This is a “preproof” accepted article for *International Journal of Technology Assessment in Health Care*.

This version may be subject to change during the production process.

DOI: 10.1017/S0266462324004823

Development and validation of the NICE artificial intelligence (AI) medical device intervention search filters for MEDLINE and Embase (Ovid)

RUNNING TITLE: NICE AI medical device intervention filters

Lynda Ayiku (Corresponding author)

National Institute for Health and Care Excellence (NICE)

Level 1a, City Tower

Piccadilly Plaza

Manchester

M1 4BT

United Kingdom

Telephone number: 0161 870 3088. Email: lynda.ayiku@nice.org.uk

Amy Finnegan

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

National Institute for Health and Care Excellence (NICE), Manchester, United Kingdom

Thomas Hudson

National Institute for Health and Care Excellence (NICE), Manchester, United Kingdom

Nicola Walsh, FCLIP (CILIP Fellow)

National Institute for Health and Care Excellence (NICE), Manchester, United Kingdom

Rachel Adams

National Institute for Health and Care Excellence (NICE), Manchester, United Kingdom

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Abstract

Objectives

Medical device interventions with an artificial intelligence (AI) component can potentially improve the effectiveness and efficiency of activities such as diagnostic screening through automation. However, inconsistent terminology for the interventions is used in research literature and it can be challenging to find evidence about them in literature searches reliably. Information specialists at the United Kingdom's National Institute for Health and Care Excellence (NICE) developed novel validated search filters to retrieve evidence about AI medical device interventions from MEDLINE and Embase (Ovid).

Methods

The filters were drafted using AI-related terminology selected from references identified in NICE AI medical device topics.

Using relative recall methodology, gold standards containing references from systematic reviews about AI medical devices were generated for each database and then divided into Development and Validation Sets. The draft filters were finalised using their Development Sets and validated externally by calculating their recall against their Validation Sets. Target recall was >90 percent.

Results

Both filters achieved 96 percent recall against their Validation Sets.

Conclusions

The NICE AI medical device intervention search filters retrieve evidence about the interventions effectively and reliably. The filters can be used by information professionals, researchers, and clinicians to find evidence about

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

AI medical device interventions in literature searches. The filters should be used as part of a ‘multi-stranded’ literature search strategy when details of brand names and/or medical device manufacturers are known.

Keywords

Artificial intelligence

Therapy, Computer-Assisted

Databases, Bibliographic

Information Storage and Retrieval

Information Science

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Introduction

The National Institute for Health and Care Excellence (NICE) and artificial intelligence (AI) medical device interventions

The United Kingdom's (UK) National Institute for Health and Care Excellence (NICE) began publishing evaluations of the effectiveness of medical device interventions with an artificial intelligence (AI) component in 2020 as part of its medical technologies and diagnostics programmes (1). In addition, NICE has published the 'Evidence standards framework for digital health technologies' which provides a description of the evidence standards that are required for digital health technologies (including AI medical device interventions) to demonstrate their value to the UK National Health Service (NHS) and social care system (2). NICE's work supports the UK government's initiatives to implement AI medical device interventions safely and effectively (3).

AI is an umbrella term that encompasses several automated machine learning tools (4). Machine learning tools are models and algorithms that are trained to learn from data (4). AI medical device interventions can automate, augment, or assist with activities that are traditionally performed by healthcare professionals (5). These activities include behavioural therapy, screening, detecting conditions, assessing the risk of developing a condition, predicting the outcomes of a condition, or assessing the relapse of a condition (5). AI medical device interventions have the potential to improve the efficiency and effectiveness of health care services. For instance, they can speed up routine tasks that are usually carried by health care professionals, assist with decision-making, target patient needs more precisely to improve outcomes, and provide convenient and broader access to healthcare services for patients (4; 5). Each of these factors help to reduce the resource costs required for health care service delivery (4; 5).

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

However, evaluating AI medical device interventions can be challenging (4). For instance, there is often a lack of rigorous research evidence available for AI interventions to evaluate their effectiveness (4). In addition, many AI interventions are frequently updated and evolve over time ("adaptive AI") which can lead to their need for re-evaluation (4).

