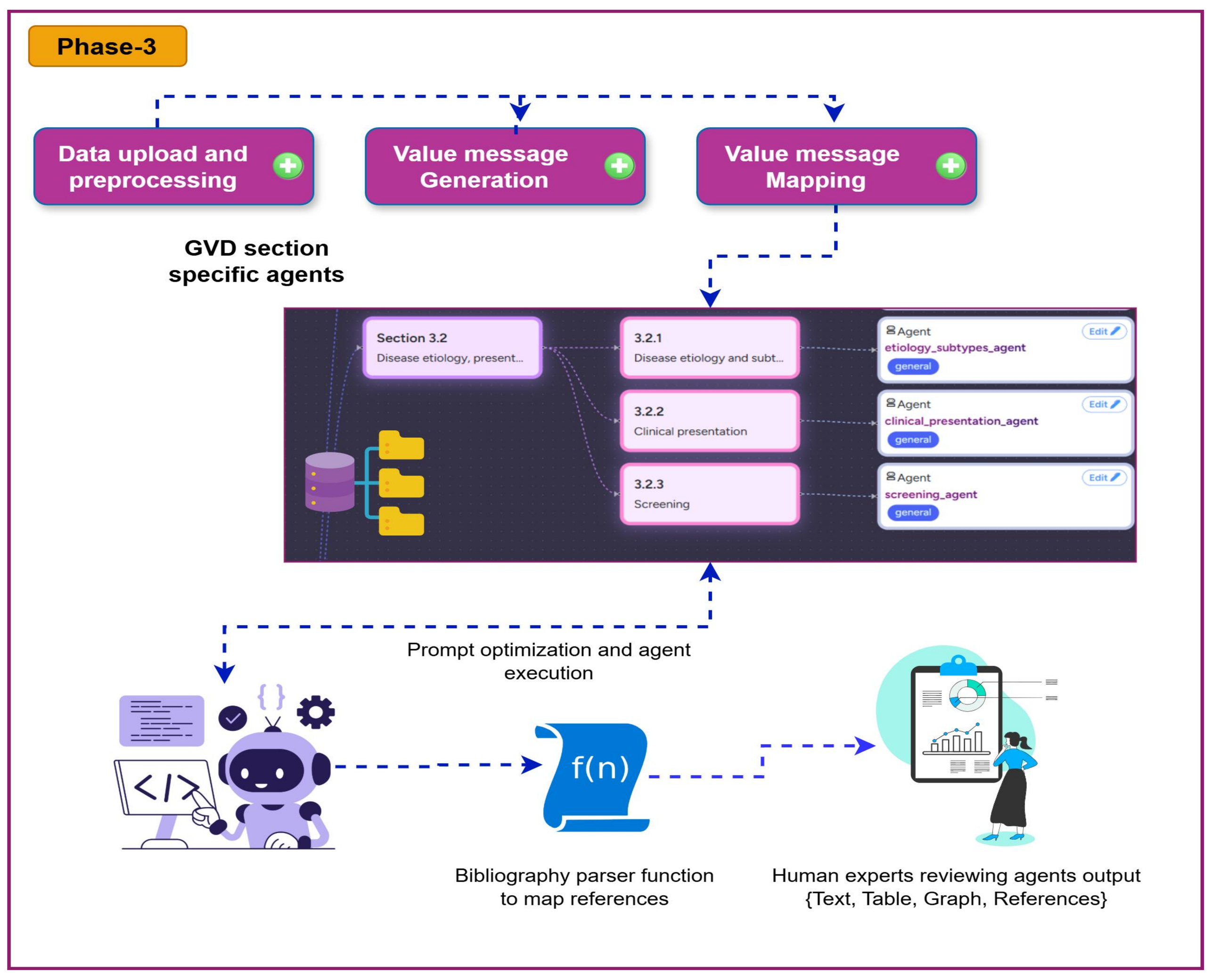
Output Parsing into GVD Template

- The validated results were integrated into the GVD template, including references
- Parser function converted AI-generated outputs into structured and formatted template
- The main parser coordinates text, table, and graph/plot parsers to ensure consistent formatting across sections
- The bibliography parser extracts document identifiers from the text generated by different agents and formats references into a specific format to ensure accuracy and traceability as shown in Figure 2

SME Validation

- The SMEs reviewed all outputs (text, tables, graphs) for completeness, clarity, accuracy, and traceability. SMEs could accept, reject, or edit the content to ensure high-quality evidence representation

Figure 2: Configuration of Section-specific agents and response generation



Results

- A total of 140 documents, including journal articles, conference abstracts, treatment guidelines, epidemiology data sources, and targeted literature reviews, were uploaded into the RAG pipeline
- This generated a 73-page GVD covering disease background, disease management, and unmet needs
- SMEs revied the outputs by considering the evaluation parameters (Relevance, Source Traceability, Language, Accuracy, Accuracy, Completeness, Overall quality) as shown in Table 1
- Outputs in tabular and graphical formats (such as pie charts, bar/line graphs) were generated without any human intervention
- Kaplan-Meier and forest plots required manual intervention due to their statistical intricacies which can mitigated in future by integrating image processing techniques with GenAI
- Approximately 5% of the disease background section required human input for specific trial details and indication.
- In the disease management section, around 10% of the content involved human input, primarily for table formatting and data calculations

