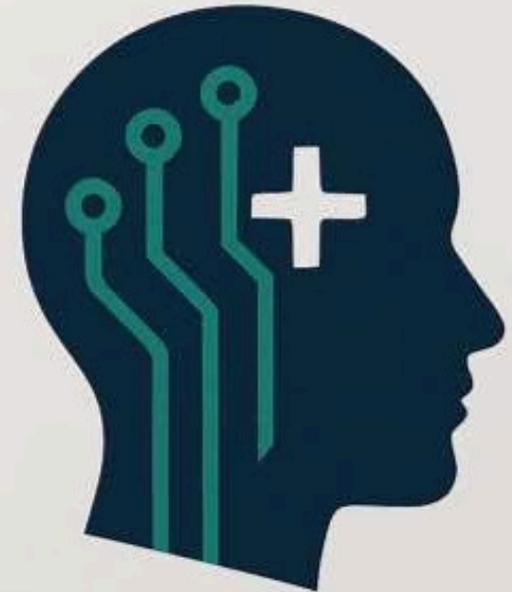


AI for Evidence Synthesis: How to Use AI Responsibly in Research

Opportunities, Risks, and Best Practices

Developed by the Centicini Research Lab, Centicini Team LTD

connect@centicini.com



CENTICIN
RESEARCH LAB

Why AI Is Entering Evidence Synthesis

The Research Challenge

The volume of published research continues to grow exponentially, creating unprecedented challenges for systematic reviewers and evidence synthesis teams worldwide.



Growing Research Volume

Over 3 million papers published annually



Time Constraints

Reviews taking 12-18 months on average



Resource Limitations

Funding pressures and staffing challenges



Automation Tools

New technologies emerging to assist researchers

What AI Is (and Is Not) in Research

Understanding the role and limitations of artificial intelligence in evidence synthesis is fundamental to using these tools responsibly and effectively in your research workflow.

AI as a Support Tool

Artificial intelligence serves as an assistant to enhance efficiency, automate repetitive tasks, and process large volumes of information. It augments human capabilities but does not replace the researcher's expertise, judgment, or decision-making authority.

Not a Replacement for Judgment

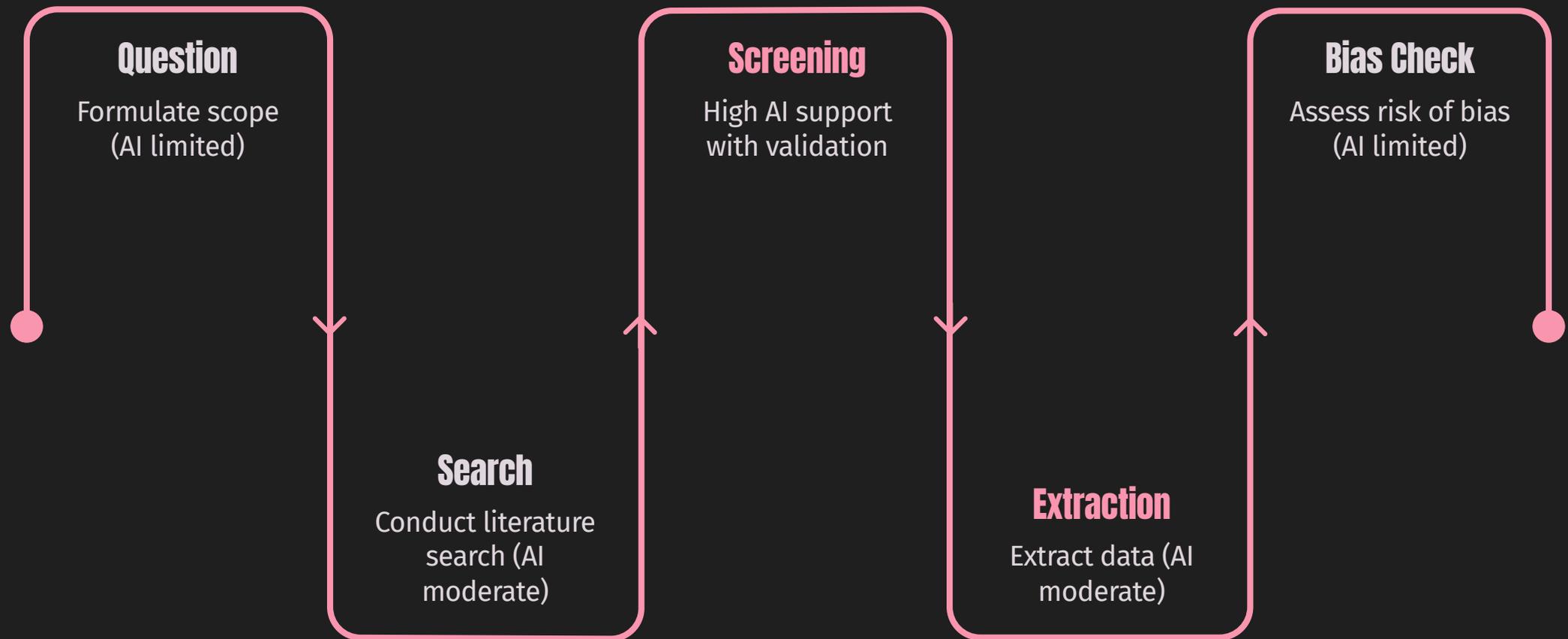
AI cannot replicate the nuanced methodological decisions, contextual understanding, and critical appraisal that experienced researchers bring to evidence synthesis. Human oversight remains essential for quality and validity.

