Pharmaco Evidence
Science-led HEOR & Tech-driven AI

Accelerating Health Economic Model Conceptualization using Multi-Agent Generative AI Systems

MSR193

Shubhram Pandey[1], Sameer Mansoori[1], Rajdeep Kaur[1], Barinder Singh[1], Nicola Waddell[2], Paul Stephen John Miller[3]
[1]Pharmacoevidence, Mohali, India; [2]Pharmacoevidence, London, UK; [3]Miller Economics, Macclesfield, UK

BACKGROUND

- Health economic modeling is critical for healthcare decision-making, yet model conceptualization remains **resource-intensive and time-consuming**[1]
- The challenge is especially pronounced in **rare diseases with limited evidence**, where health technology assessment (HTA) submissions require comprehensive model protocols built from sparse data
- Generative artificial intelligence (AI) with **Retrieval-Augmented Generation (RAG)** offers the potential to accelerate protocol development while preserving methodological rigor[2]

OBJECTIVES

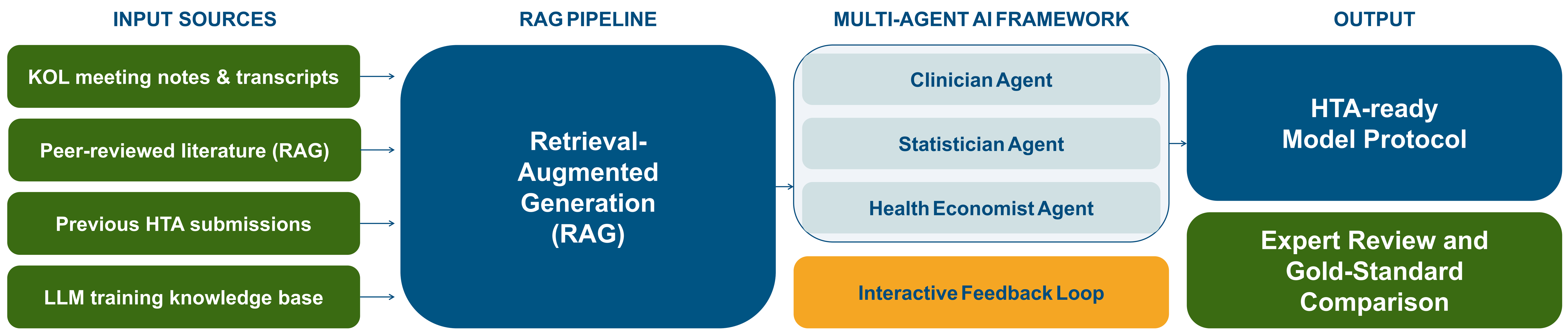
- To demonstrate the application of a multi-agent GenAI system with RAG to develop comprehensive health economic model protocols for HTA submissions
- To address two distinct scenarios: (i) de novo model development in evidence-scarce therapeutic areas, and (ii) model enhancement where model structures already exist

METHODS

- A multi-agent AI framework was developed with specialised agents representing clinicians, statisticians and health economists
- **De novo conceptualization:** leveraged key opinion leader (KOL) inputs from meeting notes, transcripts, and large language model (LLM) training data
- **Existing-model scenarios:** published peer-reviewed literature and previous HTA submissions were incorporated through a RAG architecture
- The system generated model structures (health states and transitions), identified key assumptions, specified model inputs (clinical parameters, costs, HRQoL values), and recommended relevant sensitivity & scenario analyses[1,3]
- Interactive feedback loops iteratively refined protocol outputs
- Generated protocols underwent validation through clinical expert review and comparison against gold-standard manual protocols
- Model types evaluated included markov cohort models and partitioned survival models, encompassing both cost-effectiveness and cost-utility analyses from a payer perspective
- AI-generated outputs were assessed across four protocol components: model structure, cost inputs, utility inputs, and clinical pathway
- Accuracy was defined as alignment with subject matter expert (SME) assessment using a structured scoring rubric across protocol sections
- Development time was benchmarked against a traditional manual conceptualization process conducted by experienced HEOR modellers on the same disease areas

METHODS (Continued)

Multi-Agent GenAI System Architecture for HEOR Model Conceptualization



Iterative refinement loop • Two scenarios: (1) De novo model development in evidence-scarce areas (2) Model enhancement where structures exist

