Can artificial intelligence be used to improve the efficiency of title and abstract citation screening?

Eichinger C,[a] Evuarherhe O,[a] Law L,[a] Radwan M,[a] Liew A,[a] Sellick C,[a] Cadwell T,[b] Hugo M[b] and Wager K[a]

[a]Oxford PharmaGenesis, Oxford, UK; [b]Vyasa Analytics, Boston, MA, USA

Introduction

- Literature searches performed during systematic literature reviews (SLRs) often return many studies for reviewers to screen.
- We developed an artificial intelligence (AI)-assisted, human-in-the-loop screening tool leveraging named-entity recognition, a method based on natural language processing.
- The tool leverages natural language models that identify and highlight biomedical concepts and pre-specified keywords to filter text and to guide reviewers towards relevant text in the title and abstract.
- We hypothesized that highlighting keywords and concepts may reduce the time taken to decide which references to include or to exclude, while leaving decisions in the hands of the reviewer to maintain rigour.
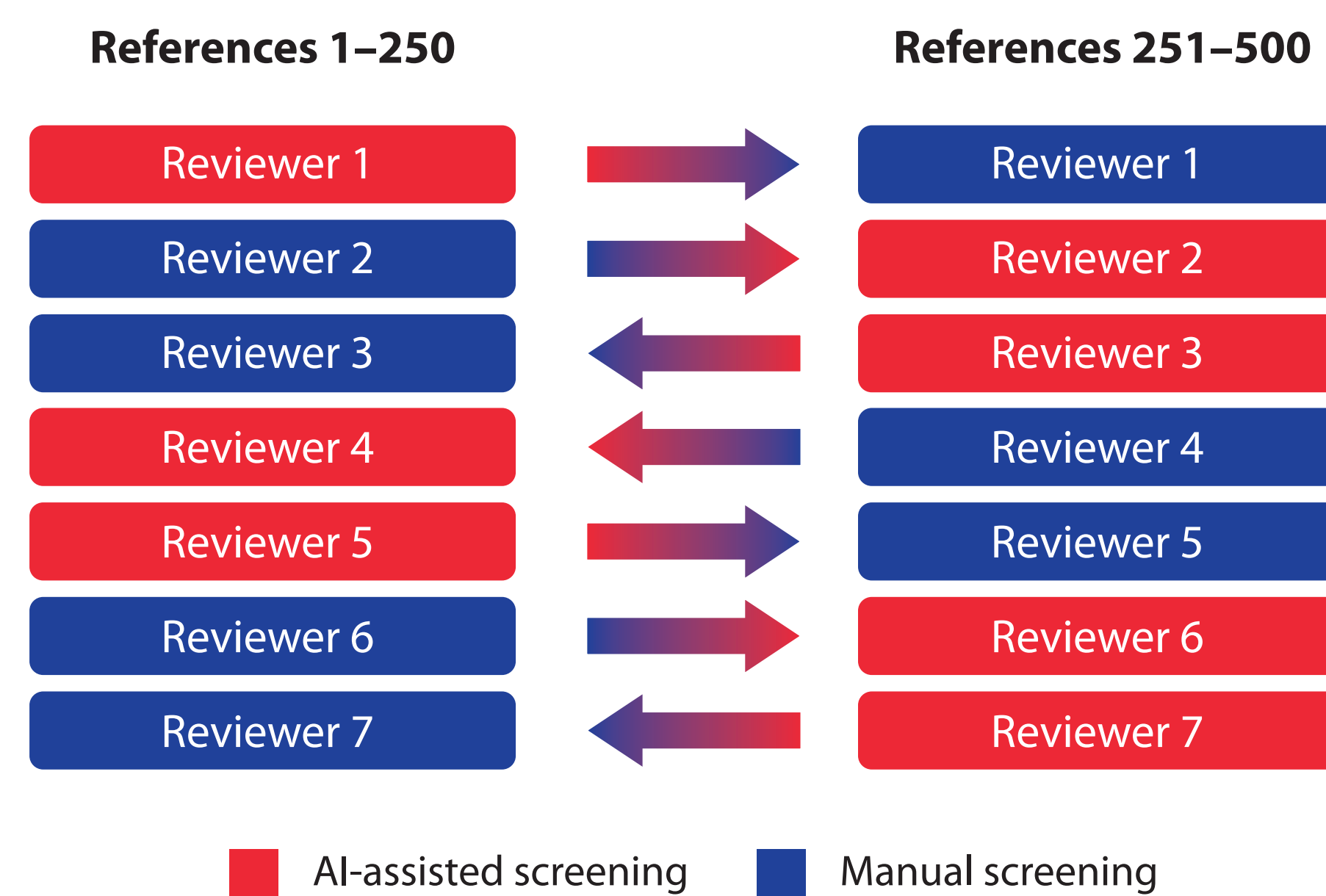
Objective

- We aimed to evaluate the time spent and accuracy of AI-assisted screening using the tool compared with unassisted screening in Microsoft Excel.

