Recommendations and guidance on responsible AI in evidence synthesis

And an introduction to the new AI Methods Group

Anna Noel-Storr, Head of Evidence Pipeline & Data Curation, Cochrane
James Thomas, Professor of Social Research & Policy, UCL
Ella Flemyng, Head of Editorial Poloicy & Research Integrity, Cochrane

Trusted evidence.
Informed decisions.
Better health.

Cochrane

Overview

1. Why we need to embrace AI and the importance of infrastructure
2. A global challenge and introducing RAISE (Responsible AI use in evidence SynthEsis)
3. Applying the RAISE recommendations in practice; what does it mean for authors?
4. Introduction to the new, joint AI Methods Group
5. Questions

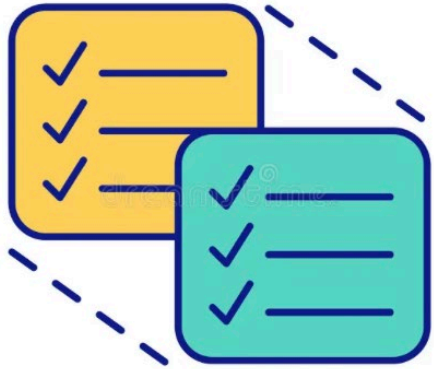
Cochrane

Why we need to embrace AI

Context: core principles of evidence synthesis

Rigor

Transparency



Replicability

The stakes are high with important decisions based on evidence syntheses

The problem!

The vast and growing amount of research produced means that we are finding it increasingly difficult to keep up!

Pinpointing needles in giant haystacks *(Shemilt 2017)*

We have a huge **specificity problem** in information retrieval for evidence synthesis.

Questions are becoming more complex.

It's hard to find what we need without retrieving large quantities of noise.

The inefficiency of isolation

Siloed working and siloed data resulting in significant duplication of effort and research waste.

The manual process of evidence synthesis has become unsustainable

Artificial intelligence

An umbrella term which means different things to different people.

Ways to automate tasks and activities currently performed by humans.