The NICE AI medical device intervention search filters

In addition to the evaluation challenges highlighted above, NICE information specialists have also encountered complexities with retrieving evidence about the interventions from bibliographic databases when literature searching. For example, AI medical device interventions can be poorly reported in the title and abstracts of research literature. In these cases, the brand names for the interventions are not reported and inconsistent, wide-ranging terminology of direct or indirect relevance to AI is used to describe the interventions. In addition, subject headings of relevance to AI technologies are applied inconsistently to the database records of the research literature in bibliographic databases. The consequence of this is that evidence for AI medical device interventions can be missed in literature searches that would typically consist of search terms for AI intervention brand names and narrow terminology that is directly relevant to AI technologies. The evidence-base for AI interventions is usually small and so the accidental non-retrieval of relevant evidence could have a major impact upon evaluation decisions. Conversely, when several search terms and subject headings of both direct and indirect relevance to AI technologies are used in literature search strategies, it can result in the retrieval of unmanageable volumes of irrelevant evidence.

In response to these searching challenges, a project team of NICE information specialists undertook a study to develop validated search filters with the aim of improving the effectiveness of literature searches to retrieve evidence on AI medical device interventions from the MEDLINE and Embase (Ovid Platform) bibliographic databases reliably. No validated search filters for

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

AI medical device interventions existed previously and a broad variety of untested search strategies were used to retrieve evidence about the interventions in literature searches. MEDLINE and Embase were chosen for the filters because these sources are two of the largest health care bibliographic databases and they are routinely used to find evidence for clinical topics at NICE (6). The project team had experience of developing search filters for digital health interventions, having developed and validated filters for retrieving evidence about health apps from MEDLINE and Embase in 2020 (7).

Search filters

A search filter is a set of pre-made, reusable search terms that are applied to literature search strategies to retrieve evidence on specific study designs or topics (8; 9). Search filters are available for a range of topics including geographic locations, population groups, conditions, and interventions (10). Search filters differ from literature search strategies because their retrieval performance has been evaluated in a process known as validation (8; 9).

There are two types of validated search filter, 'internally validated' and 'externally validated' (9):

1. **Internally validated:** Search terms for the filter are derived objectively from a set of relevant references on the topic of the filter and the development of the search filter is based on these terms. The retrieval performance of the filters is calculated against the set of relevant references from which they were derived.

2. **Externally validated:** Search terms for the filter are derived either subjectively or objectively. The retrieval performance of the filter is calculated against an independent set of relevant references that were not used in the development of the filter.

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

The external validation of search filters is important because it allows conclusions to be made about the generalisability of a search filter's recall and provides users with an indication of how successfully they work (8; 9). For both internally and externally validated search filters, a key filter retrieval performance measure is recall (also known as sensitivity). Recall is the percentage of known relevant references that a filter retrieves, and it provides an indication of how effective filters are at retrieving evidence on the topic of the filter (8; 9). Recall is calculated as the number of relevant records retrieved divided by the total number of relevant references. The resulting figure is commonly multiplied by 100 to express as a percentage (8; 9; 11).

This paper describes the development, internal validation, and external validation of the NICE AI medical device intervention filters.

Objective

The aim of the NICE AI filters for medical device interventions was to improve the retrieval of evidence for AI medical device interventions in literature searches. The objectives were to develop externally validated search filters for MEDLINE and Embase (OVID) to retrieve evidence about AI medical device interventions with high recall (defined as >90 percent).

Methods

The NICE AI medical device intervention search filters were created in two phases. In the first phase, ‘internal validation’ search filter methodology was employed to develop the initial draft filters. In the second phase, the draft filters were finalised and validated using ‘external validation’ search filter methodology.

Draft filter development: internal validation

The draft MEDLINE filter was developed first using ‘internal validation’ filter methodology (8; 9). This method involves using a set of known relevant

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

references on the filter topic to identify the search terms for the filters and then calculating the recall of the filters against the references (8; 9).

