Revolutionizing systematic reviews: the ASReview project's call for trust and transparency in AI-aided screening tools

23 March 2023

Figure 1. A pipeline of performing a systematic search resulting in a large text database, followed by either (A) a classical, time-consuming review process, or (B) fully transparent screening in interaction with advanced machine learning models. Image: © Licensed under Creative Commons License (CC BY-NC-ND 4.0)

Owing to the challenges inherent in the screening process, finding a low percentage of relevant studies at a high cost is common, making AI-aided screening tools a necessity

In today's era of information overload, the number of scientific papers and policy reports on any given topic is increasing at an unprecedented rate. This surge in textual data poses both opportunities and challenges for scholars and policy-makers. One of the most critical challenges is to write systematic reviews that provide comprehensive overviews of relevant topics.

However, a systematic search can easily result in thousands of studies that need screening to determine their relevance. The screening process is tedious and time-consuming, making it impractical to screen the entire literature on a given topic. Consequently, scholars often develop narrower searches, increasing the risk of missing relevant studies. Moreover, it is essential to ensure transparency and trustworthiness when screening literature for relevant studies, unlike a simple Google search or asking chatbots like ChatGPT a question.

This becomes even more challenging in situations that require an urgent review, such as during the COVID-19 crisis. Nevertheless, systematic reviews are critical for scholars, clinicians, policy-makers, journalists, and ultimately, the general public. Due to the challenges inherent in the screening process, finding a low percentage of relevant studies at a high cost is common, and assistance in screening literature is necessary.

The rise of Artificial Intelligence (AI)

The challenge is to build a machine that is fast (to save time) and accurate (all relevant records should be found) while learning from the human annotator. Combining these features is very useful because it interactively uses the strengths of humans (intelligent but slow) and machines (fast and accurate but not very intelligent). The rapidly evolving field of AI has allowed the development of AI-aided pipelines that assist in finding relevant texts for such search tasks. A well-established approach to increase the efficiency of title and abstract screening is determining prioritization with active learning:(2) a constant interaction between humans and machines.

It works as follows. First, the algorithm extracts information from a set of records, such as the abstracts of scientific papers, policy reports, newspaper articles, etc. Next, a machine learning algorithm is first trained on a minimal set of two records that the screener labels as relevant and irrelevant. The algorithm then estimates which other forms are most likely appropriate, ranks all unseen papers from high to low, and presents the most likely relevant record to the expert for screening. The researcher reviews the presented records and provides feedback on whether they are relevant. This feedback is then used to train the program better to estimate relevance in future iterations of the cycle. The rank order is continuously being updated, and the cycle is repeated until the annotator has seen all relevant records. The goal is to save time by screening fewer records than exist in the entire pool.

Trust and Transparency in AI-aided screening

Active learning is an effective tool for screening large amounts of textual data and can save up to 95% of screening time.(2) It has been successfully implemented in screening software, as evidenced by a curated overview. However, most existing tools are not future-proof and use closed-source applications with black-box algorithms. This lack of transparency is problematic because transparency and data ownership is essential in the era of Open Science, especially when AI systems produce output. That is, some papers may be pushed low in the ranking and are not shown to the human annotator, leading to a need for explainable AI.

During the active learning process, many iterations of the model will be trained, producing an enormous amount of data (think of many gigabytes or even terabytes of data), and machine learning models are continually becoming larger, using even more model parameters. Together, this can add up to an undesirable amount of data when naively storing all the data produced at every iteration of the active learning pipeline. Therefore, there is a tension between the desire for reproducibility and transparency on the one hand and the practical difficulties of storing and interpreting machine learning models on the other hand.(3)

To ensure the trustworthy implementation of AI in the screening process, software tools need to be fully transparent. The exact algorithms they use to produce the ranking scores should be known, or even better, the source code needs to be available so experts can

check (and adjust) the implemented algorithms. But also, the decisions made by the AI throughout the process should be transparent for the identification and correction of mistakes made by both humans and machines.

The ASReview Project

To set an example for trust and transparency in AI-aided screening, Utrecht University coordinated the open-source, community-driven ASReview Project. This project has resulted in an ecosystem that includes a web-based tool for AI-aided screening, infrastructure to add and compare different machine learning models, options to run large-scale simulation studies that mimic the screening process using AI, and a lively community discussing state-of-the-art solutions.

The ASReview Project's goal is to create a movement that demonstrates how open-source AI tools with transparent algorithms and visible decisions made by AI throughout the process can ensure trustworthy and accurate systematic reviews. Furthermore, the ASReview Project aims to challenge the industry and commercial companies to take a major leap forward in developing AI-aided screening tools that prioritize trust and transparency. By setting a high bar, Utrecht University hopes to encourage the development of reliable and trustworthy AI-aided screening tools across various fields.

AI-aided screening tools become more prevalent

In conclusion, AI-aided pipelines using active learning have the potential to save valuable time while ensuring comprehensive and accurate results. As the use of AI-aided screening tools becomes more prevalent in various fields, including healthcare and policy-making, these tools must be trustworthy and reliable. Transparency and reproducibility are crucial to maintaining the integrity of the research process. Therefore, developing open-source AI screening tools with transparent algorithms and visible decisions made by AI throughout the process is vital to ensure trustworthy and accurate systematic reviews.

References
1. Settles, B., Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2012. 6: p. 1-114.
2. Van de Schoot, R., et al., An open source machine learning framework for efficient and transparent systematic reviews. Nature Machine Intelligence, 2021. 3: pp. 125–133
3. Lombaers, P., de Bruin, J., & van de Schoot, R. (2023, January 19). Reproducibility and Data storage Checklist for Active Learning-Aided Systematic Reviews. [https://doi.org/10.31234/os]

Please Note: This is a Commercial Profile