WHITE PAPER

Accelerating Evidence Synthesis without Compromising Scientific Rigor

AUTHORS

Prosper Maposa, Director, Business Development

Masoud Pourrahmat, Executive Director, Evidence Operations

Mir Sohail Fazeli, Senior Vice President, Strategic Solutions

ubc.com

INTRODUCTION

Evidence-based medicine is only as strong as the evidence it stands on.

01

At the heart of that evidence lies the systematic literature review (SLR), the cornerstone of evidence synthesis in health and life sciences. SLR methodology has been adapted to other fit-for-purpose literature reviews including targeted, structured, and living reviews. Yet the literature review as traditionally developed is under increasing pressure: a relentless tide of publications, combined with shrinking timelines and mounting costs pressures, are making the fully manual review model increasingly difficult to sustain.

Artificial intelligence (AI) has emerged as a powerful potential solution to address these changes. Machine learning, large language models, and AI-assisted screening tools can process tens of thousands of abstracts in minutes, identify patterns invisible to fatigued human reviewers, and dramatically compress review timelines. However, AI is a tool, not a scientist and it needs responsible use and monitoring. Without the judgment of expert methodologists and subject matter experts at every critical step, AI-enabled literature reviews risk producing fast answers to the wrong questions, or worse, expert-seeming outputs that would not survive peer review or health technology assessment (HTA) scrutiny.

This white paper makes the case that the future of responsible literature review conduct is neither fully manual nor fully automated. It is expert-guided AI: a human-in-the-loop model in which trained scientists and literature review methodologists harness advances in technology processing power while applying the methodological depth, clinical insight, and strategic framing that only trained human expertise can provide. We demonstrate that our practical implementation of this model yields significantly shorter timelines, higher efficiency and significant cost savings, without compromising the accuracy, reproducibility, and defensibility that HTA submissions as well as high impact publications demand.

Limitations of Unguided Use of AI in Literature Reviews

01 Limited ability to frame the right research questions

The most consequential decisions in a literature review happen before a single abstract is screened: What is the precise clinical/research question? How is the patient population defined? What comparators reflect current standard of care? Which outcomes matter to payers or regulators? These decisions require deep disease-area knowledge, understanding of HTA conventions, and strategic alignment with the client's strategic imperatives. No AI system can fully be a substitute for an expert methodologist and/or a subject matter expert in getting these foundational decisions right.

02 Limited ability to interpret clinical or practice context

Data extraction from complex clinical trials or real-world evidence studies often requires clinical judgment: recognising that an intention-to-treat result is more appropriate than per-protocol for an HTA context; understanding that a surrogate endpoint may or may not be accepted by a specific HTA body; identifying when a study's patient population overlaps sufficiently with the submission target population. These are judgment calls that require a trained expert, not a language model.

03 AI may present reproducibility and audit challenges

Large Language Models (LLMs) are, by design, probabilistic and non-deterministic: the same query on the same document may produce different outputs on different runs. This creates fundamental challenges for the reproducibility standards required by PRISMA[1] and demanded by peer-reviewed journals and HTA bodies. Without standardised prompts, version-controlled AI configurations, and documented human adjudication at decision points, AI-enabled reviews will be difficult to reproduce and therefore lack the required transparency to be trusted. This emphasizes the need for structured human involvement to ensure quality and reproducibility of the output.

Why Human-in-the-Loop AI is Becoming the Industry Standard

As AI capabilities have advanced, the discussion has shifted from whether AI should be used in literature reviews to how it should be implemented responsibly. Rather than replacing scientific experts, the emerging consensus is that AI should accelerate repetitive tasks while experienced methodologists remain responsible for study design, clinical interpretation, quality assurance, and scientific accountability. This human-in-the-loop approach is increasingly reflected in guidance from HTA agencies worldwide, which consistently emphasizes transparency, reproducibility, and expert oversight in AI-enabled evidence generation.[2]

Figure 1 illustrates how this model can be operationalized across the literature review lifecycle. AI is applied where it delivers the greatest efficiency gains, while expert reviewers retain responsibility for scientific decisions, methodological oversight, and final accountability.

AI accelerates.

- **Title and abstract screening**
- **Full-text screening**
- **Data extraction**
- **Risk of bias assessment**
- **Report Writing**

Experts decide.

- **Study design**
- **Clinical interpretation**
- **Quality assurance**
- **Scientific accountability**

FIGURE 1

Evidinno's Expert-Guided, AI-Assisted Literature Review Workflow

Human expertise at every critical decision point. AI accelerates. Experts decide.

