The Use of Natural Language Processing in Literature Reviews

Martin C, Hood D

INTRODUCTION

- Literature reviews have many applications in health economics and outcomes research. However, they are limited in breadth and depth by the amount of time reviewers spend and are prone to human error and biases. Natural language processing (NLP) aims to address these issues.
- Axtria reviewed the use of NLP in literature reviews, assessed its benefits and detriments, administered Axtria's own test case, and developed recommendations for future researchers.

METHODS

- To identify use cases and information on the use of NLP in literature reviews, Axtria searched medical literature databases like PubMed, Science Direct, and Google Scholar; conference abstract lists; and other gray literature. The identified relevant studies are summarized herein.
- NLP was further implemented to conduct screening. Experts in systematic literature review were then consulted regarding the application of NLP to established literature review processes.

TRADITIONAL LITERATURE REVIEW

- In life sciences, literature reviews serve as the foundation for many information-oriented processes. They guide researchers and pharmaceutical companies by providing insights into unmet patient needs and current markets.

The Importance of Literature Review in Scientific Research

Building upon existing knowledge through new improvements and hypotheses is fundamental to research.

Literature reviews acquaint researchers with the existing knowledge base.

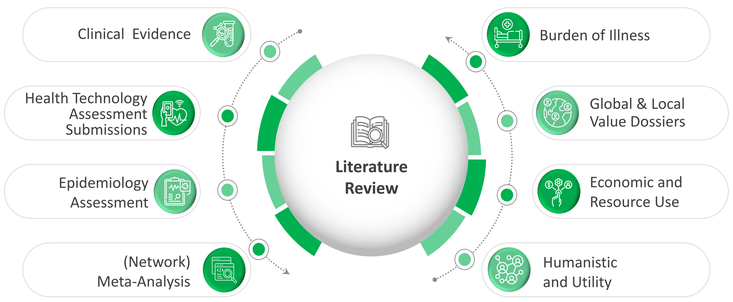
Literature reviews collate current publications and identify gaps in the knowledge.

Exploring identified gaps is essential to building and improving the knowledge base.

Literature reviews show the researchers' expertise in a given field.

- Literature reviews are essential for evaluators such as regulatory bodies and health technology assessment (HTA) bodies, preventing unintentional duplication of research and identifying gaps in existing knowledge.

Applications of Literature Review



- Targeted literature reviews (TLRs) vary in adherence to standardized processes, scope, and expertise. Systematic literature reviews (SLRs) are conducted using inclusion/exclusion criteria and databases like PubMed, Embase, Cochrane Library, etc. SLRs are considered more powerful due to their systematic and repeatable nature, while TLRs are valued because they are fast and unencumbered by strict guidelines.

CHALLENGES WITH TRADITIONAL LITERATURE REVIEWS

Screening requires significant human labor and introduces risks. Manual data extraction is time-consuming and prone to human error.

TLRs grapple with limitations in scope and depth, which are intrinsically tied to the time and effort invested by reviewers.

SLRs are often not updated frequently due to their labor-intensive nature and can be impractical for topics with rapidly evolving scopes.

SLRs are essential for compiling comprehensive evidence. Automation and NLP allow researchers to focus on critical appraisal and synthesis.

NATURAL LANGUAGE PROCESSING AND ITS APPLICATION IN LITERATURE REVIEW

- Fortunately, advances in artificial intelligence (AI) offer potential solutions to these challenges. NLP offers advantages such as prioritizing articles based on relevance and subtopics. However, its implementation requires technical expertise, and there's a risk of misclassifying studies due to rigid quality criteria. It's crucial to scrutinize discrepancies between AI and human decisions. Furthermore, third-party AI services may not always adhere to established guidelines like PRISMA.
- The key advantages and disadvantages of implementing NLP methods in literature reviews, either alone or in combination with human direction, are summarized below.

Comparison of Methods for Targeted and Systematic Literature Reviews

Targeted Literature Review Conductors

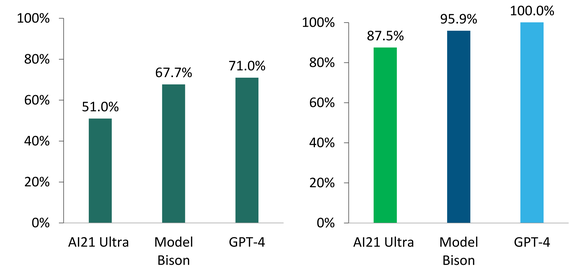
	Human	AI	Human/AI Combination
Pros	• Able to understand context and nuances • Capable of making intuitive judgments • Can handle ambiguous or poorly structured data	• Fast processing of large volumes of data • Consistent application of criteria • Less prone to fatigue	• Combines human intuition with AI efficiency • Can improve the selection process through iterative learning • Reduces bias through AI consistency while retaining human critical thinking
Cons	• Time-consuming • Subject to bias • Potentially less comprehensive due to selective focus	• May miss nuances or non-explicit connections • Requires high-quality training data • Can struggle with ambiguous information	• Requires well-defined protocols for effective collaboration • Can be costly to set up and maintain. • May have issues with the integration of human and AI decision-making

