



EVIDENCE SYNTHESIS
IRELAND



Cochrane
Ireland

Evidence Synthesis Ireland Fellowship Scheme

Project Identification Form

Mentor and Host Centre

Nadia Soliman; EPPI Centre, UCL

Project title

Agentic-mapper: Evaluating the Performance of an Agentic AI Tool for Automated Evidence Gap Mapping

Project type and methods

Applied methodology / Responsible use of AI for Evidence Synthesis

Project details

Evidence gap maps (EGMs) are a critical tool for synthesising and visualising the state of research evidence across intervention-outcome matrices. They are resource-intensive to produce and update, limiting their currency for decision-making. We have developed an agentic AI tool (<https://destiny-evidence.org/agentic-map-builder/>) that automates the process; repository search, record screening, and study coding, and now seek to rigorously evaluate its performance.

Aims

- Develop and apply a staged evaluation framework covering search recall, screening accuracy, coding fidelity and pilot across 3 EGMs varying in domain and size.
- Quantify sensitivity, specificity and workload reduction relative to manual processes
- Explore secondary measures e.g., inter-rater agreement, where time allows
- Identify systematic biases and failure modes in agent performance across literature types and topic domains
- Assess test-retest reliability and consistency of agent outputs
- Evaluate the precision-recall trade-off to identify appropriate thresholds for supervised deployment
- Produce initial guidance on where human oversight remains essential within the pipeline
- Contribute to the emerging evidence base on AI-assisted evidence synthesis

Outputs

- Validated protocol/ evaluation framework for agentic EGM tools
- Recommended performance thresholds and human-in-the-loop checkpoints for the tool
- Peer-reviewed publication(s) reporting performance benchmarks and methodology

Review current status:

We have several maps available for use as a reference e.g., climate and health, HPV, and Education. Underlying data quality will affect LLM performance, there will be a feedback loop to address human mistakes or variance when identified.

Any specific/desirable requirements for fellow (e.g. clinical expertise, methodological expertise)**Desirable**

- Experience with evidence gap maps
- Familiarity with AI/machine learning concepts
- Experience or interest in learning how to use Python

Estimated start and completion dates

October 2026 – July 2027