Artificial Intelligence

1950s

Text analytics

1960s

Rule based systems

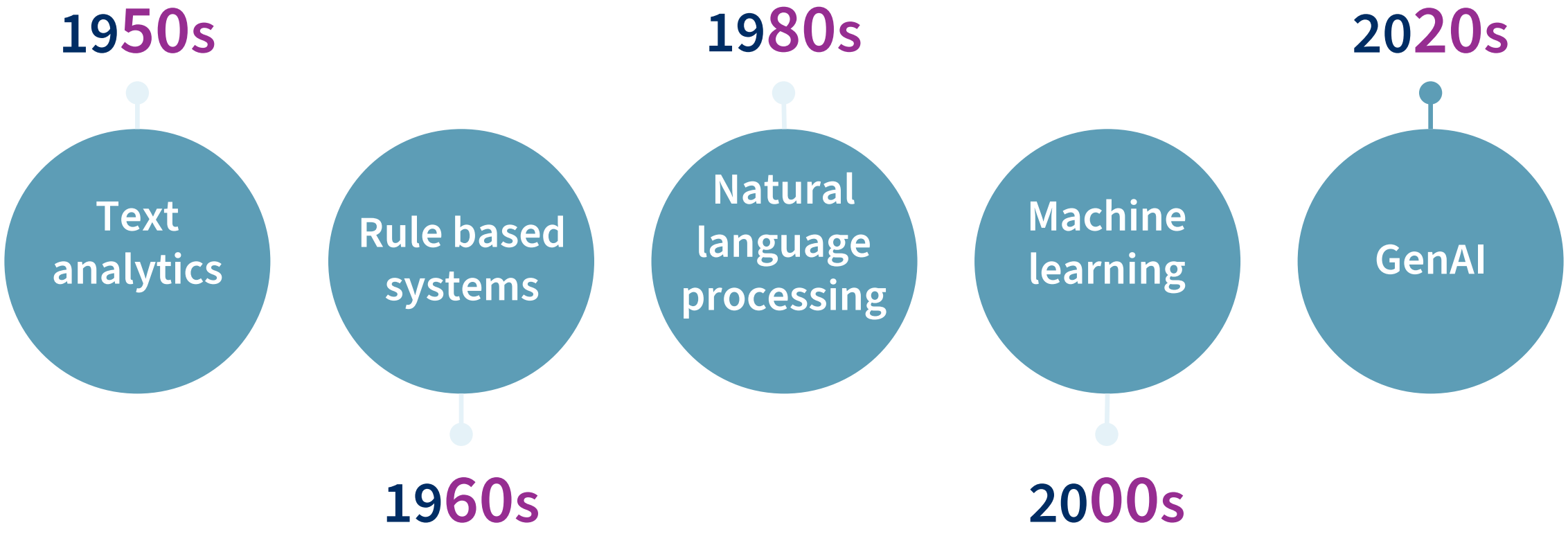
1980s

Natural language processing

2000s

Machine learning

Artificial Intelligence

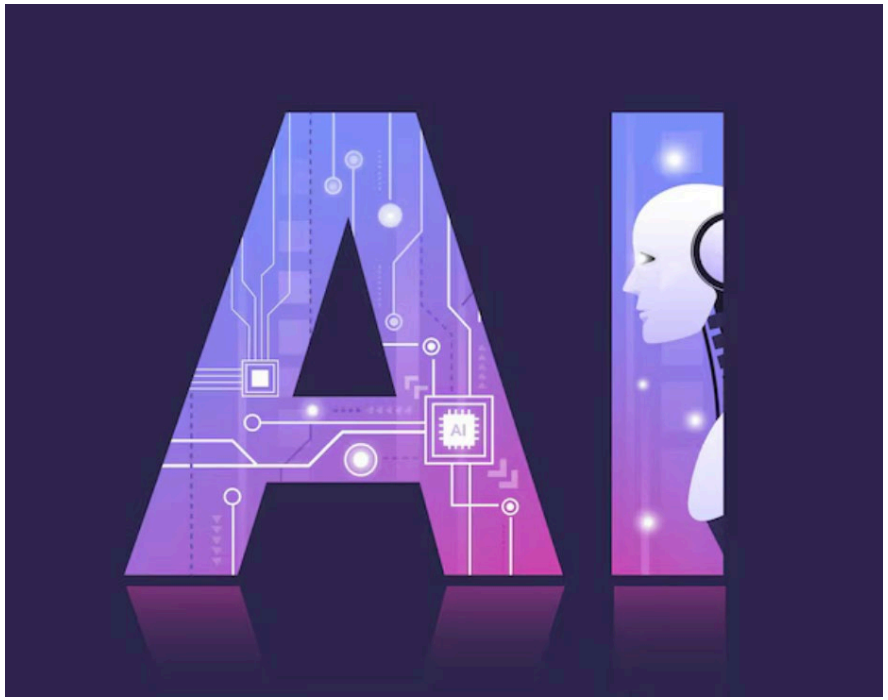


Large Language Models offer new potential

What's so special about LLMs?

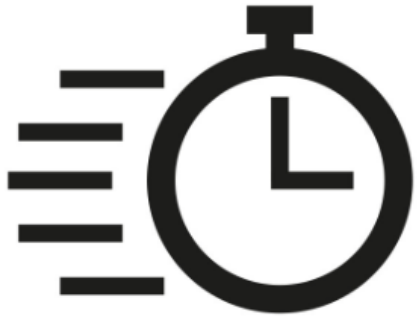
Generative AI uses Large Language Models (LLMs) which have been trained on vast datasets.

