Non-Systematic Literature Reviews: Can AI enhance current methods?

MSR49

Baisley W[1], Perriello L[1], Shoushi G[1], Nguyen K[2], Lahue BJ[1]

[1]Alkemi LLC, Manchester Center, VT, USA, [2]University of Illinois at Chicago College of Pharmacy, Chicago, IL, USA

alkemi® Unlocking Value, Achieving Access.

Background

- Artificial intelligence (AI) tools may streamline generation of narrative syntheses and sourcing of peer-reviewed and grey literature, however more evaluation is required[1]
- The objective of this methods research was to evaluate the use of AI tools in targeted literature reviews (TLRs)

Methods

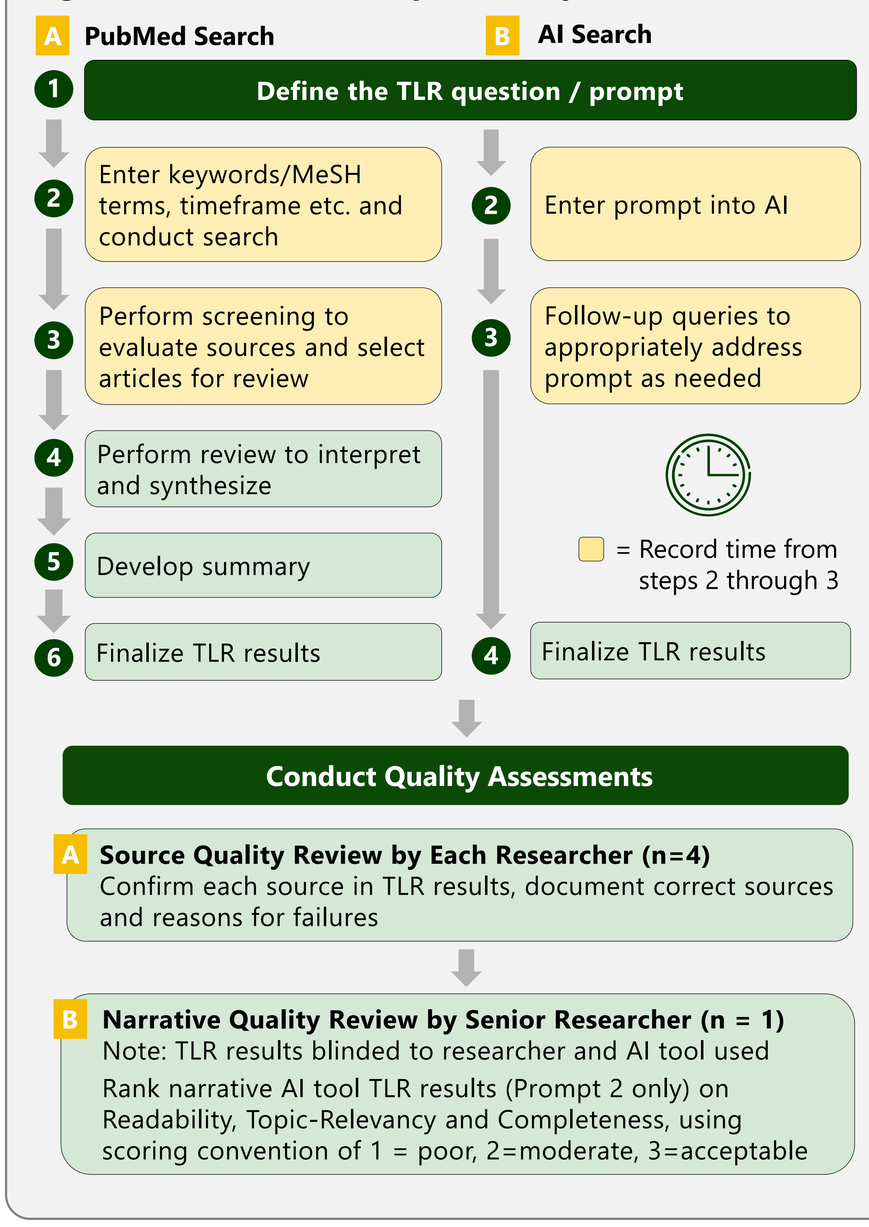
- Health economic TLR prompts and key words, as well as process and quality outcomes were pre-specified (**Table 1**)
- Study methods were mapped to design a PubMed TLR process and a new AI tool TLR process **(Figure 1)**
- TLRs were conducted using either PubMed or AI tools (ChatGPT + "*Browse with Bing Beta*" plugin, Microsoft Bing, Google Bard) and search steps timed
- Quality assessments were performed on TLR results to evaluate source and narrative quality as described below
- Outcomes were descriptively analyzed by prompt and method

