The Promise and Pitfalls of Artificial Intelligence (AI) Literature Search Tools in Medicine

Kimberly Forman (M-2), Rachel Armstrong, Allison Kowatch, Nawal Bennett, Megan Courtley, Michelle Padley, Jeffrey Jones

Michigan State University College of Human Medicine, Department of Emergency Medicine, Grand Rapids, MI

College of Human Medicine
MICHIGAN STATE UNIVERSITY

INTRODUCTION

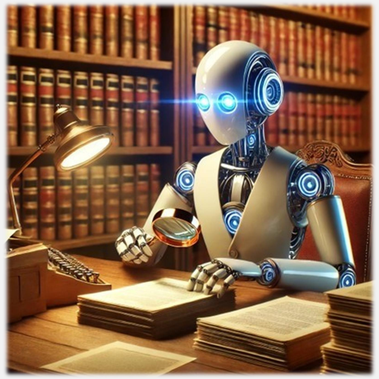
- Conducting comprehensive literature reviews often poses challenges for clinicians and medical researchers, requiring extensive reading and synthesis of existing works while advancing original insights.
- Numerous artificial intelligence (AI) research tools have emerged that aim to assist medical researchers and clinicians, but finding the best tools can be overwhelming.
- The study assessed AI tools based on their ability to accurately identify relevant research papers, extract key information, generate concise summaries, and facilitate the analysis of trends across studies.

METHODS

- This prospective descriptive analysis analyzed six semantic AI search tools that interpret the meaning of words and phrases in a query and six citation-based search tools that match keywords.
- Evaluation criteria were developed by a panel of emergency medicine faculty.
- These criteria included search functionality, data analysis and summarization, accuracy (compared to four human searches), relevance, limitations, unique features, user-friendliness, and cost.
- Descriptive statistics and frequency tables were used to describe the key quantitative and qualitative variables.

RESULTS

- The six semantics-based tools (e.g., Consensus, Perplexity) were more suitable for exploring nuanced topics and emerging research.
- In contrast, the six citation-based tools (e.g., Connected papers, Scite.ai) focused on the impact of research through citation metrics and network analysis.
- Using both approaches together can provide a more comprehensive and nuanced literature search
- Most AI search tools offered a user-friendly interface (92%), quick summaries of research findings (75%), content analysis (67%), conversational search abilities (50%), and literature mapping (42%).
- Monthly fees ranged from free to $29/month based on the features selected.
- The accuracy compared to human searches ranged from 100% to 87%.
- All of the tools flagged results as relevant when they were not (false positive rate of 3-16%).
- AI hallucinations and incorrect or mismatched citations (3-16%).



DISCUSSION

- Semantics-based AI tools are like a skilled translator who understands the meaning and context of your question and finds relevant information, even if the exact words are different (for example, recognizing that "renal" and "kidney" are related).
- Many semantics-based tools primarily use open-access sources, which means they may miss important paywalled or less-accessible research, resulting in incomplete literature coverage.
- Citation-based AI tools are like a historian who traces the network of references between documents to show you what has been influential or connected in the past. Much like a historian mapping the lineage and impact of ideas through time.

CONCLUSIONS

- While AI search tools provide rapid access to literature and can be helpful for preliminary research, they currently lack the precision and contextual understanding of human researchers.
- Users should maintain a critical perspective when using these tools, particularly when making important clinical decisions based on their findings.
- Cross-verifying data with original research is advisable to ensure accuracy.