Leveraging Artificial Intelligence to Expedite Dossier Development: A Systematic Literature Review

SA59

Ankita Sood, Gagandeep Kaur, Rajdeep Kaur, Barinder Singh

Pharmacoevidence, Mohali, India

Evidence Pharmaco®
Science-led HEOR & Tech-driven AI

INTRODUCTION

- Dossiers are comprehensive, evidence-based documents that synthesize large volumes of evidence, including clinical trial data, real-world evidence, health economics models, and patient-reported outcomes[1]
- Development of dossiers is time- and resource-intensive, and requires integrating complex data into a scientifically accurate, compelling value narrative[2,3]
- Generative artificial intelligence (GenAI) is increasingly being explored to support dossier development by enabling automated evidence synthesis and content generation, with the potential to improve efficiency and consistenc[2]

OBJECTIVE

- This systematic literature review (SLR) aimed to summarize published evidence on the application of artificial intelligence/machine learning (AI/ML) tools in dossier development

METHODOLOGY

- Key electronic databases, including EMBASE® and MEDLINE® were searched using a combination of relevant keywords related to dossier development and GenAI. In addition, ISPOR database was hand-searched to identify relevant publications
- Studies published in the English language from database inception to January 2026, evaluating the integration of AI/ML tools in the process of dossier development, were included
- The SLR followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure methodological rigor[4]
- Two independent reviewers collected data, and a third independent reviewer performed a quality check

RESULTS

- Among the 47 publications screened, 12 studies were identified and were included in the SLR. **Figure 1** depicts the PRISMA flow for the SLR

Figure 1. Flow of Studies in the SLR

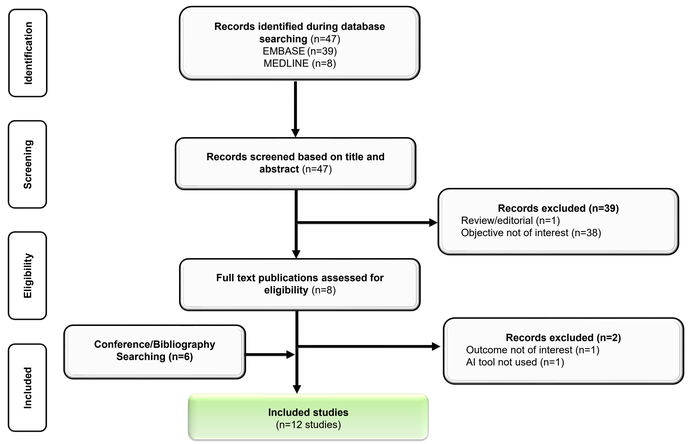
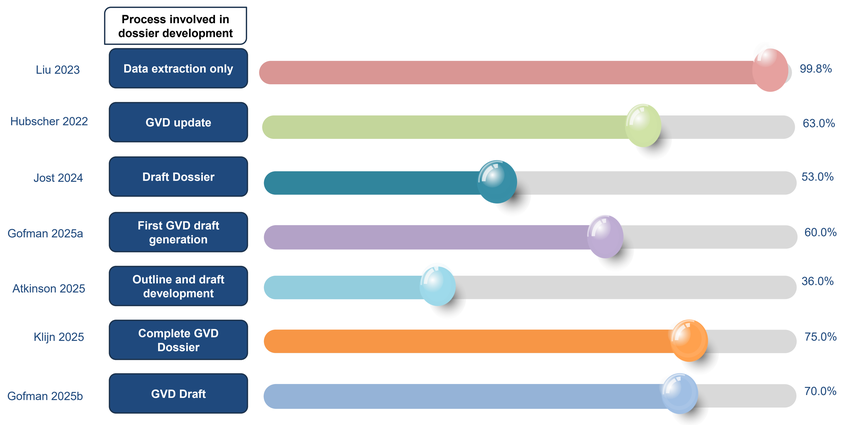


Table 1. Summary of Findings of the Included Studies

Study Name	Dossier type	Key Findings
Aggarwal 2024	GVD	533 hours saved; value messaging needed full human rewrite
Atkinson 2025	GVD	36% total time saved; 62% faster outline; 50% faster draft (Approach B)
Gofman 2025a	GVD	~85% accuracy; ~92% completeness; ~60% draft time reduction
Gofman 2025b	GVD	~95% accuracy; ~70% faster; dual-agent workflow (GVD tone and story agent; GVD evidence mapping agent)
Gofman 2025c	GVD	Shared evidence mapping across GVD & JCA streamlined; rapid iteration of JCA-compliant content
Hubscher 2022	GVD	3 weeks (LiveRef) vs. 8 weeks (conventional methodology) for GVD update; 63% effort reduction
Jost 2024	AMNOG	53%-time reduction; €78k cost saved; quality comparable to traditional dossier
Klijn 2025	GVD	73-page GVD from 140 docs; 93-95% accuracy; 70-80% time savings
Liu 2023	GVD	264.5 hours (99.8%) reduction in extraction time; 3.29 weeks saved per GVD update
Walters 2024a	CDA	High-quality drafts produced with the most relevant Canada-specific documentation; tables required human input; complexity of source documents and pre-submission form sections were found to impact output quality
Walters 2024b	GVD	92% SME agreement on overall accuracy; Clinical (100% SME strongly or somewhat agreed); Product Profile (95% strongly or somewhat Agreed); Economic (70% somewhat agreed)
Zhang 2024	GVD	Study design accuracy +97.9%; products +98.9% vs. 2023; overall lenient accuracy 94.4%