Table 1. TLR Prompts and Outcome Definitions

Prompt Question	Key Words	Goal of TLR
1. What are the 5 most recent publications on economic model considerations in rare disease?	economic modeling, rare disease	Identify 5 recent and topic-relevant sources published since 2018
2. What are economic considerations in rare disease, and provide 3 relevant citations?	economic considerations, rare disease	Create a narrative synthesis with 3 relevant citations

Outcome Measure	Units / Categories	Outcome Definition
Process Efficiency	*Minutes, seconds*	Time to complete pre-defined steps for each TLR method
Source Quality	*Valid*	Topic relevant citation, correct time frame
	Failure *- Not valid* *- Missing*	- Incorrect citation, off-topic, prior to 2018 (for Prompt 1) - Citation not provided
Narrative Quality	*3= Acceptable* *2 = Moderate* *1 = Poor*	Rank of quality from 1 to 3 on attributes

Figure 1. TLR Process Steps & Study Methods



Results

- AI method reduced standard TLR process by two steps **(Figure 1)**
- Twenty TLR responses (n=18 AI, n=2 Control) were generated for analysis

Process Efficiency: On average, AI reduced time to complete search steps compared to PubMed **(Figure 2)**, however significant additional time was required for source QC (not timed) of AI responses

Source Validity: TLRs performed with AI method were unreliable for sourcing (**Figure 3**) while PubMed method returned 100% valid references. Of note, Prompt 2 returned a higher proportion of valid sources compared to Prompt 1 and Bing had highest sourcing quality across Prompts (**Table 2**)

Narrative Quality: Prompt 2 narrative quality varied; Compared to other AI tools, ChatGPT narratives ranked highest in blinded review of 9 responses **(Figure 4)**

