

Artificial intelligence for targeted literature review screening within the Rayyan platform

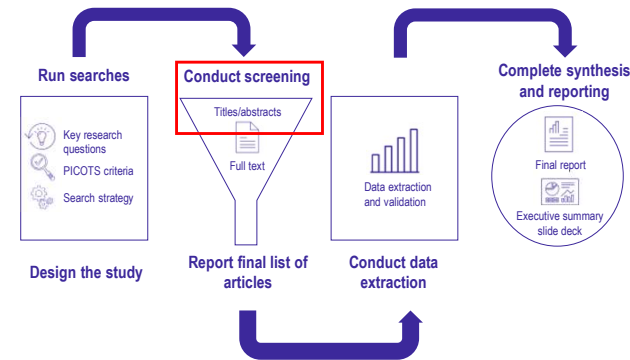
Gill M¹, Ng J¹, Szydowski N¹, Fusco N¹, Ruiz K¹

¹Cencora, Conshohocken, PA, USA

Background

- Targeted literature reviews play an important role in assessing treatment and disease landscapes and guiding strategy in early-stage drug development.
- Due to the vast amount of available scientific evidence, targeted literature reviews require significant time and effort.
- A targeted literature review requires researchers to examine hundreds, sometimes thousands, of potentially relevant publications, beginning with reading the title and abstract of each publication (title/abstract [TIAB] screening) to determine relevance for the review. Then eligible references are reviewed at the full-text level. The targeted literature review process is shown in **Figure 1**.
- Artificial intelligence (AI) is a promising technology that could be used to reduce time and workload burden by increasing the efficiency of targeted literature reviews.
- One possible application of AI is to identify relevant studies during TIAB screening at a speed considerably faster than humans.¹

Figure 1. Targeted literature review process



Key: PICOTS – population, interventions, comparators, outcomes, time, study design.

Objective

- The objective of this research is to evaluate the quantitative efficiencies and performance of the Rayyan AI tool (ie, Rayyan) for TIAB screening for targeted literature reviews.

References

1. Feng Y, Liang S, Zhang Y, et al. Automated medical literature screening using artificial intelligence: a systematic review and meta-analysis. *J Am Med Inform Assoc.* 2022;29(8):1425-1432. doi:10.1093/jamia/ocac066

Methods

- A large targeted literature review (8,755 references) previously screened by human reviewers was identified.
- Rayyan was trained using 3 subsets of the total references (5%, 10%, and 20%).
- Based on the training set, Rayyan predicted the relevance of the remaining references using a 5-level rating system ranging from “most likely to exclude” to “most likely to include.”
- Rayyan’s relevancy ratings were compared to the original targeted literature review inclusion/exclusion decisions to calculate sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) (**Table 1**).
- Screening time was compared for an AI-assisted process vs human reviewers (**Table 1**).

Table 1. Equations for calculated measurements to characterize the Rayyan AI tool

Measurement	Equation ^a
Sensitivity ^b (%)	$\frac{\text{\# of references included by both Rayyan and human reviewers}}{\text{Total \# of references included by human reviewers (excluding training set)}} \times 100$
Specificity ^b (%)	$\frac{\text{\# of references excluded by both Rayyan and human reviewers}}{\text{Total \# of references excluded by human reviewers (excluding training set)}} \times 100$
PPV ^c (%)	$\frac{\text{\# of references included by both Rayyan and human reviewers}}{\text{Total \# of references included by Rayyan}} \times 100$
NPV ^c (%)	$\frac{\text{\# of references excluded by both Rayyan and human reviewers}}{\text{Total \# of references excluded by Rayyan}} \times 100$
Accuracy ^b (%)	$\frac{\text{\# of references included by both Rayyan and human reviewers} + \text{\# of references excluded by both Rayyan and human reviewers}}{\text{Total \# of references (excluding training set)}} \times 100$
Time-savings for AI-assisted screening ^{d,e} (percentage difference, %)	$\frac{\text{Total \# of references} \times \text{hours} - (\text{\# of references in training set} + \text{total \# of references included by Rayyan} \times \text{hours})}{\text{Total \# of references} \times \text{hours}} \times 100$

Key: AI – artificial intelligence; PPV – positive predictive value; NPV – negative predictive value; TIAB – title/abstract.
^a “Human reviewers” refers to the original decisions made in the targeted literature review when TIAB was completed by humans.
^b The Rayyan inclusion category includes “no recommendation,” “likely to include,” and “most likely to include.”
^c Assumes that an experienced human reviewer screens an average of 50 title/abstract references per hour (ie, 50/hour).

Results

- The 5% training set had 438 references, the 10% training set had 876 references, and the 20% training set had 1,751 references.
- When references with Rayyan ratings of “most likely to include,” “likely to include,” and “no recommendation” were included, sensitivity was consistently high, ranging from 93% to 97% across all training sets.
- Specificity increased with training set size at 34%, 52%, and 61% for the 5%, 10%, and 20% training sets, respectively.
- Accuracy ranged from 38% to 63%, PPV ranged from 9% to 13%, and NPV was 99% for all training sets.
- Time-savings increased with greater training set size. The largest time-savings were reported for the 20% training set, where the AI-assisted process resulted in a 46% decrease in hours spent on TIAB screening.

Table 2. Performance of Rayyan for AI-assisted TIAB screening

Performance metric	5% training set	10% training set	20% training set
Number of references included by Rayyan ^a	5,633	4,021	2,946
Sensitivity	97%	96%	93%
Specificity	34%	52%	61%
PPV	9%	11%	13%
NPV	99%	99%	99%
Accuracy	38%	54%	63%

Key: AI – artificial intelligence; PPV – positive predictive value; NPV – negative predictive value; TIAB – title/abstract.
^a Rayyan inclusion relevancy ratings that are included in the calculations: “no recommendation,” “likely to include,” and “most likely to include.”

Table 3. Time-savings with AI-assisted screening

Quantitative efficiencies	5% training set	10% training set	20% training set
Screening hours for 1 human reviewer	175.1	175.1	175.1
Screening hours for 1 AI-assisted reviewer ^b	147.7	115.4	93.9
% time saved	15.6%	34.1%	46.4%

Key: AI – artificial intelligence.
^a AI-assisted reviewer used in this study is the Rayyan AI tool.
^b Screening time for AI-assisted reviewer includes time needed for a human reviewer to screen the training set.

Limitations

- The results from this analysis were generated by testing Rayyan on 1 large targeted literature review. Therefore, the results cannot necessarily be applied to other AI tools or other types of literature reviews.
- Although AI-assisted screening results in time-savings compared to human reviewers, it is recommended that some time is dedicated to quality checks of references excluded by AI to ensure accuracy.
- Only TIAB screening is possible with Rayyan. Variation in file formats and difficulty interpreting tables and figures are considerable obstacles for successful AI-assisted full-text screening.

Conclusions

- AI-assisted TIAB screening using Rayyan was highly sensitive (93%-97%) and resulted in considerable time-savings (up to 46%).
- High sensitivity (ie, ability to include relevant references accurately) is extremely important in order to produce high-quality targeted literature reviews.
- AI-assisted TIAB screening is a promising method for increasing efficiency for targeted literature reviews. However, experienced researchers are still needed to guide and validate AI processes to maintain methodological rigor and accuracy.
- Future research should confirm the performance and time-saving benefits of AI-assisted screening across targeted literature reviews that vary in size (number of references) and topics of interest.