



# Real-World Use of Generative AI Assistant Tools across Health Economics and Outcomes Research Workflows: A Scoping Review

Eniola Olatunji<sup>1</sup> MPH PhD, Yao Dang<sup>1</sup> PhD, Abimbola Williams<sup>1</sup> MSC MPH PhDc

<sup>1</sup> Boston Scientific, Marlborough, MA, USA

## BACKGROUND

- Applications of Generative Artificial Intelligence (GenAI) in health economics and outcomes research (HEOR) are emerging across evidence synthesis, economic modeling, and Health Technology Assessment (HTA).
- Published evidence on real-world use and evaluation of GenAI in HEOR remains limited, with ongoing concerns about reporting rigor and governance.

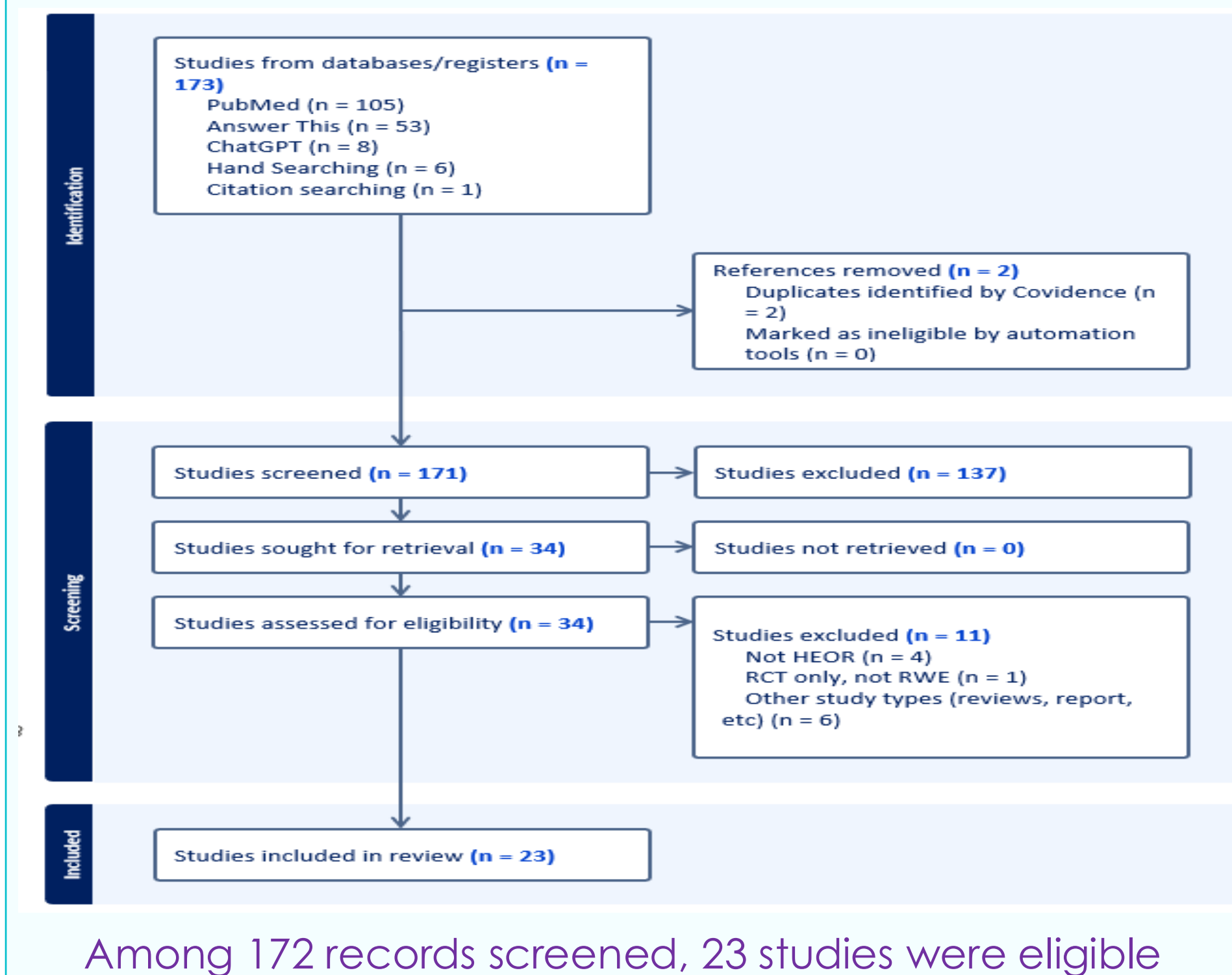
## OBJECTIVES

- To characterize the real-world use, implementation patterns, and methodological quality of GenAI assistant tools across HEOR workflows.
- To identify implementation, governance, and reporting gaps, including hallucination and reproducibility risks.