AI-assisted, Expert-led foundation

1 Research question, protocol & PICO

AI-assisted Expert review / validation

Define objectives, PICO, eligibility criteria, outcomes, and methodology; confirm scientific standards.

2 Search strategy & literature search

Expert-led

Develop and execute database-specific searches; peer review search strategy and results.

AI-assisted, expert-validated workflow

3 Title & abstract screening

AI-assisted Expert review / validation

Review AI screening decisions; resolve uncertainty; confirm inclusion/exclusion based on eligibility criteria.

4 Full-text screening

AI-assisted Expert review / validation

Validate inclusion/exclusion decisions and rationales; resolve uncertainties with documentation.

5 Data extraction

AI-assisted Expert review / validation

Verify extracted data against original publications; clarify and resolve incomplete or conflicting information.

6 Risk of bias assessment

AI-assisted Expert review / validation

Verify RoB assessments and justifications; resolve discrepancies; apply methodological judgment.

AI-assisted, Expert-led output

7 Evidence synthesis & interpretation

Expert-led

Interpret findings; assess gaps and limitations; synthesize evidence and develop the value narrative.

8 Final evidence base, reporting & approval

AI-assisted Expert review / validation

Senior scientific review, quality assurance, and final approval of the evidence base, synthesis, and GVD-ready content.

Governance & quality throughout

- Standardized workflows & SOPs
- Verification against original publications
- Documented AI use & configurations
- Audit trails & decision logs
- Source-to-claim traceability
- Human accountability at every step

No screening decision, extracted data point, methodological judgment, or scientific conclusion is included in the final evidence base without the required level of human review and approval.

Growing Acceptance of AI by HTA Agencies

The acceptance of AI in evidence generation has evolved rapidly over the past several years. In a recent comprehensive review of international HTA guidance, our team found that although adoption varies across agencies, there is a growing consensus that AI has an important role in literature reviews, evidence synthesis, health economic modeling, and real-world evidence generation when implemented with appropriate human oversight.[2]

Among global HTA organizations, NICE (United Kingdom) and Canada's Drug Agency (CDA-AMC) have published formal position statements supporting the responsible use of AI in evidence generation.[2] Other organizations, including IQWiG (Germany), HAS (France), the Norwegian Institute of Public Health (NIPH), Belgium's KCE, and EUnetHTA, have similarly acknowledged the potential benefits of AI while emphasizing transparency, validation, and expert oversight.

Formal position statements

NICE
United Kingdom

Canada's Drug Agency
CDA-AMC

Similarly acknowledged

IQWiG	Germany
HAS	France
NIPH	Norway
KCE	Belgium
EUnetHTA	

Importantly, no major HTA agency currently endorses fully autonomous AI-generated evidence reviews. Instead, guidance consistently emphasizes that AI should augment, rather than replace, experienced reviewers. This emerging consensus aligns closely with the expert-guided AI model, in which AI supports efficiency while human experts remain accountable for methodological rigor, scientific interpretation, and the quality of the final evidence package.

Joint Clinical Assessment (JCA): Accelerated Timelines Drive AI Adoption

dossier submissions expected within approximately

60 to 100

days of receiving the final assessment scope.

The implementation of the European Union Joint Clinical Assessment (JCA) has introduced significantly compressed evidence generation timelines, with dossier submissions expected within approximately 60 to 100 days of receiving the final assessment scope. These timelines place considerable pressure on traditional literature review workflows, creating a strong need for more efficient approaches to evidence synthesis.

AI-assisted literature reviews are well positioned to address this challenge by accelerating activities such as literature screening, data extraction, risk-of-bias assessment, and evidence organization, enabling teams to rapidly generate comprehensive and traceable evidence packages while maintaining methodological quality.

Recognizing this opportunity, the HTA Coordination Group (HTACG) recently published its *General Principles on the Use of Artificial Intelligence in the Preparation of Dossiers for Joint Clinical Assessments*. The guidance acknowledges the potential role of AI in supporting evidence generation while emphasizing that health technology developers remain accountable for the completeness, methodological rigor, and validity of the evidence. Human oversight, transparency, and documentation of AI-assisted activities are expected throughout the review process. Consistent with these principles, expert-guided AI should be viewed as an enhancement to scientific practice, not a replacement for expert judgment.[3]

