Development of an AI-Powered Tool to Accelerate and Enhance Systematic Literature Reviews for Evidence-Based Decision-Making in Clinical Research

Authors: Paul Loustalot[1], Boris Kopin[1], Sacha Levy[1], Basile Ferry[1], Vincent Martenot[1]
Affiliation: [1]Quinten Health, Paris, France

MSR72
ISPOR Europe
2025

INTRODUCTION

- Systematic literature reviews (SLR) are essential for synthesizing scientific evidence, particularly in clinical research. These processes remain highly time-consuming and labour-intensive due to the need to screen large volumes of publications.
- Traditional methods, such as BM25-based search engines or manual filtering, often fail to achieve satisfactory coverage without extensive manual review[1].
- This limitation underscores the need for hybrid approaches combining artificial intelligence (AI) and human validation to improve efficiency and accuracy.
- Advances in Natural Language Processing (NLP) and the emergence of Large Language Models (LLMs) have demonstrated remarkable capabilities in automating text processing tasks[2]. Several works report time savings up to 68.5% and improved precision and consistency compared with traditional approaches[3,4].
- Full automation remains limited by challenges in achieving comprehensive coverage and generalizability comparable to expert review. Consequently, hybrid approaches integrating human oversight are essential, leveraging the speed of LLMs, while preserving the interpretative rigor of experts[5].