Research design and methods

- Named-entity recognition models were trained and tested on a collection of labelled data sets across 24 life science-related entity classes.
- We used 500 abstracts from a previously completed SLR, which were screened by seven reviewers; 250 abstracts were screened manually in Microsoft Excel and 250 were screened with assistance from the AI tool (**Figure 1**).
- To mitigate interpersonal variability and an increase in screening speed over time, two factors were varied across reviewers: the order in which references were screened, and whether the reviewer used the manual or the AI-assisted approach first (**Figure 2**). Full control of decision-making remained with the reviewers.
- Screening was timed, and accuracy of decisions for inclusion or exclusion was scored against a double-blind screened data set from the previous SLR.
- A paired *t*-test and a Mann–Whitney *U* test were performed to compare the differences in time and accuracy between methods.
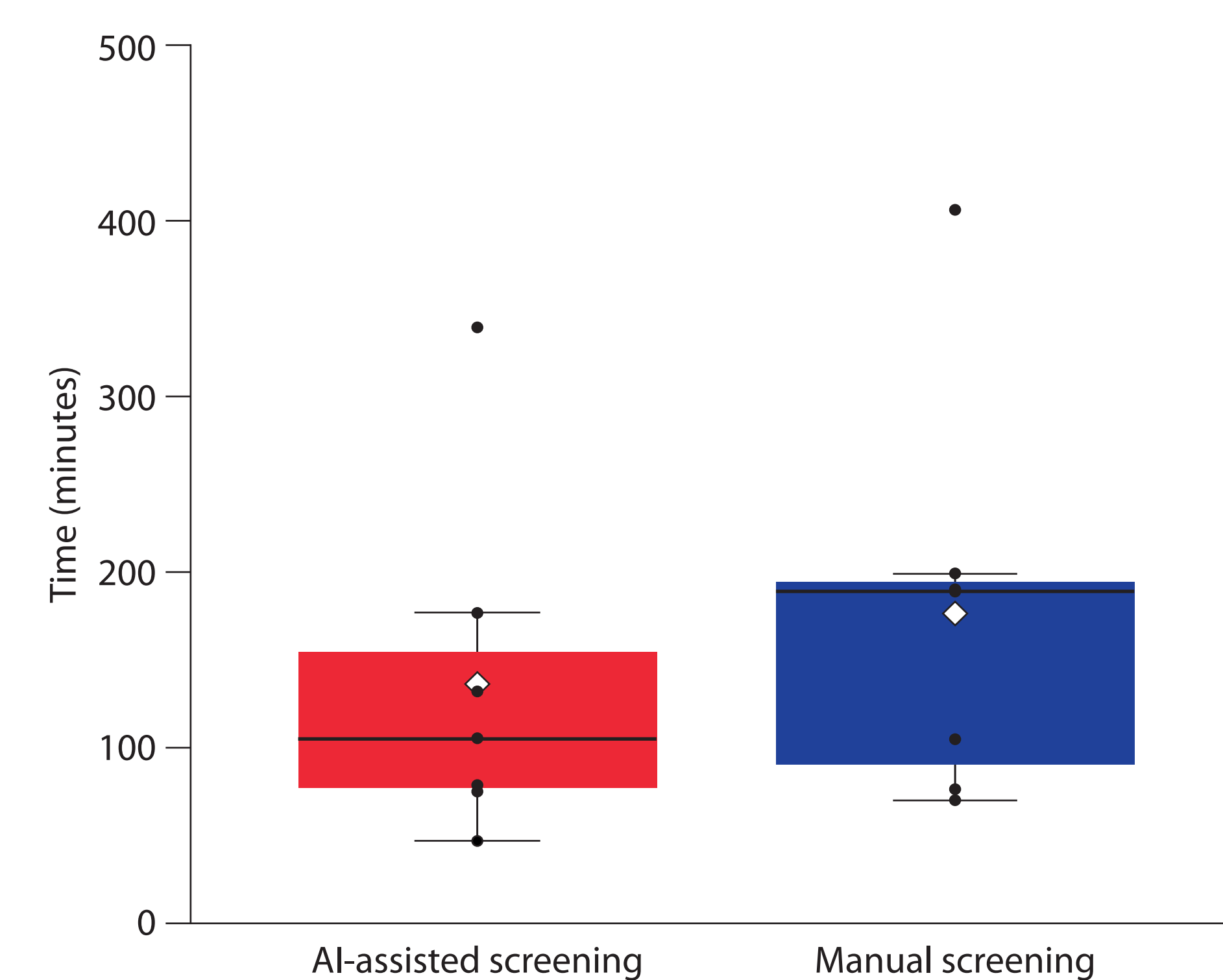
Figure 2. Allocation and screening order of references.

