Assessment of Reasoning Agents for Building Literature Search Strategies

Joshua Twaites, MS, Kevin Kallmes, MA, JD, Karl Holub, BS;

Nested Knowledge, Inc., St. Paul, MN, USA.

Correspondence to: kevin.kallmes@nested-knowledge.com

Disclosures

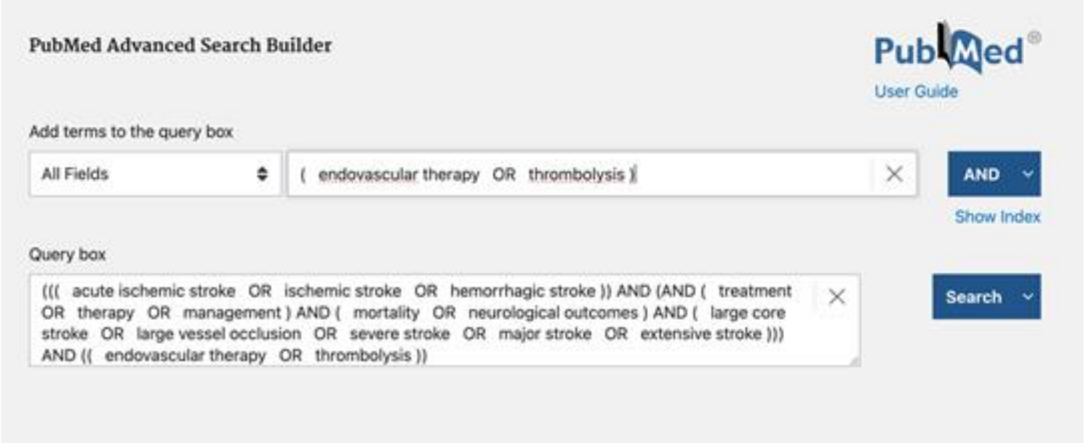
- Joshua Twaites is an equityholder in and employed by Nested Knowledge

- Kevin Kallmes is an equityholder and employed by Nested Knoweleddge, Inc. He is an equityholder and board member of Superior Medical Experts and Piraeus Medical.

- Karl Holub is an equityholder in and employed by Nested Knowledge

Background on Literature Search needs

- Comprehensive searching is inherently complex and requires specific expertise (1). These can include:
  - Knowledge of structured vocabularies, as well as Biomedical education and database-specific experience
  - Experience and training in Boolean operators (beyond “AND” and “OR”!)
  - Soft skills to balance coverage vs. time spent screening



Sample search in PubMed Advanced builder

The problem: Difficulty and complexity of crafting focused (high-precision) and comprehensive (high-recall) Boolean search strategies.

1. Cooper, C., Booth, A., Varley-Campbell, J. et al. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. BMC Med Res Methodol 18, 85 (2018). https://doi.org/10.1186/s12874-018-0545-3

Background on AI-driven Search tools

- LLMs are widely used in SLR screening and extraction, but have limitations for Search Strategies (2):
  - Current tools lack transparency and human feedback
  - Prone to hallucinations, e.g. MeSH terms
  - Embedding methods help, but reduce interpretability

PubMed Search String:

plaintext

Copy

Edit

```
(("Multiple Sclerosis, Chronic Progressive"[Mesh] OR "Multiple Sclerosis"[Mesh] OR "progr
AND
("Immunomodulation"[Mesh] OR "Immunosuppressive Agents"[Mesh] OR "immunomodulatory"[tiab]
AND
(alemtuzumab[tiab] OR azathioprine[tiab] OR cladribine[tiab] OR cyclophosphamide[tiab] OR
AND
("Randomized Controlled Trial"[Publication Type] OR "Comparative Study"[Publication Type]
```

Output from ChatGPT

Our approach: Human-in-the-loop reasoning agents for building Boolean search strings.

2. Qureshi, R., Shaughnessy, D., Gill, K.A.R. *et al.* Are ChatGPT and large language models “the answer” to bringing us closer to systematic review automation?. *Syst Rev* 12, 72 (2023). https://doi.org/10.1186/s13643-023-02243-z

AI Approach

- The “Generator” and the “Critic”:
  - Generator LLM suggests terms, Critic evaluates search results and provides changes, incl. removing errors
- Iterative loop; controls terms
  - Iteratively Scopes search results, balancing recall vs. precision,
  - Controls and heuristics to minimize hallucination
  - Human-in-the-Loop Exploration/Edits