LLMs are able to recognize, summarize, translate, predict, and generate text without any specific task training or only a few instructions as a form of prompts

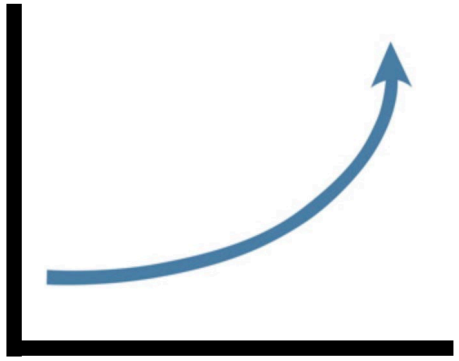


Plenty of excitement about AI

Increases capacity and brings new capability



Speed



Scale

Breadth

LLMs have dramatically opened-up the potential task space both in terms of capacity and capability

Plenty of excitement about AI

Increases capacity and brings new capability

Accuracy

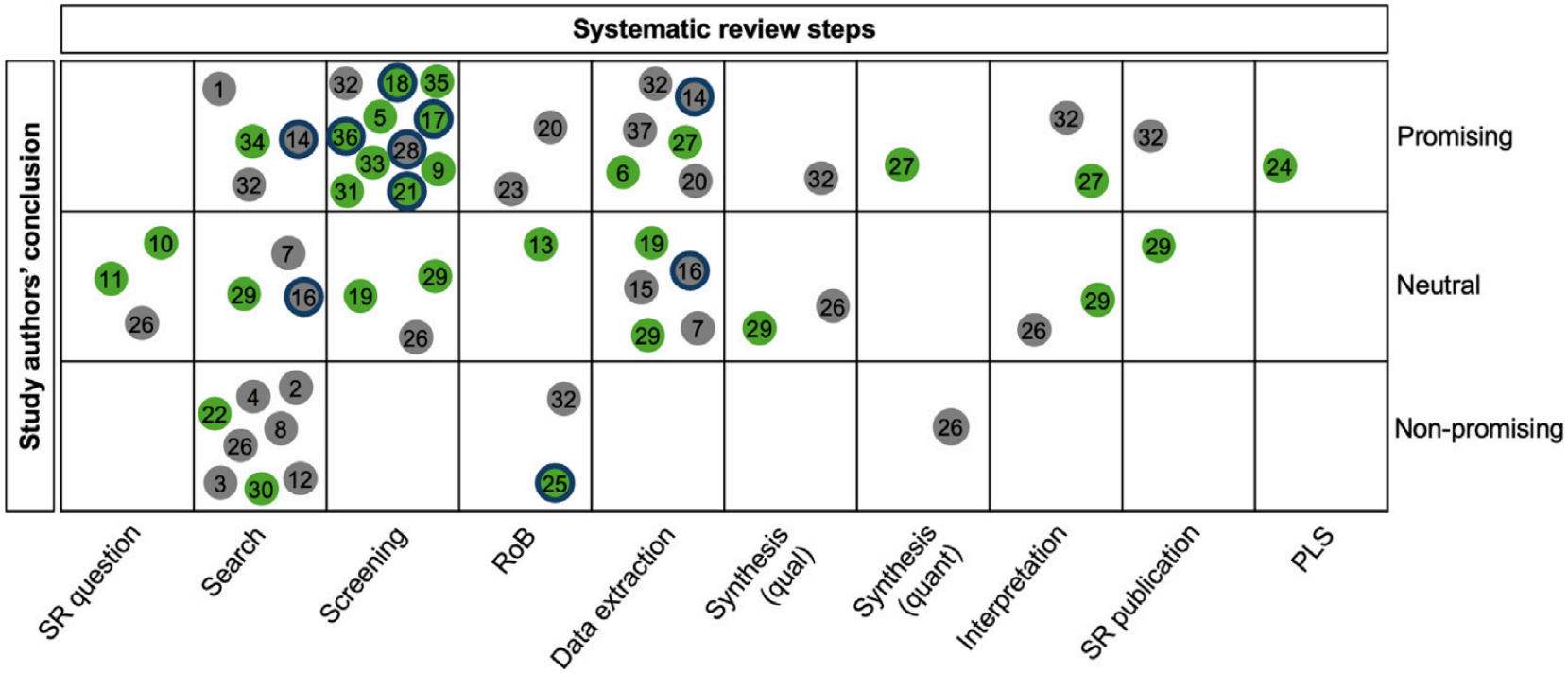
Gartlehner et al. performed a head-to-head comparison: two humans extracting data vs AI (Claude)+human.

