Pharmacoevidence

Improving patient lives by supporting evidence base for healthcare interventions

Great Place To Work. Certified

Gen AI in Systematic Literature Reviews: First Case Study on the use of Gen AI in a NICE Submission

ISPOR EU 2025 | Glasgow

Moderator
- Barinder Singh, Pharmacoevidence

Speakers:
- Sven Klijn, Bristol Myers Squibb
- Dilip Makhija, Gilead Sciences Inc.
- Raphael Sonabend-Friend, National Institute for Health and Care Excellence, UK

Pharmaco Evidence

Mohali India| London UK | Canada

Panelist and Session Outline

MODERATOR

Barinder Singh
Pharmacoevidence

Introduction

PANELISTS

Sven Klijn
Bristol Myers Squibb

AI-augmented evidence generation

Dilip Makhija
Gilead Sciences Inc.

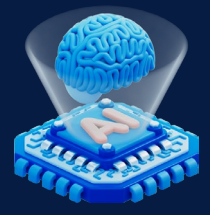
Real-world experience with automated SLRs

Dr. Raphael Sonabend-Friend
NICE, UK

HTA perspective on AI-driven evidence synthesis

SLRs lie at the top of the evidence hierarchy

Evidence Pharmaco®



What

- A structured and transparent process to identify, assess, and synthesize existing evidence on a specific research question
- Follows a transparent and replicable process
- Key features:
  - Predefined research question
  - Inclusion/exclusion criteria
  - Quality assessment of studies
  - Clear synthesis and reporting

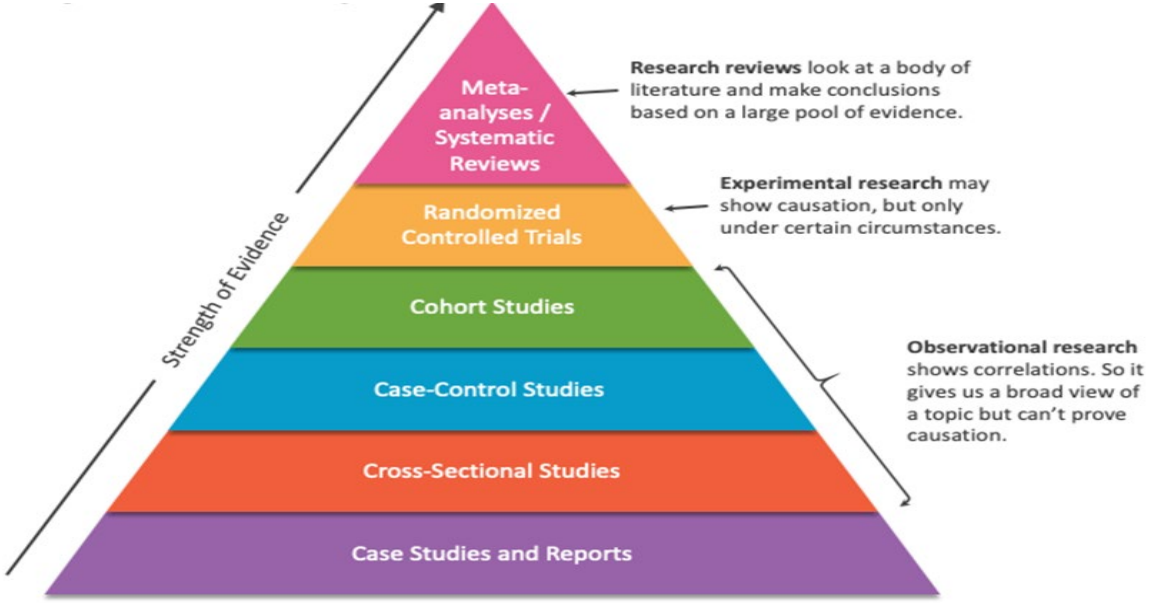
Why

- To provide a **comprehensive** and **unbiased** summary of available data to guide **clinical, regulatory, or research decision-making**
- Compare treatment efficacy and safety across multiple studies
- Synthesize high-quality data to **strengthen scientific conclusions**
- Facilitate meta-analyses by providing **structured data for statistical pooling**

How

- Define a focused research question using structured frameworks (e.g., PICO)
- Develop a review protocol with clear inclusion/exclusion criteria
- Search and screen literature across multiple databases
- Assess study quality using standardized tools
- Extract and synthesize data through narrative or statistical analysis

SLRs lie at the top of the evidence hierarchy



SLRs are The Evidence Foundation for HTA

Health Technology Assessment (HTA) relies on SLRs to ensure that all clinical and economic evidence is identified, appraised, and synthesized transparently and reproducibly

- HTA agencies (e.g., NICE, CADTH, EUnetHTA/JCA) require SLR-based evidence to support comparative effectiveness and cost-effectiveness analyses
- SLRs provide a transparent, auditable trail of how evidence was sourced and selected
- They minimize bias and duplication, ensuring that payer and regulatory decisions are grounded in complete and reliable evidence
- In the upcoming EU Joint Clinical Assessment (JCA), SLRs form the core evidence package for all new therapies

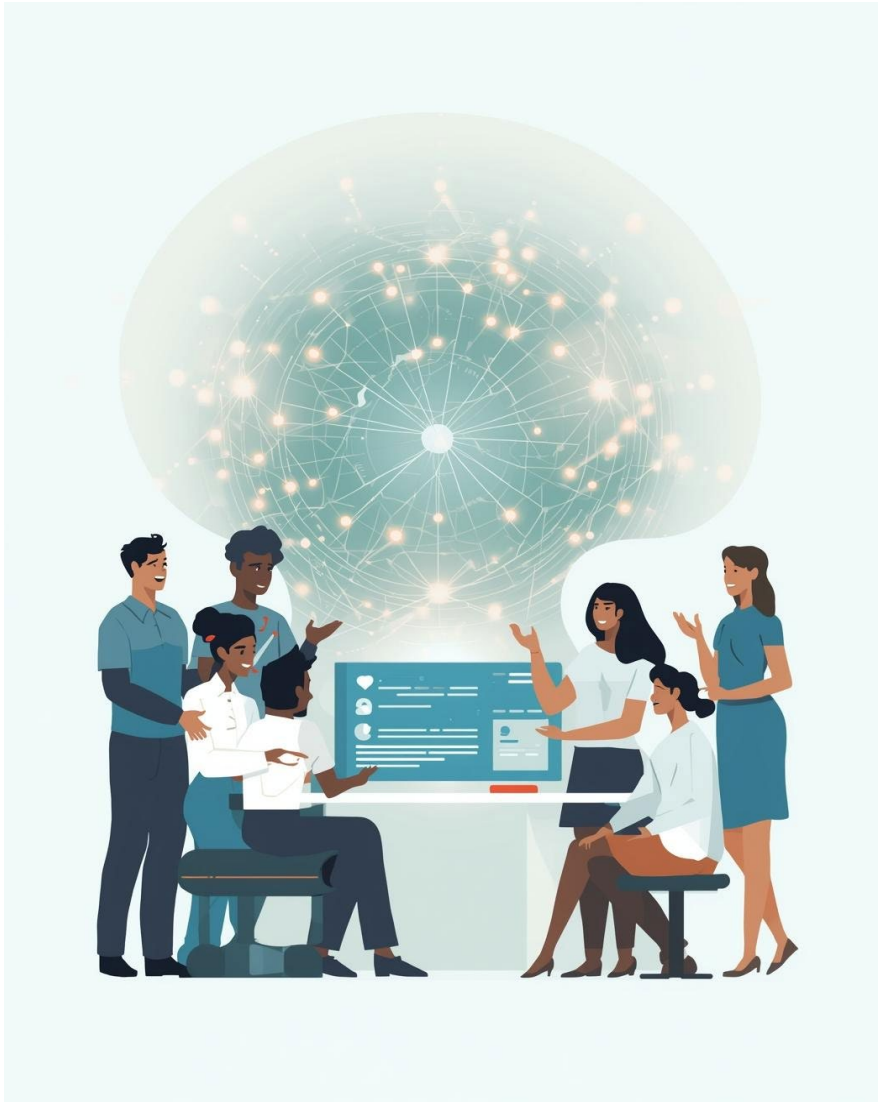
More Data. More Complexity.

- Healthcare generates nearly 30% of the world's data[1]
- The volume of healthcare data is projected to grow at a 36% Compound Annual Growth Rate (CAGR) by 2025 — faster than any other industry (Arcadia-2023) [1,2]

- “Is it possible to deliver *faster*, *scalable* outputs without compromising on methodological rigor?”

