

Integrating automation into (living) reviews

How can automation re-shape our processes?

Lena Schmidt

May 1st 2026

EPPI Centre
Institute of Education
University College London





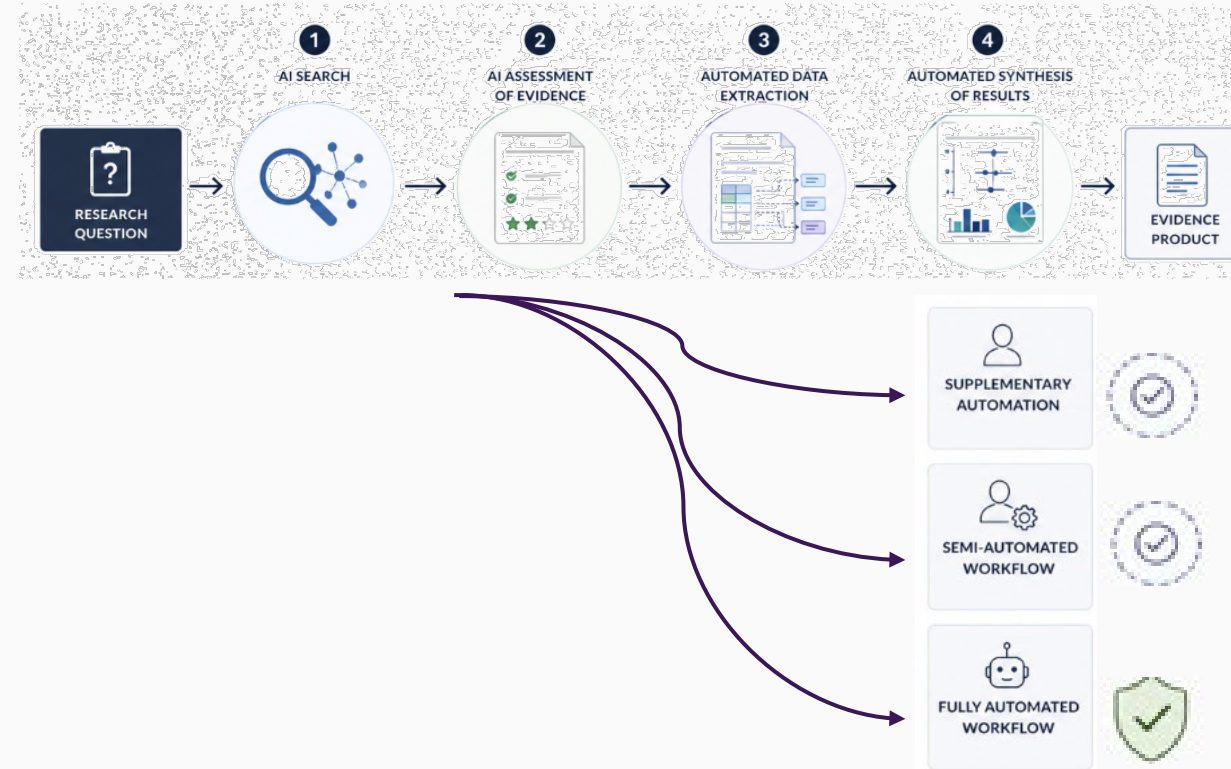
Agentic vs. step-by-step automation

Trading off 'rapidness' and rigor in static or living reviews

Shaping the Process: Agentic vs. Stepwise model



“Agentic”



“Step-by-step ‘modular’ automation”

Example: Automation of screening

Intensity of automation and type of automation depends on factors:

- Review methodology
- Impact or importance of review conclusions
- Resources (feasibility of project - > Something better than nothing?) and labelled data available
- Team's understanding of evaluation process, metrics, and their meaning