- Most studies assessed the use of diverse AI tools to support global value dossier (GVD) creation process, while two studies specifically utilized AI to develop country-specific dossiers for Germany and Canada
- The commonly used technologies were GenAI and Large Language Models (LLMs), with specific tools including LiveRef™, LiveSTART™, DO-BO®, and Coauthoring Accelerator
- These tools were primarily applied in three key areas of GVD: content generation, data extraction, and document drafting

Figure 2. Percentage of Time Saved in Dossier Development with Integration of AI/ML



- The largest time savings were observed in data extraction tasks: Liu 2023 reported a 99.8% reduction (264.5 hours saved) using LiveSTART™ for abstract-level data extraction, covering objective, indication, population, country, study category, study design, treatment products, sample size, data source, and reported variables in GVD update
- For full GVD pipelines encompassing disease background, disease management, and unmet needs, Klijn (2025) reported the greatest time savings (70-80%) using a retrieval-augmented generation (RAG) and multi-agent system, while Gofman (2025) reported approximately 60-70% reductions for the disease overview and disease burden chapters of a GVD (**Figure 2**)
- Content accuracy across AI-generated GVD sections ranged from approximately 85% to 95%, with the highest rates reported by tools using well-curated RAG pipelines and structured prompt engineering (**Figure 3**)
- Completeness scores were similarly high, with Gofman 2025a reporting ~92% completeness for disease overview content and Klijn 2025 noting only minor formatting refinements were needed (**Figure 4**)
- In the study conducted by Walters 2024b, 92% of subject matter experts (SMEs) agreed that generated content was sufficiently relevant and accurate for incorporation into de novo GVD development, though completeness agreement was lower at 75%, with room for improvement through prompt refinement and RAG enhancements
- Traceability to source references was reported as strong in studies using RAG frameworks (n=8; **Figure 5**)
- All the included studies required SME or expert review, with no AI-generated outputs deemed submission-ready without human validation

Figure 3. Percentage of Accuracy Across Different Stages of Dossier Development with Integration of AI/ML

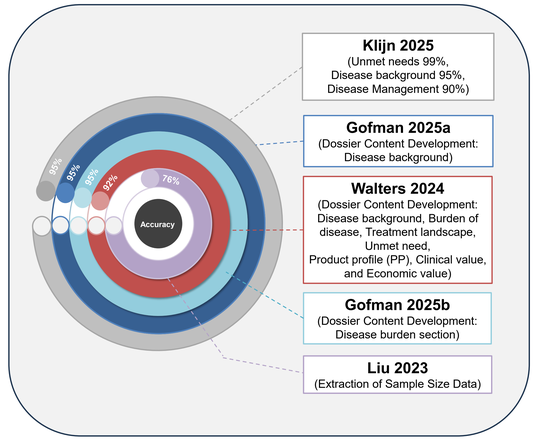


Figure 4. Completeness Scores

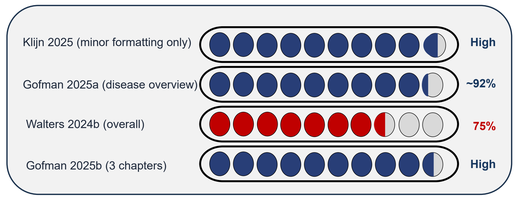
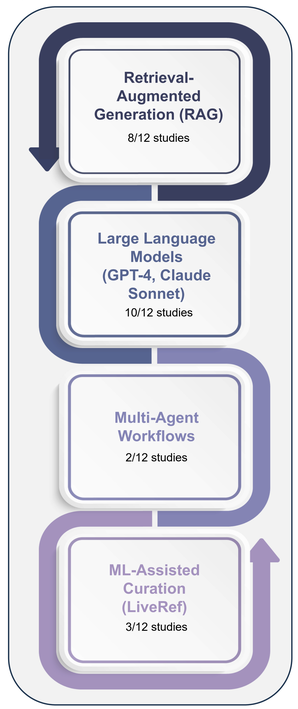


Figure 5. Technology Types Across Studies



CONCLUSIONS

Time-savings
AI-assisted GVD development consistently delivers significant time savings compared to traditional manual workflows across all study designs and dossier types

Enhanced accuracy
Accuracy of AI-generated content reached 76-95%, with high completeness and strong traceability to source references when RAG frameworks are properly configured

Hybrid approach
Hybrid approaches (AI + experienced HEOR experts) produce the highest quality outputs

Role of SME
SME review is universally non-negotiable: all 12 studies identify expert oversight as essential for clinical accuracy, regulatory alignment, and value messaging

Future Directions
Exploration of country-specific customization for diverse HTA frameworks (NICE, AMCP, JCA)

References:
1. Atkinson et al. Value in Health, 28(12), S437; **2.** Klijn et al. Value in Health, 28(12), S521; **3.** Jost et al. Value in Health, 27(12), S375; **4.** Moher D et al., PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015; 4(1):1

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

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

Abbreviations: AI/ML: Artificial Intelligence/Machine Learning; AMNOG: German Pharmaceuticals Market Reorganization Act; CDA: Canada's Drug Agency; GenAI: Generative GPT: Generative Artificial Intelligence; Pre- Trained Transformer; GVD: Global Value Dossier; JCA: Joint Clinical Assessment; LLM: Large Language Models; MLT: Machine Learning Technologies; RAG: Retrieval-Augmented Generation; SLR: Systematic Literature Review; SME: Subject Matter Expert

ISPOR, May 17-20, 2026, Philadelphia, PA, United States