“The AI-assisted approach had fewer incorrect extractions (9.0% vs. 11.0%) and similar risks of major errors (2.5% vs. 2.7%) compared to the traditional human-only method, with a median time saving of 41 minutes per study”

Lieberum JL, Töws M, Metzendorf MI, Heilmeyer F, Siemens W, Haverkamp C, Böhringer D, Meerpohl JJ, Eisele-Metzger A. Large language models for conducting systematic reviews: on the rise, but not yet ready for use-a scoping review. J Clin Epidemiol. 2025 Feb 26;181:111746

Large Language Models offer new potential

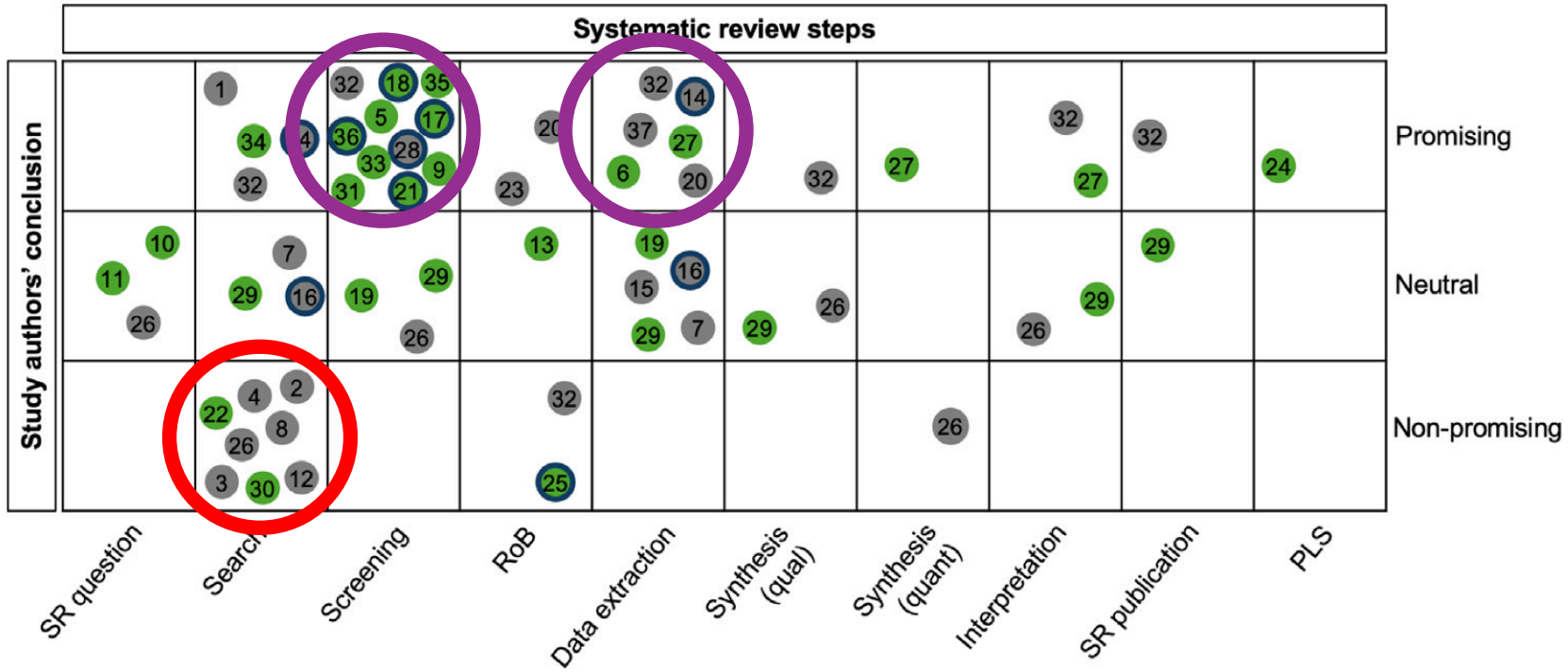
Methodological studies are emerging and have covered



Lieberum 2025.