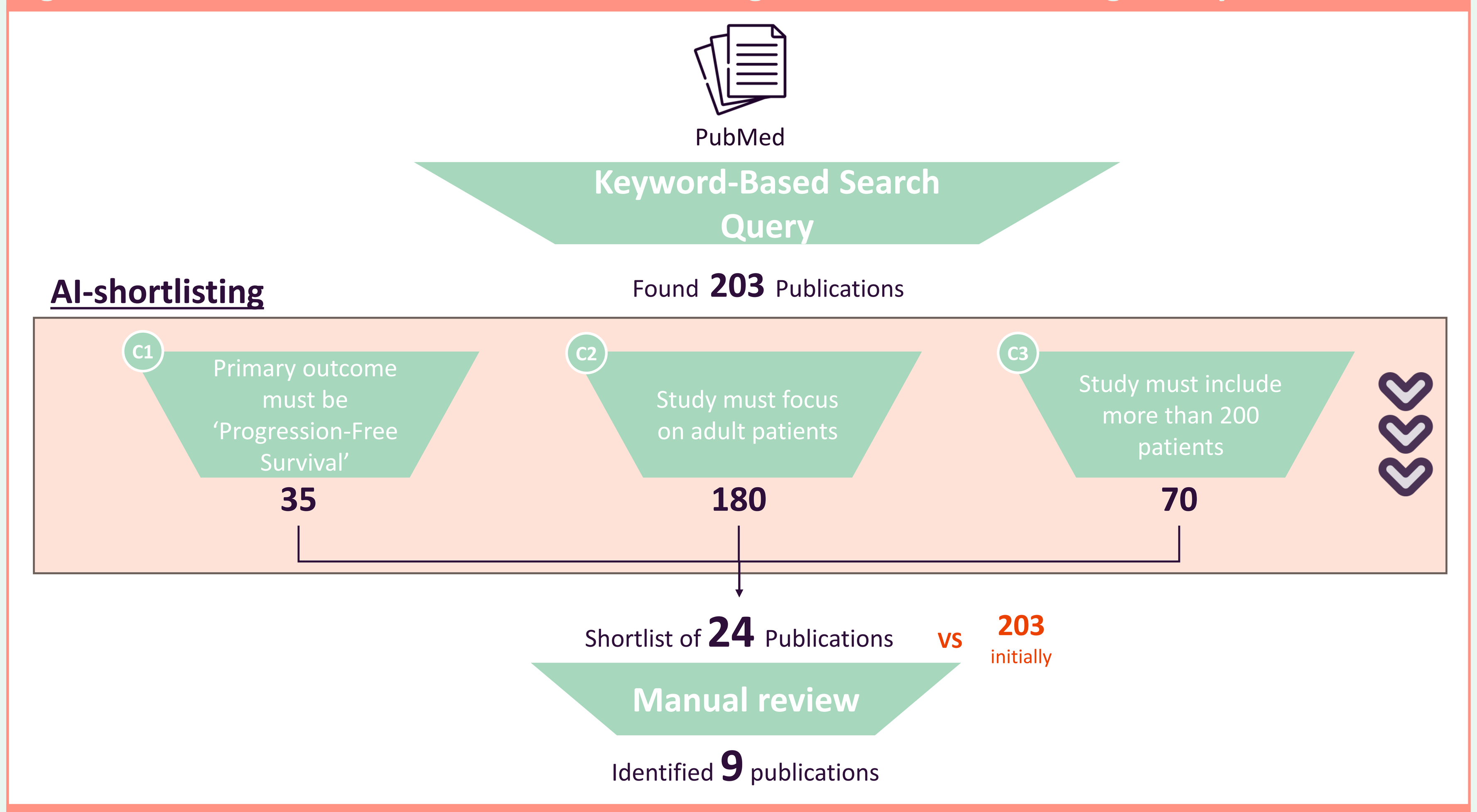
OBJECTIVE

This study aimed to develop and validate a semi-automated **AI-powered tool** to **accelerate** and **enhance the SLR process** by semi-automating **title/abstract screening**, enhancing **efficacy** and curtailing **time consumption**, while preserving methodological **rigor** and adherence to **established systematic review methodologies**.

METHODS

The AI assistance in the developed tool intervenes at the second step of the literature review process - the title/abstract screening based on **reviewer-defined YES/NO criteria** - following the initial keyword-based search. The reviewer then manually reviews the AI shortlisted publications.

Figure 1: Semi-automated Literature review using LLM to assess criteria given by reviewer



The AI-shortlisting of Title/Abstract reduced manual review of publications from 203 to 24.

The **LLM prompting framework** is based on two steps:

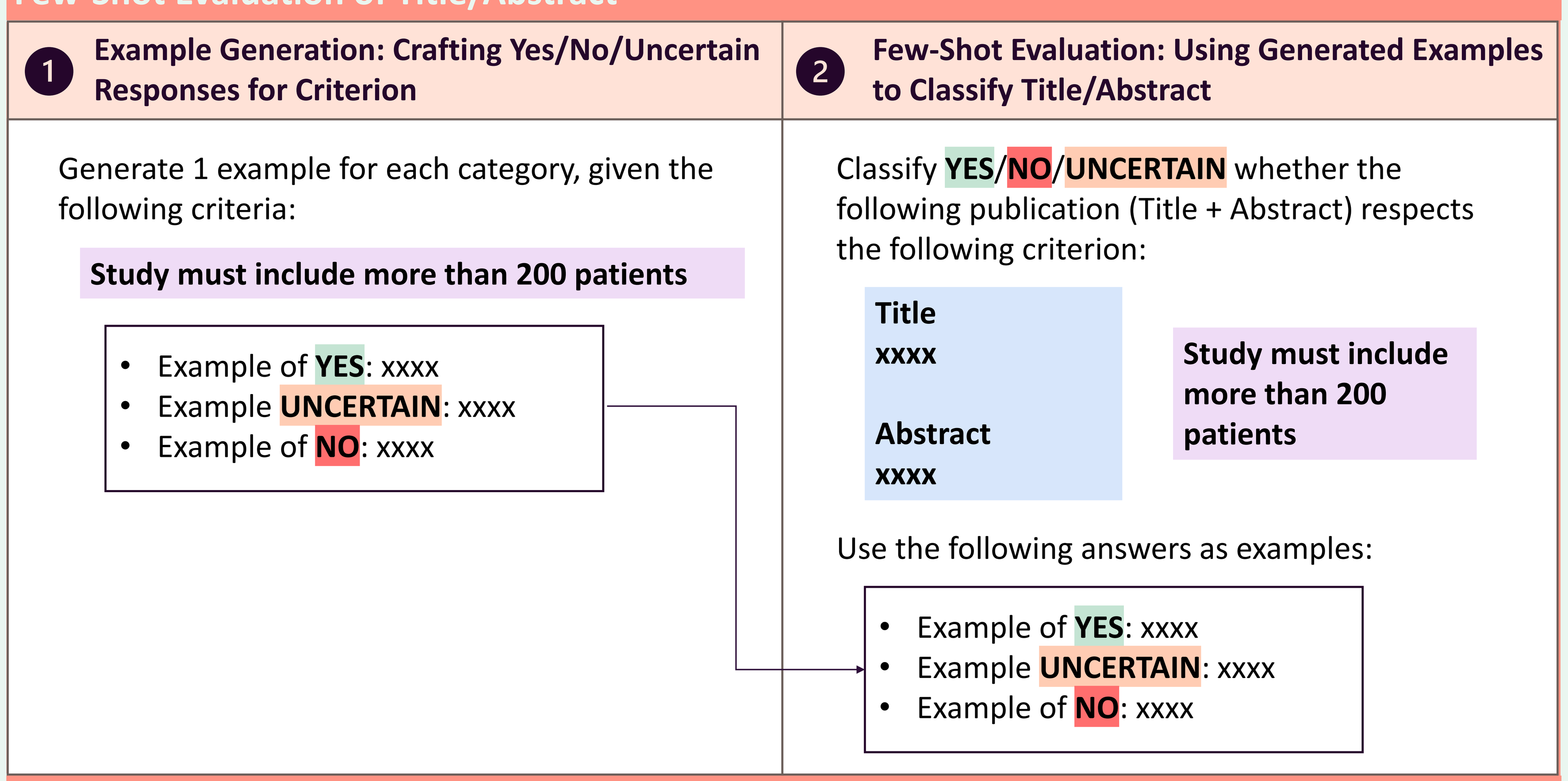
1. Uncertainty-aware prompting

Tackles hallucinations by forcing explicit triage: YES (meets criteria), NO (does not meet criteria), or **UNCERTAIN** (requires reviewer input). This ensures no false positives/negatives slip through unchecked.

2. Few-shot learning

Provide examples of YES/NO/UNCERTAIN answers to the LLM, to help it internalize criterion and prevent misaligned responses.

Figure 2: Two-Step Automated Classification of Criterion: From LLM-Generated Examples to Few-Shot Evaluation of Title/Abstract



