GenAI Goes To ISPOR: Exploratory, Descriptive Analysis of Generative AI Performance for Summarizing and Synthesizing ISPOR Research Abstracts as Sources

Cynthia Drake Morrow PhD[1,2], Ellen Thiel[1], Jennifer P. Wisdom[1,3], Blaine Reeder[4]

1. Knowledge Resolution, 2. Michigan State University, 3. Wisdom Consulting, 4. University of Missouri

KNOWLEDGE RESOLUTION

CONNECT WITH US

POSTER NUMBER MSR38

OBJECTIVE: To assess two publicly available GenAI tools for summary and synthesis of real-world evidence sources

Background

- Generative artificial intelligence (GenAI) tools hold promise to streamline evidence synthesis in life sciences
- GenAI accuracy and traceability are unknown
- Evidence verification steps for GenAI are missing
- ISPOR established official Working Group to explore key implementation areas and limitations of GenAI[1]

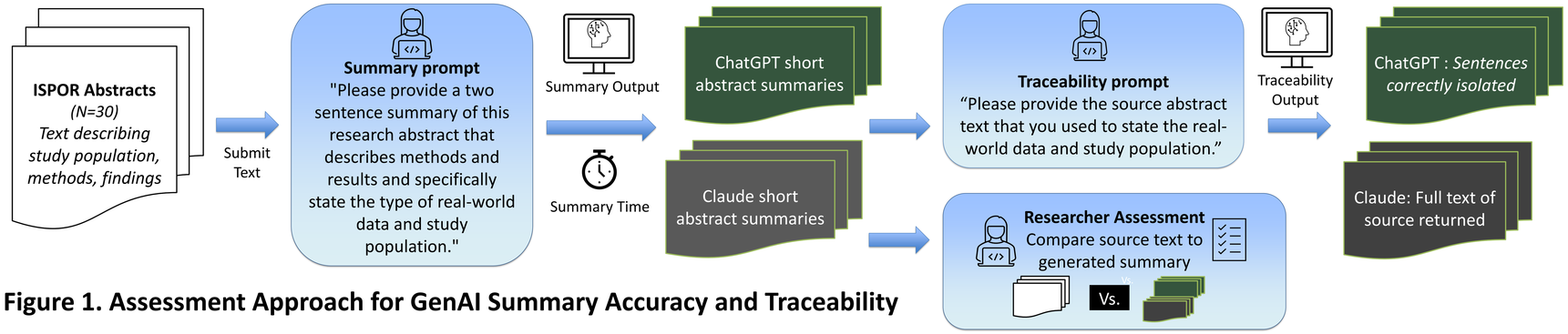

Figure 1. Assessment Approach for GenAI Summary Accuracy and Traceability

Methods

SAMPLE AND TOOLS

- Test corpus consisted of a random sample of ISPOR abstracts (N=30) from the June 2023 issue *Value in Health*
- ChatGPT v2 and Claude 3.5 Sonnet were assessed by two researchers for summarization performance and output using the prompt and approach as shown (**Figure 1**).

ASSESSMENT METRICS

- **Speed of Single Source Summary:** Time to generate summary from a source
- **Accuracy:** Binary score of accurate/not accurate for summary to source for 4 categories:
  - Data type (if missing, excluded from denominator)
  - Study population
  - Methodology
  - Primary research findings
- **Traceability:** Researchers prompted tools to identify source text that was used for the creation of the short summary.
- **Synthesis Across Multiple Sources:** Full body of 30 abstracts uploaded and tools were prompted for counts of abstracts that included electronic health record (EHR), claims data, or multiple real-world data (RWD) sources.

	ChatGPT v2	Claude 3.5 Sonnet
Mean (SD) time, in seconds	7.4 (2.2)	4.5 (0.7)

Figure 2. Time to Generate 2-Second Summary

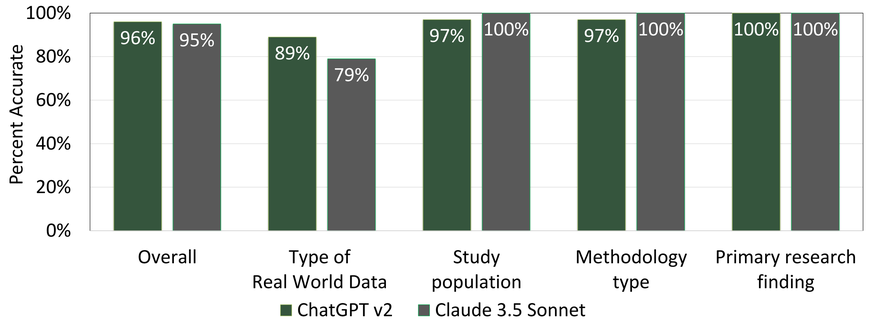

Figure 3. Summary Accuracy, Overall and by Specific Category

	ChatGPT v2	Claude 3.5 Sonnet
SPEED	A+	A+
ACCURACY	A	A
TRACEABILITY	B	D
SYNTHESIS	F	F

Figure 4. Summary of Conclusions

DISCLOSURE. *Dr. Morrow is an independent contractor/consultant in life sciences who receives payment for data analysis and writing service. Dr. Reeder is co-founder and Chair of the Board of Directors for Hekademeia Research Solutions, a non-profit organization whose mission is to support technology-based academic research. This role is an unpaid volunteer position.*

GenAI tools were NOT used to generate abstract or poster text.

Results and Conclusion

Speed. Both ChatGPT and Claude generated summaries from a single source in under 10 seconds (**Figure 2**).

Accuracy. Both tools had high over all summary accuracy scores with ChatGPT at 96% and Claude at 95% (**Figure 3**).

Traceability. ChatGPT accurately returned relevant source sentences from the methodology or results sections (30 out of 30 prompts). Claude was *especially limited in its ability to isolate source text* for traceability purposes and simply returned the full text of each abstract.

Synthesis Across Sources. ChatGPT reported no abstracts utilized EHR or claims data; Claude provided inaccurate counts (2 for EHR, 13 for claims). Manual review revealed that abstracts utilized EHR data (N=7), claims (N=12), and other/multiple sources (N=11).

CONCLUSION

GenAI is promisingly fast for accurately summarizing research findings from a single source (**Figure 4**). Improvements are need for traceability and accuracy when synthesizing multiple sources. Rapidly evolving technology may quickly date these GenAI performance assessment results from December 2024.

REFERENCES

1. Fleurence RL, Bian J, Wang X, Xu H, Dawoud D, Higashi M, Chhatwal J; ISPOR Working Group on Generative AI. Generative Artificial Intelligence for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations: An ISPOR Working Group Report. Value Health. 2025 Feb;28(2):175-183. doi: 10.1016/j.jval.2024.10.3846. Epub 2024 Nov 12. PMID: 39536966; PMCID: PMC11786987.