- "Can generative AI truly be trusted to conduct systematic literature reviews for HTA submissions?“

Growing confidence in GenAI across HTA bodies



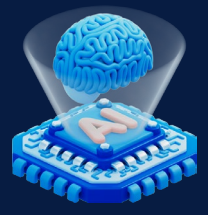
- HTA bodies now recognize GenAI as an emerging contributor to evidence generation and evaluation
- However, concerns remain around the appropriateness, transparency, and reliability of AI

It is important to consider the use of AI methods carefully to ensure the anticipated benefits are balanced against the known concerns[1]

Use of AI in evidence generation: NICE position statement | Our research work | What we do | About | NICE

Current regulatory landscape

Evidence Pharmaco®



NICE UK position statement

- ✓ Engage early with NICE and inform them of your plans to use AI
- ✓ Apply AI only when it adds demonstrable value
- ✓ Keep human reviewers in the loop; AI assists, never replaces
- ✓ Disclose algorithms, data sources, and validation tools.
- ✓ Perform and document validation, bias, and security tests.
- ✓ Comply with UK AI rules, licenses, and IP requirements

CDA position statement

- ✓ Engage early with CDA-AMC to discuss plans for using AI
- ✓ All uses of AI should align with Canada's Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems
- ✓ Principle of augmentation, not replacement of humans
- ✓ Ensure compliance with licensing agreements, maintain explainability of AI methods, and promote transparency in applications
- ✓ Ensure that all algorithms, models, datasets, and data processing pipelines used are in alignment with relevant Canadian regulations and guidance for AI use
- ✓ Any potential risks should be mitigated by adhering to established guidance and checklists
- ✓ Must adhere to the ethical principles for the use of AI in health

Use of AI in evidence generation: NICE position statement | Our research work | What we do | About | NICE; CDA position statement – Position_Statement_AI_Renumbered.pdf

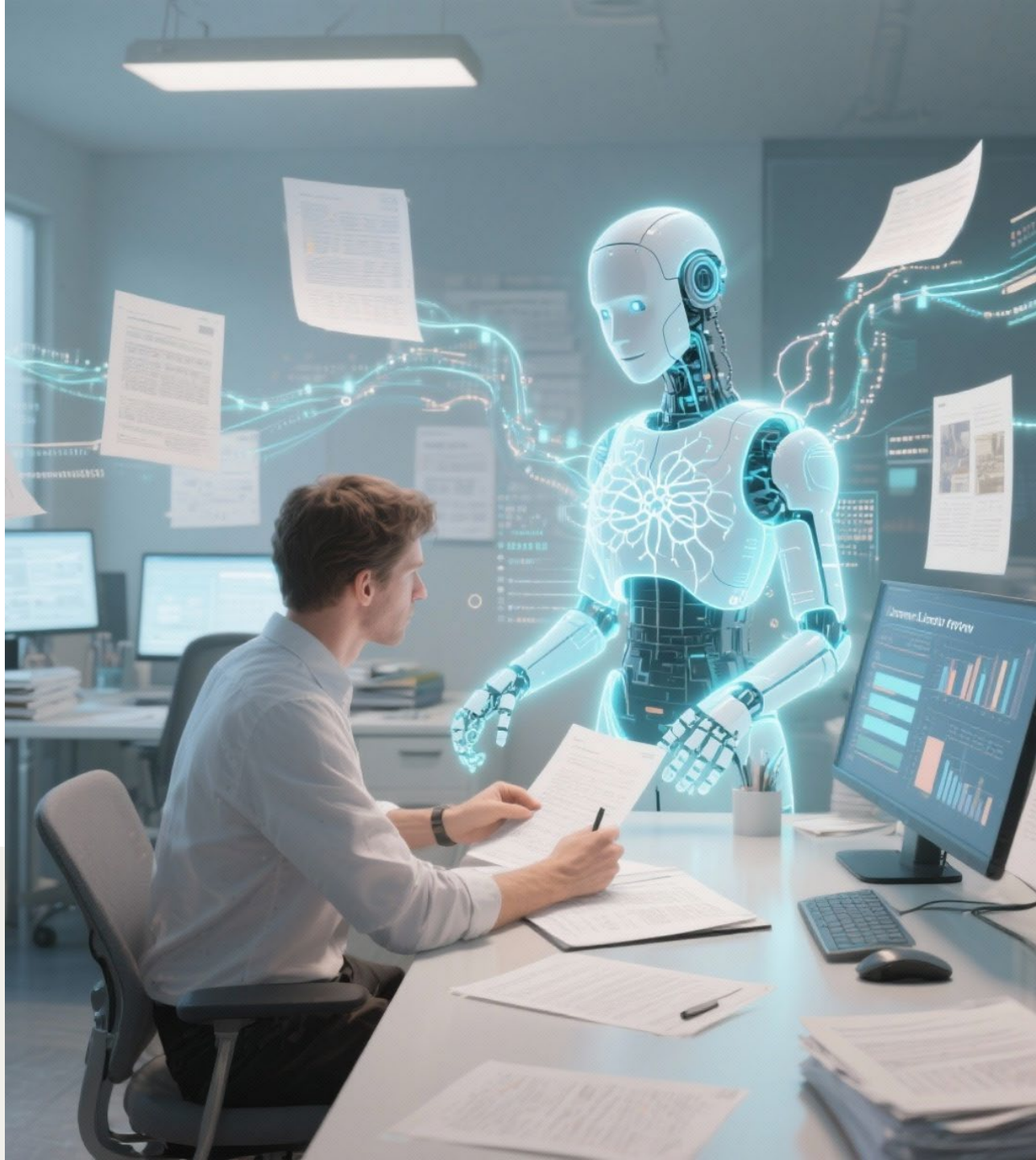
- "Have we reached a point where AI-assisted SLRs could stand up to regulatory scrutiny in real-world HTA submissions?“
- “Are there any real-world examples of Gen AI-assisted SLRs in HTA submissions?”

GenAI in SLRs: From Dual Review to Human–AI Collaboration

November 11, 2025

ISPOR Glasgow

Sven Klijn



The opinions expressed in this presentation and on the following slides are solely those of the presenter and do not necessarily represent opinions of Bristol Myers Squibb.

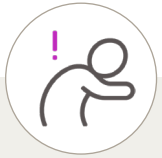
Dual human review

Human 1	Human 2	Human 3
Formulating the research question		
PICO definition		
Search query		
Title/abstract screening	Title/abstract screening	Arbitration
Full text screening	Full text screening	Arbitration
Data extraction	Data extraction	Arbitration
Reporting		

The gold standard is not perfect

Time-consuming

Limited scalability

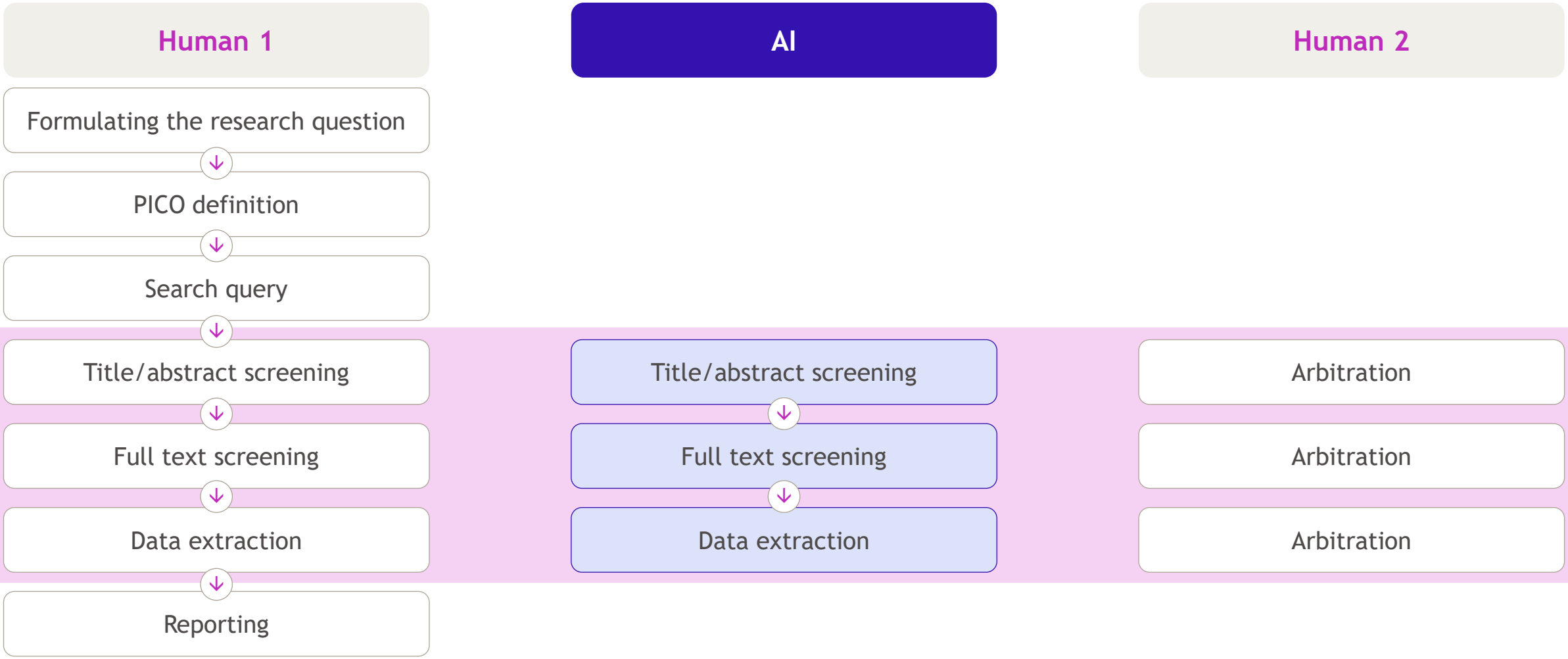


Screening fatigue

“The total error rate
(false inclusion and false exclusion)
was 10.76% (95% CI: 7.43% to 14.09%)”