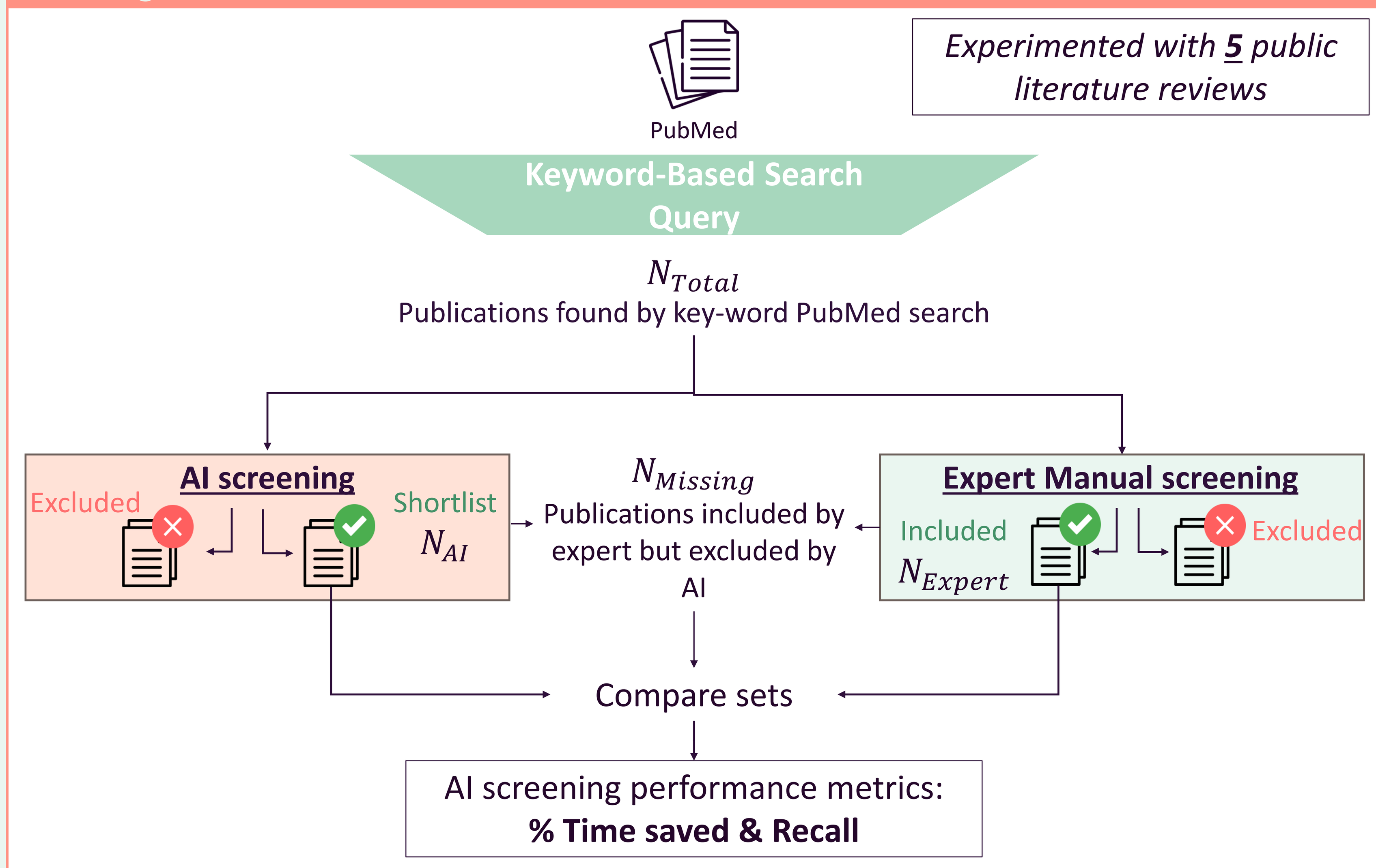
Literature review criterion. Must be a yes/no question

Publication studied (Title and abstract)

Experiments to evaluate the tool

- To evaluate the performance of the tool in pre-selecting the right articles for the reviewers, we identified several published systematic reviews and tried to reproduce them.
- Using the public PubMed queries, the original sets of publications were retrieved. Then the sets of publications manually screened by experts were compared to the shortlist automatically screened by our AI-screening tool.
- **Mistral-Large 2024** was the model used for all inferences.

Figure 3: Process to evaluate the performance of the AI-Powered tool in Title/Abstract screening



