Leveraging artificial intelligence for evidence synthesis

Well-conducted evidence syntheses are the cornerstone of evidence-based health care decisions.[1] Whether systematic reviews, scoping reviews, or qualitative evidence syntheses, these approaches are designed to provide decision-makers with the best available evidence through systematic, rigorous, and transparent methods.[2] Over recent decades, the scientific community has advanced methods for every step of the evidence synthesis process to increase trustworthiness and rigor. Yet these gains have also increased the workload: evidence syntheses are resource-intensive and can take a year or longer to complete.[3] For those making health care decisions – in clinical practice or at organizational and policy levels – this is often too long.

Developments in artificial intelligence (AI; specifically large language models [LLMs]) offer hope that we can increase efficiency without sacrificing rigor. Using AI for evidence synthesis is not new; machine learning was incorporated into systematic review software before the introduction of LLMs.[4] However, LLMs open up new possibilities: they do not require large task-specific training datasets, nor do they require review authors to have programming skills, which makes them very attractive to use for a wide audience of evidence synthesists.

Since the introduction of LLMs to the public in 2023, the way we conduct systematic reviews has already changed, though not to the extent some predicted. Full automation is not yet possible and, according to experts, should not be the aim. Human oversight is essential to guarantee rigor and trustworthiness.[5,6] This is especially relevant for 2 reasons. First, LLMs are not designed to conduct comprehensive evidence syntheses. Their primary objective is to predict and generate text,[13] rather than systematically identify, critically appraise, and synthesize all relevant evidence. Consequently, they may draw on information from sources of varying quality, fail to retrieve important studies, lack access to subscription-only content, or provide limited transparency regarding how conclusions were reached.[5] Second, the results of evidence synthesis ultimately affect people's lives. It is our responsibility to make methodological decisions carefully, weighing speed against rigor and considering potential biases and downstream consequences.

While discussions of AI in evidence synthesis often focus on efficiency gains, the more important question is whether AI can support trustworthy evidence synthesis.[7] Evidence syntheses are not merely exercises in information retrieval, they are scientific endeavors that depend on transparency, reproducibility, methodological rigor, and accountability.[2] Any assessment of AI-enabled approaches should, therefore, consider not only speed and resource savings, but also their impact on the trustworthiness of review findings and the confidence decision-makers can place in them.

Keeping our responsibility in mind, it is encouraging to see the enthusiasm and willingness of the evidence synthesis community to use AI responsibly and, therefore, to evaluate its applications. A living evidence map of AI use in evidence synthesis included more than 1000 studies by May 2026,[8] reflecting how quickly this research is evolving. Yet the evidence base is highly fragmented, with studies using different AI technologies, models, and versions applied at different review steps, often with small sample sizes.

There is an urgent need for greater collaboration and concerted methods research on AI in evidence synthesis. We also need prospective evaluations under real-world conditions to learn about actual efficiency gains and practical challenges, such as data extraction failing for qualitative data.[6] One promising avenue is the Study Within A Review (SWAR), which evaluates new methods alongside conventional review conduct.[9] When the same protocol is applied across multiple reviews and teams, generalizability improves. Collaboration with AI tool developers is equally important to ensure tools meet the demands of rigorous evidence synthesis — the Cochrane AI study is a good example of how to do this.[10]

Importantly, not all review tasks are equally amenable to AI support. AI has demonstrated promise for procedural tasks such as citation screening, deduplication, and structured data extraction.[4] However, many aspects of evidence synthesis require higher-order interpretive judgment, including synthesizing qualitative findings and translating evidence into actionable recommendations.[11] These activities depend on contextual understanding, methodological expertise, and transparent reasoning. Determining how AI can best support such interpretive processes while preserving scientific rigor represents one of the most important methodological challenges facing our field.

When applying and testing AI, some types of evidence synthesis might be better testing grounds than others. In rapid evidence syntheses, for example, commissioners accept higher uncertainty in exchange for faster results,[12] so the additional uncertainty introduced by using AI may be more tolerable. In rapid evidence syntheses, AI could even improve methodological quality by mitigating risks associated with abbreviated methods, such as single-reviewer workflows.[12]

Evidence syntheses demand high accuracy and transparency because their results directly affect people's health and lives. This makes them a particularly demanding use case for AI, as providing transparency is especially difficult for LLMs. Explainable AI will, therefore, play a major role in the future and warrants continued attention from computer scientists and tool developers.[14]

While many methods studies evaluate AI tools on single tasks of the evidence synthesis process, we may need to rethink the evidence synthesis process altogether. Applying innovative tools to outdated processes is rarely efficient. The future will likely involve agentic AI, meaning multiple agents performing distinct tasks and working together. Moving forward, human involvement can go beyond mere oversight. AI and human scientists can truly collaborate to navigate complex workflows, leveraging the strengths of both. In these mixed-initiative systems, agents can reach out to humans for clarification and guidance. Evaluation studies of such approaches and mechanisms to ensure provenance and transparency will be essential.