Identifying the search terms and search fields for the draft MEDLINE filter

The candidate AI technology search terms for the draft MEDLINE filter were identified using the titles, abstracts, keywords, and subject headings from included references from 11 published NICE AI medical technologies and diagnostics topics. See Supplementary File 1 for details of the NICE AI topics. A total of 66 included references from the published NICE AI topics were available in the MEDLINE All database and these were used to create a Test Set to aid the development of the draft filter. See Supplementary File 2 for details of the 66 included references in the Test Set.

Title, abstract, and keyword terms of most relevance to AI medical device interventions were selected as candidate search terms for the draft filter from the references in the Test Set. Terms for artificial intelligence-related technologies such as ‘machine learning’ and terms of wider relevance to artificial intelligence such as ‘algorithm’ were deemed to be relevant. See Supplementary File 3 for the candidate search terms. Subject headings were not selected because the project team made a judgement based on previous experience that the application of AI-related subject headings to database records is inconsistent as well as being imprecise and that their inclusion in literature search strategies can lead to the over retrieval of irrelevant evidence.

To test the usefulness of each candidate search term for the draft filter, the terms were translated into MEDLINE search syntax in the MEDLINE ALL bibliographic database and searched on separate search lines in the following search fields: title, abstract, keyword heading and keyword heading word. The purpose of the testing was to find the most efficient combination of terms, eliminating those that did not contribute to the overall recall. The primary

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

author and title for each included reference was used to form individual search lines in MEDLINE ALL to create a search strategy for the 66 included references in the Test Set.

To test the performance of each candidate search term for the draft filter, single search lines and combinations of the search lines were run against the search strategy for the Test Set and their recall of the references was examined. The combination of search terms and search fields plus the use of proximity operators for the search lines that were judged to be the most appropriate were investigated to assess their impact on the recall of the 66 included references in the Test Set. Superfluous search terms were removed from the filter. The aim of the investigations was to reduce the number of irrelevant results found by each individual search term or combination of search terms while ensuring that they retrieved as many of the Test Set references as possible. Examples of the investigations can be found in Supplementary File 4.

The draft MEDLINE AI medical interventions filter was created using what was judged to be the most appropriate combination of candidate search terms, search fields, and use of proximity operators for retrieving the Test Set. The draft MEDLINE AI medical interventions filter is presented below:

1 algorithm*.ti,kf.
2 (algorithm* adj1 (learn* or automate* or detect* or treatment* or therap* or radiolog* or AI or DL or data or dataset* or base*)).ab.
3 artificial intelligen*.ti,ab,kf.
4 AI.ti,kf.
5 machine learning*.ti,ab,kf.
6 deep learn*.ti,ab,kf.
7 convolutional neural network*.ti,ab,kf.
8 automate*.ti.
9 (automate* adj3 (system* or score* or software* or analysis* or analyse* or risk* or evaluat* or tool* or detect* or process*)).ab,kf.
10 or/1-9

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Draft filter recall

The draft MEDLINE filter retrieved 53 out of the 66 included references from the Test Set in MEDLINE ALL, achieving 80 percent recall. See Table 1 for further details.

Table 1: Recall of the draft MEDLINE filter against the Test Set

Upon investigation, it was found that only the 53 references that were retrieved by the filter had AI terminology in the searchable fields of their MEDLINE records. The missing 13 references did not contain any terminology of relevance to AI technologies their records. Therefore, the MEDLINE filter achieved 100 percent recall of the included references with AI-related terminology in the searchable fields of their database records.

Upon further examination, it was found that 12 of the missing 13 references would have been retrieved in a 'real life' literature search because the device brand names for the interventions were available in the searchable fields of their database records. It was also found that the remaining single missing reference was not about an AI device topic. See Supplementary File 5 for more detailed information about the 13 missing references.

Creating the draft Embase filter

Upon completion, the draft MEDLINE filter was translated into Embase search syntax to create a draft Embase filter for use in the Embase bibliographic database. The Embase draft filter is identical to the draft MEDLINE filter that is presented above.

