NICE National Institute for Health and Care Excellence

Use of AI in evidence generation: NICE position statement

Corporate document
Published: 15 August 2024

www.nice.org.uk/corporate/ecd11

Contents

This corporate should be read in conjunction with ECD12.

Background

NICE helps the NHS and wider health and care system deliver the best care to people, fast, while ensuring value for the taxpayer. We do this by developing guidance, advice and information through a diverse range of programmes, that share the same core process of identification, assessment and interpretation of evidence.

The use of artificial intelligence (AI) methods, from relatively well-established machine learning approaches to newer and more complex generative AI, offers several potential benefits for this process.

AI methods can efficiently process and analyse large datasets to reveal patterns and relationships that may not be readily apparent to human analysts. And increasingly, generative AI can create novel outputs based upon what it learns from data.

Such capabilities may offer superior approaches to evidence generation for health technologies. It's highly likely that, in the near future, evidence considered by NICE will be informed by AI methods.

However, concerns about the appropriateness, transparency and trustworthiness of AI do exist. It's important to consider the use of AI methods carefully to ensure the anticipated benefits are balanced against the known concerns. This means guidance is needed on how to present evidence that has been informed by AI methods. Any consideration of AI methods should support our commitment to better meeting the needs of our users by creating guidance that is more relevant, timely, useable and impactful.

This position statement sets out our view on the use of AI methods in the generation and reporting of evidence considered by its evaluation programmes. It aims to:

- 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.

This statement relates to the potential use of AI in the generation and reporting of evidence considered by NICE. It does not consider the evaluation of health technologies that use AI methods to perform their function (AI-enabled technologies).

We have not included a detailed description of methodological aspects, process considerations and NICE's methods research relating to AI, because future papers will cover these in detail.

Terminology

There is no generally accepted definition of AI. In this position statement, we consider AI methods as those that exhibit some level of adaptivity and autonomy (DSIT 2023). UK Parliament definitions for specific terms referred to or covered by this statement are listed below (Gajjar 2024).

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 text, 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).

Natural language processing

An approach to programming computer systems to understand and generate human speech and text, by looking at linguistic patterns and word and sentence structures.

Our position on the use of AI in evidence generation and reporting

1.1   There are several potential benefits to using AI methods in health technology assessment (HTA). These must be balanced against potential risks, for example: algorithmic bias, cybersecurity, and reduced human oversight, transparency and accessibility to non-experts (Gervasi et al. 2022; Zemplényi et al. 2023). In light of the potential risks and the rapidly evolving nature of AI methods, they should only be used when there is demonstrable value from doing so.

1.2   The use of AI methods may introduce added complexity. It is important that submitting organisations (including manufacturers and external assessment groups) considering using them for evidence generation and reporting ensure the rationale for doing so is clear. If more explainable and common methods are potentially robust, those should be presented in the first instance, with supplementary use of less transparent approaches. Submitting organisations should clearly justify the use of these methods and outline assumptions (using, for example, the PALISADE checklist) and consider the plausibility of their results.

1.3   Submitting organisations considering using AI methods should engage with NICE to discuss their plans. When appropriate, early engagement could be sought through NICE Advice. At later stages of evidence development, organisations should discuss their plans with appropriate NICE technical teams.

1.4   Requests to use NICE content for AI purposes are subject to an approval process, licensing arrangement and a fee (for international use). If you would like to use our content for AI purposes, you can find more information on NICE's webpage on reusing our content.

1.5   All use of AI should align to the UK Government framework for regulating AI and the key principles should be referenced when considering the value of AI use cases. In the UK, the public sector has embraced various AI ethical frameworks to guide the development and deployment of AI systems (Cabinet Office 2018; DHSC 2021; DSIT 2019; DSIT et al. 2019; Leslie 2019). It is the submitting organisation's responsibility to determine which legislation applies, including data

protection laws and ethical standards. When relevant these should be clearly documented.

1.6   In alignment with the Medicines and Healthcare products Regulatory Agency (MHRA) and European Parliament guidance for AI use in a medicinal product lifecycle, it is the submitting organisation's responsibility to ensure that all algorithms, models, datasets, and data processing pipelines used are fit for purpose and are consistent with ethical, technical, scientific, and regulatory standards (MHRA 2024; EU 2024; European Medicines Agency 2023).

1.7   There remains a need to build trust in the application and use of AI in decision making (Zemplényi et al. 2023). Therefore, any use of AI methods should be based on the principle of augmentation, not replacement, of human involvement (that is, having a capable and informed human in the loop; Fleurence et al. 2024). For example, submitting organisations should conduct careful technical and external validation when AI methods are used, and present the results.