Automation vs. Intelligence

Most "AI" tools in research are sophisticated automation systems rather than truly intelligent agents. They follow programmed rules and patterns but lack genuine understanding, reasoning capability, or the ability to contextualize findings within broader scientific discourse.

Overview of the Evidence Synthesis Workflow

The systematic review process involves multiple stages, each with distinct opportunities and limitations for AI assistance. Understanding where AI can—and cannot—be appropriately applied is crucial for maintaining methodological rigor.



Green shading indicates stages where AI can provide substantial assistance with appropriate human oversight. Yellow indicates moderate support roles. Red indicates areas where AI should play minimal or no role due to the need for expert judgment.

AI in Research Question Development

Potential Applications

AI tools can assist in the early stages of question development by scanning existing literature, identifying potential gaps, and helping researchers refine their focus areas. These technologies can quickly survey broad topic areas and suggest unexplored angles.

- **Literature Landscape Mapping**

Rapid identification of research trends and patterns

- **Gap Identification**

Highlighting understudied areas within a field

- **Concept Refinement**

Exploring related terminology and frameworks

Critical Limitations

- ❏ **Caution:** AI-generated suggestions may reflect biases present in training data, potentially steering research toward overrepresented populations, settings, or topics while neglecting marginalized or underresourced areas.

The fundamental research question must emerge from human expertise, clinical insight, and understanding of population needs—not from algorithmic pattern detection. AI lacks the contextual awareness to identify truly important clinical or public health questions.

AI in Literature Searching

Search strategy development is a specialized skill requiring deep understanding of databases, controlled vocabularies, and information science principles. AI tools are beginning to play a supportive role, but with important caveats.

Query Expansion

AI can suggest alternative keywords, synonyms, and related terms to broaden search coverage and capture relevant variations in terminology across different studies and databases.

1

2

3

Human Validation

All AI-generated search strategies must be reviewed and refined by information specialists or experienced researchers. Translation across databases and final strategy approval remain human responsibilities.

Keyword Refinement

Natural language processing can help identify the most relevant search terms from a sample of known relevant studies, though this requires careful validation.

"AI cannot replace the expertise of information specialists in developing comprehensive, reproducible search strategies that meet systematic review standards."

AI in Screening Studies

Title and abstract screening represents one of the most promising applications of AI in evidence synthesis, offering potential efficiency gains while maintaining scientific rigor when properly implemented.



AI Assistant

- Free AI tier in JetBrains IDEs
- Latest cloud AI models and support for local models
- Improved AI Chat with multi-file Edit mode
- Codebase context, web search in chat, MCP, and more

01

Initial Prioritization

AI ranks studies by relevance probability, allowing reviewers to see likely relevant studies first

02

Semi-Automated Screening

Machine learning tools can exclude clearly irrelevant studies with high confidence after human training

03

Dual Human Review

All included studies and borderline cases must undergo independent review by two human screeners

04

Missed Study Analysis

Regular checks to ensure AI hasn't systematically excluded relevant studies

Important: Never allow AI to make final inclusion/exclusion decisions without human oversight. The risk of missed relevant studies can compromise the entire review's validity and comprehensiveness.