AI holds considerable potential to transform not only how evidence syntheses are conducted, but also how evidence is translated and delivered to decision-makers. Future applications may extend beyond review production to support evidence surveillance, living evidence systems, the development of recommendations, and the translation of evidence to practice. Realizing this potential will require robust evaluation, transparent reporting, clear provenance of AI-generated outputs and continued human oversight. Ultimately, the success of AI in evidence synthesis will not be judged by how much faster reviews can be produced, but by whether these technologies help us generate, translate and apply trustworthy evidence while maintaining the standards upon which evidence-based health care depends.

Funding

The authors did not receive any funding to write this manuscript.

Declaration of using generative AI

The authors declare the use of the LLM Claude (Opus 4.7) for editing/proofreading of the final manuscript.

Barbara Nussbaumer-Streit,[1,2]
Marc Streit,[3] Kylie Porritt[4]

[1]*JBI Austrian Centre for Evidence-based Healthcare, University for Continuing Education Krems, Austria*
[2]*Cochrane Austria, University for Continuing Education Krems, Austria*
[3]*Visual Data Science Lab, Johannes Kepler University Linz, Linz, Austria*
[4]*JBI, School of Public Health, Adelaide University, Adelaide, SA, Australia*

BNS is a co-convenor of the Cochrane Rapid Reviews Methods Group. MS is a co-founder and shareholder of datavisyn. KP is an associate editor of JBI Evidence Synthesis *and a paid employee of JBI, Adelaide University.*

Correspondence: Barbara Nussbaumer-Streit,
Barbara.nussbaumer-streit@donau-uni.ac.at

DOI: 10.11124/JBIES-26-00371

References

1. Dusin J, Melanson A, Mische-Lawson L. Evidence-based practice models and frameworks in the healthcare setting: a scoping review. BMJ Open 2023;13(5):e071188.
2. Nussbaumer-Streit B, Booth A, Garritty C, Hamel C, Munn Z, Tricco AC, *et al*. Overview of evidence synthesis types and modes. J Clin Epidemiol 2025;187:111970.
3. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545.
4. Affengruber L, van der Maten MM, Spiero I, Nussbaumer-Streit B, Mahmić-Kaknjo M, Ellen ME, *et al*. An exploration of available methods and tools to improve the efficiency of systematic review production: a scoping review. BMC Med Res Methodol 2024;24(1):210.
5. Flemyng E, Noel-Storr A, Macura B, Gartlehner G, Thomas J, Meerpohl JJ, *et al*. Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI, and the Collaboration for Environmental Evidence 2025. JBI Evid Synth 2025;23(11):2162-6.
6. Thomas J, Flemyng E, Noel-Storr A, Moy W, Marshall IJ, Hajji R, *et al*. Responsible use of AI in evidence SynthEsis (RAISE): recommendations for practice [internet]. OSF, 2026 [cited 2026 May 21]. Available from: https://osf.io/fwaud/files/cqa82.

7. Fletcher IK. Advocating for trust in and trustworthy AI to transform evidence synthesis. Environ Evid 2025;14(1):19.
8. Shokraneh F. Living Evidence Map for Automation of Systematic Reviews (LEMASyR) [internet]. Nested Knowledge; 2026 [cited 2026 May 18]. Available from: https://nested-knowledge.com/nest/21035.
9. Devane D, Burke NN, Treweek S, Clarke M, Thomas J, Booth A, *et al*. Study within a review (SWAR). J Evid Based Med 2022; 15(4):328-32.
10. Gartlehner G, Banda S, Callaghan M, Chase J-A, Dobrescu A, Eisele-Metzger A, *et al*. Cochrane Evaluation of (Semi-) Automated Review Methods (CESAR): protocol for an adaptive platform study within reviews. J Clin Epidemiol 2026:112390.
11. Porritt K, Evans C, Bennett C, Loveday H, Bjerrum M, Salmond S, *et al*. Systematic reviews of qualitative evidence. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis [internet]. JBI; 2024 [cited 2026 May 21]. Available from: https://synthesismanual.jbi.global.
12. Gartlehner G, Nussbaumer-Streit B, Hamel C, Garritty C, Griebler U, King VJ, Devane D, Kamel C. Responsible Integration of Artificial Intelligence in Rapid Reviews: A Position Statement From the Cochrane Rapid Reviews Methods Group. Cochrane Evid Synth Methods 2025 Nov 24;3(6):e70063.
13. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. EBioMedicine 2023;90:104512.
14. Thomas J, Hair K, Noel-Storr Aea, *et al*. Responsible use of AI in evidence SynthEsis (RAISE2026): building and evaluating AI evidence synthesis tools [internet]. OSF; 2026 [cited 2026 May 21]. Available from: https://osf.io/.