Artificial intelligence bias in systematic literature reviews (SLRs) for health technology assessment (HTA)

Mangat G[1]; Sharma, S[1]; Bergemann, R[2]

[1]Parexel International, Mohali, India;
[2]Parexel International, Basel, Switzerland

Background

Systematic literature reviews (SLRs) are an important tool for evaluating the effectiveness of health technologies and informing policy decisions. However, bias can occur at various stages of the review process, including in the selection of studies for inclusion, the interpretation and synthesis of the evidence, and the reporting of the results. We aim to explore the issue of AI bias in SLRs for health technology assessment (HTA).

There can be multiple sources of bias in SLRs. These include the search strategy for identifying relevant studies, the selection of studies for inclusion in the review, the interpretation and synthesis of the identified evidence and reporting of the results. AI tools may exacerbate these biases if the algorithms used to identify and synthesize the evidence are trained on biased data or if they are not transparent in their decision-making processes. They also need to be programmed to consider all relevant factors for interpretation, synthesis and reporting.

Introduction

Systematic literature reviews (SLRs) have become an essential component in the evaluation of health technologies and the formulation of evidence-based policy decisions. These reviews offer a rigorous and comprehensive synthesis of the existing body of knowledge on a specific topic, helping to identify the most effective interventions and guide resource allocation in healthcare (Moher et al., 2009). However, as technology continues to advance, artificial intelligence (AI) tools have begun to play an increasingly prominent role in the SLR process, automating various stages of the review to improve efficiency and reduce human error (O'Mara-Eves et al., 2015). While AI has the potential to revolutionize the way SLRs are conducted, concerns have arisen regarding the potential introduction of bias through the use of these tools (Challen et al., 2019).

Bias in SLRs can undermine the validity and reliability of the review, leading to inaccurate conclusions and potentially influencing policy decisions that affect patient outcomes (Ioannidis, 2016). AI bias in SLRs for health technology assessment (HTA) can occur at multiple stages of the review process, including the selection of studies for inclusion, the interpretation and synthesis of the evidence, and the reporting of the results (Garg et al., 2018). The potential for AI tools to exacerbate bias has raised concerns about the overall integrity of the review process, as well as the need to ensure the transparency and accountability of AI algorithms (Gibson et al., 2020).

Given the growing reliance on AI tools in SLRs for HTA, it is critical to explore the issue of AI bias and its implications on the quality and trustworthiness of these reviews. This research aims to provide a comprehensive examination of the different stages in the SLR process where AI bias may occur, the potential consequences of such biases, and the approaches that can be implemented to mitigate these effects. By gaining a better understanding of AI bias in SLRs for HTA, researchers, policymakers, and healthcare providers can work together to develop more robust and reliable methodologies, ensuring that the evidence generated from these reviews is of the highest quality and can effectively inform healthcare decisions.

Conclusion

In conclusion, addressing AI bias in SLRs for HTA is of paramount importance to ensure the reliability and accuracy of these reviews, which are crucial for informing evidence-based policy decisions. AI bias can manifest in various stages of the SLR process, from the initial title and abstract screening to full-text review, data extraction, synthesis, and reporting of results, all of which can impact the validity and credibility of the findings. Furthermore, maintaining transparency and rigor in each stage of the SLR process is critical for reducing bias and enhancing the overall quality of SLRs.

To effectively counter AI bias, a multifaceted approach is necessary.

- By employing strategies such as more robust training of the AI tool with diverse data set, combining the expertise of human reviewers with the efficiency of AI tools, utilizing multiple AI tools with diverse algorithms and features, and applying established methodologies for validation, potential biases can be identified and addressed systematically.
- Engaging in collaborative efforts and interdisciplinary research can also contribute to the development of AI tools with improved performance, reduced biases, and increased transparency. These collaborations can promote the sharing of knowledge and expertise, allowing for the refinement and optimization of AI tools to better serve the needs of the scientific community.
- As AI tools continue to develop and become more integrated into the SLR process, ongoing research and evaluation are necessary to monitor their performance and establish best practices for minimizing bias. This proactive approach will ensure that SLRs for HTA contribute effectively to evidence-based policy decisions that ultimately enhance public health outcomes.

In summary, by adopting a comprehensive approach to addressing AI bias in SLRs for HTA and continually refining our methodologies, we can improve the reliability, accuracy, and impact of these reviews, paving the way for more informed, evidence-based policy decisions that promote better healthcare and improved quality of life for all.