Wang et al. (2020). Error rates of human reviewers during abstract screening in systematic reviews. DOI: 10.1371/journal.pone.0227742

Human-AI Collaboration



AI Governance

Model Documentation

- Architecture
- Training data
- Performance metrics
- Logging

Life Cycle Maintenance

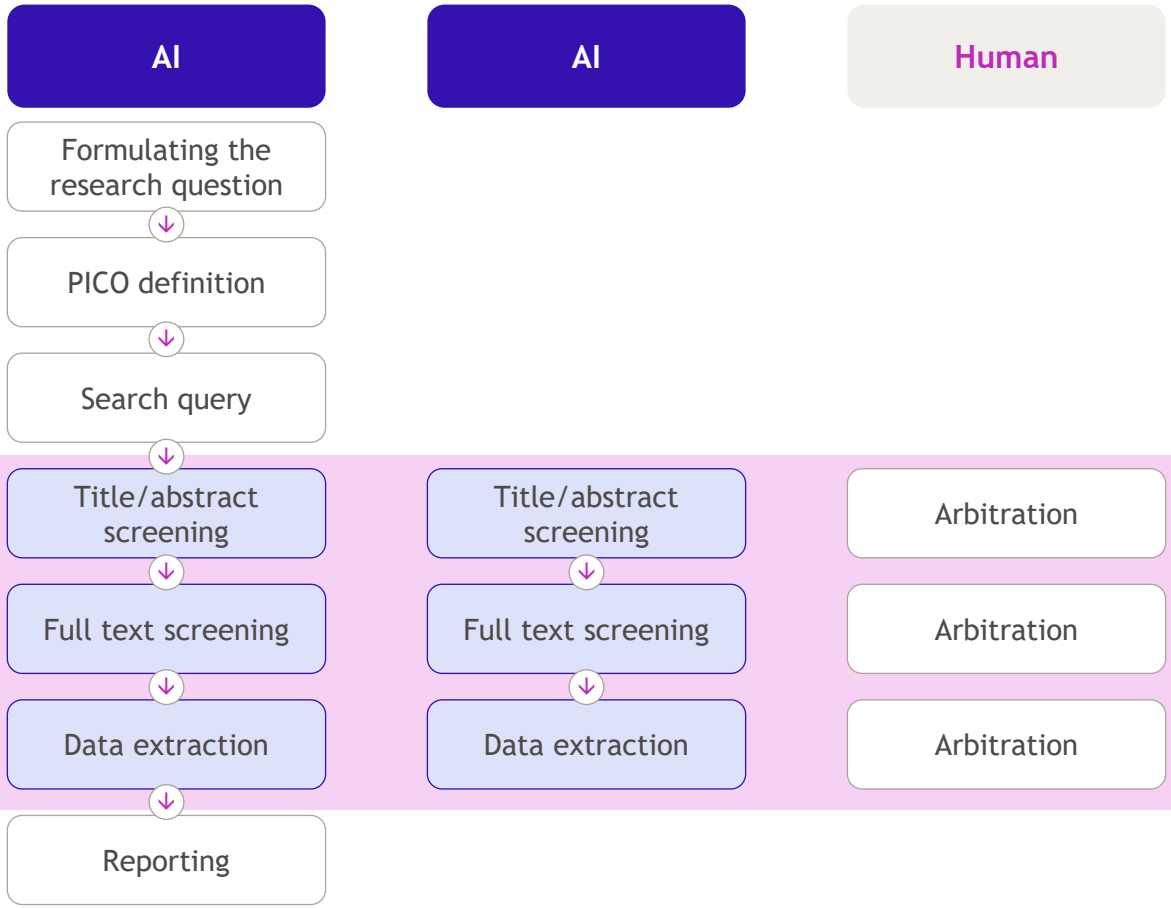
- Version control
- Bias & fairness assessments: Systematic evaluation of outputs
- Reproducibility

Best practice principle 6:
"Ensure GenAI-assisted analyses are reproducible and auditable where possible."

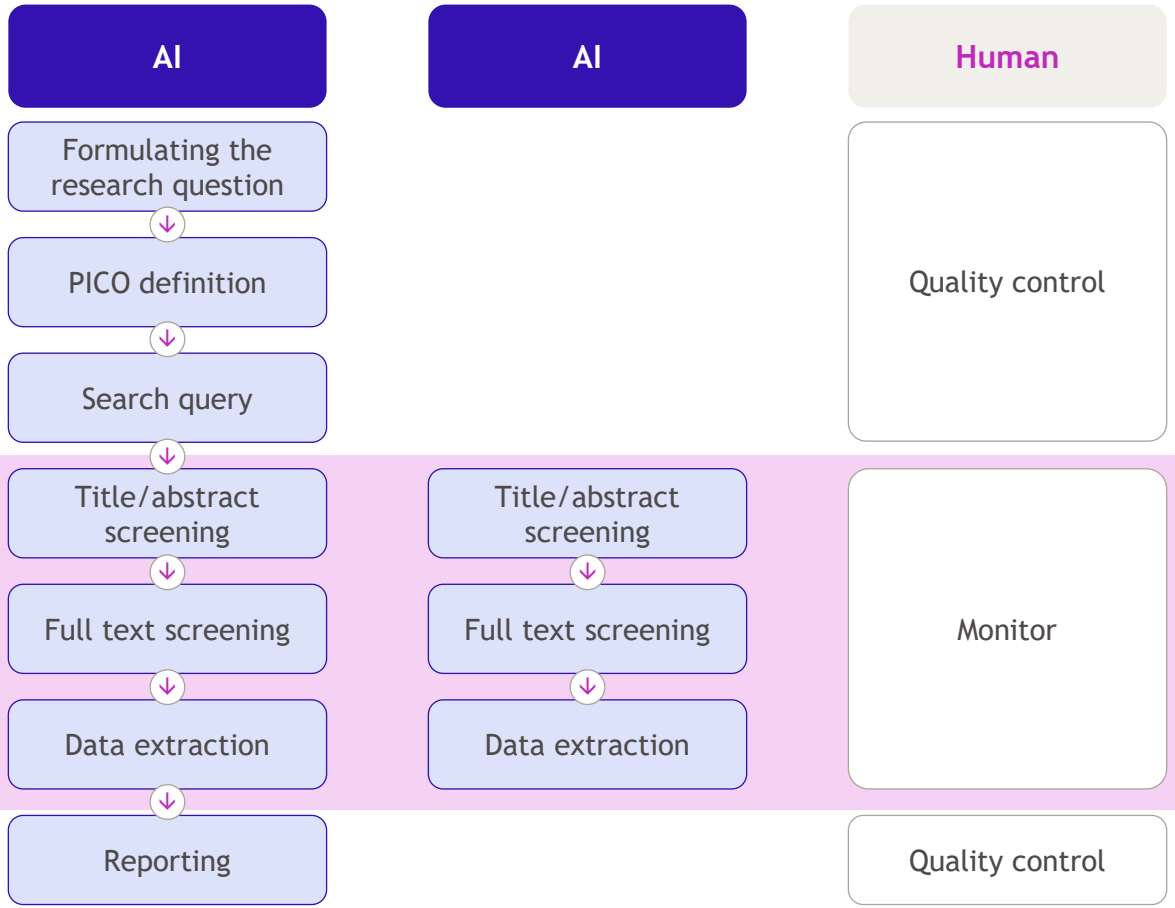
NICE (2025). Generative Artificial Intelligence (GenAI) best practice principles

Alternative models

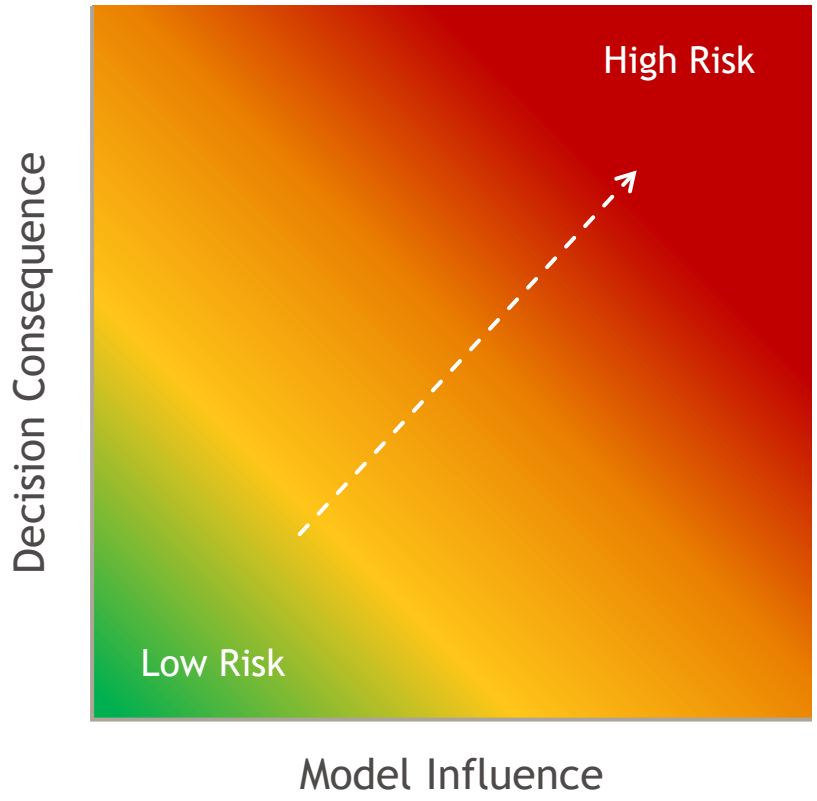
Human Arbitration



Human Monitoring



Model risk matrix



Decision consequence depends on:

- Type of publication
- Type of error made:
  false inclusion vs. false exclusion

Model influence depends on:

- Model used:
  Human-AI / Human Arbitration / Human Monitoring
- Guardrails and mitigation strategies

Image adapted from: FDA (2025). Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products. Guidance for Industry and Other Interested Parties

Reflection

- Discussion required on where the bar should be set
- Strong model governance is critical
- Role of the human is essential, but differs by implementation model

GILEAD
Creating Possible

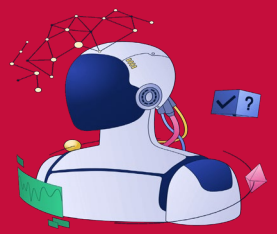
Implementation of AI tools: A real-world example for the use of an automated SLR in HTA submissions

Dilip Makhija| ISPOR EU Glasgow Panel 2025

Disclaimer

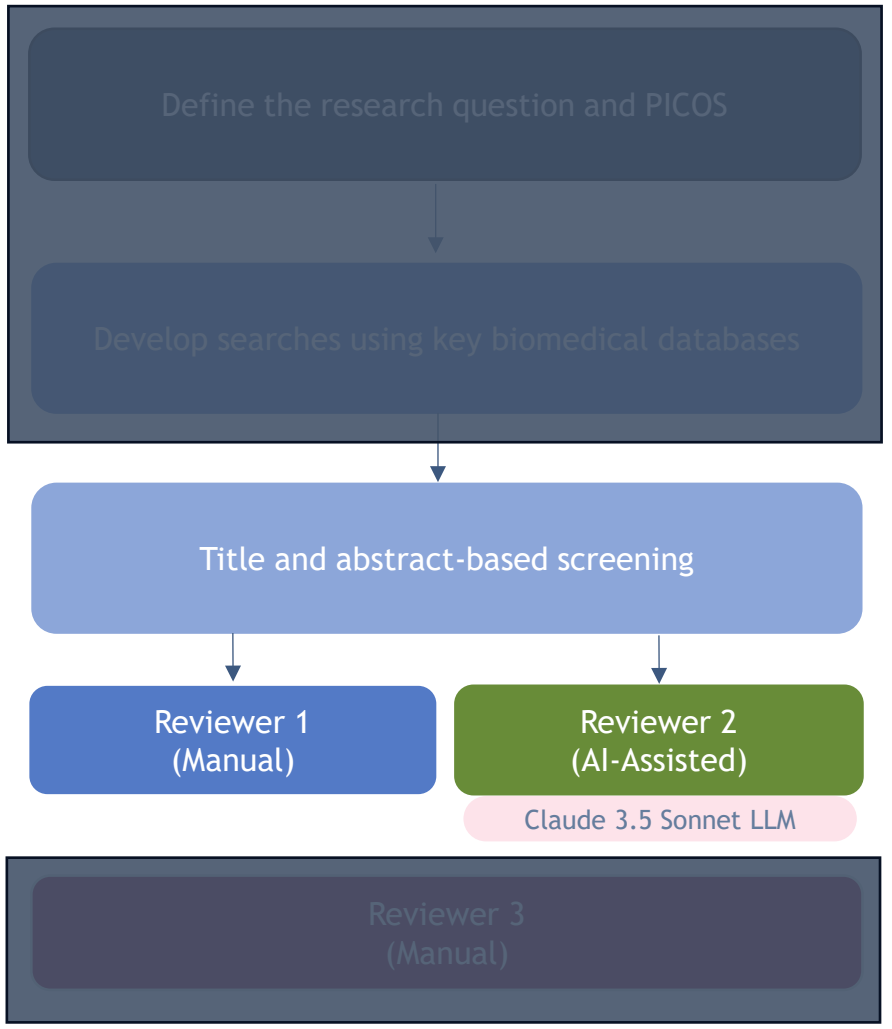
The opinions expressed on the following slides are solely those of the presenter and not those of Gilead Sciences, Inc.

An AI tool was employed as a second reviewer to manage the large volume of citations retrieved from the literature search, helping to adhere to the tight timelines for the UK NICE submission



- AI was implemented as a complementary second reviewer, in alignment with NICE and CDA position statement, and the Cochrane guidance, i.e., human-in-the-loop process

- The approach adhered to NICE methodological standards as outlined in the technical support document (TSDs) on identifying, selecting, and synthesizing clinical evidence with a two-reviewer screening and robust quality control to reduce bias and errors



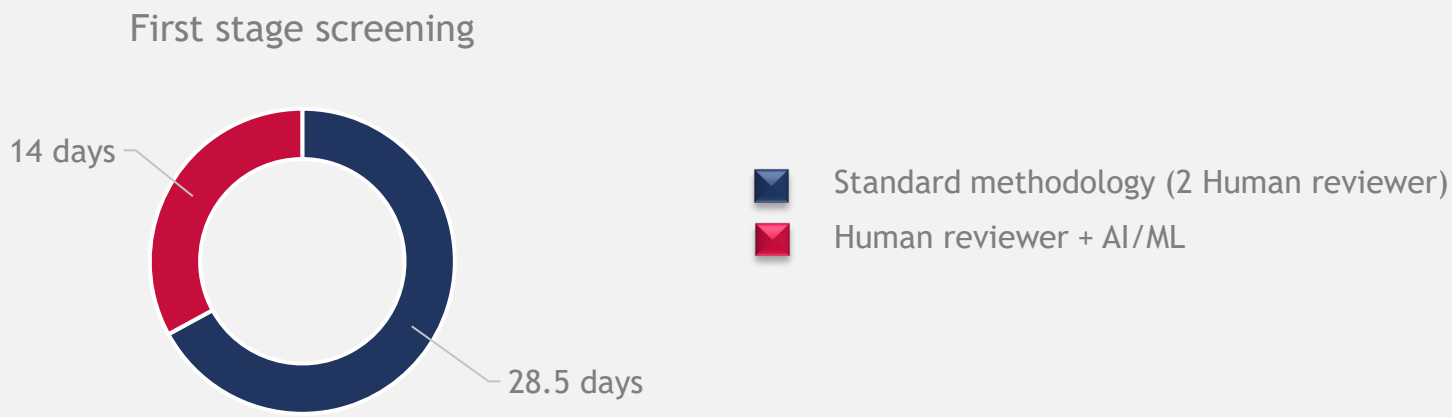
Integration of the AI tool helped meet ambitious timelines and save ~50% time

Rationale and value of using Gen AI for SLR

Published evidence on the effectiveness of the tool

Level of agreement between Human and AI

- ✓ The rationale for integrating the AI tool was driven by the **substantial volume of citations** retrieved during literature searches, **presenting significant timeline challenges for submission**
- ✓ AI-based screening **reduced the time** required for the Title/Abstract-based screening by **~2 weeks**



Over 10 published studies that had previously employed the automated tool were provided to supports its validation and justification for use

Rationale and value of using Gen AI for SLR

Published evidence on the effectiveness of the tool and Early engagement with HTA

Level of agreement between Human and AI

- Early engagement with NICE regarding the use of AI is ideal, but we couldn't do that this time because of tight submission timelines
- We submitted comprehensive evidence supporting the tool's effectiveness, including over 10 published studies demonstrating its accuracy, specificity, sensitivity, and reviewer agreement rates

Human-AI agreement was high, and the AI tool did not exclude any relevant citation

