

Application of Artificial Intelligence as a Decision Support Tool for Abstract Screening: Implications for Time and Cost Savings

Allie Cichewicz, Ananth Kadambi, Louis Lavoie, Lalith Mittal, Vicki Pierre, Renuka Raorane

Evidera, a part of Thermo Fisher Scientific

Introduction

- With the rise of artificial intelligence (AI) to support the labor-intensive screening phase of literature reviews, much focus has been given to the accuracy of AI compared to human reviewers. However, the downstream impact on cost and resources associated with employing AI to replace human reviewers for title and abstract screening with AI remains largely undefined.
- This study is ancillary to previous assessments reported in Cichewicz et al. 2022¹ and quantifies cost savings and improved efficiencies with the use of AI to support screening.

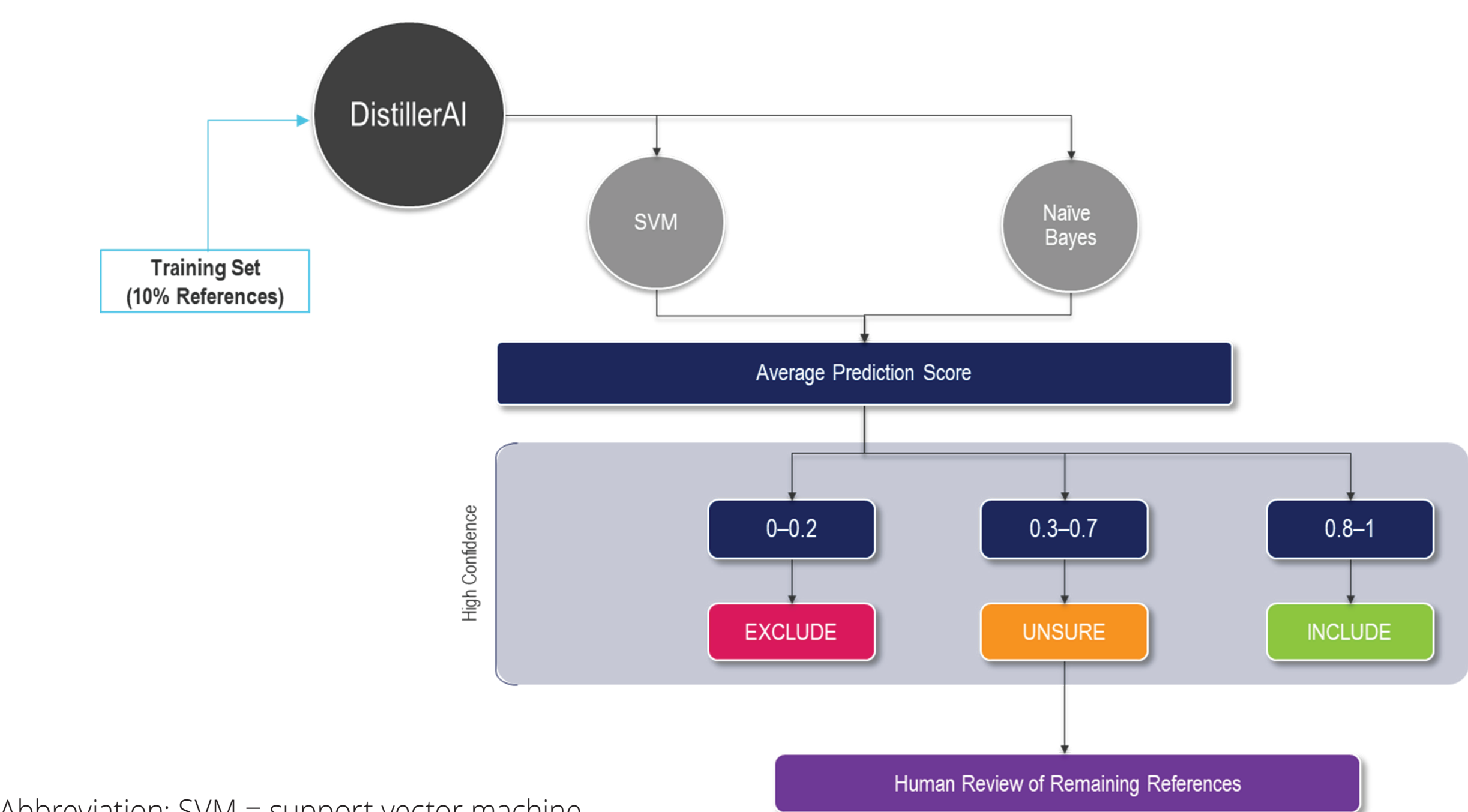
Objective

- To determine the potential time and cost savings when implementing AI tools in the conduct of title and abstract screening for literature reviews.

