Poster #21

Exploring the utility of diverse and innovative artificial intelligence (AI) platforms for literature searches in medical communications

Johnson Ying, Maura C Belanger, Stephanie Justice-Bitner, Ryan P Reynolds, Clotilde Robinson, Danielle L Rubin-Shepherd, and Meryl Gersh

Red Nucleus, Yardley, PA, USA

Can AI platforms with literature search capabilities identify accurate and appropriate references for scientific writing?

Background

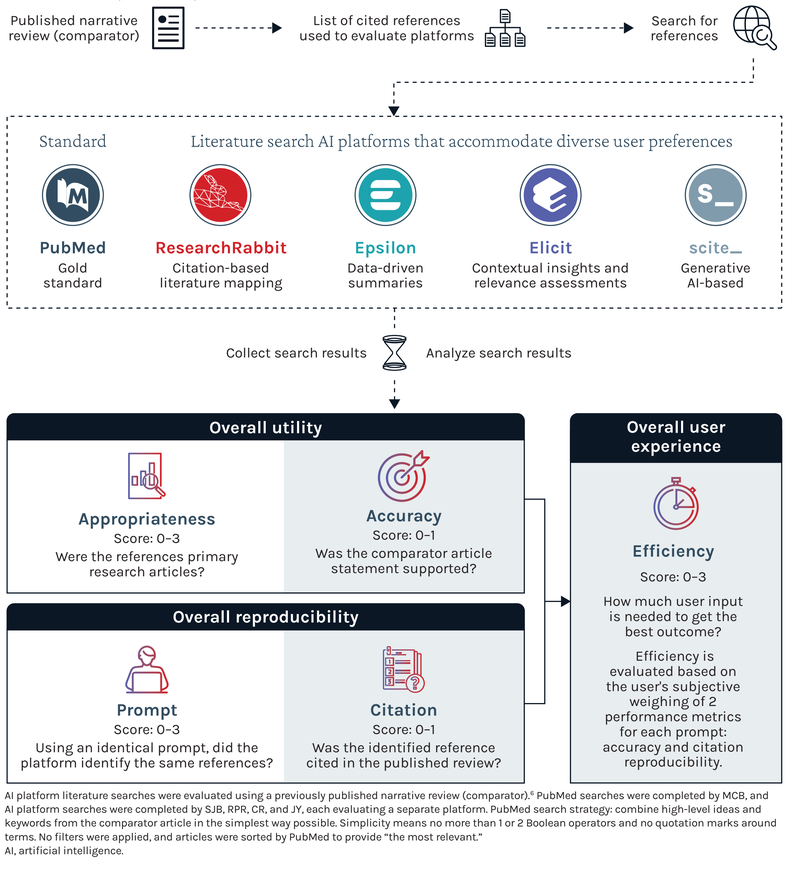
- Artificial intelligence (AI) platforms are increasingly being adopted into the scientific process to streamline multiple facets of medical communications[1,2,3] but are also known to generate unreliable references for scientific writing[4,5]
- With the growing availability of diverse AI platforms specialized for literature searches, carefully assessing their advantages and limitations is crucial

Objective

- To evaluate the accuracy and appropriateness of AI-identified references compared with human-conducted PubMed searches

Research Design and Methods

- Four AI platforms specialized to conduct literature searches (ResearchRabbit, Epsilon, Elicit, and scite_) were prompted to find references supporting statements from a published comparator narrative review article[6]
  - ResearchRabbit is powered by PubMed and Semantic Scholar. Epsilon and Elicit are powered by Semantic Scholar. scite_ indexes all documents with a DOI
- Performance metrics related to the overall utility, reproducibility, and user experience of each AI platform were scored and compared with human-conducted literature searches (PubMed)



AI platform literature searches were evaluated using a previously published narrative review (comparator).[6] PubMed searches were completed by MCB, and AI platform searches were completed by SJB, RPR, CR, and JY, each evaluating a separate platform. PubMed search strategy: combine high-level ideas and keywords from the comparator article in the simplest way possible. Simplicity means no more than 1 or 2 Boolean operators and no quotation marks around terms. No filters were applied, and articles were sorted by PubMed to provide "the most relevant."
AI, artificial intelligence.

Conclusions

Standard

PubMed

Traditional search strings and Boolean operators

The gold standard for literature searches, using customizable prompts and filters

Literature search AI platforms that accommodate diverse user preferences

ResearchRabbit

Interactive, graphical, citation-based literature mapping exploration

Supports literature exploration through visualization of citation relationships

Epsilon

Recommends papers relevant to question-based prompts

Concise summarization of relevant papers that is heavily influenced by prompt quality

Elicit

Summarizes key insights of papers relevant to question-based prompts

Streamlines extraction of specific information from relevant papers

scite_

Generative AI-based interface with prompt-based search

Generative summary of relevant papers that is heavily influenced by prompt quality

Findings

- AI-identified references were less accurate than human-conducted PubMed searches
- AI-identified references were generally appropriate but often included review articles instead of primary research articles compared with PubMed searches
- Neither PubMed nor the AI platforms reliably identified all citations used in the comparator article
- AI platforms were less efficient compared with PubMed, but user experience is partly dependent on the learning curve for each AI platform