## METHODS

- **Design:** Scoping review of GenAI applications in HEOR.
- **Scope:** Evidence describing or evaluating GenAI assistants, including use of AI agents to support HEOR/HTA deliverables.
- **Sources:** PubMed, HTA agency databases, GenAI tools, and ISPOR database.
- **Review window:** 2024-2025.
- **Eligible study types:** Peer-reviewed studies, conference abstracts, grey literature, and HTA documents/submissions.
- **Data extracted:** HEOR task, GenAI tool used, integration pattern, reported benefits, and evaluation metrics.
- **Analysis:** Studies were characterized by workflow, tool type, integration pattern, and evaluation approach.
- **Quality assessment:** ISPOR GenAI-HEOR Quality Assessment Framework across 10 domains scored from 0 to 2.

## PRISMA STUDY SELECTION FLOW



## RESULTS

### Implementation Patterns

- Most implementations (91%) relied on prompting GenAI models, usually with human-in-the-loop.

### Quality and Efficiency

- Human-only workflows were the predominant comparator to GenAI models (83%).

### Performance and Governance

- Majority of GenAI tools (83%) reported moderate-to-high accuracy for structured extraction or screening tasks.
- Factuality and comprehensiveness were frequently evaluated (96%) (Fig 1).
- Security, fairness, and bias were infrequently assessed.

### Productivity Gains

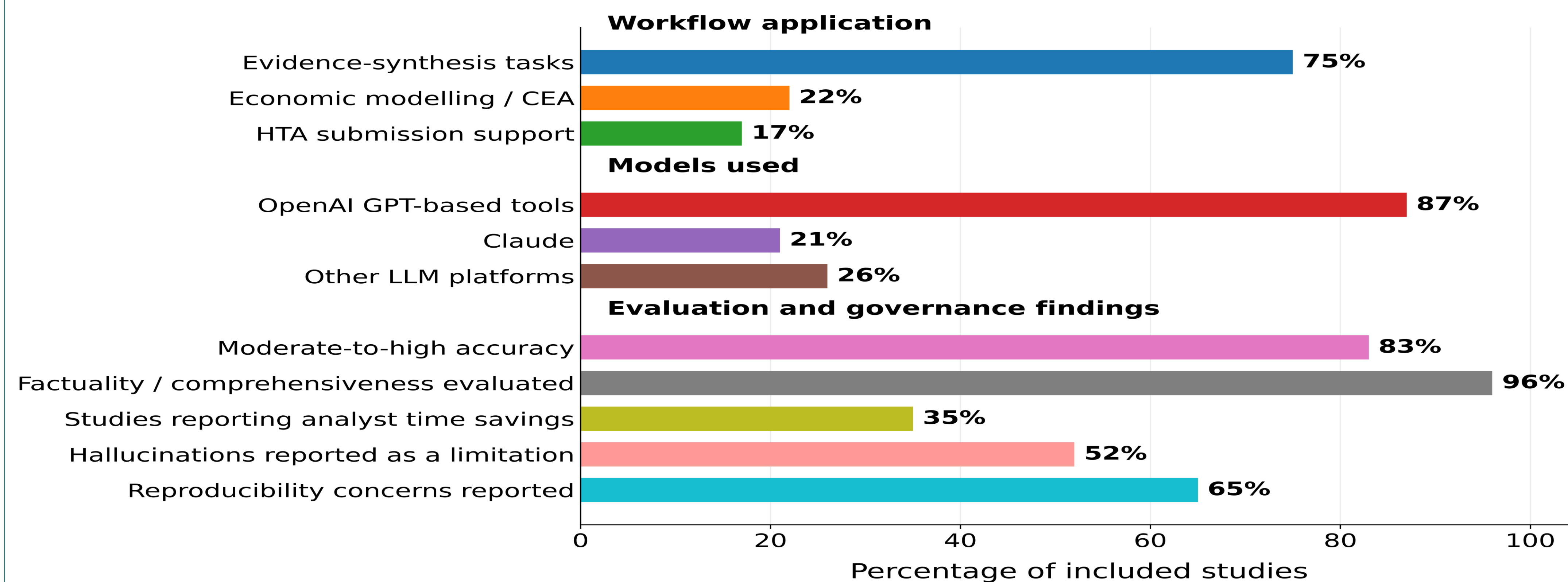
- Analyst time savings were reported in 35% of studies, range: ~15 minutes per study to 350 minutes per reviewer, depending on task complexity.

## CONCLUSIONS

- GenAI use in HEOR is currently concentrated in supervised assistant roles within evidence-synthesis workflows, with empirical evaluation in economic modelling, HTA submissions, and quality assessment.
- The evidence base is most mature for evidence-synthesis, while broader validation is still needed for economic modeling and HTA-related applications.
- Future evaluations should assess accuracy and efficiency alongside governance-oriented outcomes, including transparency, reproducibility, privacy/security, fairness, and bias.

## RESULTS

Figure 1. GenAI in HEOR: Key Results (n=23 studies)



## LIMITATIONS

- Review limited to English-language studies published or presented in 2024-2025.
- HEOR task categories were not mutually exclusive, and included studies varied by task, GenAI tool, comparator, and evaluation metrics.
- Heterogeneity in GenAI tools, comparators, and outcome definitions limited direct cross-study comparison.

## DISCLOSURES

This study was funded by Boston Scientific. All authors are full-time employee of, and shareholders in Boston Scientific.