Methods

- Five previously completed systematic literature reviews (SLRs) on epidemiology, treatment patterns, health utilities, treatment guidelines, and SLRs and/or meta-analyses were replicated using DistillerSR AI as a second reviewer. Ten percent of the total volume of references in each SLR (representing 525–1,883 references) were hand-selected to train the AI reviewer.
- Once trained on a set of references, DistillerSR's AI generates a prediction score for probability of inclusion which ranges from 0–1, with 0 depicting a definite exclude and 1 depicting a definite include. For this study, prediction thresholds were set at ≤0.2 for excludes and ≥0.8 for includes (**Figure 1**).
- The proportion of references screened by the AI reviewer within the preset threshold was used to estimate the number of hours and associated cost for title and abstract screening in the following two scenarios:
 - The AI reviewer is used as a second screener (with reviewer #1 being a human reviewer) for systematic reviews that require two independent reviewers, with disagreements resolved by a second, human reviewer.
 - The AI reviewer is used as the sole screener for narrative or targeted reviews that may only require a single reviewer, with quality assurance measures taken wherein humans validate a proportion of references screened by the AI.

Figure 1. Flow Diagram of the AI Reviewer Training and Decision Process



Abbreviation: SVM = support vector machine

Methods (Cont'd)

- The estimated time differential between literature reviews conducted entirely by humans was then compared with reviews employing AI as an independent or second screener. The time differential was then used to approximate potential cost savings for the title/abstract review step of a literature review assuming a mix of junior and mid-level staff carry out the screening.

Results

- After training the AI, 472 references for SLRs/NMAs, 643 for epidemiology, 1,087 for utilities, 1,604 for treatment guidelines, and 1695 for treatment patterns were to be title/abstract screened (**Table 1**). Without the use of AI, this screening task was estimated to require approximately 17–62 hours across SLR topics for a review conducted with dual screening (two human reviewers) and 8.5–31 hours for a review conducted with single screening (one human reviewer). *Note: these estimates do not include additional time for conflict resolution or quality control of screening decisions.*
- As suggested in Cichewicz et al. 2022¹, title and abstract screening decisions (include or exclude) would be made by the AI reviewer for approximately 75% (range: 73%–77%) of references for most SLR topics, and up to 91% of these references for the treatment guidelines SLR when using a pre-defined prediction threshold of ≤0.2 for excludes and ≥0.8 for includes.

Table 1: Estimate of the Screening Burden for AI and Human Reviewers

SLR Topic	Total number of records to be screened	Records screened by AI	Remaining records to be screened by human reviewers
Epidemiology	643	471 (73.3%)	172 (26.7%)
SLRs/NMAs	472	343 (72.7%)	129 (27.3%)
Treatment Guidelines	1,604	1,452 (90.5%)	152 (9.5%)
Treatment Patterns	1,695	1,277 (75.3%)	418 (24.7%)
Utilities	1,087	838 (77.1%)	249 (22.9%)

Abbreviations: AI = artificial intelligence; NMA = network meta-analysis; SLR = systematic literature review

Dual-screened Reviews

- Based on the anticipated volume of references to be screened by the AI, employing it as one of the two reviewers could save an average of 16 (range: 6-27) hours of a human reviewer's time across SLR topics while maintaining high confidence in the accuracy of the AI decisions (**Figure 2**).

On average, the use of AI has the potential to reduce the title and abstract screening burden for human reviewers by more than one-third. This also results in an approximate cost savings of 40% (range: 36%–45%) of the budget required for title and abstract screening (**Figure 3**).

Single-screened Reviews

- For reviews where it may be appropriate to employ AI on its own as a single screener, it is expected to offset the same number of screening hours based on volume of literature. However, this represents a much larger proportion of the total time needed to complete the screening. In this setting, the use of AI can reduce the title and abstract screening burden for human reviewers by more than three-quarters (Figure 4), which also results in an approximate cost savings of 80% (range: 73%–91%) of the total screening costs (**Figure 5**).

