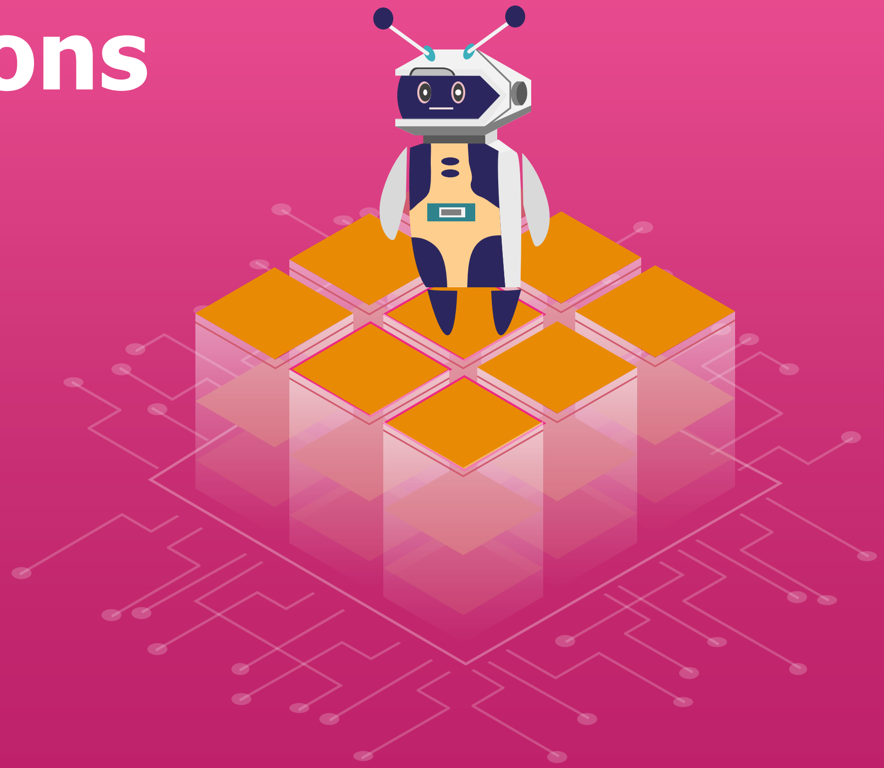
Cytel

Finally It's Out! What the NICE Position on the AI Use in Evidence Generation and Synthesis Means for HTA Submissions



The Do's and Don'ts

20th November, 2024

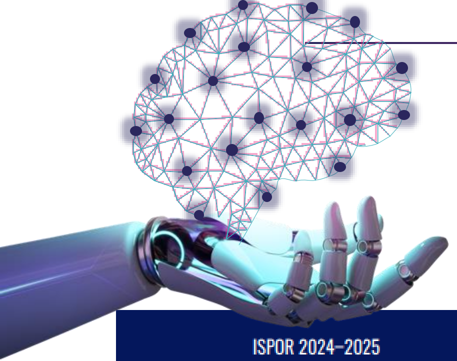
Who we are

Grammati Sarri, PhD
Executive Research Principal

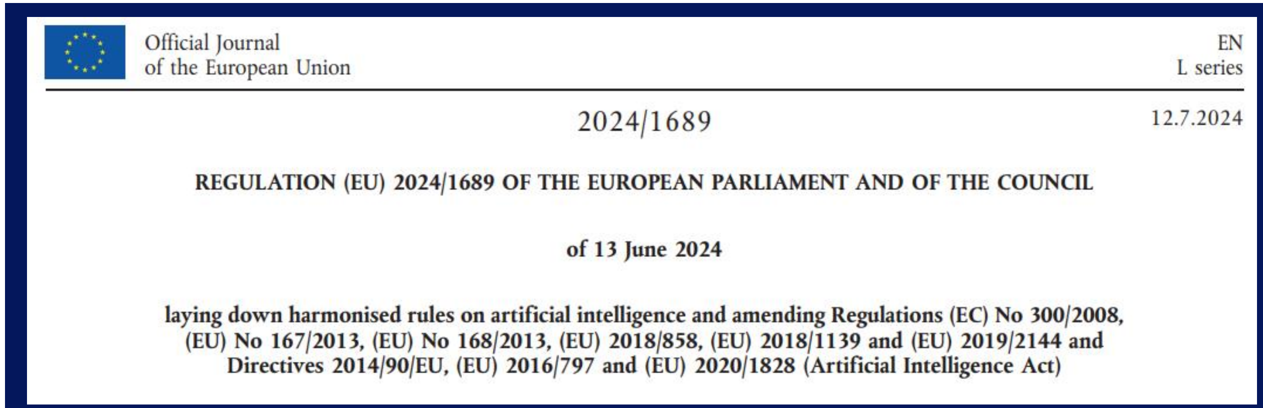
Lydia Vinals, PhD
Associate Director & Research Principal, Evidence Review and Software