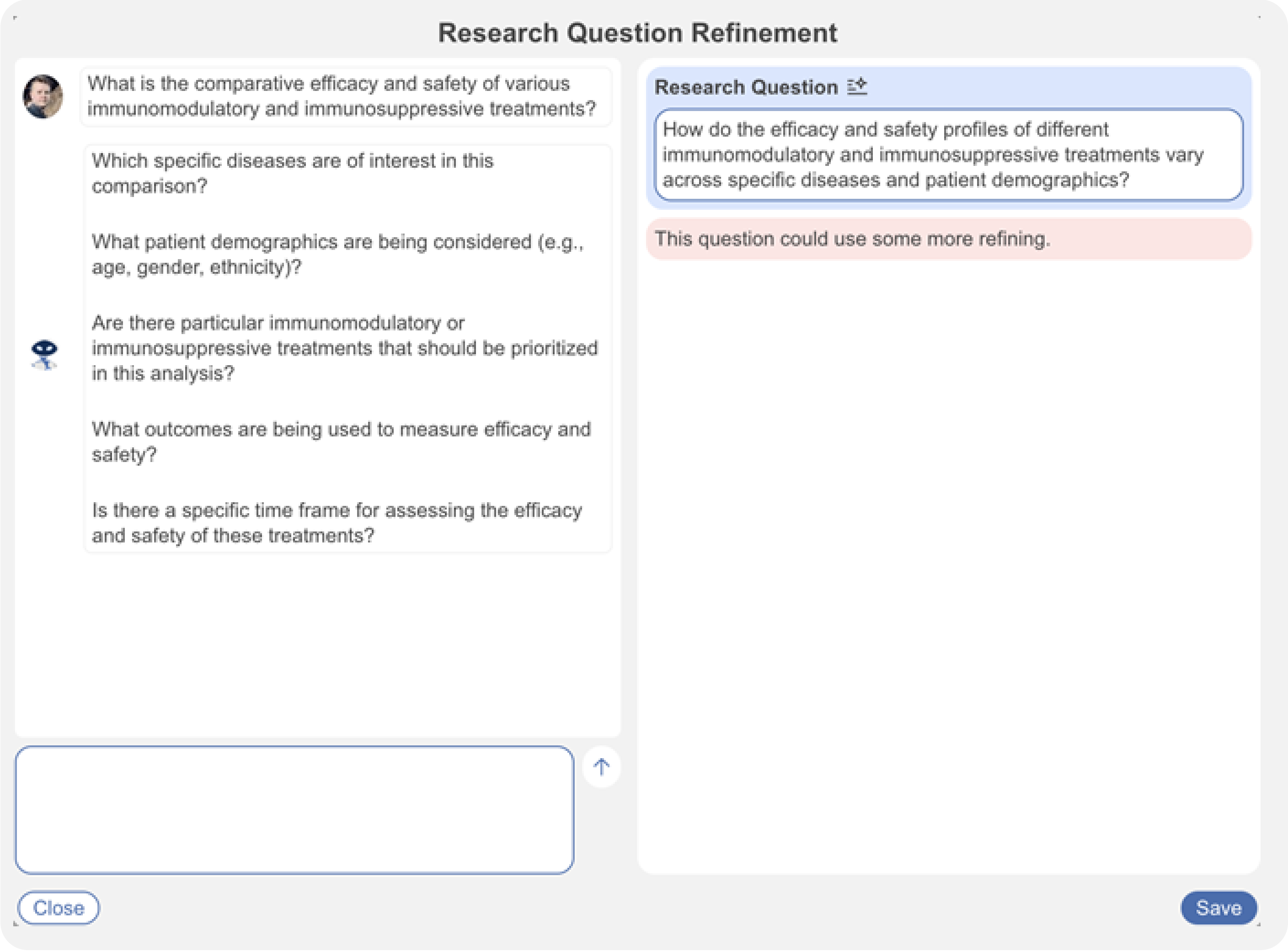
Conceptual process underlying Smart Search (simplified)

PALISADE compliance: Fully transparent methods disclosed online (3).

3. Twaites, J, Holub K. Model Card: Smart Search. Nested Knowledge 2024 (accessed May 12, 2025). https://about.nested-knowledge.com/docs/smart-search-2/

Feature Overview

- In November 2024, Launched **Smart Search**, an LLM-based reasoning agent in the Nested Knowledge SLR platform.
- Uses chain-of-thought reasoning and a generator-critic loop.
- Builds Boolean strings from user Research Questions and chat-based clarifications.



Output from Smart Search

NESTED KNOWLEDGE.

Validation Methods

- Validated using 10 Cochrane SLRs—measured Recall vs. gold-standard PubMed records.
- Repeated on 20 SLRs in Nested Knowledge and compared with black-box LLMs (e.g., GPT).

Topic	Cochrane full text	Cochrane Includes
Multiple Sclerosis	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD015443.pub2/epdf/full	28
Immunotherapy for NSCLC	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011300.pub3/full	11
Renal Cell Carcinoma	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012796.pub2/full	87
Drugs for Subfertile Obese women	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012650.pub2/full	10
Drugs for NAFLD	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011640.pub2/full	85
Drugs for Epilepsy	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD008781.pub3/full	5
HIV	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD006495.pub5/full	11
Statins	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013673.pub2/full	113
Ischemic Conditioning	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011719.pub3/full	39
Prostate Cancer	https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012548.pub2/full	33

Ten Cochrane reviews chosen across various fields of study

Results

Cochrane reviews covered topics including MS, NSCLC, RCC, subfertility, NAFLD, epilepsy, HIV, statins, ischemic conditioning, and prostate cancer.

Smart Search Recall:

- 76.8% vs. Cochrane reviews
- 79.6% vs. Nested Knowledge reviews

Black-box LLM Recall: 13.0%

Recall:
76.8-79.6%

Results

Cochrane reviews covered topics including MS, NSCLC, RCC, subfertility, NAFLD, epilepsy, HIV, statins, ischemic conditioning, and prostate cancer.

Smart Search “inclusion rate”:

- 1.81% of results were included in Cochrane reviews,
- 0.47% of results were included in Nested Knowledge reviews

Typical in expert reviews: 1%-5%

Wang et al. (4): Across 139,467 records, inclusion rate of 5.3%

Interpretation:

Smart Search slightly prioritizes Recall over Precision

4. Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. PLoS One. 2020 Jan 14;15(1):e0227742. doi: 10.1371/journal.pone.0227742. PMID: 31935267; PMCID: PMC6959565.

Conclusion

- Human-in-the-loop reasoning agents using “generator-critic” methods can effectively generate SLR search strategies.
- Smart Search outperformed black-box LLMs and achieved acceptable Recall across diverse clinical topics.
- Further research is needed to compare reasoning agent searches with expert-crafted strategies.

Output from Smart Search in the Human-in-the-loop Exploration module

{NESTED} KNOWLEDGE.