Results (Cont'd)

Figure 2. Time Savings when Employing AI as a Second Reviewer

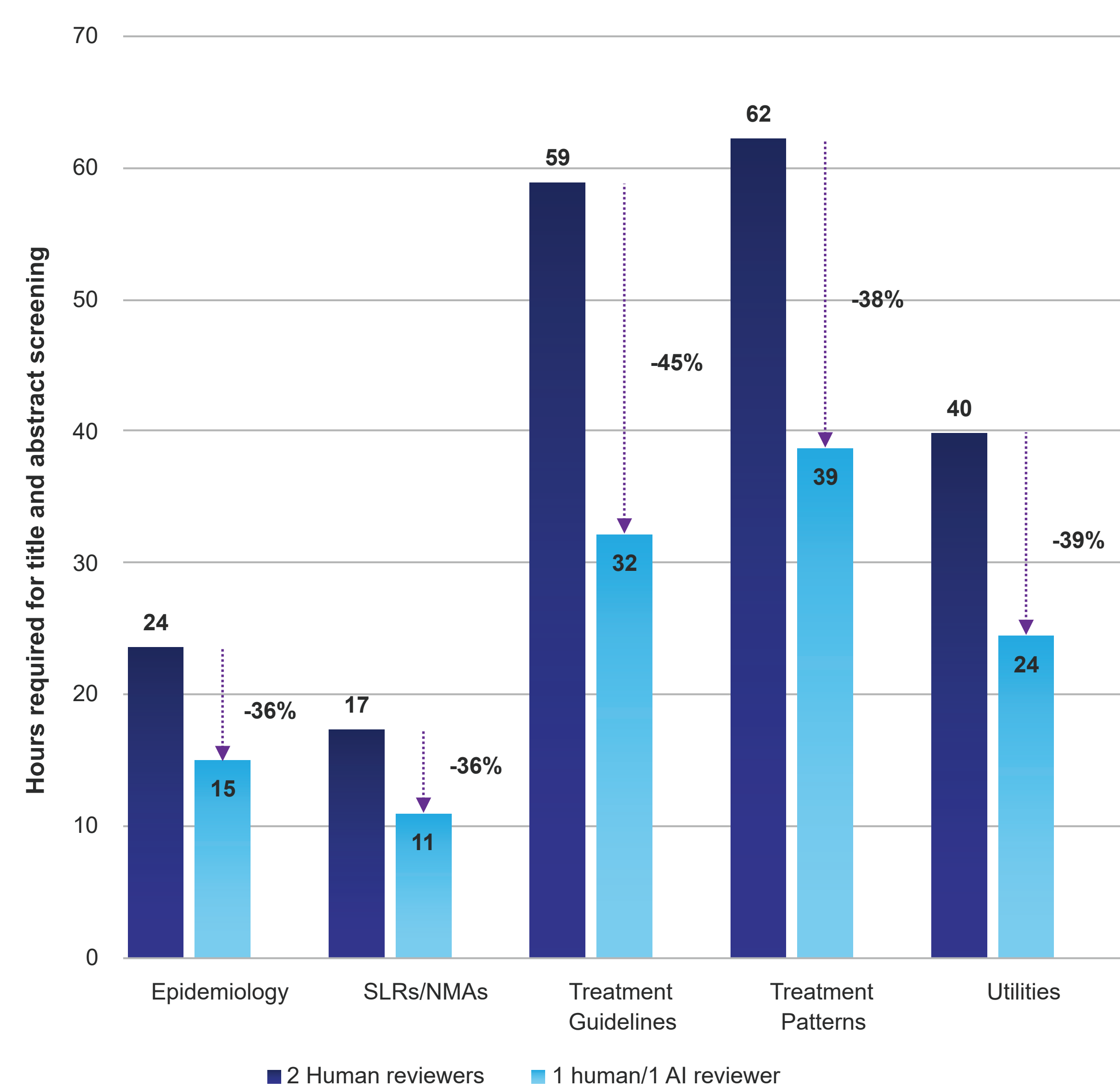
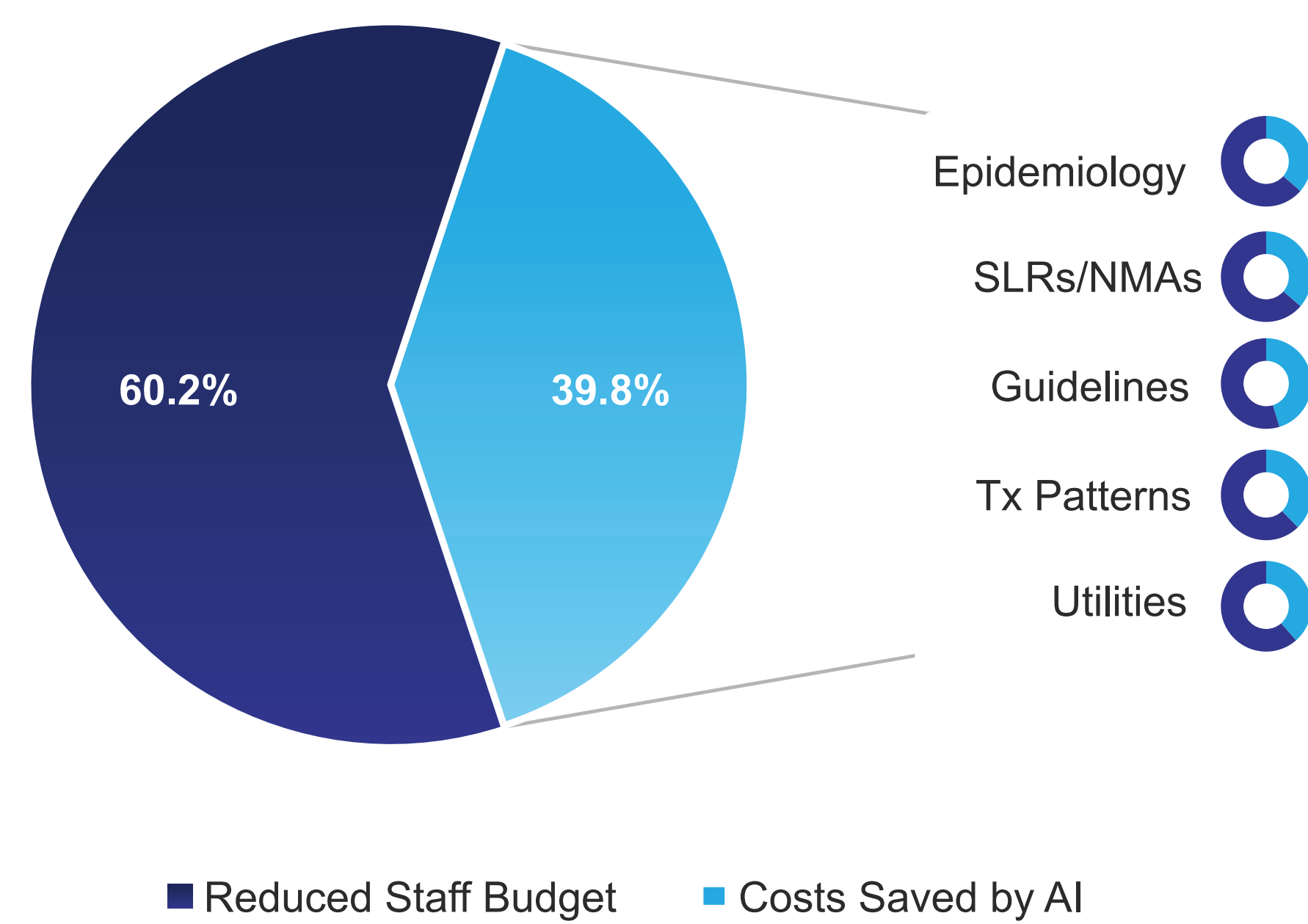