Sophie van Beekhuizen, MSc, Mphil
Senior Director, EVA Health Economics

Grace Hsu, MSc
Director, Real World Evidence



Generative AI is transforming the healthcare landscape including health technology submissions and assessments

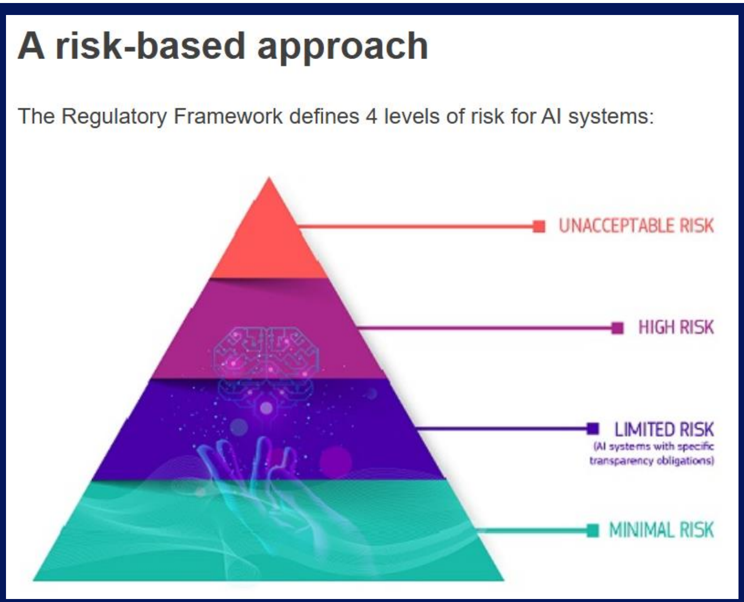

Official Journal of the European Union — EN L series

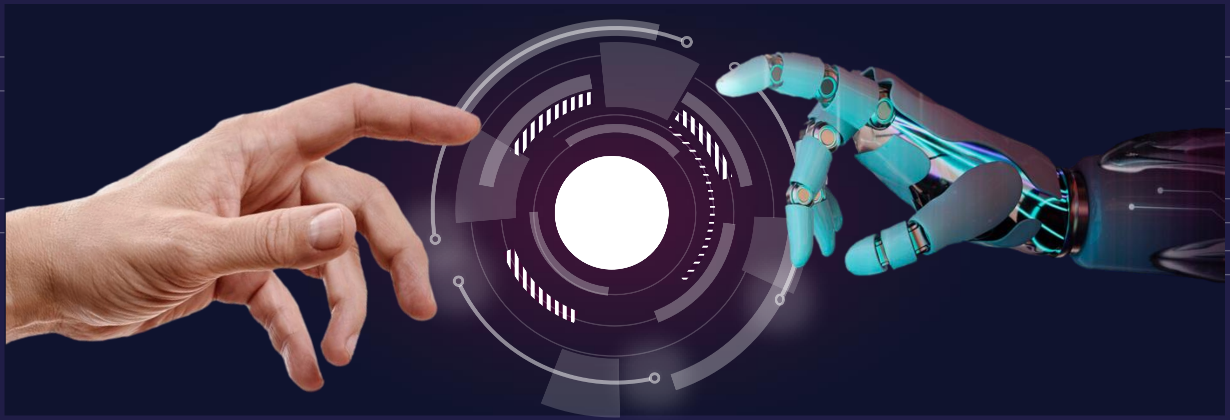
2024/1689 — 12.7.2024

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 13 June 2024

laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)





