Responsible AI Adoption in HEOR Through Human-in-the-Loop (HITL) Frameworks aligning with Global HTA Expectations

MSR184

Inderpreet Singh-Marwaha, Rajdeep Kaur, Shubhram Pandey, Barinder Singh, Gagandeep Kaur
Pharmacoevidence, Mohali, India

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

INTRODUCTION

- Despite emerging Artificial Intelligence (AI) position statements and initial momentum, adoption of Generative AI (GenAI) in HEOR remains cautious and fragmented, driven by persistent uncertainties around methodological soundness, transparency, regulatory expectations, and the reliability of GenAI-generated evidence[1-3]
- To address this gap in practical guidance, we present the implementation of a comprehensive Human-in-the-Loop (HITL) framework for HEOR workstreams, designed to align with global Health Technology Assessment (HTA) expectations

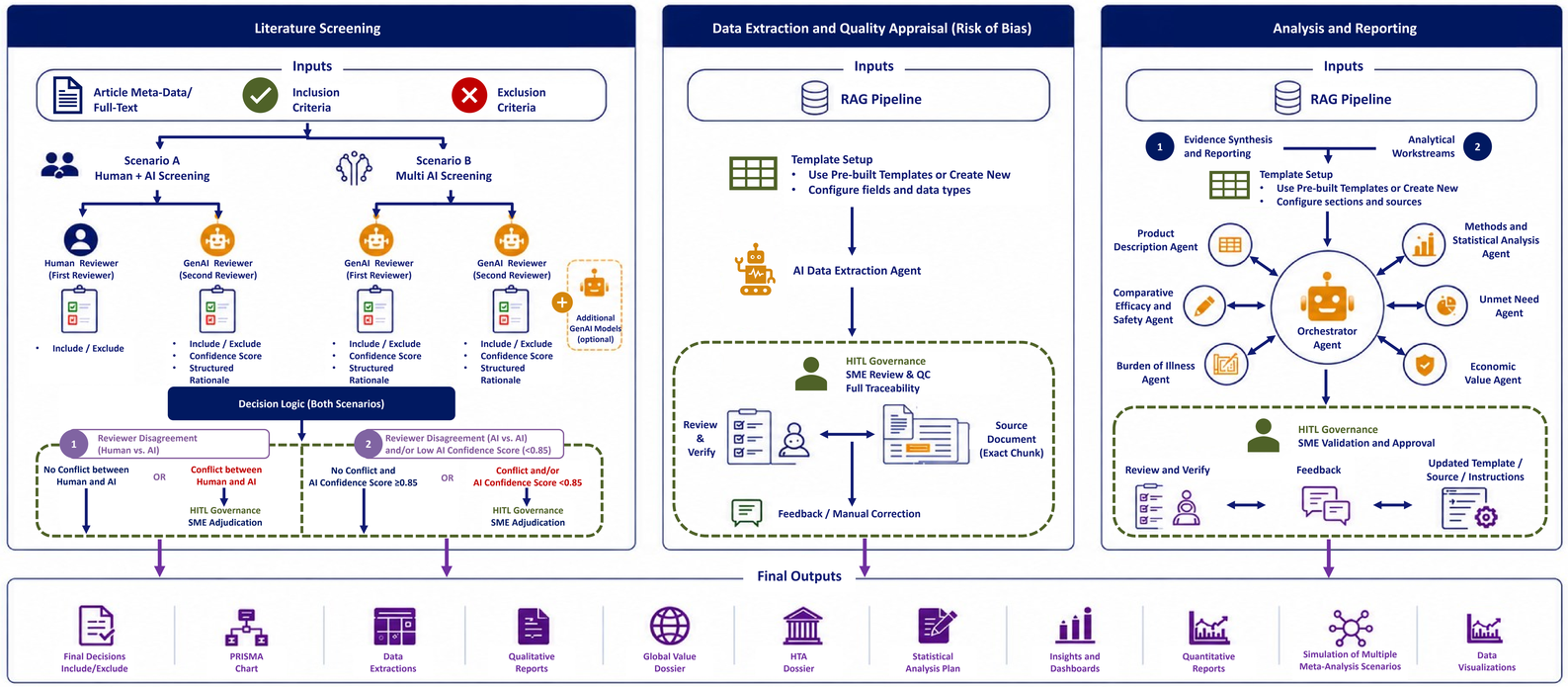
METHODS

- We developed and implemented a GenAI governance architecture integrating mandatory HITL checkpoints across four core evidence generation workflows (**Figure 1**):
  1. **Literature screening:** GenAI functioned as a second reviewer to support inclusion/exclusion decisions by generating diagnostic confidence scores and structured rationales
  2. **Data extraction and quality appraisal:** GenAI was deployed as a second reviewer or as an augmentation tool to enhance reviewer efficiency
  3. **Evidence synthesis and reporting:** GenAI-automated literature review reports, global value dossiers (GVDs), and HTA dossiers supported by rigorous subject matter expert (SME) oversight and control to ensure accurate synthesis and citation provenance for all generated outputs
  4. **Analytical workstreams:** GenAI-automated insights, data visualizations, and statistical analyses (including meta-analyses) with rigorous SME validation layers to ensure methodological appropriateness, code integrity, and accurate interpretation
- Across all workflows, AI outputs were restricted to an advisory role. Predefined decision rules were applied to trigger expert validation at critical checkpoints to ensure traceability and methodological integrity prior to integration into the evidence chain.

RESULTS

- Across multiple reviews, literature screening achieved a mean accuracy of 95% (range: 89% to 99%), with mandatory human arbitration of low confidence outputs and discordant decisions affecting 5% of records (**Figure 2**)
- While the AI successfully matched experts in data extraction (95%) and Risk-of-Bias assessment (78%), human review was still required to verify subjective decisions, interpret context, and correct the AI's tendency to be too conservative (**Figure 2**)
- GenAI workflows produced ~90% submission-ready GVDs and Academy of Managed Care Pharmacy (AMCP) dossiers with 92% expert alignment, reducing traditional development time by ~80% and costs by 70–75% (**Figure 2**)