1.8   When AI is used, the submitting organisation and authors should clearly declare its use, explain the choice of method and report how it was used, including human input (see paragraph 1.7). The submitting organisation remains accountable for the content included in any submission.

1.9   It is the submitting organisation's responsibility to ensure that it is compliant with any licensing agreements. This includes, but is not limited to:

- copyright considerations, such as whether the organisation is authorised to use copyrighted or licensed materials in the AI tool, how the AI tool handles copyrighted or licensed materials, and its compliance with copyright law or user licences

- whether a business licence is required for any third-party AI tools that are used

- who owns the intellectual property produced by the AI tool, and is the organisation authorised to share it with NICE.

1.10   The use of AI methods, particularly 'black box' models, can introduce challenges for transparent reporting of evidence (Fleurence et al. 2024). When their use is

justified, submitting organisations should consider how these methods can be accessibly presented, including appropriate referencing (for example, of AI tools used and suitability assessment) and the use of lay language. When available, consider using tools to support the explainability of AI methods and increase transparency of their application (Amann et al. 2020).

1.11   The use of AI can introduce new risks. These risks should be mitigated by adhering to established guidance and checklists (such as Cochrane, PALISADE, TRIPOD+AI and the Algorithmic Transparency Reporting Standard) during the development, application and reporting of AI, and using AI only in the context of following other relevant best practice guidance recommended by NICE (such as NICE's real-world evidence framework).

1.12   When using AI methods, submitting organisations should report the risks they identified with doing so (for example, regarding concerns about transparency and bias) and steps they took to address those risks (Fleurence et al. 2024).

1.13   The use of novel AI methods presents cybersecurity risks, such as manipulation of data (data poisoning) or injecting malicious content into prompts (prompt injection attacks; Branch et al. 2022; National Cyber Security Centre 2024). These risks should be considered alongside other risks posed by AI systems. When using AI methods, submitting organisations should provide evidence of the steps taken to ensure robust security measures are in place to prevent such unauthorised access and manipulation.

1.14   The use of AI methods in the context of estimating comparative treatment effects (causal inference), represents a potentially very influential and therefore higher-risk application of AI. Their use should be accompanied by considered sensitivity analysis, checked against other suitable methods, and results presented in the context of available clinical evidence ('triangulation').

1.15   Ideally, the use of machine-learning methods should be accompanied by pre-specified outcome-blind simulations, conducted independently, to demonstrate their statistical properties in similar settings (for example, different data types or populations) and the correctness of their implementation.

1.16   AI methods used for real-world data extraction and curation must be reported, in

detail, as part of the data suitability assessment outlined in NICE's real-world evidence framework, making use of reporting tools when possible.

How this position statement was developed

This position statement has been developed through consideration of how AI methods could be applied to all aspects of evidence considered by NICE. For information, the following sections of this document outline this work. They summarise potential uses of AI methods that NICE is aware have been applied to, or are being researched for, HTA-related purposes.

We understand that the use of AI is a rapidly developing field, and so the uses listed in the following sections are not expected to be exhaustive. Additionally, our reflections on the potential uses do not represent endorsement or acceptance of these methods in evidence considered by NICE. Organisations considering using AI methods in their evidence should discuss this with NICE (see paragraph 1.3 in the section on our position on the use of AI in evidence generation and reporting).

Systematic review and evidence synthesis

1.17   Conventional literature search and review processes are largely undertaken manually, and typically require substantial time and resources. AI methods have the potential to automate various steps in these processes.

1.18   Machine learning methods and large language model prompts may be able to 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 (Cochrane 2024; ISPOR 2024; Fleurence et al. 2024; NICE's guidelines manual).

1.19   Large language models could be used to automate data extraction from published quantitative and qualitative studies by inputting prompts into the AI tool to generate the preferred output (Cochrane 2024; ISPOR 2024; Fleurence et al. 2024; Reason et al. 2024a). This is less well established than the uses described in paragraph 1.18.

1.20   Large language models could be provided with prompts to generate the code required to synthesise extracted data in the form of a (network) meta-analysis (Reason et al. 2024b). This is less well established than the uses described in paragraph 1.18.

1.21   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.

Clinical evidence

1.22   Clinical-effectiveness evidence is typically informed by clinical trials on the intervention, clinical trials on the comparator(s) and real-world data. It may include evidence to quantify a treatment effect, establish a side effect profile, or assess the generalisability of trial data to the NHS population.