See RAISE [1] for more detailed guidance focussed on systematic

Cochrane Evidence Synthesis and Methods

RESEARCH ARTICLE OPEN ACCESS

Don't Stop Me Now, 'Cause I'm Having a Good Time Screening: Evaluation of Stopping Methods for Safe Use of Priority Screening in Systematic Reviews

Tim Repke^{1,2} | Francesca Tinsdeall³ | Diana Danilenko⁴ | Sergio Graziosi⁵ | Finn Müller-Hansen¹ | Lena Schmidt^{2,4} | James Thomas⁶ | Gert van Valkenhoef⁶

¹Climate Economics and Policy, Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany | ²Evidence for Policy and Practice Information Center, UCL Social Research Institute, Institute of Education, University College London, London, UK | ³Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK | ⁴NIHR Innovation Observatory, Newcastle University, Newcastle upon Tyne, UK | ⁵Central Executive Team, Cochrane Collaboration, London, UK

Correspondence: Tim Repke (tim.repke@pik-potsdam.de)

Received: 20 June 2025 | Revised: 7 November 2025 | Accepted: 25 December 2025

Funding: Wellcome Trust, Grant/Award Numbers: 313586/Z/24/Z, 227165/Z/23/Z; European Research Council, Grant/Award Number: 951542; Ministry of Research, Science and Culture (MWFK) of Land Brandenburg, Grant/Award Number: 22-Z105-05/002/001

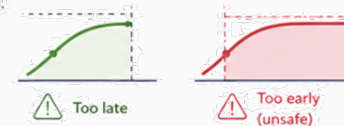
1. Classic machine-learning and priority screening

Helps to identify records earlier, but does not necessarily save time.



Statistical stopping rules:

Often too conservative (stop too late) or unsafe (stop too early):



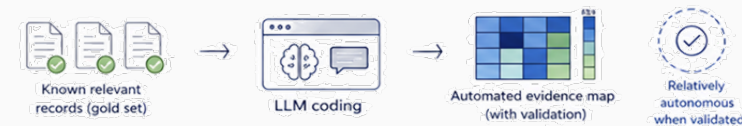
2. LLM screening

Requires some human labelling of data to develop prompts and validate them.



3. Evidence maps

If known relevant records exist, LLMs can automate coding for evidence maps.



4. Rapid reviews

Validated prompts can replace one reviewer.



5. Full systematic reviews

LLM can act as a third reviewer and increase methodological rigor.



Tools and how to use them (small selection from RAISE 3)



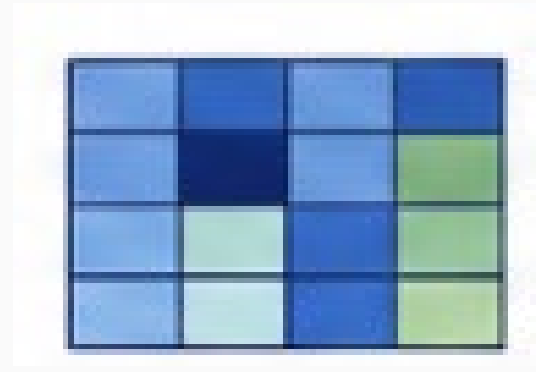
Agentic, Semantic Search:

- Elicit
- NestedKnowledge
- Asta Find Papers



Screening (prioritised):

- EPPI Reviewer
- Abstrackr
- Covidence
- Rayyan



Screening and mapping (generative LLM):