Figure 3. Cost Savings when Employing AI as a Second Reviewer



Abbreviations: AI = artificial intelligence; NMA = network meta-analysis; SLR = systematic literature review

Figure 4. Time Savings when Employing AI as a Single Reviewer

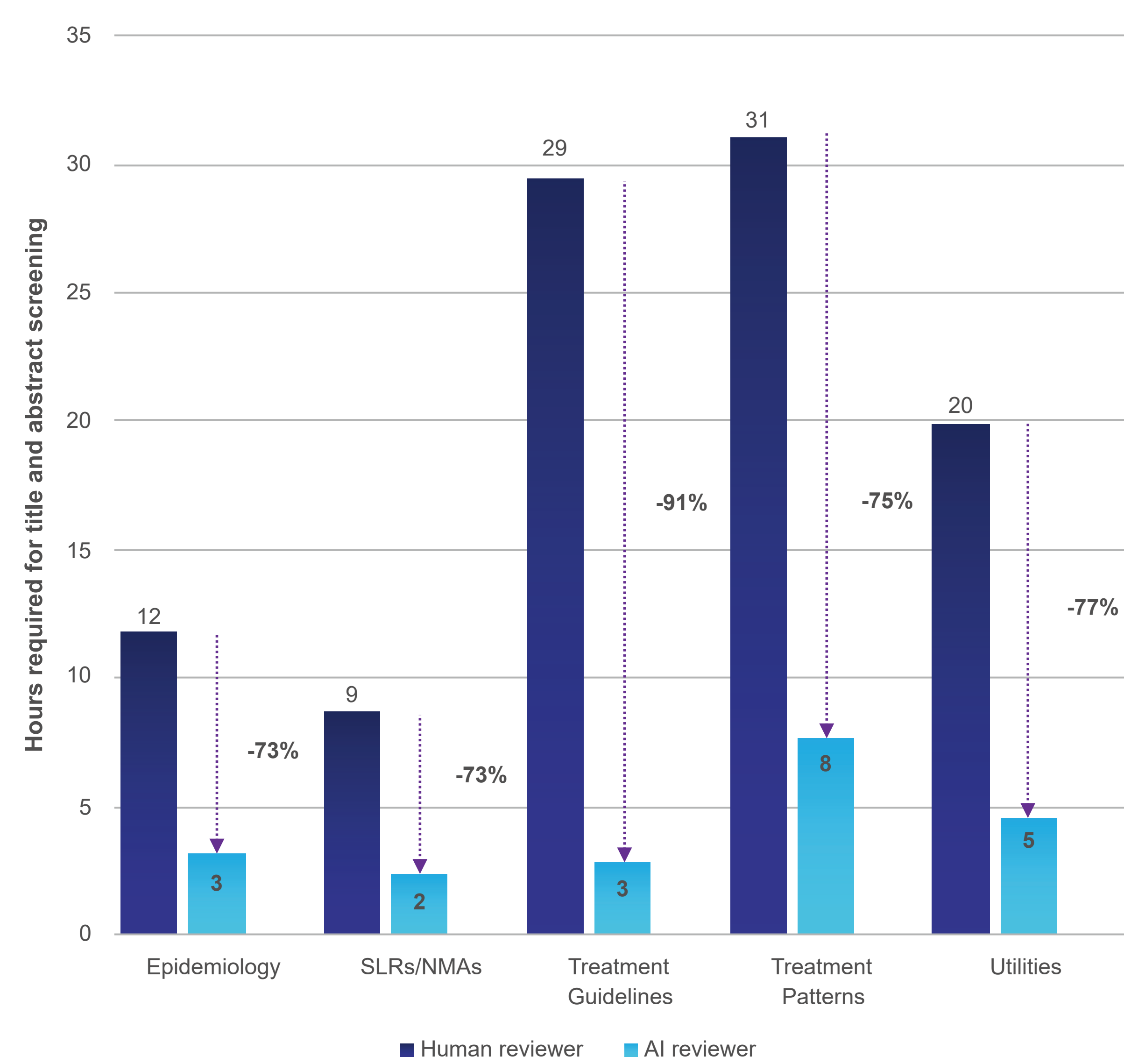
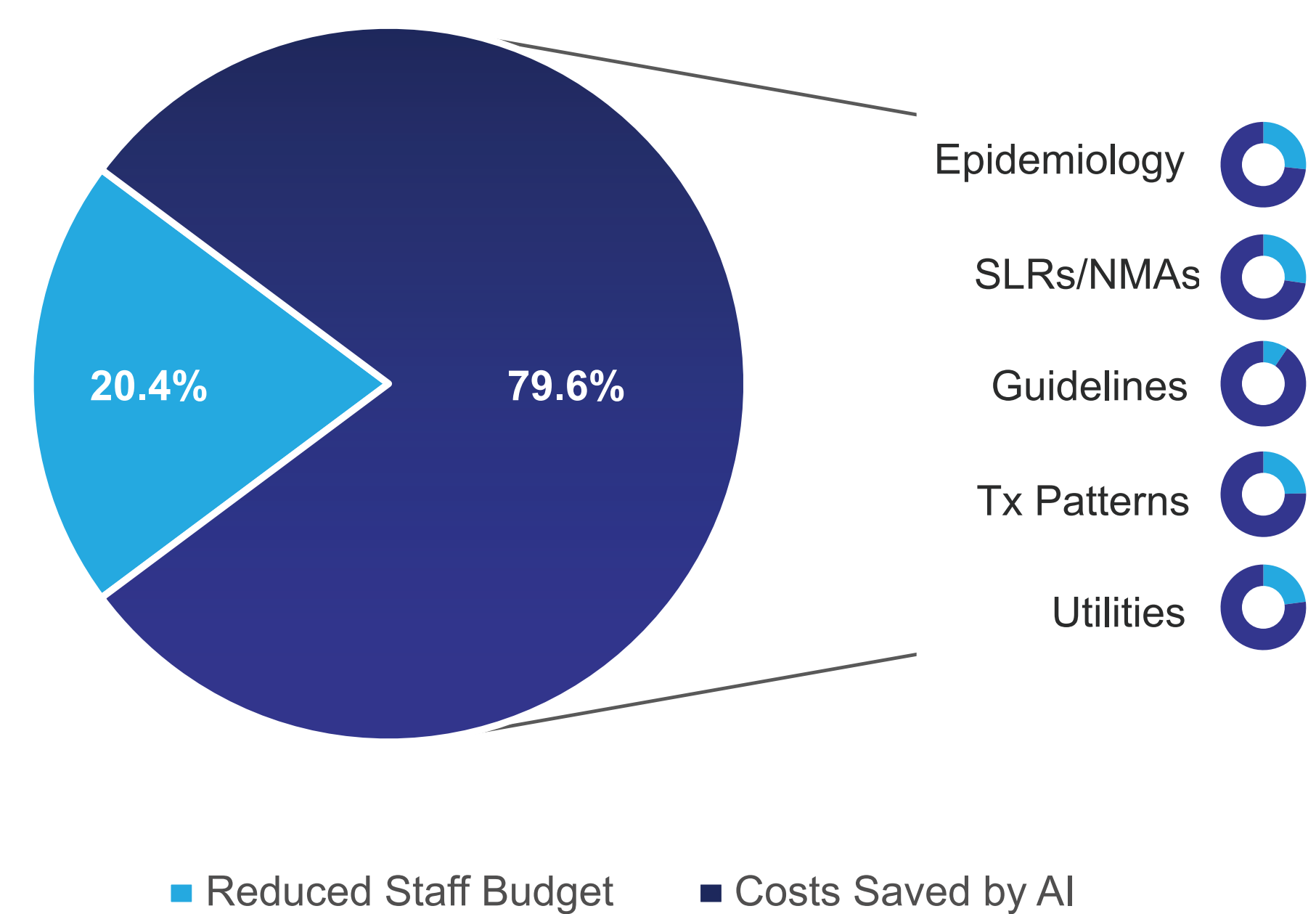