'Watchful waiting' period

Following the development of the draft MEDLINE and Embase AI medical device intervention filters, a 'watchful waiting' period of 6 months was implemented. The purpose of this was to further assess the retrieval performance of the filters by testing them on active and retrospective NICE AI

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

topics and to allow for the potential identification of new relevant AI terminology from the NICE AI topics and the field of AI medical device intervention technologies. However, no changes were required for the draft filters following the 'watchful waiting' period.

Gold Standard set

A Gold Standard is a set of known relevant references that is used to develop and validate search filters (8; 9; 11). The relative recall method was used to identify references for the Gold Standard set (11). This method involves identifying the Gold Standard set from included references of evidence reviews such as systematic reviews, literature reviews, guidelines, and reports (11).

The Gold Standard set for the AI medical device intervention filters was identified from included references from systematic reviews on AI medical device intervention topics. The following search strategy was run using the 'Advanced search' function in Google Scholar between Tuesday 21st June 2022 to Thursday 30th June 2022 to find the systematic reviews: allintitle: "artificial intelligence" and "systematic review". A variety of literature search approaches was used to find evidence on AI interventions in these systematic reviews.

The first 12 relevant systematic reviews about AI medical device interventions that were published in 2022 and had full-text access available with no upfront cost were selected. The details of the 12 systematic reviews are available in Supplementary File 6. The primary author and title for each of the included references from the 12 systematic reviews were used to form individual search lines in MEDLINE ALL to create a search strategy for the included references. The references that were available in MEDLINE ALL and Embase

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

formed a Gold Standard set for each of the databases. There were 464 references for the MEDLINE Gold Standard set and 446 references for the Embase Gold Standard set.

Development Sets

A portion of each Gold Standard set was used to create Development Sets for MEDLINE All and Embase. The Development Sets were used to further test the recall of the draft filters and to identify if the filters required revisions to their search terms, search fields, or proximity operators. To create the Development Sets, every 3rd reference from each Gold Standard was selected after the results had been listed in alphabetical order by the surname of the primary author. There were 155 MEDLINE Development Set references (see Supplementary File 7 for details of the MEDLINE Development Set references) and 146 Embase Development Set references (see Supplementary File 8 for details of the Embase Development Set references).

When run against their respective Development Sets, the draft MEDLINE filter achieved 74percent recall (115 out of 155 references retrieved) and the draft Embase filter achieved 80percent recall (117 out of 146 references retrieved). These results fell short of the >90percent recall target for the filters and so the references that the filters did not retrieve were investigated with the aim of identifying additional AI search terms, search fields, or proximity operators for the filters to increase their recall performance.

Upon investigation, it was found that some of the references that were missed by the filters contained AI technology terminology such as ‘vector machine’, ‘radiomics’, and ‘supervised classifier’ which were not included in the search terms of the draft filters. It was also found that certain search lines in the draft filters were too specific to find some of the missing references. For instance, in search line 7, the search phrase ‘convolutional neural network*.ti,ab,kf’ was used in the draft filters but this phrase did not retrieve references that

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

contained related technology terms such as ‘artificial neural network’ or ‘deep neural network’. In addition, it was found that proximity operators were required for some lines of the draft filters to retrieve the missing references. For instance, one missing reference had referred to ‘deep metric learning’ and this was not retrieved because search line 6 of the draft filters used ‘deep learn*.ti,ab,kf’ to retrieve references about deep learning technologies. See Supplementary File 9 for a detailed examination of missing reference using the process used for the draft Embase filter as an example.

In accordance with the examination findings, the draft filters were revised by adding the additional AI technology terms and proximity operators with the aim of retrieving as many of the missing references as possible. Their recall against the Development Set was then calculated. The revised MEDLINE draft filter now achieved 96 percent recall and the revised Embase draft filter now achieved 98 percent recall. See Table 2 for further details.

Table 2: Recall of the revised filters against their Development Sets

These results met the >90 percent recall target, and the revised filters became the final filters that were deemed to be ready for the validation process.