Use of AI in evidence generation NICE position statement

NICE National Institute for Health and Care Excellence

Cytel

Theatre's Aim

- To introduce the main points in the NICE AI Statement
- To present specific opportunities and challenges in translating the NICE AI position for HTA preparatory activities (systematic literature reviews, comparative clinical and cost-effectiveness analysis, RWE data analytics)
- To present a roadmap when using AI tools in NICE technology submissions

This session is not about AI technical properties, building capabilities etc…

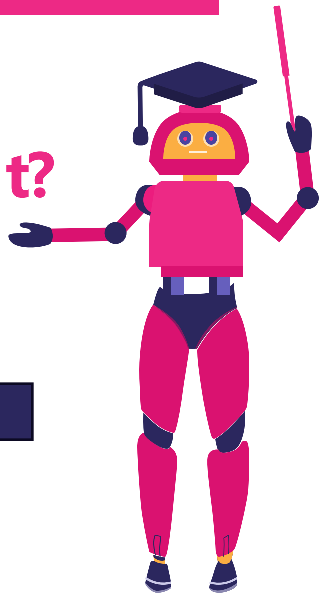
What’s In and What’s Out in NICE AI Statement

What’s in?

- Outline what NICE expects when AI methods are considered or used for evidence generation and reporting
- Indicate existing regulations, good practices, standards and guidelines to follow when using AI methods, where appropriate
- Support our committee members and external assessment groups to understand and critique the potential uses of AI methods

What’s out?

- Evaluation criteria of AI-enabled health technologies
- Detailed methodologies, process considerations related to AI; rapidly evolving field

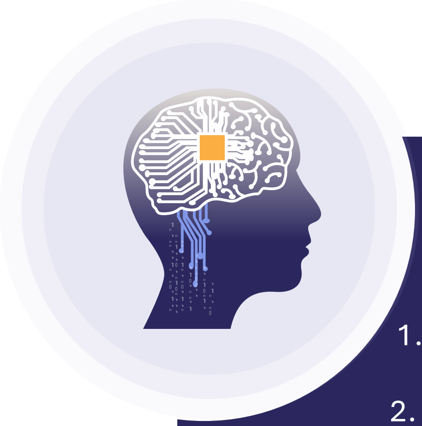


Definitions

Deep learning: A subset of machine learning that uses artificial neural networks for complex learning tasks, such as recognising patterns in data and providing an output (for example, a prediction).
Generative AI: An AI model that generates data, such as text, in response to user prompts.
Large language models: A type of model that is trained on vast amounts of text to understand and generate human speech and tex, and infer new content.
Machine learning: A type of AI that allows a system to learn and improve from examples without all its instructions being explicitly programmed. They learn by finding patterns in training datasets and translating those findings into a model (or algorithm).

Systematic review and evidence synthesis

AI opportunities for systematic reviews



NICE National Institute for Health and Care Excellence

'AI methods have the potential to automate various steps in these [literature search and review] processes'

Machine learning methods and large language models may be able to...

1. support evidence identification by generating search strategies, automating the classification of studies (for example, by study design), the primary and full-text screening of records to identify eligible studies, and the visualisation of search results
2. automate data extraction from published quantitative and qualitative studies by inputting prompts into the AI tool to generate the preferred output
3. Generate the code required to synthesise extracted data in the form of a (network) meta-analysis (less well established)

Opportunities

- Substantially reduce the time and resources required for SLRs by automating steps in the process especially
  - Screening papers for relevant data
  - Extraction of relevant data
- Potential to increase rigor and reduce risk of human error; LLMs + human reviewer can perform better than 2 human reviewers
- Potential to reduce the time it takes to assess novel treatments so that patient access is expedited

Challenges

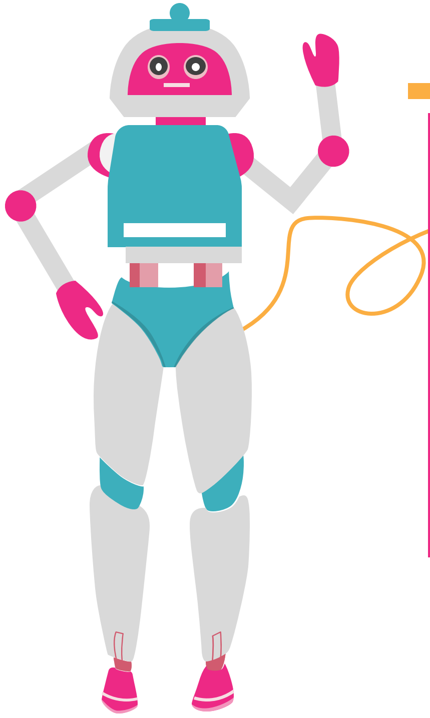
- Not all LLMs are equal – selecting an appropriately trained LLM is of prime importance. E.g. ChatGPT is a broad model, LiveSTART specific to medical topics.
- Effectiveness of AI for screening and data extraction is likely to vary by disease area – less certainty around data extraction than screening
- Lack of clear guidance/precedent – potential for reduced rigor and transparency if not managed well.
- Setting the precedent will be important to understand the EAGs assessment of methods using AI