AI in Data Extraction

Data extraction requires careful attention to detail, understanding of study designs, and ability to interpret reported findings. AI can provide structured support but cannot replace human judgment in this critical phase.



Structured Extraction Support

AI tools can help populate standardized data extraction forms by identifying and extracting clearly reported numerical data, study characteristics, and outcome measures from structured sections of papers.



High Error Risk Areas

AI frequently misinterprets complex statistical reporting, incorrectly extracts data from tables, confuses outcome measures, or fails to recognize when data are adjusted versus unadjusted. Error rates can exceed 20-30% even for straightforward extractions.



Mandatory Verification

Every data point extracted by AI must be verified against the original source by a human reviewer. This includes checking calculations, units, sample sizes, and ensuring data correspond to the correct outcome and time point.



When Not to Use AI

Avoid AI extraction for complex study designs, narrative data, qualitative findings, nuanced intervention descriptions, or any data requiring contextual interpretation. The time spent correcting errors often exceeds manual extraction time.

AI in Risk of Bias Assessment

Risk of bias assessment is fundamentally a judgment-based process requiring methodological expertise, understanding of study context, and ability to evaluate reporting quality. This represents a significant limitation for current AI capabilities.

Why Judgment Matters

Assessing risk of bias requires evaluating whether study design features were appropriate for the research question, whether conduct was adequate, and whether reporting suggests potential problems. These are inherently subjective decisions requiring expertise.

AI's Critical Limitations

AI cannot understand methodological nuances, recognize when unclear reporting suggests actual bias versus incomplete description, or contextualize findings within study settings. It often mistakes reporting quality for methodological quality.

Acceptable Supportive Roles

AI may help flag potentially relevant text sections or extract clearly stated methodological details. However, all bias assessments must be completed independently by trained reviewers with methodological expertise. AI should never assign bias ratings.

AI in Narrative Synthesis and Writing

Appropriate Uses

→ Drafting Support

AI can help generate initial text structures or reorganize existing content for clarity and flow

→ Language Refinement

Grammar checking, sentence restructuring, and clarity improvements for non-native English speakers

→ Format Conversion

Adapting text to meet journal requirements or presentation formats

Critical Risks

Hallucination: AI frequently generates plausible-sounding but factually incorrect statements, fabricates citations, or misrepresents study findings. These errors can be subtle and difficult to detect.

Loss of Voice: Over-reliance on AI writing tools can result in generic, impersonal text that lacks the critical analysis and expert interpretation expected in systematic reviews.

Full Human Control Required: Authors must write or substantially revise all content themselves. AI-generated text should never be used without thorough verification and personalization. Every sentence must reflect the authors' genuine understanding and analysis.

AI in Meta-Analysis

Meta-analysis involves sophisticated statistical techniques requiring understanding of heterogeneity, appropriate models, and interpretation of results. AI's role here is limited but evolving in specific supportive capacities.



Data Preparation

AI may assist with data formatting, conversion between effect measures, or calculating standard errors from reported statistics—though all calculations require verification



Visualization

Automated generation of forest plots and other meta-analysis visualizations based on extracted data



Cannot Calculate Independently

AI lacks the methodological understanding to select appropriate models, assess heterogeneity, or conduct sensitivity analyses without expert guidance



Human-Led Interpretation

Statistical biostatisticians or methodologists must design analyses, interpret results, assess certainty of evidence, and draw conclusions—AI cannot perform these expert functions

❏ The statistical expertise required for rigorous meta-analysis cannot be replaced by automation. AI tools may speed certain processes but cannot substitute for biostatistical training and methodological judgment.

Ethical Risks of AI in Evidence Synthesis

The integration of AI into research workflows introduces significant ethical concerns that can compromise research integrity, reproducibility, and trustworthiness if not carefully managed.



Hallucinations and Fabrication

AI systems confidently generate false information, non-existent citations, or misrepresented study findings. These fabrications can be sophisticated and difficult to detect without careful verification.



Bias Amplification

AI models trained on existing literature can perpetuate and amplify historical biases related to race, gender, geography, and socioeconomic status, potentially skewing evidence synthesis toward privileged populations.



Plagiarism Concerns

AI tools may reproduce copyrighted text or ideas without attribution, creating inadvertent plagiarism. Researchers remain legally and ethically responsible for all submitted content.