% of time saved:
% of articles excluded by AI shortlist

$$TS = 1 - \frac{N_{AI}}{N_{Total}}$$

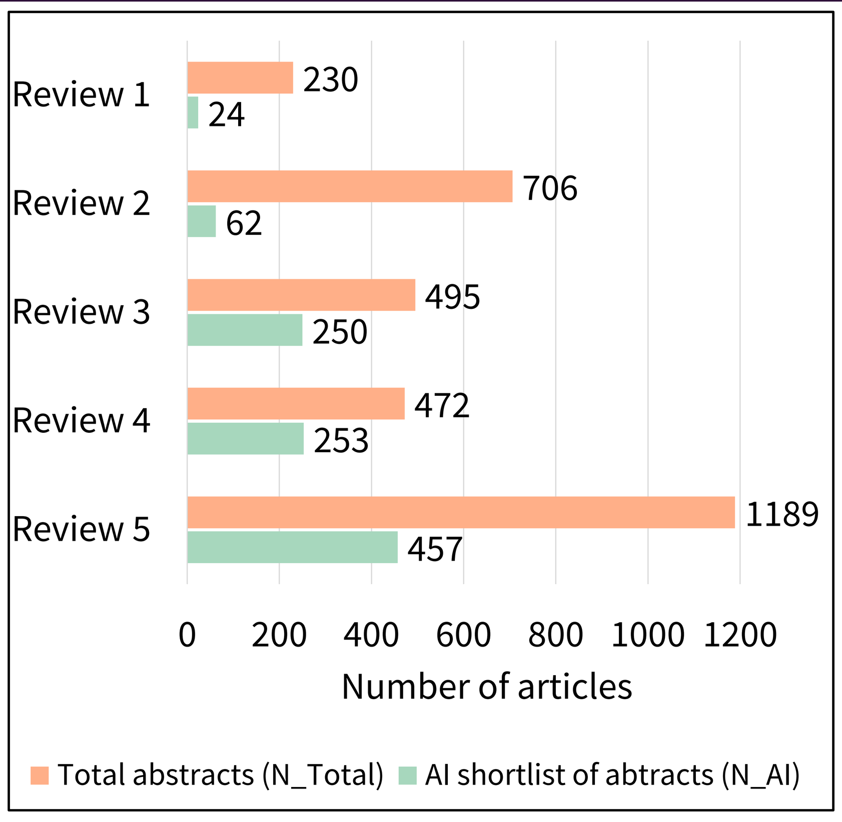
Recall:
AI shortlist's overlap with manual selections (%)

$$R = \frac{N_{Expert} - N_{Missing}}{N_{Expert}}$$

RESULTS & DISCUSSION

- As shown in Figure 4, across the five replicated literature reviews used to test the semi-automated framework, the AI shortlists reduced the number of abstracts requiring manual screening by **46% to 91%.**
- Nearly all articles identified by human experts were included in the AI shortlists, with only two missed articles (reviews 2 and 5), corresponding to a **recall above 96%.**
- By designing a conservative prompting strategy, encouraging the model to answer "UNCERTAIN" when appropriate, a high recall was maintained while further decreasing the number of articles requiring manual assessment.

Figure 4: Results of evaluation on 5 different literature reviews



Review	% of Time saved (TS)	Expert selection	Missing in shortlist $N_{Missing}$	Recall (R)
Review 1	**90%**	9	0	**100%**
Review 2	**91%**	25	1	**96%**
Review 3	**50%**	20	0	**100%**
Review 4	**46%**	13	0	**100%**
Review 5	**62%**	32	1	**97%**

CONCLUSION

The AI-assisted shortlisting method reduced abstract screening effort by 46–91% (2–10× faster), while missing less than 5% of relevant studies, demonstrating both efficiency and methodological rigor.

REFERENCES:

1. Jaspers et al., *Machine learning techniques for the automation of literature reviews and systematic reviews in EFSA,* 2018. DOI:10.2903/sp.efsa.2018.EN-1427
2. Khraisha et al., *Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages*, 2024. DOI: 10.1002/jrsm.1715
3. Edwards et al., *ADVISE: AI-accelerated Design of Evidence Synthesis for Global Development*, 2023. https://arxiv.org/abs/2305.01145
4. Choe et al., Short Review on srBERT: Automatic Article Classification Model for Systematic Review Using BERT, 2022. DOI: 10.1186/s13643-021-01763-w
5. Coelho et al., Enhancing Industrial Productivity through AI-Driven Systematic Literature Reviews, 2023. DOI: 10.5220/0012235000003584

ABBREVIATIONS:

AI, Artificial Intelligence; BM25, Best Matching 25; LLM, Large Language Model; NLP, Natural Language Processing; SLR, Systematic Literature Review

ACKNOWLEDGEMENT: We would like to thank Gaëtan Pinon, and Sonia Zebachi for their support in drafting and reviewing the submitted abstract and poster.
CONFLICT OF INTEREST: NA

CONTACT INFO: Paul Loustalot, p.loustalot@quinten-health.com