NICE National Institute for Health and Care Excellence

NICE process and methods

Published: 31 October 2014; last updated: 29 May 2024

Summary: Justification of methods (large language model used) in the same way all methods should be described and justified.



Key points:
- Pay attention to performance characteristics of the model
- Model used should be 'trained' on similar data to data being used for (not GPT!)

Use of machine learning-based classifiers

Machine learning-based classification software has been developed for some study types (for example the Cochrane RCT classifier, Thomas et al. 2020). These classifiers apply a probability weighting to each bibliographical reference within a set of search results. The weighting relates to the reference's likelihood to be a particular study type, based on a model created from analysis of known, relevant papers. The weightings can then be used to either order references for screening or be used with a fixed cut-off value to divide a list of references into those more likely to be included, and those that can be excluded without manual screening.

We support the use of machine classifiers if their performance characteristics are known, and if they improve efficiency in the search and screening process. However, caution is needed when using classifiers, because they may not be as effective if used on data that is different to the type of data for which they were originally developed. For example, the Cochrane RCT classifier is reported to have over 99% recall for health studies but showed "unacceptably low" recall for educational research (Stansfield et al. 2022).

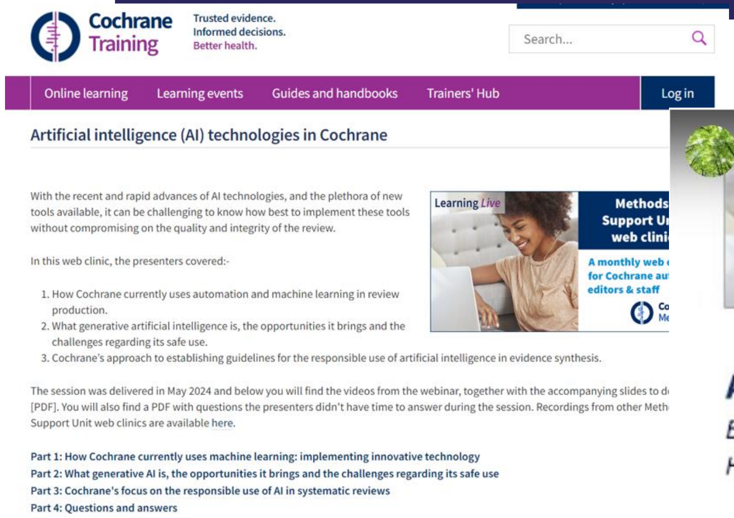
Priority screening, a type of machine classifier that orders references for manual sifting based on previous sifting decisions, is considered in the chapter on reviewing evidence.

Cytel

NICE National Institute for Health and Care Excellence

NICE will be guided by upcoming guidance from Cochrane and the guidelines working group

“ We are aware that Cochrane is developing guidance on the responsible use of AI in evidence synthesis (Cochrane 2024), and the Guidelines International Network has established a working group that will produce guidance and resources (GIN 2024). These are likely to be useful sources of good practices for submitting organisations seeking to use such methods.’


Cochrane Training Trusted evidence. Informed decisions. Better health.

Search...

Online learning | Learning events | Guides and handbooks | Trainers' Hub | Log in

Artificial intelligence (AI) technologies in Cochrane

With the recent and rapid advances of AI technologies, and the plethora of new tools available, it can be challenging to know how best to implement these tools without compromising on the quality and integrity of the review.

In this web clinic, the presenters covered:-

1. How Cochrane currently uses automation and machine learning in review production.
2. What generative artificial intelligence is, the opportunities it brings and the challenges regarding its safe use.
3. Cochrane’s approach to establishing guidelines for the responsible use of artificial intelligence in evidence synthesis.