- Expert review was utilized to finalize contextual framing, interpretive accuracy, and adherence to regulatory standards
- For analytical workstreams, expert review of AI-generated insights, visual outputs, and statistical code was required by design to ensure defensible claims, logical consistency and alignment with the underlying research methodology

Figure 1: Human-in-the-Loop (HITL) Governance Framework for Responsible GenAI Adoption Across HEOR and HTA Evidence Workflows



Note: AI-generated outputs are intended to support, not replace, expert judgment and remain subject to human-in-the-loop (HITL) review, validation, and final approval; confidence thresholds, adjudication rules, and workflow configurations may be adapted based on study design, evidence complexity, and organizational requirements; all outputs should remain fully traceable to source documents and underlying evidence through the retrieval-augmented generation (RAG) architecture; statistical analyses and evidence syntheses require appropriate methodological oversight and quality control prior to regulatory or HTA submission
Abbreviations: AI: artificial intelligence; GenAI: generative artificial intelligence; HITL: human-in-the-loop; HTA: health technology assessment; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; QC: quality control; QA: quality assurance; RAG: retrieval-augmented generation; RoB: risk of bias; SME: subject matter expert

Figure 2: Responsible GenAI Performance Across HEOR Workflows: SME Alignment, Quality, and Efficiency

Abbreviations: AI: Artificial intelligence; AMCP: Academy of Managed Care Pharmacy; GVD: global value dossier; RoB: Risk of Bias; SME: subject matter expert; SQ: signalling question

CONCLUSIONS

- ❖ This implementation confirms that HITL-enabled GenAI workflows assist experts instead of replacing them, allowing HEOR teams to harness efficiency through mandatory checkpoints while preserving decision provenance and alignment with global HTA expectations
- ❖ Ultimately, this supports the responsible AI adoption, ensuring evidence integrity and enabling researchers to focus on high-value activities

References:
1. National Institute for Clinical Excellence. Use of AI in evidence generation: NICE position statement [Internet]. 2024
2. Fleurence, R.L., et al. Generative AI for health technology assessment: opportunities, challenges, and policy considerations-an ISPOR working group report. Value in health, 2024: 28(2), 175
3. Fleurence R, et al. ELEVATE-GenAI: Reporting Guidelines for the Use of Large Language Models in Health Economics and Outcomes Research: An ISPOR Working Group Report. Value in Health, 2025; 28, 1611-1625

Correspondence: Barinder Singh; barinder.singh@pharmacoevidence.com
Disclosure: ISM, RK, SP, BS, and GK the authors declare that they have no conflict of interest