Specialized Agent Roles & Generated Protocol Components

CLINICIAN AGENT *Defines disease biology & patient pathways*	**📊 STATISTICIAN AGENT** *Ensures analytical rigour & uncertainty quantification*	**HEALTH ECONOMIST AGENT** *Structures economic value & HTA-relevant parameters*
Health States *Disease progression pathways* Identifies clinically meaningful stages from diagnosis to end-of-life	**Model Structure** *Markov / partitioned survival* Selects optimal modelling framework based on indication & data	**Cost Inputs** *Drug, admin, monitoring* Sources country-specific unit costs from national tariffs & databases
Transition Probabilities *Evidence-based rates* Extracts rates from trial data, registries & clinical literature	**Sensitivity Analyses** *One-way, PSA, scenario* Specifies analyses to test robustness across parameter ranges	**Utility Values (QoL)** *EQ-5D / utility mapping* Derives health-state utilities for QALY calculation per HTA guidance
Clinical Pathway *Patient journey mapping* Maps treatment sequences, lines of therapy & care settings	**Statistical Parameters** *Distributions & uncertainty* Assigns prior distributions for probabilistic sensitivity analysis	**Resource Use** *Healthcare utilization* Quantifies visits, procedures & resource consumption per health state
Key Assumptions *Clinically valid inputs* Validates biological plausibility & standard-of-care alignment	**Validation Checks** *Face & internal validity* Cross-checks outputs against clinical expectations & benchmarks	**Discount Rates** *Perspective-specific* Applies jurisdiction-appropriate rates for costs & outcomes (e.g. 3.5%)

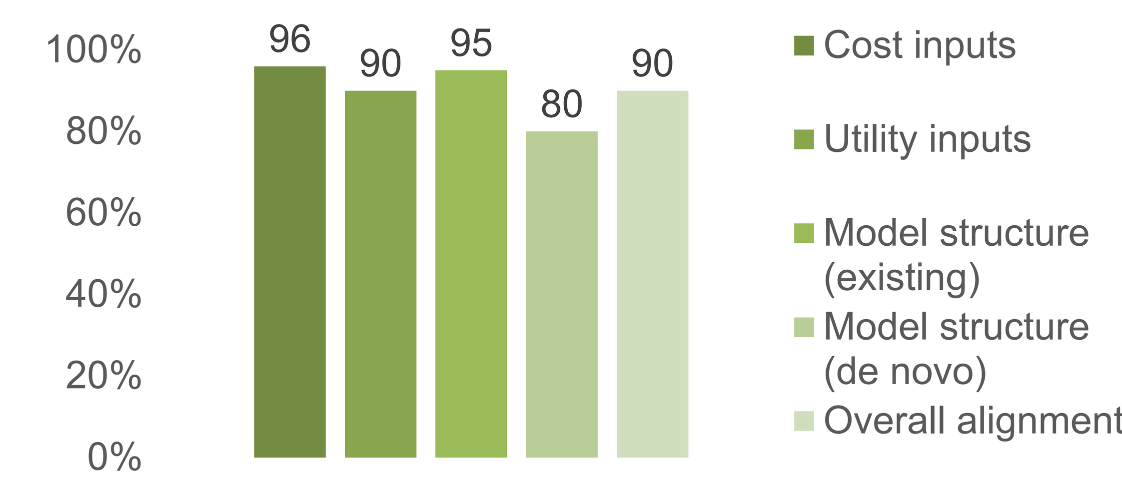
⚙ **Integrated Output** → Comprehensive HTA-ready model protocol: parameters, assumptions, sensitivity analyses & economic framework

RESULTS

90%

Overall alignment with subject matter expert assessments **85% reduction in development time**

Section-Specific Accuracy of AI-Generated Protocols



Section-Level Accuracy

96%	90%	85%	80%
Cost Inputs	Utility Inputs	Model Structure	De Novo

Assessed against subject matter expert benchmarks

Validation Process

1. **AI Generation** – Multi-agent system produces model output
2. **SME Review** – SME evaluates each section blinded to AI vs. human origin
3. **Scoring** – Binary accuracy + qualitative feedback

12 models tested across Markov cohort & partitioned survival; 3 independent SME reviewers

LIMITATIONS

- Lower accuracy in de novo model structure conceptualization (80%) requires continued expert oversight
- Current scope is limited to cohort-level state-transition and partitioned-survival models restricting applicability to more complex modeling approaches
- Dependence on quality of input evidence; outputs require validation by subject matter experts before use

CONCLUSIONS

- Accuracy was highest for well-defined, data-driven inputs; Cost inputs (96%) and utility inputs (90%) showed the strongest agreement with expert assessments, supported by the system's ability to retrieve structured evidence from national cost databases and published quality-of-life literature
- Model structure and clinical pathway development achieved 85% overall accuracy, with stronger performance when refining existing structures (95%) than when building de novo (80%). This gap reflects the greater complexity of reasoning about disease progression and clinical pathways from limited evidence - and underscores where expert oversight remains most valuable
- An 85% reduction in development time enabled significant resource savings, with future efforts focused on improving performance, expanding model scope, and validating outputs for routine use across disease areas

References:
1. Briggs AH, Sculpher MJ, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford University Press; 2006.
2. Lewis P, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems. 2020;33:9459-9474.
3. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford University Press; 2015.

Correspondence: Shubhram Pandey; shubhram.pandey@pharmacoevidence.com
Disclosure: SP, SM, BS, RK, NW and PSJ, the authors declare that they have no conflict of interest.