1.23   AI can be used in trial design, such as defining the inclusion and exclusion criteria, and retention. Pattern recognition and machine learning may be used to avoid excluding people based on factors which do not affect the treatment response. Dosage levels, sample size and trial duration can also be optimised using AI approaches. Natural language processing can be used to mine electronic health records, for example, to identify people who meet the trial criteria and have highest potential for benefit, and side effect reporting (Datta et al. 2024; Fleurence et al. 2024; Harrer et al. 2019; Padula et al. 2022).

1.24   AI approaches can also be used to identify and adjust for limitations in clinical data. For example, pattern recognition can identify relevant covariates that influence treatment response and adjust for these in the statistical analyses (Padula et al. 2022).

1.25   AI methods can take into account complex, non-linear relationships between covariates, producing models that have fewer structural assumptions compared with parametric models. This may be particularly useful for improving performance of predictions and reducing bias in causal inference, with benefits

also for precision of the effect estimate and estimates of uncertainty (Padula et al. 2022).

1.26 AI approaches can be used to produce synthetic data and generate external control arms (Fleurence et al. 2024). This may be applicable when it is unethical to include a placebo arm in a trial. It can also be used to predict clinical effectiveness in different populations, for example, by applying data from a clinical study to a population with different characteristics (Ling et al. 2023; Piciocchi et al. 2024).

1.27 Natural language processing may be used to analyse large amounts of information. This could be applied, for example, to generate an executive summary of the clinical evidence. It may also be used to simplify technical language for the purpose of creating a lay summary (Velupillai et al. 2018; Saleh et al. 2024).

1.28 When AI is used in clinical evidence generation, reporting should be transparent and make use of relevant checklists such as PALISADE to justify its use (Padula et al. 2022) and TRIPOD+AI to explain AI model development (Collins et al. 2024). Any AI approach used (for example, a cohort selection model) should be considered part of the clinical trial and full details provided within a submission.

Real-world data and analysis

1.29 Real-world data may lend itself increasingly to the use of AI approaches as accessibility and standardisation of large datasets reflecting routine care and real-world populations improve. AI approaches have several potential roles for supporting real-world evidence across numerous stages of evidence generation.

1.30 AI methods may have a role to play in data processing before the development of real-world evidence. For example, natural language processing approaches are being used to generate structured data from unstructured real-world data (Keloth et al. 2023; Fleurence et al. 2024; Li et al. 2022; Soroush et al. 2024). Approaches such as multimodal data integration can combine different data sources into a cohesive dataset (for example, Boehm et al. 2022), and aspects such as data matching and linkage, deduplication, standardisation, data cleaning and quality

improvement (for example, error detection and imputation of missing data) are increasingly being automated and are scalable to process large volumes of data (Elliott et al. 2022; Pfitzner et al. 2021; OHDSI 2024).

1.31   AI approaches may support the efficient selection of relevant populations and observations from large datasets for the purposes of addressing specific research questions (Shivade et al. 2014).

1.32   AI methods can support estimation of comparative treatment effects (causal inference), primarily through using feature selection methods which select from a subset of relevant features for use in model construction. Additionally, analytical methods using AI approaches can provide more targeted estimates of causal effects, sometimes harnessing predictive capabilities of multiple valid machine learning algorithms (Padula et al. 2022).

Cost-effectiveness evidence

1.33   Cost-effectiveness evidence is typically informed by economic models. Developing an economic model is a resource-intensive multi-step process involving model conceptualisation, parameter estimation, construction, validation, analysis and reporting steps. AI methods may have a role in several of these steps (Fleurence et al. 2024).

1.34   AI methods have the capability of interrogating complex or different datasets in new ways (for example, Boehm et al. 2022 and Cowley et al. 2023). This may generate new or deeper insights into cost drivers and health outcomes, such as disease progression, surrogate relationships, and clinical pathways. This information could inform the conceptualisation and parameterisation of an economic model (for example, in terms of included health states, transitions, and events; Chhatwal et al. 2024), reducing structural uncertainty.

1.35   Methods using large language models could be used to automate the construction and calibration of new economic models, and generation of model reports (Reason et al. 2024a). Following human-led model conceptualisation and parameter estimation steps, large language model prompts can be designed to generate the code for the economic model. This may permit the construction and

comparison of multiple models to assess structural uncertainty (Fleurence et al. 2024).