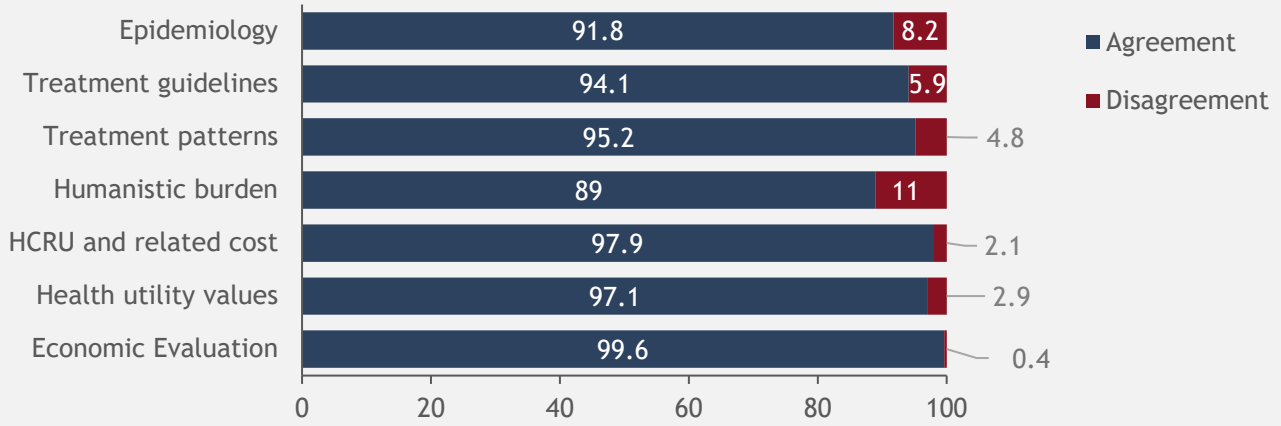
Rationale and value of using Gen AI for SLR

Published evidence on the effectiveness of the tool

Level of agreement between Human and AI

- ✓ In standard practice for SLRs, an **ideal agreement** between two human reviewers is **approximately 95%.** However, **real-world experience** typically demonstrates around or **below 90% agreement**, mainly due to reviewer expertise and judgment variations
- ✓ In our AI-assisted review process, the overall average agreement between AI and human reviewers was **94.97% (median: 95.24%)**, ranging from 89.01% (Humanistic Burden Review) to 99.59% (Economic Evaluation Review)

Decision alignment between human and AI



Did AI exclude any relevant citations? No, it didn't

NICE was in agreement and aligned on the use of the AI-assisted tool

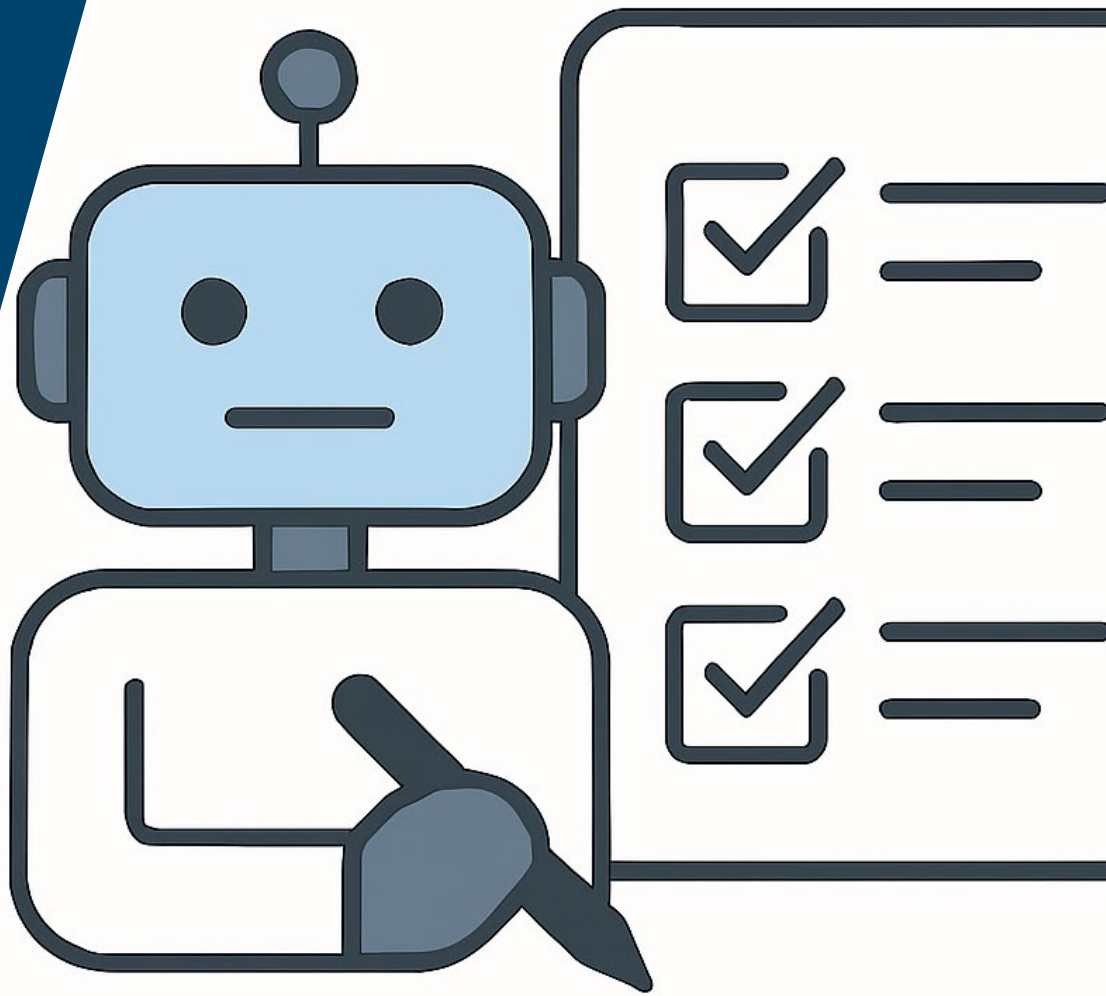
- ✓ The AI-assisted reviewer functioned as a complementary second reviewer
- ✓ Any disagreements between the human and AI reviewers (e.g., one includes, one excludes) were resolved by an independent QC human reviewer
- ✓ To ensure robust quality assurance, all excluded citations, even those where both reviewers agreed, were rechecked by a separate QA reviewer to confirm no relevant studies were missed
- ✓ AI as second reviewer and layered QC/QA framework ensured accuracy, transparency, and methodological rigor, maintaining alignment with NICE's "human-in-the-loop" guidance and confirming the reliability of AI-assisted screening
- ✓ This submission established an important benchmark for the integration of Generative AI in evidence generation and represents the first AI-assisted SLR accepted by NICE (Singh 2025[1])

[1]Singh B, Diaby K, Makhija D, Thakur D, Dubey R, Rai P. Where Is AI in Evidence Synthesis for NICE HTA Submissions? ISPOR Europe 2025, Glasgow; 10 November 2025.

AI in systematic reviews

Raphael Sonabend-Friend

Scientific Adviser



NICE National Institute for Health and Care Excellence

AI-assisted citation screening accepted in a NICE submission [ID6429]

1. Pharmacoevidence AI tool used as a second reviewer for screening citations with an independent human expert checking discrepancies.

2. EAG asked for the rationale behind AI integration and for validation of the AI's effectiveness.

3. "Given the QA processes described by the company, the EAG believed that the use of the AI/ML tool was appropriate". [ID6429]

AI Statement of Intent and Position Statement

Statement of Intent

Intention to develop our approach to: providing guidance on best practice development; evaluation; internal use.

Evaluation

In-house experimentation; living systematic reviews; external consultations (HTA Lab).

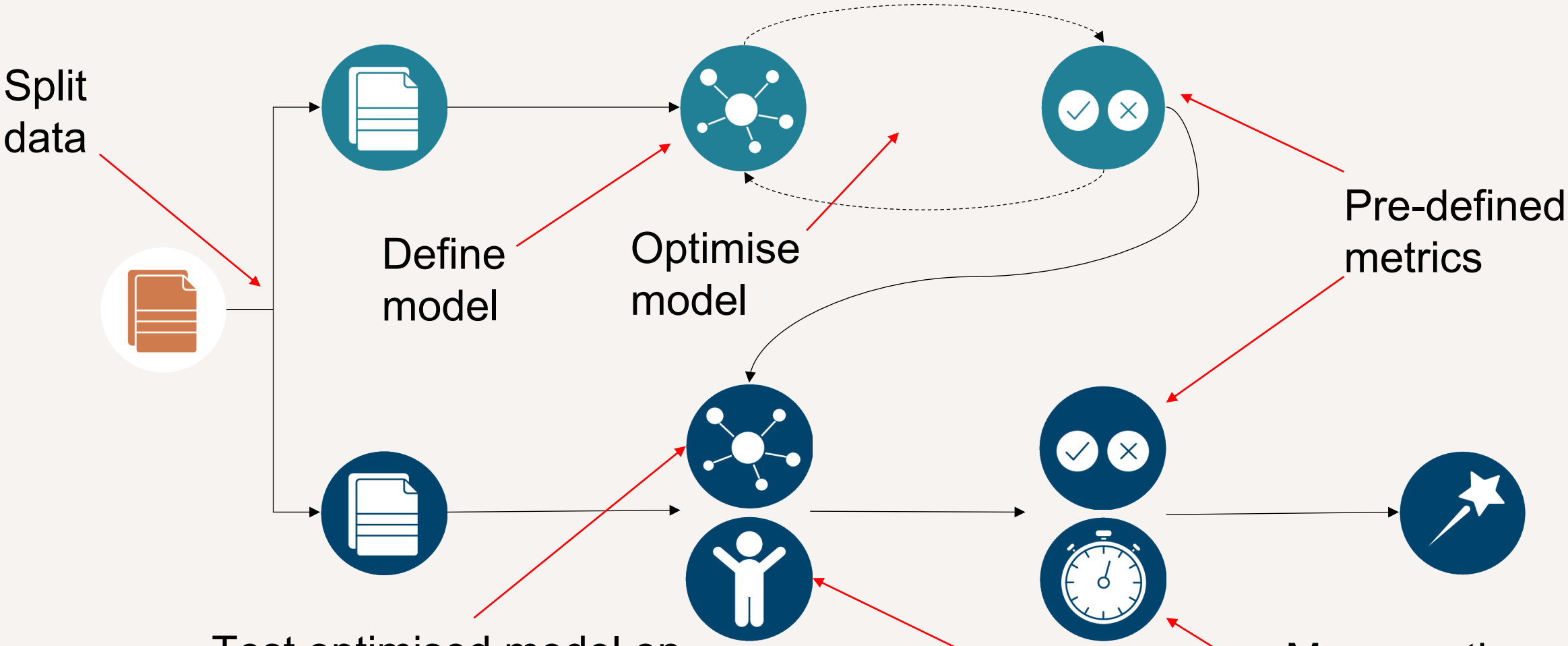
Position Statement

Living document outlining NICE's position on use of AI in generation and reporting of evidence.