Figure 3. Timing of AI-assisted and manual screening.

The white diamond represents the mean; the black horizontal line in each box represents the median; the black whiskers represent the furthest data point within 1.5× interquartile range.
AI, artificial intelligence.

Results

- The average (interquartile range) time taken to screen 250 abstracts manually was 189.0 (90.5–194.5) minutes compared with 105.0 (77.0–154.5) minutes using the tool; $p = 0.02$ (**Figures 3** and **4**).
- Accuracy was 88% for both approaches; $p = 0.75$ (**Figure 5**).
- Reviewers provided positive feedback regarding the text-highlighting function of the tool, stating that it made screening easier than when using Microsoft Excel.

Figure 4. Time spent on screening by each reviewer and for each method.

AI, artificial intelligence.
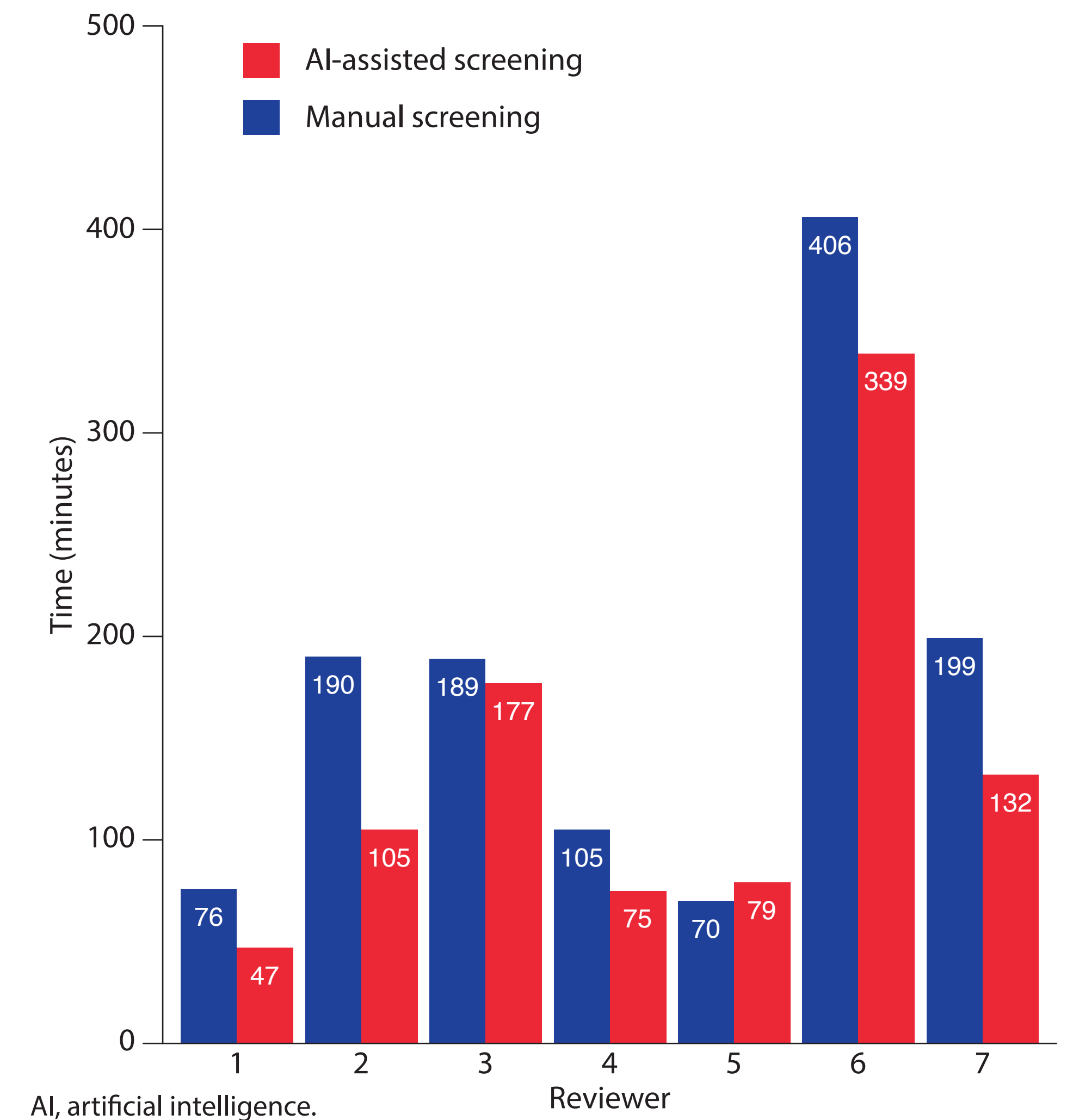
Figure 5. Accuracy of the decisions per reviewer and method.