Lack of Transparency

Many AI systems operate as "black boxes" with opaque decision-making processes, making it impossible to understand why certain recommendations were made or studies were prioritized.

Responsible AI Use Principles

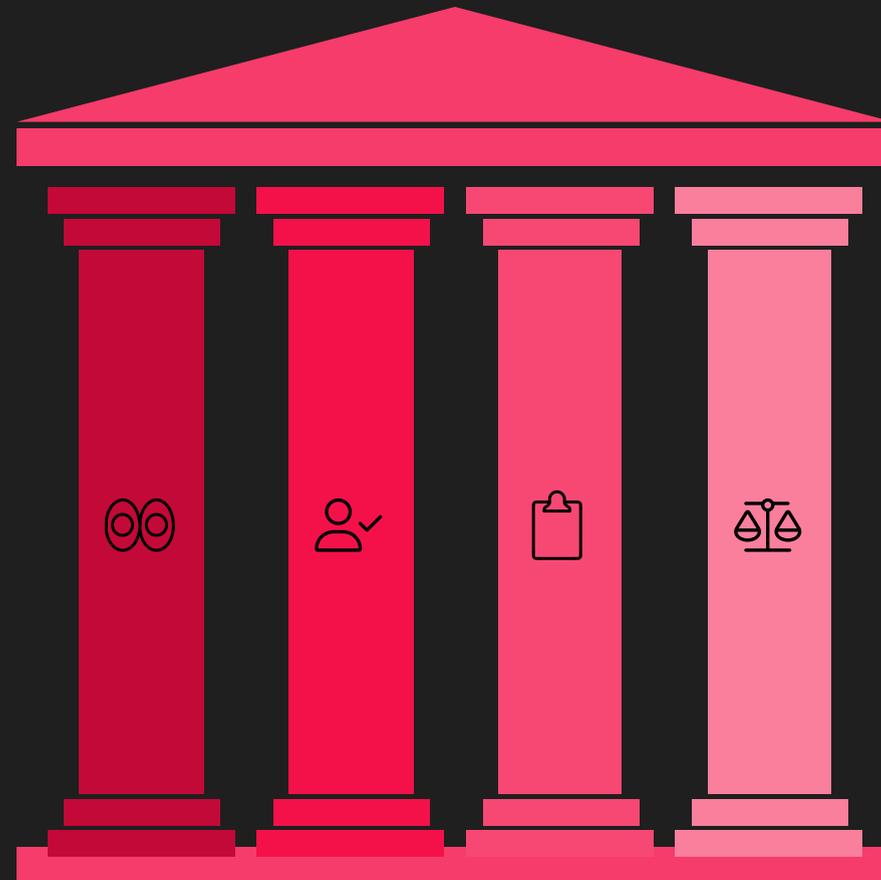
Ethical AI integration in evidence synthesis requires adherence to core principles that protect research integrity and maintain public trust in scientific findings.

Transparency

Disclose all AI use in research

Reproducibility

Document methods, tools, and settings



Human Oversight

Experts validate and review outputs

Accountability

Authors take responsibility for accuracy

These principles form the foundation for ethical AI use in systematic reviews and meta-analyses. Each principle reinforces the others, creating a framework that protects both research quality and scientific integrity.

Transparency: Researchers must clearly document where, when, and how AI tools were used, including tool names, versions, and prompts given. This enables peer reviewers and readers to assess potential impacts on findings.

Reproducibility: Sufficient detail must be provided to allow other researchers to understand and potentially replicate the AI-assisted processes. This includes settings, parameters, and validation procedures.

Human Oversight: Expert researchers must review, validate, and take responsibility for all AI-generated outputs. Automation does not reduce the need for methodological expertise—it increases it.

Accountability: Authors retain full responsibility for accuracy, interpretation, and ethical conduct regardless of AI involvement. "The AI made an error" is not an acceptable defense for research misconduct.

Reporting AI Use in Systematic Reviews

What to Disclose

Comprehensive documentation of AI use is becoming a requirement for publication in major journals. Your methods section should clearly describe AI involvement with sufficient detail for readers to evaluate potential impacts.

