LASERAI

Case study

Automation of Literature Reviews

Presenter: Artur Nowak

Potential COI:
Chief Technology Officer, Evidence Prime

April 27th, 2026

Where Laser AI is used in practice

Core review stages supported

1. **Set up**
   criteria, forms, vocabularies
2. **Screen**
   dedupe, prioritize, resolve
3. **Extract**
   PDF/table suggestions + source links
4. **QA / report**
   conflicts, audit trail, exports

Examples across evidence ecosystems

- **HTA / reimbursement**
- **Medical devices**
- **Clinical practice guidelines**
- **Toxicology / hazard assessment**

CER uptake is not absent—it is less uniform

“Anna Karenina principle”: pharmaceutical/HEOR pipelines are more repeatable; comparative effectiveness reviews vary by population, comparators, outcomes, study designs, decision context and workflow.

LASERAI

Where Laser AI is used in practice

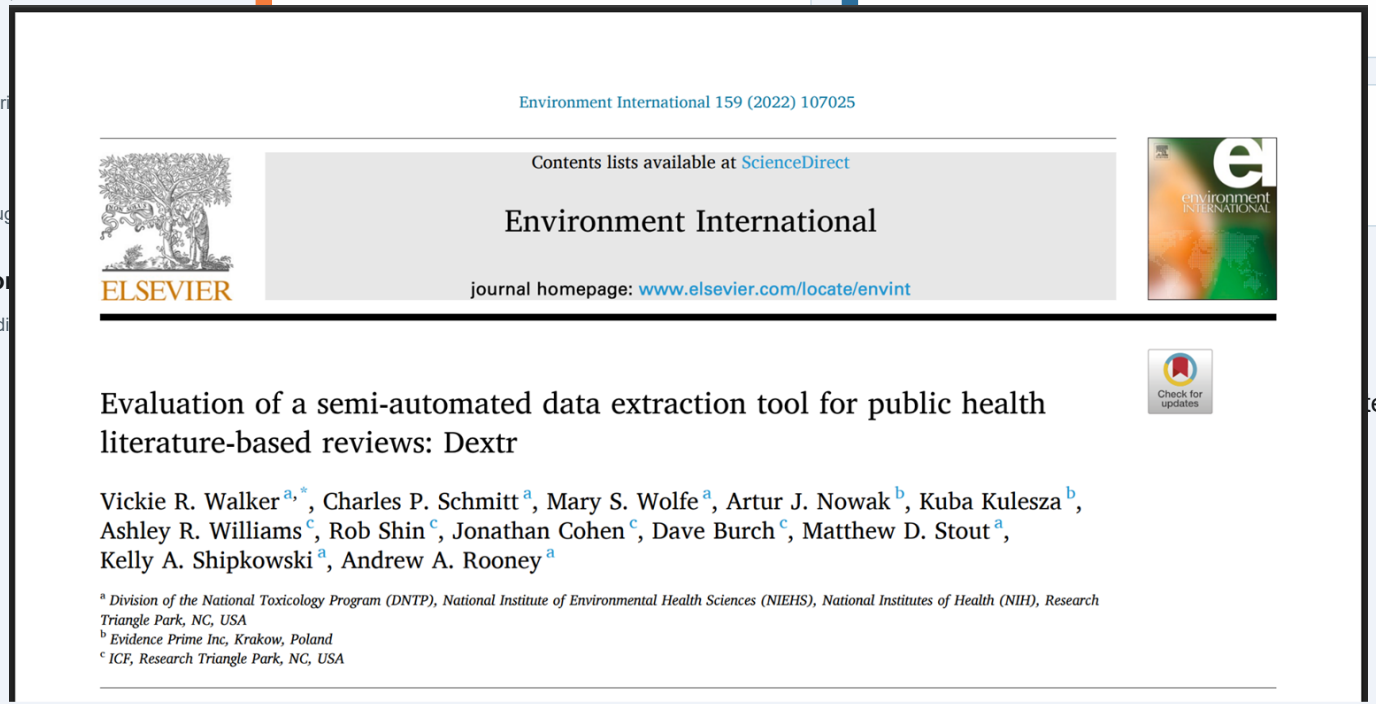
Core review stages supported

1. **Set up**
   criteria, forms, vocabularies
2. **Screen**
   dedupe, priori
3. **Extract**
   PDF/table su
4. **QA / repo**
   conflicts, audi