Systematic Literature Review Conductors

	Human	AI	Human/AI Combination
Pros	• Deep comprehension of study quality and relevance • Capable of sophisticated synthesis of findings • Can navigate complex methodologies within studies	• Can rapidly screen titles and abstracts • Can apply consistent I/E criteria • Manages large datasets and multiple databases with ease	• AI increases the speed of initial screening processes, while humans ensure the final selection's quality and relevance • Allows for a more robust and reproducible review process with high decision-match rates
Cons	• Very labor-intensive • May be influenced by individual or group biases • Time constraints can affect comprehensiveness and depth	• Limited ability to assess the quality of studies beyond predefined parameters • May require extensive customization to handle different research questions • Potential for missing relevant studies due to rigid criteria • Lack of exclusion decision reporting and non-adherence to PRISMA guidelines	• Can be challenging to balance the input from AI and humans • Iterative process of refining AI parameters can be time-consuming • Integration of qualitative data may still rely heavily on human analysis

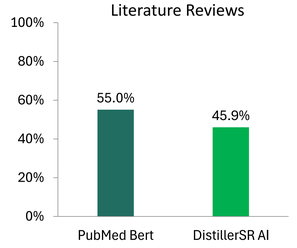
PREVIOUS APPLICATIONS AND AXTRIA'S USE CASE

Accuracy of NLP in Literature Reviews



- A systematic review analyzed the potential of NLPs to assist with the SLR process using 3 LLMs. GPT-4 achieved a decision match rate of **71.0%**, outperforming AI21 Ultra (**51.0%**) and Model Bison (**67.7%**).[1]
- Another study validated the AI classifier tool used by the DistillerSR platform, and found an **87.5%** match rate with human decisions.[2]
- A pilot study of the cancer-specific, NLP-based information retrieval software, RetriLite, achieved **95.9%** accuracy in identifying relevant articles.[3]
- In Axtria's use case, GPT-4 attained a **100%** accuracy rate in screening a sample of **80** articles for relevance by therapeutic area.

Time Savings of NLP in Literature Reviews



A recent study validating DistillerSR's AI classifier for systematic reviews estimated time savings at 55% of human working hours.[2] Another study using PubMed BERT demonstrated a 45.9% decrease in screening time per abstract.[4]

An SLR examined how AIs ran or assisted with SLRs. The authors evaluated the use of AI in the MEDLINE and Embase databases, which revealed **no concrete evidence** of AI adhering to **PRISMA** guidelines.[5]

However, in Axtria's application of GPT-4 (above), we maintained alignment with **PRISMA** guidelines, reporting reasons for exclusion while following predefined **PICOS** criteria.

CONCLUSIONS

- There are many promising developments in NLP applications for literature review synthesis, which can improve the depth and breadth of literature reviews while reducing human labor and the risk of bias and error.
- However, since the NLP models' accuracy and reliability vary, literature reviewers should implement NLP cautiously, giving precise instructions and sufficient training, verifying NLP decisions, and following practice guidelines where possible.
- Last, optimal NLP usage in literature review synthesis necessitates a certain degree of technical expertise in conjunction with a strong knowledge of scientific and medical writing. Therefore, Axtria recommends consultation with technical and clinical experts when implementing NLP in literature reviews.

REFERENCES

1. Hemant R, Malik A, Behera DC, Kamboj G. A comparative analysis of large language models utilized in systematic literature review. Podium presentation at ISPOR Annual European Meeting; 13 Nov 2023; Copenhagen, Denmark. Accessed December 16, 2023.
2. Bhagat A, Moon D, Khan H, Kochar P, Kanakagiri S, Kaur R, Bhalla Y, Kaur K, Singh R, Goyal R, Aggarwal A. MSR42 Adoption of artificial intelligence in systematic reviews. Value Health. 2022 Dec;25(12):S358. Doi: 10.1016/j.jval.2022.09.1773
3. Zeng J, Cruz-Pico CX, Saridogan T, Shufean MA, Kahle M, Yang D, Shaw K, Meric-Bernstam F. Natural language processing-assisted literature retrieval and analysis for combination therapy in cancer. JCO Clin Cancer Inform. 2022 Jan;6:e2100109. doi: 10.1200/CCI.21.00109
4. Perlman-Arrow S, Loo N, Bobrovitz N, Yan T, Arora RK. A real-world evaluation of the implementation of NLP technology in abstract screening of a systematic review. Res Synth Methods. 2023 Jul;14(4):608-621. doi: 10.1002/jrsm.1636
5. Queiros, L., Mearns, E. S., Ademisoye, E., McCarvil M, Alarcao J, Garcia MJ, Abogunrin S. MSR22 Is artificial intelligence replacing humans in systematic literature reviews? A systematic literature review. Value Health. 2022 Jul;25(7):S522. doi: 10.1016/j.jval.2022.04.1229

DISCLOSURES

- CM & DH are employees of Axtria Inc., NJ, USA.
- This study was funded by Axtria Inc.

POSTER PRESENTED AT

ISPOR US, Atlanta, Georgia, USA; 5 May – 8 May 2024.