AI updates to NICE submission templates

1. Has AI been used in the generation of data or research presented in this submission?
2. Has AI been used in the reporting or synthesis of research presented in this evidence submission?
3. Please indicate where AI was used, any potential risks with its use and what steps were taken to address these risks.
4. Please indicate the level of human oversight for these uses.

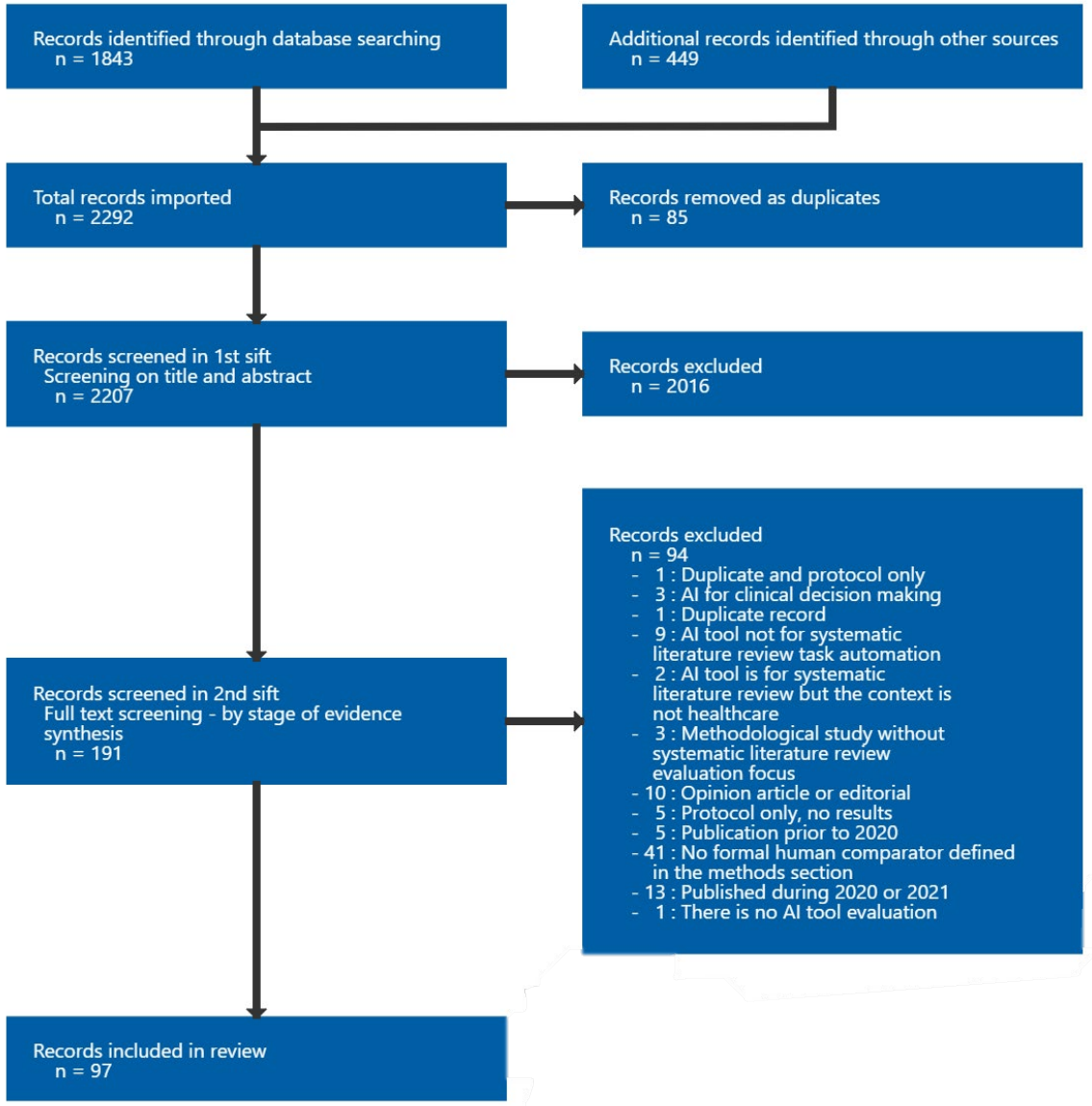
Best practice evaluation



AI in SLR

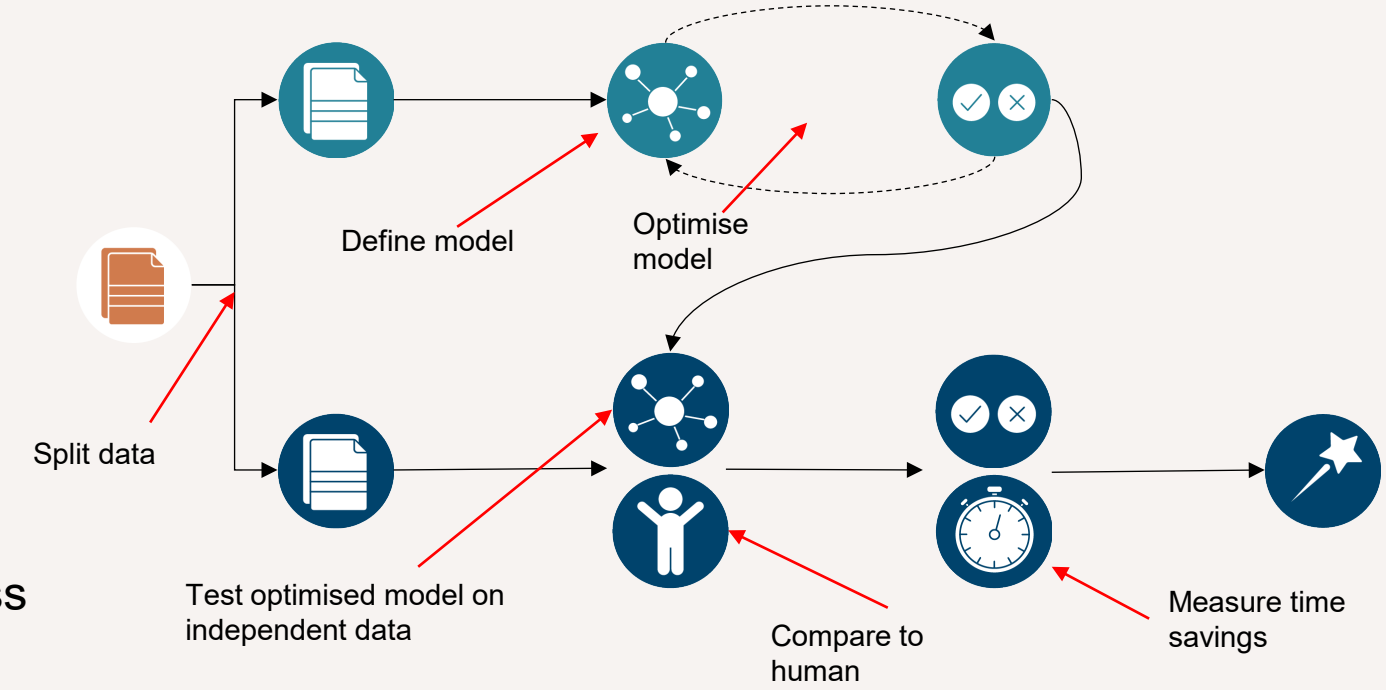
Inclusion criteria

1. **English** language studies
2. **Healthcare** context
3. Published from **2022**
4. Evaluation of **AI** performance on **systematic literature review** tasks
5. Formal **comparison to human** reviewers.



Common problems

1. No consistent approach to model testing.
2. No measure of time savings and/or human comparator.
3. No model description.
4. No model optimisation.
5. No data splitting.
6. No interoperability or standardisation across tasks.