Quality, Efficiency, and Reproducibility

The value of AI-enabled literature reviews should be measured not only by speed, but also by the quality, reproducibility, and transparency of the outputs they generate. At Evidinno, we have developed and validated proprietary AI modules across multiple stages of the literature review process, including title and abstract screening, full-text screening, data extraction, and risk of bias assessment. Collectively, these evaluations demonstrate that AI can substantially reduce manual effort while maintaining the scientific rigor required for evidence synthesis.[4–7]

Across these validation studies, AI-assisted workflows consistently achieved substantial reductions in review timelines. For example, our title and abstract screening model reduced screening time by up to 95% while maintaining high sensitivity for identifying relevant studies. Similarly, our full-text screening model demonstrated high sensitivity and negative predictive value while reducing screening time by approximately 96%, supporting its use as an efficient first-pass screening tool when coupled with expert review.[4,5]

Title and abstract screening

95%

reduction in screening time

Fast-pass full-text screening

96%

reduction in screening time

Beyond study selection, our AI-assisted data extraction module substantially reduces the time required to extract structured study data while maintaining high agreement with expert reviewers. Likewise, our retrieval-augmented risk of bias assessment module demonstrated encouraging performance in supporting structured quality assessments across multiple appraisal tools, while also reinforcing that expert validation remains essential for nuanced methodological judgments.[6,7]

Extending beyond individual review tasks, Evidinno has AI-enabled tools that support both protocol development and report generation. The protocol developer creates structured first-draft review protocols from project-specific requirements, while the report writer helps transform validated evidence into literature review reports and evidence summaries. Consistent with our expert-guided, human-in-the-loop approach, these tools streamline documentation activities while preserving expert responsibility for study design, methodological decisions, interpretation, and quality assurance.

Taken together, these findings support a consistent conclusion: AI delivers the greatest value when integrated within an expert-guided, human-in-the-loop workflow. By accelerating repetitive tasks while preserving scientific oversight, organizations can achieve meaningful gains in efficiency without compromising the quality or credibility of the final evidence synthesis.

Expert-Guided AI: Our Approach

At Evidinno, we integrate proprietary AI modules across the literature review lifecycle, including title and abstract screening, full-text screening, data extraction, and risk of bias assessment. Each module has been developed and validated within standardized, expert-guided workflows, where AI accelerates repetitive tasks while experienced methodologists retain responsibility for scientific judgment, quality assurance, and final accountability.

As illustrated in **Figure 2**, our human-in-the-loop approach combines the efficiency of AI with the quality, reproducibility, and transparency required for high-quality evidence generation. The result is a scalable evidence generation process that delivers faster timelines and lower costs while supporting the methodological rigor, transparency, and reproducibility required for HTA-quality evidence synthesis and peer-reviewed publications.