1 Specific Tools and Versions

Name the AI systems used, including version numbers and access dates, as capabilities change over time

2 Workflow Integration Points

Specify exactly which review stages involved AI assistance and which remained fully manual

3 Validation Procedures

Describe how AI outputs were checked, who performed validation, and what error rates were detected

4 Limitations and Rationale

Explain why AI was used, acknowledge its limitations, and describe mitigation strategies

II. EXPERIMENTAL SET-UP

Alcator C-Mod is a high field (2.1 – 8.1 T), compact ($a = 0.22$ m, $R = 0.68$ m) high performance, diverted tokamak, with high-Z metal plasma facing components. Data from a single plasma discharge with an L-I transition (shot 1101209029) are used in this paper for comparisons with nonlinear gyrokinetic simulations, and data from similar plasmas are used to examine trends across L-I transitions.

The selected discharge has been described in detail previously.^[8] The plasma parameters are $\langle n_e \rangle = 2.1 \times 10^{20}$ m⁻³, $I_p = 1.31$ MA, $B_z = 5.66$ T, $q_{95} = 3.25$. The discharge is operated with the ion ∇B -drift direction pointing away from the active x-point (unfavorable direction), which raises the power threshold for the L-H mode transition. Auxiliary input power from ion cyclotron range of frequency (ICRF) heating is applied, with $P_{ICRF} = 5$ MW.

This particular plasma is selected for comparisons with gyrokinetic codes, because there is very steady density across the L-I transition which allowed for measurements of the evolution of core and edge density fluctuations across the transition with a multi-channel reflectometer. The time history for the discharge is shown in Fig. 1. Across the L-I transition, which begins at $t = 0.875$ s, the core temperature increases as the edge temperature pedestal forms. Edge turbulence measurements with a reflectometer at $r/a = 0.99$ showed reduced broadband turbulent density fluctuations, and the appearance of the Weakly Coherent Mode (WCM), across the L-I transition. Core turbulence measured with a reflectometer at $r/a = 0.55$, showed reduced fluctuations across the L-I transition. The changes in edge and core turbulence persist throughout the I-mode; the changes are not transient. Detailed descriptions of the reduction in core fluctuations are presented in Ref. 8.

Excerpt from White, A.E., et al. Physics of Plasmas 22, 056109 (2015). Annotations by the NSE Communication Lab

Reasons for using these methods — Passive voice

Details of the methodology — Citation of standard methods

Reasons for using these methods — Active voice but the subject isn't the experimenter

Details of the methodology — Citation of standard methods

Many journals now require a dedicated statement on AI use or integration within methods sections. The PRISMA guidelines are evolving to include AI-specific reporting items. Check target journal requirements early in your project.

AI and Research Integrity

Research integrity extends beyond avoiding fabrication and falsification—it encompasses honest representation of contributions, proper attribution, and taking responsibility for all aspects of published work.

Avoiding Plagiarism

AI-generated text may inadvertently reproduce copyrighted material or closely paraphrase existing publications. All content must be checked through plagiarism detection software and substantially revised to reflect original synthesis and interpretation by the authors.

Intellectual Ownership

Authors must be able to explain and defend every aspect of their systematic review. If AI assistance was used, authors should still demonstrate deep understanding of the topic, methods, and findings. Superficial knowledge resulting from over-reliance on AI is ethically problematic.

Proper Citation Practices

AI tools that suggest references must be meticulously verified, as fabricated citations are a common problem. Never cite a paper without confirming its existence and relevance. Missing or incorrect citations constitute research misconduct.

Author Accountability

Listed authors must meet standard authorship criteria, including substantial intellectual contributions. AI cannot be listed as an author. All authors remain jointly responsible for the accuracy and integrity of the entire work, regardless of which sections involved AI assistance.

Common Misuses of AI in Research

Understanding what not to do is as important as knowing appropriate applications. These misuses can result in manuscript rejection, retraction, or damage to professional reputation.



Writing Methods Sections

AI-generated methods often contain generic descriptions, methodological errors, or misrepresent what was actually done. Methods must authentically describe your specific procedures.



Generating Results

Using AI to interpret data, describe findings, or generate results sections risks fabrication and misrepresentation. Results must come directly from your analysis of extracted data.



Replacing Screening Decisions

Allowing AI to make final inclusion/exclusion decisions without human review can miss relevant studies and compromise review comprehensiveness and validity.



Fabricating References

Using AI-suggested citations without verification results in non-existent or irrelevant references—a serious form of research misconduct that is increasingly detected by editors.