- EPPI Reviewer
- Directly within Copilot, CHatGPT



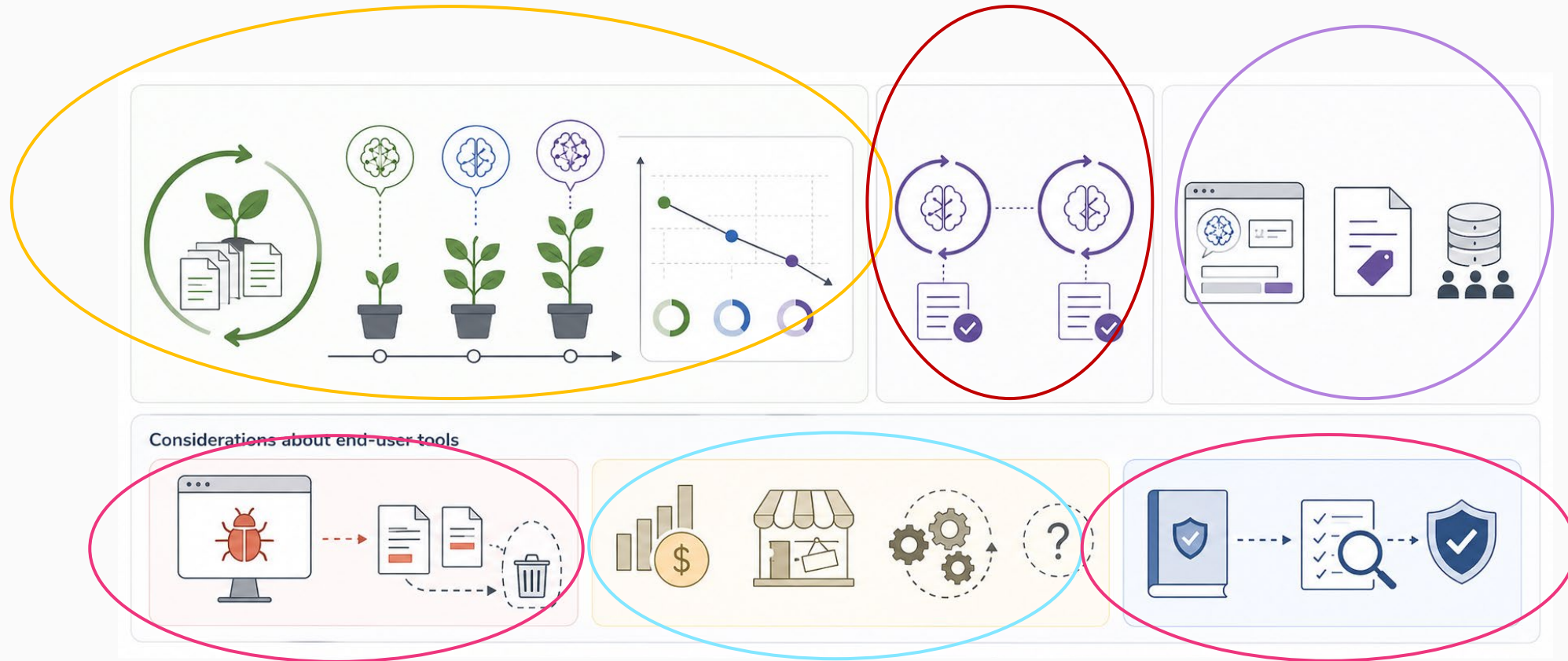
Synthesis (qualitative):

- The agentic tools

Visual synthesis:

- EPPI Mapper
- EviAtlas

Practical and 'living' considerations



Sources

[1] Thomas J, Hair K, Noel Storr, A. et al. **Responsible use of AI in evidence SynthEsis (RAISE): recommendations for practice (version 3; updated 13 March 2026).** In: Open Science Framework [<https://osf.io/>], Washington DC: Center for Open Science. DOI 10.17605/OSF.IO/FWAUD (accessed 30/04/2026 via <https://osf.io/cqa82>)

Repke T, Tinsdale F, Danilenko D, Graziosi S, Müller-Hansen F, Schmidt L, Thomas J, van Valkenhoef G. **Don't Stop Me Now, `Cause I'm Having a Good Time Screening: Evaluation of Stopping Methods for Safe Use of Priority Screening in Systematic Reviews.** *Cochrane Evid Synth Methods.* 2026 Jan 21;4(1):e70068. doi: 10.1002/cesm.70068. PMID: 41583534; PMCID: PMC12825451.