1.36 Large language models can be provided with prompts to reflect new information in an economic model, such as clinical data or comparators, facilitating updates and adaptations (Reason et al. 2024a). In the extreme, AI methods may support economic models being updated in real time, though we are yet to see a published example of this in healthcare.

1.37 Large language model methods can support the replication and cross-validation of existing economic models (Reason et al. 2024a).

1.38 Machine learning methods can be used for simulation optimisation (Amaran et al. 2015). In the context of economic modelling, this could reduce, and ideally minimise, the computational time that a simulation model takes to run. By increasing the efficiency of economic models, more complex models that use fewer simplifying assumptions may become more practical to use, including for probabilistic sensitivity analysis (Fleurence et al. 2024).

Next steps

1.39 NICE will review this position statement if significant new evidence becomes available that might require a change to the use of AI methods in evidence submissions as outlined in this statement.

1.40 If an update to NICE's health technology evaluation manual concerning AI methods is identified as being potentially appropriate, it will be considered as part of the framework for modular updates to NICE's methods manual.

1.41 NICE will monitor any future use of AI methods in NICE evaluations, and consider whether their use poses challenges or opportunities to NICE's guidance development processes. NICE will continue to assess operational considerations around the upskilling and training of staff and committee members. In the future there may be a need to commit to expanding capacity and capabilities across related disciplines that support the adoption of AI in our evaluations (Zemplényi et al. 2023).

1.42   NICE will continue its ongoing policy and research activities examining the use of AI methods in HTA.

1.43   If you have any questions or comments about this position statement, please contact us using the following email address: htalab@nice.org.uk.

Summary

1.44   NICE is aware that the use of AI methods is increasingly being explored for purposes relating to HTA, and there is likely to be increasing use or consideration in evidence generation and reporting in the near future.

1.45   This position statement provides clarity on how NICE will consider the use of AI methods in the generation and reporting of evidence to be evaluated within its guidance production programmes.

1.46   NICE requires that when AI methods are used, the transparency, rigour and trust in our guidance production is maintained. Therefore, any use of AI should be done judiciously, leveraging the strengths of AI to support and enhance decision making only when it is suitable and when it adds value. The use of AI methods should augment human involvement, not replace it.

References

- Amann J, Blasimme A, Vayena E et al. (2020) Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Medical Informatics and Decision Making 20(1): 310

- Amaran S, Sahinidis NV, Sharda B, Bury SJ (2015) Simulation optimization: a review of algorithms and applications. Annals of operational research 240: 351–80

- Boehm KM, Aherne EA, Ellenson L et al. (2022) Multimodal data integration using machine learning improved risk stratification of high-grade serous ovarian cancer. Nature 3: 723–33

- Branch HJ, Rodriguez Cefalu J, McHugh J, Hujer L, Bahl A et al. (2022) Evaluating the susceptibility of pre-trained language models via handcrafted adversarial examples [online; accessed 15 July 2024]

- Cabinet Office (2018) Data Ethics Framework [online; accessed 3 June 2024]

- Central Digital and Data Office and Department for Science, Innovation and Technology (DSIT; 2021) Algorithmic Transparency Recording Standard [online; accessed 1 July 2024]

- Chhatwal J, Yildrim IF, Balta D et al. (2024). Can large language models generate conceptual health economic models? [online; accessed 25 June 2024]

- Cochrane (2024) Artificial intelligence (AI) technologies in Cochrane. Webinar (9 May 2024) [online; accessed 29 May 2024]

- Collins GS, Moons KGM, Dhiman P et al. (2024) TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385: e078378

- Coombs L, Orlando A, Wang X et al. (2022) A machine learning framework supporting prospective clinical decisions applied to risk prediction in oncology. NPJ Digital Medicine 5(1): 117

- Cowley HP, Robinette MS, Matelsky JK et al. (2023) Using machine learning on clinical data to identify unexpected patterns in groups of COVID-19 patients. Nature Scientific Reports 13(1): 2236

- Datta S, Lee K, Paek H et al. (2024) AutoCriteria: a generalizable clinical trial eligibility criteria extraction system powered by large language models. Journal of the American Medical Informatics Association 31(2): 375–85

- Department for Health & Social Care (DHSC; 2021) A guide to good practice for digital and data-driven health technologies [online; accessed 3 June 2024]

- Department for Science, Innovation and Technology (DSIT; 2023) A pro-innovation approach to AI regulation [online; accessed 26 June 2024]

- Department for Science, Innovation and Technology (DSIT; 2019) The Centre for Data Ethics and Innovation's approach to the governance of data-driven technology [online; accessed 3 June 2024]