Table 1: Evaluation parameters to validate the agents' response

Parameter	Evaluation Focus
Relevance	Content alignment with study objectives
Language	Clarity, tone, and appropriateness of language
Accuracy	Correctness and factual precision
Completeness	Coverage and sufficiency of content
Source Traceability	Verifiability and documentation of sources
Overall Quality	Comprehensive quality across all parameters

- The unmet needs section required minimal human input (about 1%), limited to minor language modifications
- The AI-generated GVD was assessed by SMEs for completeness, formatting, and traceability of data points, confirming accuracy of the output. The AI + human process resulted in 70-80% time savings compared to a human-only process
- Overall, the AI-assisted GVD generation achieved an accuracy of approximately 93-95% across all evaluated domains (disease background, management, and unmet needs), as validated by SMEs, suggesting high concordance between AI-generated and reference human-curated outputs

Figure 4: SME validation results demonstrating GVD development.

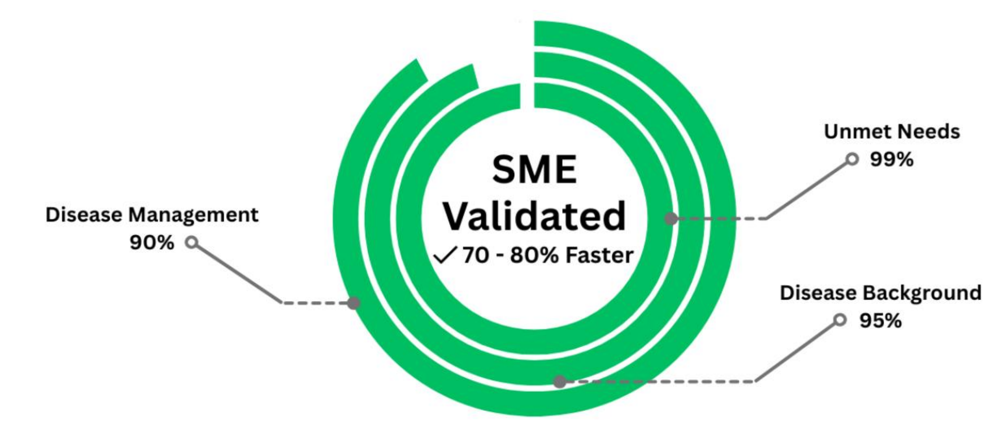
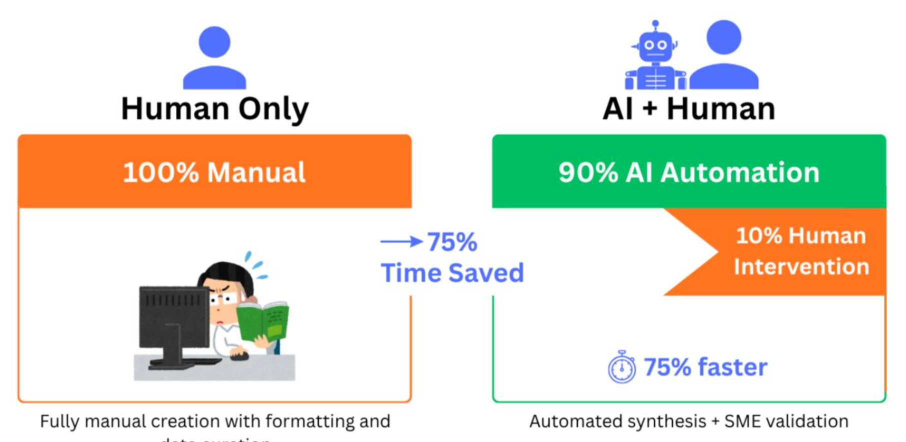


Figure 5: Manual versus AI-assisted workflow with human oversight



Conclusion

- The AI-driven system demonstrated strong capabilities in generating well-structured outputs across multiple sections of the GVD (epidemiological, humanistic burden, economic burden, and treatment), delivering text, tables, and plots
- The AI-generated GVD underwent review by SMEs to evaluate completeness, formatting, and traceability of data points
- This study demonstrates the feasibility of leveraging GenAI for parts of the GVD creation process, changing the GVD development timeline from weeks/months to days, while retaining accuracy and traceability. Further research is needed to assess generalizability across broader use cases
- Future improvements will focus on reducing variability observed when agents are rerun and strengthening systematic SME feedback loops

Presented at ISPOR Europe 2025; November 09-12, 2025; Glasgow, Scotland, UK

Copies of this poster are for personal use only and may not be reproduced without written permission of the authors.