Critical Reminder: If you cannot explain exactly how something was done or defend a specific decision, that content should not be in your manuscript. AI assistance does not excuse lack of understanding or accountability.

Best-Practice Workflow for AI-Assisted Reviews

A responsible integration model balances efficiency gains with methodological rigor, ensuring AI supports rather than compromises research quality.



Human-Led Design

Experts develop the research question, protocol, and methodological approach without AI involvement. This ensures the review addresses genuinely important questions and uses appropriate methods.



AI-Supported Efficiency

AI tools assist with time-intensive tasks like screening prioritization, data extraction support, or literature searching—always with clear boundaries and validation plans in place.



Human Validation

Trained reviewers verify all AI outputs, check for errors, assess quality, and make final decisions. Validation rates and error detection should be documented and reported.



Transparent Reporting

Methods section clearly describes AI use, validation procedures, and limitations. Readers can assess potential impacts on findings and reproduce the approach.

"The goal is not to use AI for the sake of innovation, but to enhance efficiency while maintaining or improving the quality and trustworthiness of evidence synthesis."

Special Considerations for LMIC Researchers

Researchers in low- and middle-income countries face unique challenges and ethical considerations when adopting AI tools for evidence synthesis.

Resource Constraints

Limited budgets may restrict access to premium AI tools, requiring careful evaluation of free alternatives and their limitations. Cost should not compromise methodological rigor.

Tool Access Issues

Some AI services have geographic restrictions, require payment methods unavailable in certain countries, or have limited functionality in non-English languages.

Training Data Bias

Most AI systems are trained predominantly on Western literature and may perform poorly with research from underrepresented regions, potentially marginalizing LMIC evidence.

Ethical Responsibility

LMIC researchers have particular obligations to ensure AI does not further entrench inequities or exclude evidence from resource-limited settings in their reviews.

Recommended Approach

Prioritize open-source and freely available tools where possible. Collaborate internationally to access resources and expertise. Be explicit about tool limitations in your reporting.

Consider that manual methods with expert review may sometimes be more reliable than using inadequate AI tools. Quality matters more than adopting trendy technologies.

Advocate for AI tool developers to consider global research needs, include diverse training data, and ensure equitable access to technologies.

Case Example: Responsible vs Irresponsible Use

Good Practice Example

01

Screening assistance: Team used Rayyan AI to prioritize 5,000 abstracts by relevance, reviewing them in ranked order

02

Validation: All included studies and a 10% random sample of excluded studies underwent independent dual human review

03

Documentation: Methods section specified the tool name, version, training approach, and validation error rate (1.2%)

04

Limitations: Authors acknowledged AI assistance and discussed potential for missed studies in their limitations section

ACCEPTED

Poor Practice Example

01

ChatGPT methods: Authors used AI to write methods section, which contained generic language and didn't match what was actually done

02

Unverified citations: Discussion included 12 AI-suggested references, 4 of which were fabricated and 3 misrepresented

03

No disclosure: AI use was not mentioned anywhere in the manuscript

04

Data errors: AI-extracted data contained calculation errors that affected the meta-analysis results

REJECTED

📌 **Consequences:** The poorly conducted review was rejected by three journals. One editor flagged the authors' names for heightened scrutiny on future submissions. This damaged the researchers' professional reputations and delayed important research by months.

Future Directions and Final Reflections

Artificial intelligence will continue evolving, offering new capabilities and raising new ethical questions for the evidence synthesis community. The fundamental principles of responsible research, however, remain constant.

"The future of evidence synthesis is human-led, AI-supported, and ethically grounded. Technology serves researchers and society—not the other way around."

AI Supports, Never Replaces

Expert judgment, methodological rigor, and scientific integrity cannot be automated. AI tools should enhance researcher capabilities while preserving the essential human elements of critical thinking and contextual understanding that make evidence synthesis valuable.

Ethics and Transparency Are Non-Negotiable

As AI capabilities expand, the imperative for clear documentation, honest reporting, and responsible use becomes even more critical. Researchers must prioritize research integrity over technological convenience or efficiency pressures.

Responsible Use Strengthens Credibility

Thoughtful, transparent AI integration can improve efficiency without compromising quality—enhancing rather than threatening the trustworthiness of systematic reviews. The research community must collectively establish and uphold high standards for AI-assisted evidence synthesis.