Examples across evidence ecosystems

HTA / reimbursement

Medical devices



Environment International 159 (2022) 107025

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr

Vickie R. Walker [a,*], Charles P. Schmitt [a], Mary S. Wolfe [a], Artur J. Nowak [b], Kuba Kulesza [b], Ashley R. Williams [c], Rob Shin [c], Jonathan Cohen [c], Dave Burch [c], Matthew D. Stout [a], Kelly A. Shipkowski [a], Andrew A. Rooney [a]

[a] Division of the National Toxicology Program (DNTP), National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), Research Triangle Park, NC, USA
[b] Evidence Prime Inc, Krakow, Poland
[c] ICF, Research Triangle Park, NC, USA

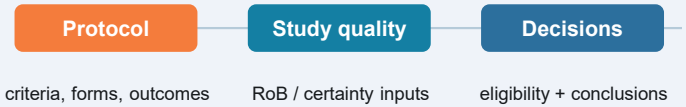
LASERAI

What Laser AI is not used for—and what slows adoption

Not positioned as a fully automated reviewer

Users can automate parts of the workflow, but prefer to stay in the loop where judgments affect validity.

AI can assist… reviewer finalizes



Protocol is king

Automation works only when the task is well specified. If the protocol is underspecified, the ideal system should ask clarifying questions.

Self-validating ≠ one-size-fits-all: every review needs its own checks.

Barriers facing AI tools in SRs

Tool-side

- validation in each review context
- generalizability across topics & study types
- transparency of errors, model versions, explainability
- data protection, copyright, workflow integration

User / institution-side

- trust, procurement, AI literacy
- no accepted “good enough” thresholds
- limited guidance from journals, guideline groups and HTA bodies

LASERAI

Updating SR standards: make human oversight specific

Standards should be task - and risk - based — not simply “AI allowed” or “AI prohibited.”

AI- assisted, human - finalized

We cannot automate what we cannot standardize

1 **Transparent reporting**

tool, version/model, date, task, settings, prompts, training examples, stopping rules, validation

2 **Human accountability**

AI may suggest, rank, extract or summarize; named reviewers remain responsible

3 **Source traceability**

values and AI-generated judgments link to citation, PDF passage, table or dataset

4 **Pre-specified validation**

metrics, error thresholds, sampling checks and conflict-resolution procedures

5 **Auditability + governance**

decision logs, change history, data protection, copyright compliance, limitations

Practical implication: lower-risk support tasks can be AI-assisted with logging; higher-risk judgment tasks require explicit reviewer verification.

LASERAI

Received: 13 December 2023 | Accepted: 10 June 2024

DOI: 10.1002/cesm.12094

Cochrane
Evidence Synthesis and Methods

RESEARCH ARTICLE

Real-world evaluation of interconsensus agreement of risk of bias tools: A case study using risk of bias in nonrandomized studies-of interventions (ROBINS-I)

Samer Saadi[1,2] | Bashar Hasan[1,2] | Adel Kanaan[1] | Mohamed Abusalih[1,2] | Zin Tarakji[1,2] | Mustafa Sadek[1] | Ayla Shamsi Basha[1] | Mohammed Firwana[1,2] | Zhen Wang[1,2] | M. Hassan Murad[1,2]

[1]Evidence-Based Practice Center, Mayo Clinic, Rochester, Minnesota, USA

[2]Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota, USA

Abstract

Background: Risk of bias (RoB) tools are critical in systematic reviews and affect subsequent decision-making. RoB tools should have adequate interrater reliability

Results: We identified 903 systematic reviews that used the tool with 51,676 cited references, from which we eventually analyzed 171 duplicated studies assessed using ROBINS-I by different systematic reviewers. The observed agreement on ROBINS-I domains ranged from 54.9% (missing data domain) to 70.3% (deviations from intended interventions domain), and was 63.0% for overall RoB assessment of the study. Kappa coefficient ranged from 0.131 (measurement of outcome domain) to 0.396 (domains of confounding and deviations from intended interventions), and was 0.404 for overall RoB assessment of the study.

Conclusion: A post hoc evaluation of RoB tools is feasible by focusing on duplicated studies that overlap systematic review. ROBINS-I assessments demonstrated

[2]Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota, USA

Background: Risk of bias (RoB) tools are critical in systematic reviews and affect subsequent decision-making. RoB tools should have adequate interrater reliability

LASERAI

Thank you!

artur.nowak@evidenceprime.com
@artur_nowak
https://laser.ai

Evidence Prime