Risks and biases

Risk/bias	Description	Potential Mitigations
Context bias	Occurs when LLMs are biased towards the prompter, often occurring when the prompt template is too complex and the LLM learns the ‘voice’ of the user.	Independently verify prompts.
Benchmark bias	Occurs when the benchmark standard itself has errors. For example, making mistakes when using K-shot prompting.	Independently verify prompts.
Explainability	LLMs are ‘black boxes’, the results should not only be plausible but also verifiable and justifiable.	Output justifications and exact quotes.
Hallucinations	When LLMs ‘make up’ plausible but incorrect results.	Prompt engineering: RAG, self-consistency
Algorithmic bias	Occurs when a model favours one subset of data over another, for example medical domains.	Present performance across multiple domains.

NICE National Institute for Health and Care Excellence

Thank you.

Stay up to date

- Sign up to our newsletters - scan the QR code or visit nice.org.uk/newsletters
- Follow us on social media - visit link.tr.ee/nicecomms

- “Is it possible to deliver *faster*, *scalable* outputs without compromising on methodological rigor?”
  - ✓ **Yes, using Gen AI assisted evidence synthesis frameworks**
- "Can generative AI truly be trusted to conduct systematic literature reviews for HTA submissions?“
  - ✓ **Yes, by keeping human-in-loop and adhering to HTA guidance on the use of Gen AI**
- “Are there any real-world examples of Gen AI-assisted SLRs in HTA submissions?”
  - ✓ **Yes, Singh 2025 and Makhija 2025 posters provide the NICE landscape assessment on the use of AI-assisted SLRs and a detailed example**
- "Have we reached a point where AI-assisted SLRs can stand up to regulatory scrutiny and truly save time and effort in real-world HTA submissions?“
  - ✓ **Yes, early implementations show around 50% savings in time and cost**

AI-Assisted Systematic Literature Screening According to NICE, and CDA Position Papers: An HTA Submission Use Case

MSR21

Dilip Makhija[1], Marvin Rock[1], Chong H Kim[1], Mirko von Hein[2], Rajdeep Kaur[3], Sumeet Attri[3], Barinder Singh[3]

[1]Gilead Sciences, Inc., Foster City, CA, USA, [2]Gilead Sciences, London, United Kingdom, [3]Pharmacoevidence, Mohali, India

CONCLUSIONS

- This study represents an AI-assisted SLR submitted for seladelpar to NICE, UK, which followed the NICE and CDA guidance on human-in-the-loop processes.
- The concordance between human- and AI-screened citations approached the ideal benchmark of 95%, notably surpassing the <90% inter-reviewer agreement typically observed between human reviewers in conventional systematic literature reviews
- This high agreement rate highlights the feasibility and rigor of integrating AI into evidence synthesis and HTA workflows, providing a scalable, rapid, and cost-efficient solution
- Further, the AI-assisted SLR tool reduced screening time and cost by approximately 50% while maintaining methodological rigor

PLAIN LANGUAGE SUMMARY

- An AI-based screening tool was used to review titles and abstracts across multiple systematic literature reviews (SLRs) in primary biliary cholangitis (PBC) , which is aligned to the position statement for AI by NICE and CDA.. One human and the AI tool screened the title and abstracts in parallel, with an independent human reviewer resolving any conflicts. The quality assurance confirmed that no relevant citations were missed by AI
- AI tool achieved an ideal agreement rate of 95% with ~50% reduction in screening time and costs, making the review process faster and economical compared to standard methods
- Evidence from these SLRs supported the NICE submission of seladelpar for previously treated PBC (ID6429), the first AI-assisted NICE submission to date. The Evidence Assessment Group (EAG) acknowledged the appropriate and effective use of the AI/ML tool

INTRODUCTION

- Systematic literature reviews (SLRs) are the gold standard in evidence-based medicine to inform policy and practice, and support clinical decision-making; however, the process of conducting them is very resource-intensive, costly, and time-consuming[1,2]
- The artificial intelligence (AI)-based Large Language Model (LLM) has the potential to enhance the efficiency of SLRs by rapidly and accurately analyzing substantial volumes of textual data[3]
- The National Institute for Health and Care Excellence (NICE), UK[4], and Canada's Drug Agency (CDA)[5], Canada, acknowledge the supportive use of AI in evidence generation, requiring a transparent and clear approach, sustained human involvement, and appropriate external validation

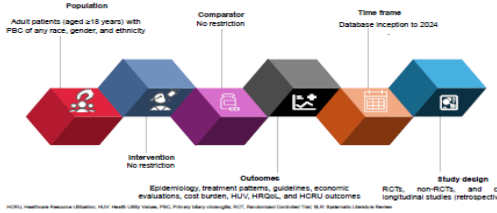
OBJECTIVE

- The current study aimed to implement a hybrid review framework integrating AI-assisted screening with expert human oversight to conduct multiple SLRs for Primary Biliary Cholangitis (PBC)
- This approach aligns with guidance from NICE, UK, and CDA, Canada, position papers
- The reviews focused on seven key objectives: epidemiology, treatment patterns, treatment guidelines, economic evaluation, healthcare resource utilization and related costs, health utility values, and humanistic outcomes

METHODS

- A comprehensive literature search was conducted across key biomedical databases, including EMBASE®, MEDLINE®, NHS EED, CENTRAL, and CDSR from inception through 2024
- A subject matter expert (SME) with a decade of experience in SLRs prepared and optimized the prompt based on the results obtained from a small subset of the citations, delivered via a Python Application Programming Interface, to identify evidence aligning with the key inclusion and exclusion criteria. The optimized prompt was then applied to the complete dataset (**Figure 1**)
- Data was collected using the inclusion/exclusion criteria guided by the PICOS approach
- A parallel dual-review and quality control process was used for data collection, where citations were independently screened by a human reviewer and a Claude 3.7 Sonnet-based AI tool applying predefined inclusion and exclusion criteria
- Further, the SME resolved the conflict between human and AI-screened citations
- A comprehensive depiction of the entire process is presented in **Figure 2**
- To ensure robust quality check (QC), citations excluded by the AI were checked at two distinct levels:
  - **QC Level:** Applied when there was disagreement between the human reviewer and AI tool (e.g., human included vs. AI excluded, or vice versa). An SME resolved disagreements at this stage
  - **Quality assurance Level:** Applied when both the human and AI reviewers agreed to exclude the citation, ensuring that exclusion decisions were consistently accurate
- Additional QC measures included bibliographic validation against published SLRs and technology appraisals, confirming comprehensiveness. Further, AI-assisted results were internally benchmarked against traditional manual screening, confirming consistent performance and sensitivity:

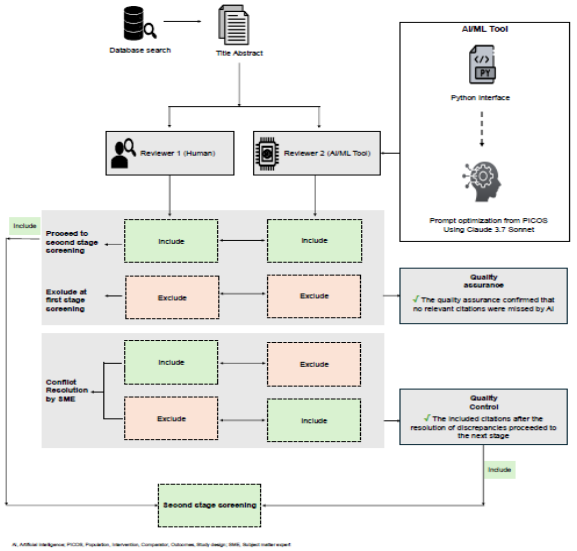
Figure 1: Eligibility criteria for the SLR



RESULTS

- A total of 6,309 citations were screened across the burden domains of PBC using the AI-human hybrid framework in the NICE submission of seladelpar for previously treated PBC (ID6429)[6]
- In this review process, the overall average agreement between AI and human reviewers was 94.97% (median: 95.24%), ranging from 89.01% in humanistic burden review to 99.59% in economic evaluation review
- This demonstrates that the AI-assisted process meets or exceeds expected human reviewer agreement levels, reinforcing the reliability and validity of our screening approach
- High concordance was also observed in other domains: 97.1% in health utility values, 97.9% in healthcare resource utilization and related costs, 95.2% in treatment patterns, 94.1% in treatment guidelines, and 91.8% in epidemiology (**Figure 3**)

Figure 2: Schematic diagram of the AI-assisted screening process

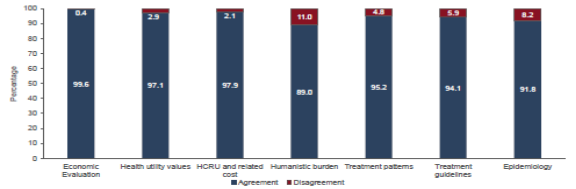


RESULTS (CONT.)