- Department for Science, Innovation and Technology (DSIT), Office for Artificial Intelligence and Centre for Data Ethics and Innovation (2019) A guide to using artificial intelligence in the public sector [online; accessed 3 June 2024]

- European Medicines Agency (2023) Draft reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle [online; accessed 3 June 2024]

- European Parliament (2024) Briefing: Artificial intelligence act [online; accessed 13 June 2024]

- Elliott A (2022) Better, broader, safer: using health data for research and analysis (the Goldacre review). Journal of Radiological Protection 42(3)

- Fleurence RL, Bian J, Wang X et al. (2024) Generative AI for health technology assessment: opportunities, challenges, and policy considerations [online; accessed 17 July 2024]

- Gajjar D (2024) The Parliamentary Office of Science and Technology artificial intelligence (AI) glossary [online; accessed 26 June 2024]

- Gervasi SS, Chen IY, Smith-McLallen A et al. (2022) The potential for bias in machine learning and opportunities for health insurers to address it. Health Affairs 41(2): 212–18

- Guidelines International Network (GIN, 2024) Working groups: artificial intelligence. [online, accessed 30 May 2024]

- Harrer S, Shah P, Antony B, Hu J (2019) Artificial Intelligence for Clinical Trial Design. Trends in Pharmacological Sciences 40(8): 577–91

- ISPOR (2024) Revolutionizing systematic reviews: harnessing the power of AI. Webinar (21 May 2024) [online: accessed 29 May 2024]

- Keloth VP, Banda JM, Gurley M et al. (2023) Representing and utilizing clinical textual data for real world studies: an OHDSI approach. Journal of Biomedical Informatics 142: 104343

- Leslie D (2019) Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute [online: accessed 19 July 2024]

- Li I, Pan J, Goldwasser J et al. (2022) Neural natural language processing for unstructured data in electronic health records: a review. Computer Science Review 46

- Ling AY, Montez-Rath ME, Carita P et al (2023) An overview of current methods for real-world applications to generalize or transport clinical trial findings to target populations of interest. Epidemiology 34(5): 627–36

- MHRA (2024) Policy paper: Impact of AI on the regulation of medical products. [online; accessed 13 June 2024]

- National Cyber Security Centre (2024) AI and cyber security: what you need to know [online; accessed 15 July 2024]

- OHDSI (Observational Health Data Sciences and Informatics) Standardised data: The OMOP Common Data Model [online; accessed 4 June 2024]

- Padula WV, Kreif N, Vanness DJ et al. (2022) Machine learning methods in health economics and outcomes research—the PALISADE checklist: a good practices report of an ISPOR task force. Value In Health 25(7): 1063–80

- Piciocchi A, Cipriani M, Messina M et al. (2024) Unlocking the potential of synthetic patients for accelerating clinical trials: results of the first GIMEMA experience on acute myeloid leukemia patients. EJHaem 5(2): 353–9

- Pfitzner B, Steckhan N, Arnrich B (2021) Federated learning in a medical context: a systematic literature review. ACM Transactions on Internet Technology 21(2): 1–31

- Reason T, Rawlinson W, Langham J et al. (2024a) Artificial intelligence to automate health economic modelling: a case study to evaluate the potential application of large language models. Pharmacoeconomics Open 8(2): 191–203

- Reason T, Benbow E, Langham J et al. (2024b) Artificial intelligence to automate network meta-analyses: four case studies to evaluate the potential application of large language models. Pharmacoeconomics Open 8(2): 205–20

- Saleh M, Wazery Y, Ali A (2024) A systematic literature review of deep learning-based text summarization: Techniques, input representation, training strategies, mechanisms, datasets, evaluation, and challenges. Expert Systems with Applications 252(A): 124153

- Shivade C, Raghavan P, Fosler-Lussier E et al. (2014) A review of approaches to identifying patient phenotype cohorts using electronic health records. Journal of the American Medical Informatics Association 21(2): 221–30

- Soroush A, Glicksberg BS, Zimlichman E et al. (2024) Large language models are poor medical coders – benchmarking of medical code querying. NEJM AI 1(5)

- Velupillai S, Suominen H, Liakata M et al. (2018) Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. Journal of Biomedical Informatics 88: 11–9

- Zemplényi A, Tachkov K, Balkanyi L et al. (2023) Recommendations to overcome barriers to the use of artificial intelligence-driven evidence in health technology assessment. Frontiers in Public Health 11: 1088121