Figure 5. Cost Savings when Employing AI as a Single Reviewer



Discussion

- The potential of AI to expedite title and abstract screening and reduce human labor and costs associated with conducting narrative or targeted reviews and traditional SLRs appears promising.
- Results suggest that substituting a human reviewer with AI is most valuable when conducting reviews of clinical treatment guidelines, with maximum efficiencies gained with targeted reviews where more rigorous, dual-screening methodologies are not necessary.
- AI may have a more significant impact on time and cost of conducting SLRs with a higher yield of references to screen than those assessed herein. This, in turn, will allow reviewers to conduct more comprehensive searches to and employ AI to reduce the number of records to be screened by humans.
- While Health Technology Assessment bodies currently do not allow for AI to be used in place of humans when conducting SLRs, further investigations to validate AI algorithms and define limits or thresholds around its appropriate use could improve the utility of these tools and facilitate faster evidence generation practices to keep up-to-date with advancements in the healthcare industry.

Limitations

- A key limitation of this study is that a small number of previously completed SLRs were tested as case studies to calculate time and cost savings, which subsequently limits the generalizability of results to a wide range of literature reviews conducted in health economics and outcomes research.
 - AI was employed in a static manner wherein it was only trained once. However, AI may be best utilized in a more dynamic setting where it is able to learn as more references are screened during the review process to improve accuracy, resulting in subsequent time and cost savings.
- AI was employed in a static manner wherein it was only trained once. However, AI may be best utilized in a more dynamic setting where it is able to learn as more references are screened during the review process to improve accuracy, resulting in subsequent time and cost savings.

Conclusions

- While this study does not account for the other burdensome aspects of conducting literature reviews such as data extraction and synthesis, the use of AI has the potential to substantially reduce the volume of resources and timeframe in which reviews can be conducted and minimize labor costs. These savings may be more impactful when employing AI in literature reviews with higher volumes of references.

References

- Cichewicz, A., Kadambi, A., Lavoie, L., Mittal, L., Pierre, V., Raorane, R. Automation of Abstract Screening: A Case Study of Multiple Systematic Literature Reviews. Value In Health. 2022; 25(12 Suppl):S362.