AI, artificial intelligence.

Figure 1. Study design.
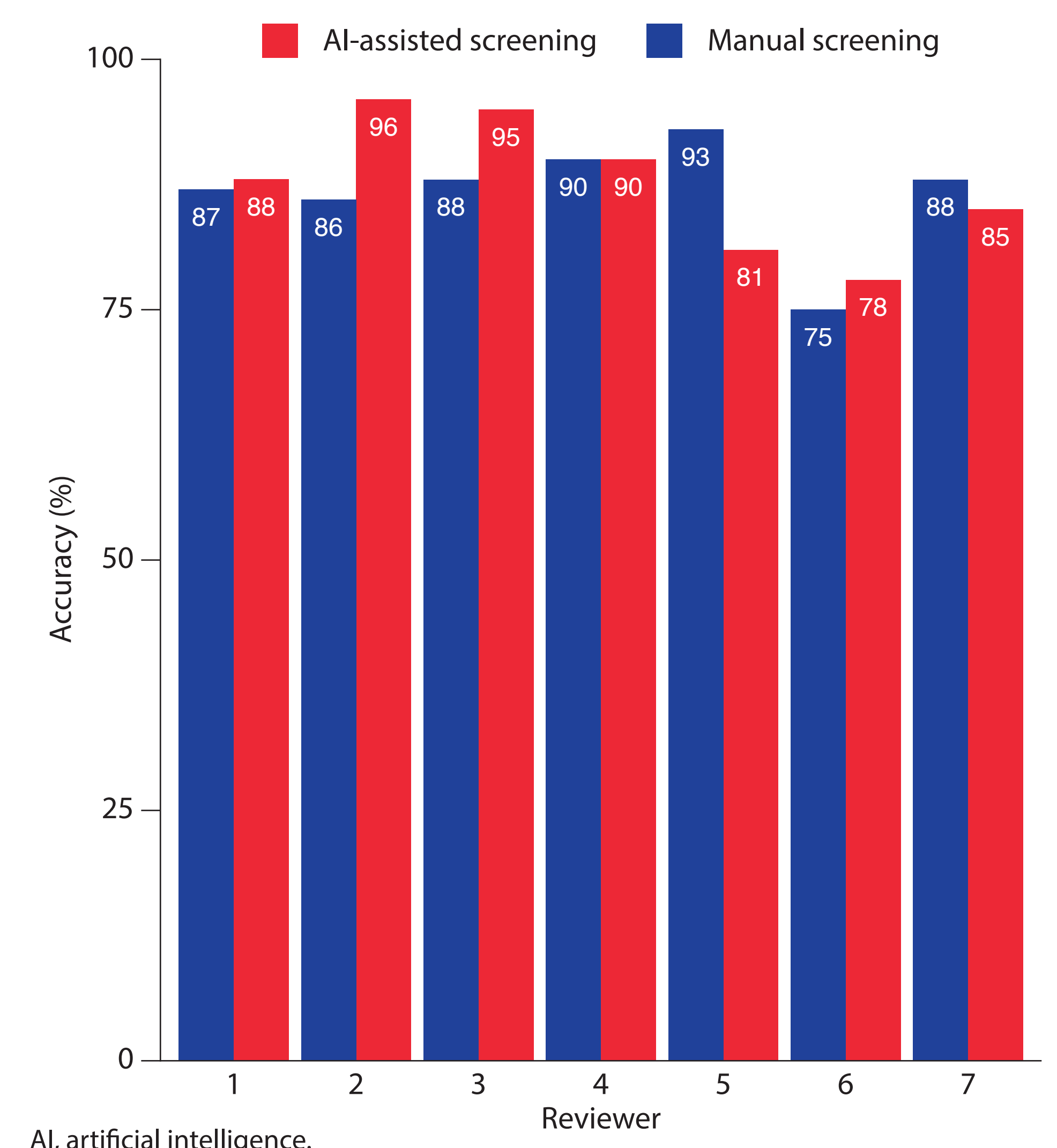
AI, artificial intelligence.

Conclusions

- Citation screening was 44% faster when assisted by the AI tool than when performed manually with Microsoft Excel; accuracy was similar for both approaches, and decisions were fully controlled by the reviewers in both cases.
- These findings support further development of the tool to reduce manual screening burden when developing SLRs.