- Compared to a human reviewer, the use of an AI-based tool reduces the time spent on the first stage screening by 110 hours, leading to an approximate 50% cost savings
- Moreover, in alignment with the NICE and CDA position statement, transparency and rigor were ensured during the whole process
- NICE's Evidence Assessment Group (EAG) evaluated the SLR methodology and results and considered them appropriate

- ✓ Based on a recent NICE UK landscape assessment, this was the first NICE submission to incorporate an AI-assisted SLR
- ✓ NICE EAG inquiries focused on concordance between human and AI screening, AI-related exclusions, quality control processes, and prior publications on the AI tool
- ✓ Following the review, the EAG deemed the use of the automated SLR tool appropriate

Figure 3: Decision alignment between human and AI reviewers at the QC stage



References

1. Atkinson CF. *Soc Sci Comput Rev.* 2023; 42(2): 376–393
2. Bolaños F. et al. *Artif Intell Rev.* 2024; 57(10):259
3. Bernard N et al. *BMC Med Res Methodol.* 2025; 25(1):75
4. NICE position statement. *Use of AI in evidence generation.* Accessed 29 May 2025
5. CDA-AMC. *New Position Statement.* Accessed 29 May 2025
6. Seladelpar for previously treated primary biliary cholangitis [ID6429]. . Accessed 30 July 2025

Acknowledgements

We extend our thanks to the patients, their families, and all participating investigators. This study was funded by Gilead Sciences, Inc. Editing and Pharmacoevidence provided production assistance

Disclosures

DM, MR, CK, and MH are employees of Gilead Sciences, Inc., RK, SA, and BS are employees of Pharmacoevidence

Evidence Pharmaco
Science-led HEOR & Tech-driven AI

Where is AI in Evidence Synthesis for NICE HTA Submissions?

SA105

Barinder Singh, RPh[1] , Karam Diaby, PhD[2], Dilip Makhija, MS[3], Deepika Thakur, PhD[4], Amit Kulkarni, PhD[5,6] , Ritesh Dubey, PharmD[1] , Pankaj Rai, MS[1]

[1]Pharmacoevidence, Mohali, India, [2]Department of Pharmaceutical Sciences, University of Florida, Florida, FL, USA, [3]Gilead Sciences Inc., Foster City, CA, USA, [4]Hoffman La Roche AG, Mississauga, ON, Canada, [5]Consultant, [6]Formerly with Otsuka Pharmaceutical Development & Commercialization Inc., Princeton, NJ, USA

GET POSTER PDF

INTRODUCTION

- Interest in applying Artificial Intelligence (AI) in medical research and evidence generation has grown significantly in recent years, offering new tools for automating literature reviews, classifying studies, extracting data, and visualizing complex results[1,2]
- These advances have the potential to overcome the limitations of conventional review processes, enhancing efficiency, reducing resource burden, and improving the rigour, accuracy, and transparency of evidence generation and assessment[2,3]
- In recognition of this growing trend and the transformative role of AI, the National Institute for Health and Care Excellence (NICE) issued a comprehensive position statement in August 2024, which outlines the key principles and expectations for the use of AI-driven approaches in evidence generation and submissions across its evaluation programs[4]
- The NICE statement underscores the need for transparency, methodological rigor, and trust in the application of AI to evidence generation. It advocates the careful use of AI to support, not replace, human expertise, and provides guidance on regulatory standards and best practices to promote responsible adoption within Health Technology Assessment (HTA)[4]
- This study aimed to investigate the extent to which AI technologies have been utilized in NICE UK HTA submissions, specifically for conducting systematic literature reviews (SLRs), over recent years

METHODS

- A comprehensive review was carried out covering all NICE HTA submissions published within the last three years, i.e., June 2022 to June 2025, ensuring a complete coverage of the most recent trends and practices. **Figure 1** illustrates the workflow used to identify AI utilization in NICE HTA submissions in recent years
- Committee papers, committee slides, and associated guidance documents were retrieved and systematically reviewed manually
- A structured keyword strategy was used to identify HTA submissions, focusing on terms related to AI
- Through this screening process, documents that contained explicit references to the use of AI or automation technologies in the conduct of SLRs were identified and shortlisted
- The selected documents were then reviewed in detail to capture insights on:
  - ☐ The specific scope of AI application (e.g., screening, data extraction, synthesis)
  - ☐ The objectives of AI integration (e.g., efficiency, consistency, scalability)
  - ☐ The context of use, including how AI supported or complemented existing SLR methodologies in HTA submissions

RESULTS

- A total of 288 technology appraisals (TAs) were identified between June 2022 and June 2025 across all indications
- Approximately 800 documents, including committee papers, presentation slides, and associated guidance, were downloaded and systematically reviewed to identify the applications of AI within the evidence synthesis process
- A comprehensive search was employed using a predefined list of AI-related keywords to ensure both sensitivity and specificity.
- The keywords included AI, automation, automated, machine learning, large language models (LLM), generative pre-trained transformer (GPT), natural language processing (NLP), etc., in the context of evidence synthesis
- By June 2025, only two TAs had reported using AI at intermediate stages of the SLR process and were therefore included. **Figure 2** provides a detailed overview of the stages and methods of AI application across the included TAs
- Until June 2025, no TA had reported the comprehensive use of AI in conducting SLRs. However, TA11540, expected to be published in November 2025 (committee papers available from July 2025), post-search period, demonstrated the use of the Pharmacoevidence tool as a secondary screener to identify relevant publications, which was deemed acceptable by the evidence assessment group (EAG)

Figure 1: Workflow for identifying AI use in NICE HTA submissions

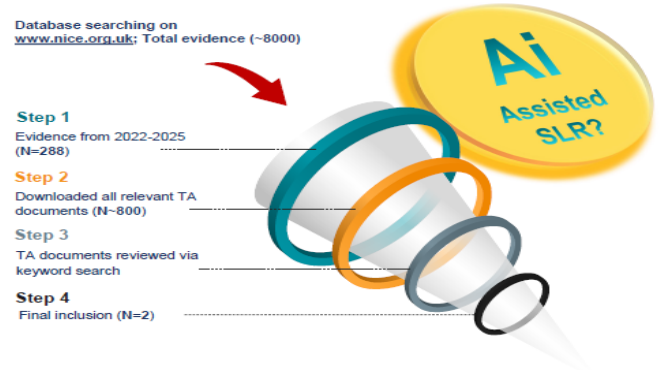
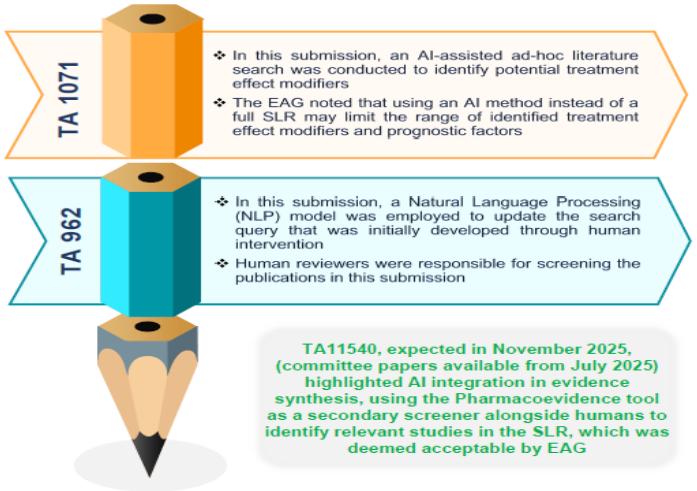


Figure 2: Overview of AI use across TAs



TA11540, expected in November 2025, (committee papers available from July 2025) highlighted AI integration in evidence synthesis, using the Pharmacoevidence tool as a secondary screener alongside humans to identify relevant studies in the SLR, which was deemed acceptable by EAG

Benchmark submission for the use of Gen AI in HTA Assessments

CONCLUSIONS

- Despite NICE 2024 position paper emphasizing the importance of responsible use of AI in evidence generation, its integration and application in NICE HTA submissions remain limited
- Between 2022 and 2025, only two NICE appraisals reported the use of AI, both in supplementary capacity. The ongoing TA11540, expected publication in November 2025, highlighted the use of the Pharmacoevidence® AI/ML tool as a secondary reviewer to manage large citation volumes, while maintaining compliance with NICE and Cochrane standards
- This indicates a significant opportunity to leverage AI in evidence generation, to improve the efficiency, rigour, and timeliness future submissions, supporting more transparent, informed, and accelerated decision-making in NICE HTA

References

1. Ge L et. al. Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. Systematic reviews,13(1-2024), 269.
2. Khalifa M et al. Using artificial intelligence in academic writing and research: An essential productivity tool, 5 (2024) 100145.
3. Andersen TH et al. Using Artificial Intelligence Tools as Second Reviewers for Data Extraction in Systematic Reviews: A Performance Comparison of Two AI Tools Against Human Reviewers 3 (4, 2025), e730036
4. NICE. Use of AI in Evidence Generation: NICE position statement; Version: 1.0; Thursday 15 August 2024

Correspondence: Barinder Singh; barinder.singh@pharmacoevidence.com

Disclosure: BS, RD, and PR are employees of Pharmacoevidence. DM is affiliated with Gilead Sciences, Inc. KD is affiliated with Otsuka Pharmaceutical Development & Commercialization, Inc. (Princeton, NJ, USA). DT is affiliated with Hoffmann-La Roche AG (Mississauga, ON, Canada) and AK is formerly associated with Otsuka Pharmaceutical Development & Commercialization, Inc. (Princeton, NJ, USA), at the time of conducting this research

ISPOR 2025; November 09-12; Glasgow, Scotland, UK

Thank you