The session was delivered in May 2024 and below you will find the videos from the webinar, together with the accompanying slides to d [PDF]. You will also find a PDF with questions the presenters didn't have time to answer during the session. Recordings from other Meth Support Unit web clinics are available here.

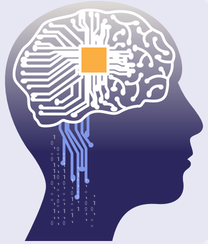
Part 1: How Cochrane currently uses machine learning: implementing innovative technology
Part 2: What generative AI is, the opportunities it brings and the challenges regarding its safe use
Part 3: Cochrane's focus on the responsible use of AI in systematic reviews
Part 4: Questions and answers

Generative AI for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations

Rachael L. Fleurence, PhD, MSc[1], Jiang Bian, PhD[2 3 4], Xiaoyan Wang, PhD[5 6], Hua Xu, PhD[7], Dalia Dawoud, PhD[8 9], Mitch Higashi, PhD[10], Jagpreet Chhatwal, PhD[11 12]

Cost-effectiveness evidence

AI in health economic modelling



NICE National Institute for Health and Care Excellence

- Health economic modelling is a resource-intensive multi-step process - **AI can support in these steps**
- AI may automate construction of economic models after **human-led conceptualization** & **parameter estimation**
- LLM can support the replication and cross-validation of existing economic models
- More complex models may be more feasible through **AI model optimisation** and **increased efficiency**

NICE is supportive of the complementary role of AI in health economics

Opportunities

- **All-round support:** conceptualization through evidence review and taking learnings from previous developed models; reduce burden of cumbersome economic modelling tasks; identifying model drivers and opportunities for model optimization
- **Efficiency**: More time to spend on finding novel methodological solutions
- **Transparency**: AI may help navigating through complex models

Challenges

- **AI hallucinations** are a great risk: you will need an expert to know it's wrong. Prompting engineering is crucial. Developing automated prompts may help but requires expertise.
- **No need to understand** the 'how and why' if AI automatically provides a solution with confidence.
- No **standard methods/guidelines** for using AI, reproducibility in question. Reinforcing bias.

Real-world data and analysis

AI in real-world data and analysis

NICE National Institute for Health and Care Excellence

- Supports using AI to ***enhance trial design and RWD analysis***
- Acknowledges AI's role in ***addressing data complexity***
- Values AI's role in ***enhancing causal inference***
- Requires ***transparency*** in AI applications, emphasising comprehensive reporting standards to ensure AI-driven results are reliable, reproducible, and thoroughly documented

Opportunities

- **Enhancing trials**: AI may help improve sample selection, eligibility criteria, etc., thus making evidence more applicable to target populations
- **Addressing data complexity**: Supporting structured and unstructured information using AI can improve data quality, standardisation, and integration
- **Causal inference and bias reduction**: Can potentially reduce bias and improving treatment effect estimation through models capable of handling non-linear relationships among variables
- **Broaden evidence sources**: By having AI support generation of synthetic control arms and creating ethical, robust real-world evidence, this broadens the range of evidence supporting decision-making

Challenges

- **Validating the AI model** used for identifying eligible patients or other trial settings—how can we avoid potential selection bias?
- Integrating multimodal data (e.g., EHRs, imaging, and genomics) can be challenging and **computationally intensive**—if custom AI models are required, when will they be worth the effort of building and validation?
- **Transparent reporting** must follow standards like PALISADE and TRIPOD+AI—are these standards flexible enough for most applications?
- How can AI models with few structural assumptions help with **causal interpretation**? To what extent and how will clinical expertise be included?

Roadmap

Do’s and Don’ts in AI use for NICE TA submissions

Do’s

- **Early engagement** with NICE
- Only use AI tools when with a **demonstrable value**; robust justification , use dedicated checklists
- Follow UK governmental and regulatory **AI standards** (MHRA 2024, EU 2024, EMA 2023)
- **Fit for purpose** assessment (technical and external validation, pre-specified outcome-blind simulations, licenses and copyrights)

Don’ts

- Do not hide risks identified through the AI use (e.g. concerns about transparency, biases)
- Do not do data manipulation or prompt injection attacks
- Do not downplay security concerns
- Do not use highly technical language to describe model’s components and results

Thank you