Figure 2. Researcher Time Recorded on TLR Timed Steps

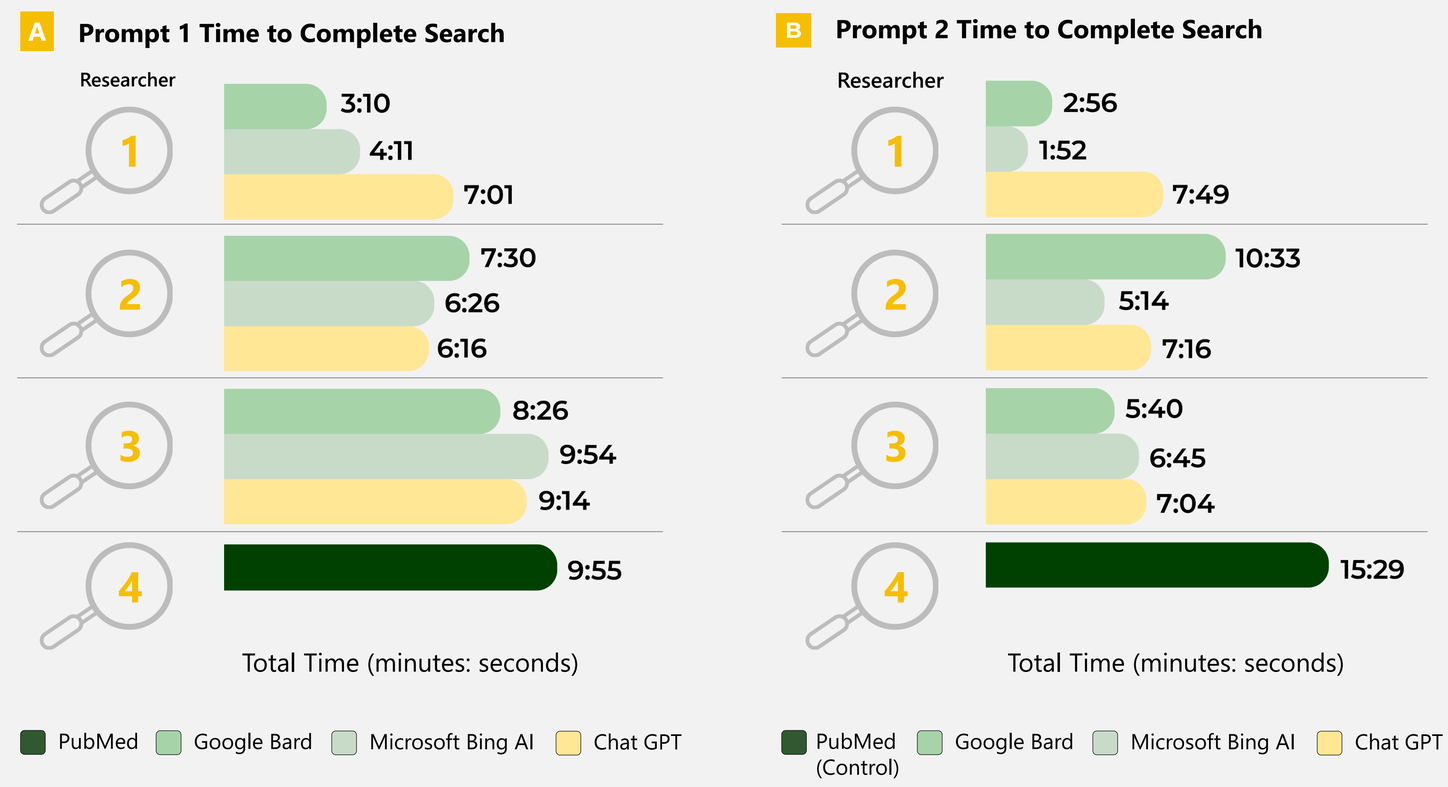


Figure 3. AI TLR Source Quality

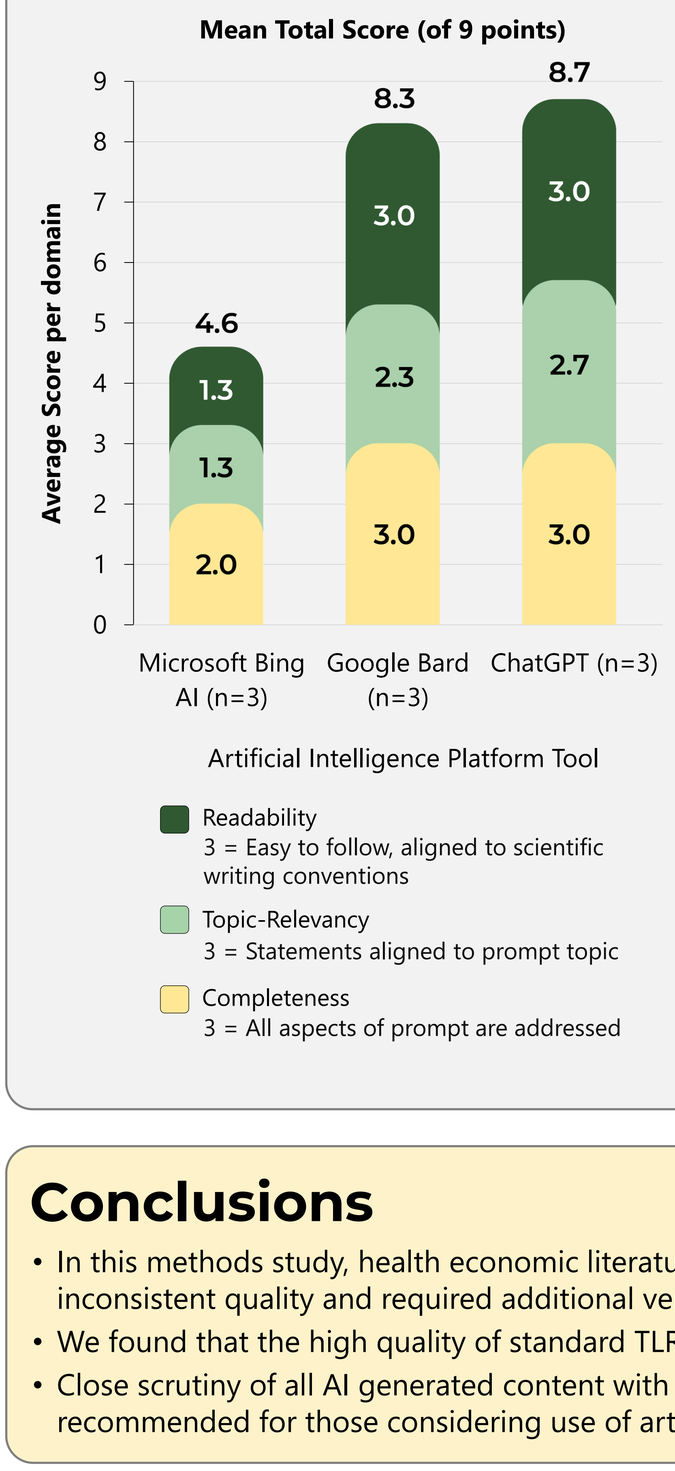
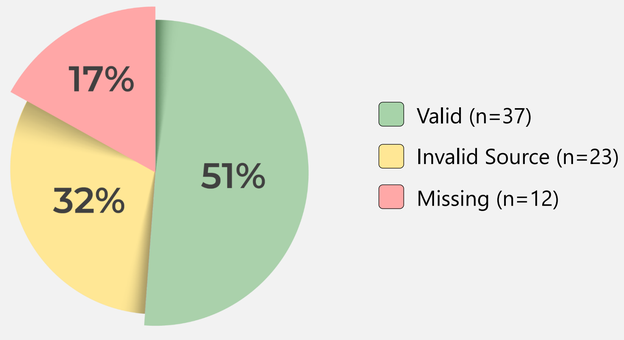


Table 2. Source Quality by Prompt and AI Tool

Prompt, Tool	Source Failures/ Total Requested (%)	Reasons for Failure
All Prompt 1	**30/45 (67%)**	**Majority of summaries failed source validity test**
ChatGPT	13/15 (87%)	Unable to source literature more recent than 2021 despite plugin), site crash (20% of searches)
Google Bard	13/15 (87%)	Repeated citation within the same prompt, missing citations
Bing AI	4/15 (27%)	Repeated citation within same prompt
All Prompt 2	**5/27 (18.5%)**	**Minority of summaries failed source validity test**
ChatGPT	3/9 (33%)	Not peer-reviewed, tool crashed, missing citations
Google Bard	2/9 (22%)	Title and DOI did not match
Bing AI	0/9 (0%)	N/A

Figure 4. AI TLR Narrative Quality



Conclusions

- In this methods study, health economic literature searches executed with the AI were associated with inconsistent quality and required additional verification steps that offset anticipated time savings
- We found that the high quality of standard TLR methods for sourcing was not matched by the tested AI tools
- Close scrutiny of all AI generated content with extra quality control procedures to verify sources are strongly recommended for those considering use of artificial intelligence to support literature research

Acknowledgements: Alkemi LLC thanks Lucas Blackmore and Alexa Abrenilla for their assistance in creating this poster.

Disclosures: WB, GS, LP, and BJL were employees of Alkemi at the time of research.

References: 1. Wagner G, Lukyanenko R, Paré G. Artificial Intelligence and the conduct of literature reviews. *Journal of Information Technology*. 2022;37(2):209-226. doi:10.1177/02683962211048201

Presented at ISPOR EU, November 2023, Copenhagen, Denmark.
Please contact betsy.lahue@alkemihealth.com for more information.
www.alkemihealth.com