Large Language Models offer new potential

Methodological studies are emerging and have covered



But… (there's always a but!)

Challenges that might impact the utility of LLMs in evidence synthesis

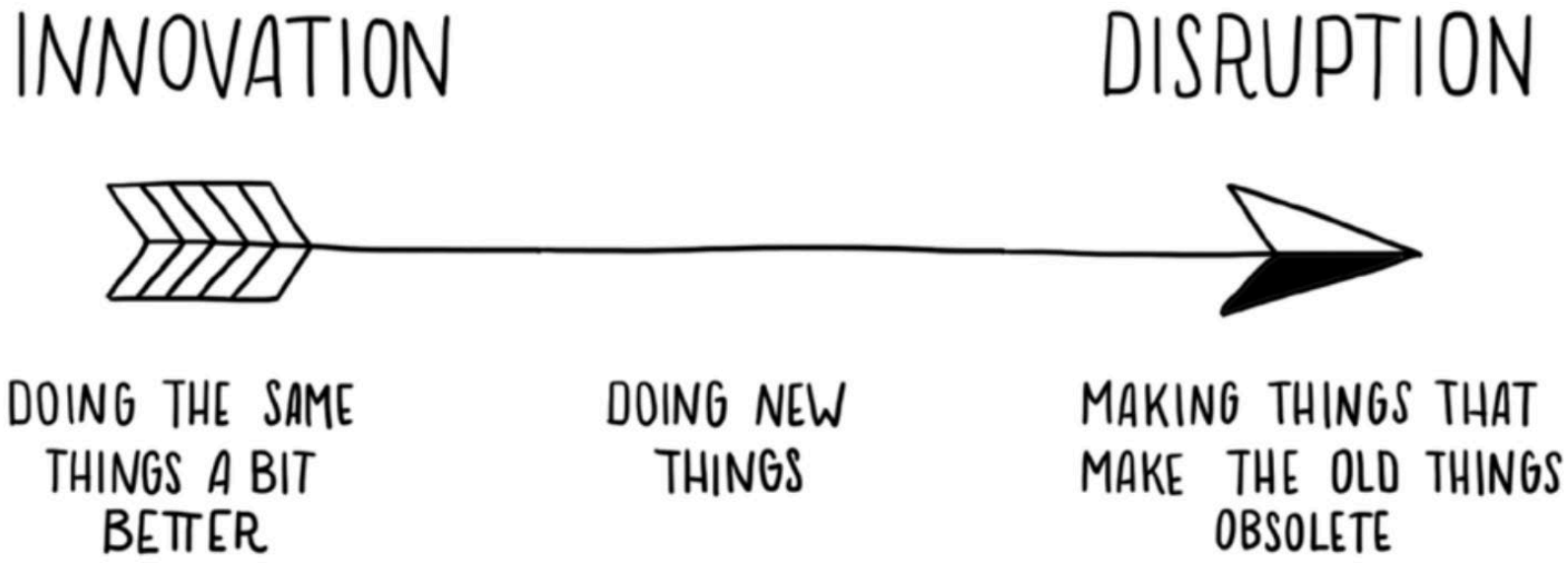
- Over-fitting
- Algorithmic bias
- Black box predictions
- Hallucination

Barriers to effective implementation

Some significant non-technical challenges

- AI literacy
- Access
- Ethical
- Fast pace of change

From innovation to disruption



Cochrane

Cross-organization collaboration and infrastructure

We have a responsibility to use AI, but only if we can use it safely and responsibly.

We'll be more effective at understanding and overcoming the challenges together: both at a methodological level and at a data level.

To do this we need guidance, frameworks and infrastructure that enables us to learn and evolve as the technology itself learns and evolves.