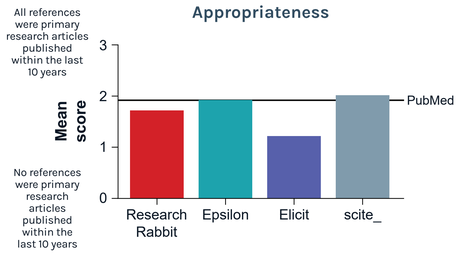
Perspectives

- While AI platforms have the potential to identify review articles that facilitate early-stage literature exploration, human verification of AI-identified references remains necessary, thereby limiting the overall utility of AI platforms for literature searches
- PubMed and AI platforms can facilitate reference selection, but the final choice is primarily shaped by informational context and remains a human-driven decision open to diverse interpretation
- AI platform interfaces offered diverse user experiences, suggesting that a platform's value emerges from the synergy between its core capabilities and human preference

Results

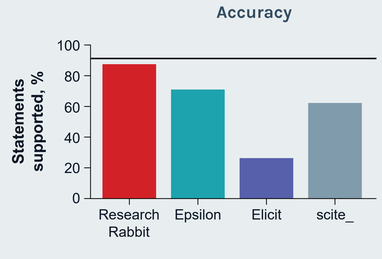
Overall utility

AI platforms generally identified appropriate primary research articles for scientific statements from the comparator article compared with PubMed



Appropriateness evaluated the ability to identify appropriate references, irrespective of whether they were cited in the comparator article. Appropriate references were primary research articles published within the last 10 years.

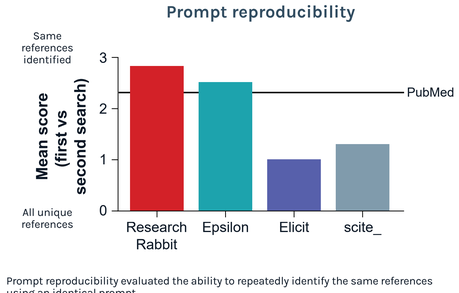
AI platforms were less accurate than PubMed in identifying references that supported comparator article statements



Accuracy evaluated the ability to support a given scientific statement in the comparator article with identified references.

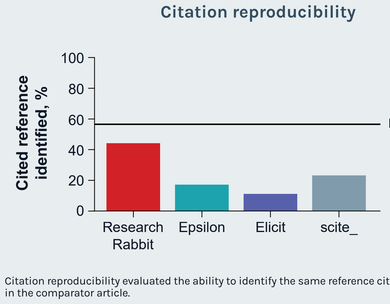
Overall reproducibility

AI platforms were diverse in their ability to repeatedly identify the same references using an identical prompt



Prompt reproducibility evaluated the ability to repeatedly identify the same references using an identical prompt.

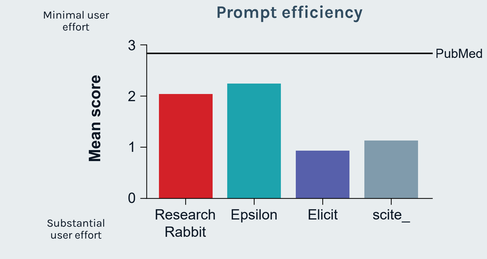
Neither AI platforms nor PubMed identified >60% of the references used in the comparator article, indicating that reference choice is a human-driven step open to diverse interpretation



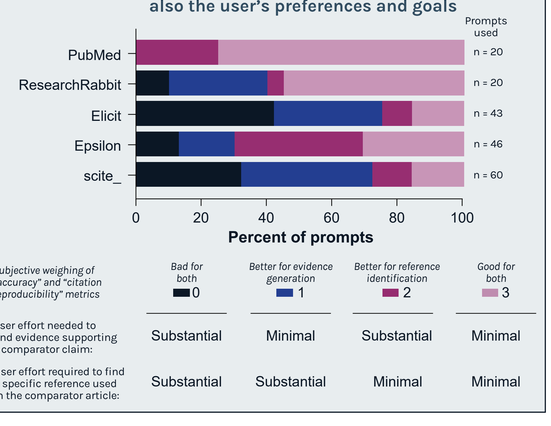
Citation reproducibility evaluated the ability to identify the same reference cited in the comparator article.

Overall user experience

AI platforms were less efficient and demanded greater user effort in identifying references that supported comparator claims and specific references used in the comparator article compared with PubMed



Efficiency is evaluated based on the user's subjective weighing of 2 performance metrics for each prompt: accuracy and citation reproducibility.



Limitations

- Free versions of AI platforms were used
- Results are based on the first 25 references identified from each prompt
- The study design required experimenters to be unblinded
- Platforms were scored retrospectively based on a published article

Acknowledgements

This study was sponsored by Red Nucleus.

Disclosures

All authors have no conflicts of interest to disclose.

Corresponding author

Corresponding author: Johnson Ying (jying@rednucleus.com)

References

1. LeCun Y, et al. *Nature.* 2015;521(7553):436-44.
2. Wang H, et al. *Nature.* 2023;620(7972):47-60.
3. Martin K. *Medical Writing.* 2023;32(3):22-7.
4. Kacena MA, et al. *Curr Osteoporos Rep.* 2024;22(1):115-21.
5. Athaluri SA, et al. *Cureus.* 2023;15(4):e37432.
6. Kushner T, et al. *Curr Med Res Opin.* 2024;40(5):813-20.