Key hurdles/challenges with the use of AI tools for conducting SLRs

Traditionally, SLRs have employed a dual-reviewer process, to minimize the risk of bias and errors in the evidence generated from an SLR (Higgins & Green, 2011). To preserve the rigor and quality of the review process along with making it more efficient, it is important to understand the potential bias that can be introduced with the used of AI tools. Key areas where AI tools can introduce bias are discussed below:

Screening (title, abstract and full-text based)

- Selection of studies at title/abstract is a critical step in an SLR. Bias in this stage can lead to the exclusion of relevant studies, which can ultimately impact the validity and generalizability of the review's findings (Ioannidis, 2016).
- Evaluation of the study based on the full-text is even more challenging as it requires thorough understanding of the disease area and the project objective. Bias can be introduced at this stage by an AI tool not trained properly and can also impact the validity of the review findings (Ioannidis, 2016).

Data extraction

- This step requires collection of relevant information from the text and tables of the relevant publications, and it is very important that the appropriate information is picked from the appropriate tables and text. Any bias at this stage such as missing key evidence or identifying inappropriate evidence will also impact the results of quantitative analysis, if conducted. Similar to screening, bias at this stage can lead to erroneous data and impact the validity of the SLR (Ioannidis, 2016).

Report generation

- The compilation and summarization of results during the SLR process involve synthesizing and interpreting extracted data from the selected articles to address the review's research question. This step also includes drawing appropriate and relevant insights in line with the SLR's objectives. Errors at this stage can lead to a poor quality SLR even with appropriate study selection and extracted data. This can lead to inappropriate value messages and positioning of any product as well as rejection of the dossier by the HTA bodies.

Potential solutions to mitigate bias due to AI at different stages of an SLR

It is imperative to identify ways to make AI tools more robust as it will assist in generating more reliable evidence with the use of AI, enhance adoption of AI tools for SLR in the industry and increase acceptability of evidence generated with the help of AI tools by HTA bodies. Some key ideas are discussed below:

Training the AI tool

- It is important to train the AI tool on diverse set of data from different independent reviewers to reduce bias in their assessment and decision-making process. The tool should be trained with diverse data across all steps of screening, extraction and reporting to make it robust for the entire process of literature review.

Use of two AI tools

- Using two different AI tools, each designed with distinct algorithms, training data, and features, will allow for the comparison of the two AI tools' outputs, potentially highlighting discrepancies and reducing the risk of bias stemming from a single AI system (Naudet et al., 2017). Additionally, employing two AI tools with diverse underlying mechanisms can help identify a broader range of relevant articles and minimize the impact of biases that may be present in one tool (Marshall et al., 2018). This approach can also increase the efficiency of the screening process, particularly when dealing with a large volume of articles.

Combination of AI tool and human reviewer

- This strategy can benefit from the nuanced understanding and expertise of the human reviewer while leveraging the efficiency and consistency of the AI tool (Gates et al., 2019). The human reviewer can help mitigate potential bias introduced by the AI tool, while the AI tool can assist in managing the increasing volume of literature and reducing the workload for the human reviewer (O'Connor et al., 2020, Tsafnat et al., 2014).

REFERENCES

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. BMJ Quality & Safety, 28(3), 231-237.
Garg, A. X., Hackam, D., & Tonelli, M. (2018). Systematic review and meta-analysis: when one study is just not enough. Clinical Journal of the American Society of Nephrology, 3(1), 253-260.
Gates, A., Guitard, S., Pillay, J., Elliott, S. A., Dyson, M. P., Newton, A. S., & Hartling, L. (2019). Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Systematic Reviews, 8(1), 278.
Gibson, A., Ouzzani, M., & Ghassibe-Sabbagh, M. (2020). Artificial intelligence in evidence-based medicine: A systematic review. Medical Decision Making, 40(8), 944-957.
Higgins, J. P., & Green, S. (Eds.). (2011). Cochrane handbook for systematic reviews of interventions (Vol. 4). John Wiley & Sons.
Ioannidis, J. P. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. The Milbank Quarterly, 94(3), 485-514.
Marshall, I. J., Kuiper, J., & Wallace, B. C. (2018). RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. Journal of the American Medical Informatics Association, 25(1), 95-102.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med, 6(7), e1000097.
Naudet, F., Sakarovitch, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D., & Ioannidis, J. P. (2018). Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies
O'Connor, A. M., Tsafnat, G., Gilbert, S. B., Thayer, K. A., & Wolfe, M. S. (2020). Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of the International Collaboration for the Automation of Systematic Reviews (ICASR). Systematic Reviews, 9(1), 1-5.
O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Reviews, 4(1), 5. https://doi.org/10.1186/2046-4053-4-5
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F., & Coiera, E. (2014). Systematic review automation technologies. Systematic Reviews, 3, 74. https://doi.org/10.1186/2046-4053-3-74

Parexel International
2520 Meridian Pkwy
Durham, NC 27713, USA
+1 919 544-3170
www.parexel.com