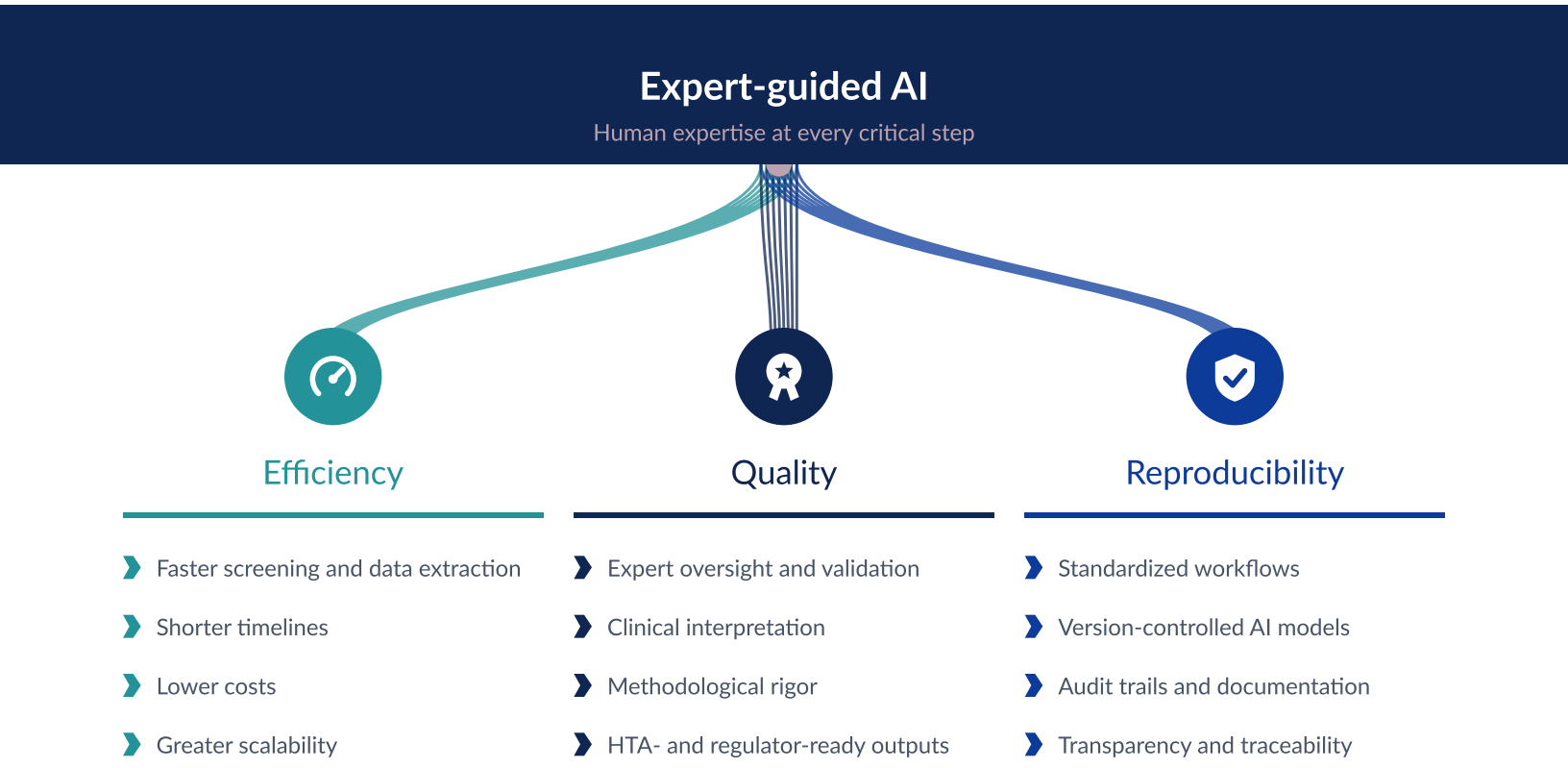

Figure 2. The Pillars of Responsible AI in Literature Reviews.

Human-in-the-loop AI delivers the optimal balance of speed, scientific rigor, and trust for high-stakes decisions.

The future of literature reviews is not about replacing experts with AI.

It is about enabling experts to do more, faster, and with greater confidence.

The demands placed on evidence generation are changing rapidly. Growing volumes of biomedical literature, increasingly compressed timelines, and evolving evidence requirements mean that traditional literature review approaches alone are no longer sufficient.

Expert-guided AI offers a practical path forward. By combining validated AI technologies with the scientific judgment of experienced methodologists, organizations can accelerate literature reviews, improve efficiency, and scale evidence generation without compromising methodological rigor.

References

1. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.

2. Gaind N, Badin M, Jacob J, et al. Artificial Intelligence in Health Technology Assessment Submissions: A Targeted Review of Global Policy and Practice. *Value in Health Regional Issues.* 2026:101658.

3. HTA Coordination Group. General Principles on the Use of Artificial Intelligence in the Preparation of Dossiers for Joint Clinical Assessments. Adopted July 15, 2026. Accessed July 31, 2026. https://health.ec.europa.eu/latest-updates/general-principles-use-artificial-intelligence-preparation-jca-dossier-2026-07-16_en.

4. Fazeli MS, Kasireddy E, Pourrahmat M-M, Chow C, Collet J-P. Automating Screening of Titles and Abstracts in Systematic Reviews: An Assessment of GPT-4o mini. *medRxiv.* 2026:2026.2005.2015.26353334.

5. Huang WH, Poojary V, Hofer K, Fazeli MS. MSR217 Evaluating a Large Language Model Approach for Full-Text Screening Task in Systematic Literature Reviews With Domain Expert Input. *Value in Health.* 2024;27(12):S481.

6. Kasireddy E, Chow C, Collet J, Pourrahmat MM, Fazeli MS. Evaluating the Performance of Claude 3.7 Sonnet in Data Extraction Automation for Systematic Literature Reviews. *Value Health Reg Issues.* 2026;53:101539.

7. Kasireddy E, Chow C, Pourrahmat M-M, Collet J-P, Fazeli MS. MSR22 Evaluating the Application of GPT-4o and Retrieval-Augmented Generation (RAG) for Assessing Risk of Bias and Study Quality in Systematic Literature Reviews (SLRs): Preliminary Findings from a Comparative Study. *Value in Health.* 2025;28(6):S278.

UBC