Validation Sets

The remaining references in the MEDLINE and Embase Gold Standard sets formed the Validation Sets for each database. There were 309 MEDLINE Validation Set references (See Supplementary File 10 for details of the MEDLINE Validation Set references) and 300 Embase Validation Set references (See Supplementary File 11 for details of the Embase Validation Set references).

The purpose of the Validation Sets was to validate the filters externally using an independent set of references that had not been used in their

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

development. The final MEDLINE and Embase AI medical device filters were validated by calculating their recall against their respective Validation Sets.

A sample size of at least 100 relevant references is suggested to provide a reasonable confidence interval for a filter that aims to retrieve at least 90 percent of all relevant references (11). Both the MEDLINE and Embase Validation Sets exceeded this minimum specification.

Results

The final NICE AI medical device intervention filters are presented below in Figure 1:

Figure 1: Final NICE MEDLINE and Embase AI medical device intervention filters

Validation

Both the final MEDLINE and Embase filters achieved 96 percent recall against their respective Validation Sets (as shown in Table 3), exceeding the >90 percent recall target.

Table 3: Validation of the NICE MEDLINE and Embase AI medical device intervention filters

The majority of the missing Validation Set references were found to be poorly reported and they contained vague terminology to describe AI technologies. Two of the references were not about AI interventions. Detailed information about the missing references can be found in Supplementary File 12.

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Discussion

The NICE AI medical device intervention filters for MEDLINE and Embase are the first validated filters for retrieving evidence about AI medical device interventions. As indicated in the results, AI technology can be poorly described in the searchable fields of database records. However, where AI technology is described clearly, the filters retrieve evidence about AI medical device interventions effectively with high recall.

The filters are also likely to be more efficient than previously used untested search strategies for AI medical device interventions that included subject headings plus a wide range of indirect terminology for AI technologies.

The NICE AI medical device intervention filters do not contain subject headings and so they can be used in any bibliographic database with comparable search fields and search features to MEDLINE and Embase. However, it is important to note that the filters' retrieval performance in databases other than MEDLINE and Embase has not been examined.

Proposed usage of the filters

It is intended that the NICE AI medical device intervention filters will be applied to literature search strategies that describe the type of AI medical device intervention technologies required for the research question using the 'AND' Boolean operator. For instance, for a topic on AI-aided chest x-rays for detecting lung cancer, a literature search strategy would be conducted to find results about x-rays for lung cancer and then the NICE AI medical device intervention filters would be applied to the search strategy using the 'AND' Boolean operator. The literature search structure for this example is as follows:

1. Description of intervention (chest x-rays)

AND

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

2. Condition (lung cancer)

AND

3. NICE AI medical device search filter

The filters should be used as part of a 'multi-stranded' literature search strategy when details of brand names and/or medical device manufacturers are known as advised by Cooper, Dawson, & Lefebvre (12) in their practical guide on literature searching for medical device interventions. In this approach a search concept to identify AI medical device interventions using known AI device brand names and/or device manufacturers would be created when these details are available (12). This search concept would then be combined with the example search strategy structure above with the 'OR' Boolean operator. The reason for this is that the first phase of the filter development study revealed that 20percent of included references for AI medical interventions from NICE AI topics did not provide terminology for AI technologies in the searchable fields of their database records. Instead, these database records for these papers only provided the brand name of device manufacturer name for the technologies. This finding is likely to be an underestimate because the sample is based on known relevant references for AI medical device interventions.

When AI-related terminology is available in the searchable fields of their database records, the filters can be used to retrieve evidence on the efficacy and effectiveness of specific AI medical device interventions, discover AI medical device interventions for the purpose of 'horizon scanning', find evidence about evaluation methods for the interventions, or to retrieve evidence about safety and regulatory issues related to the technologies.

Limitations

As with any search filter, a limitation of the NICE AI medical device filters is that they are unable to retrieve references that have poorly reported details

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

about AI technologies in the searchable fields of their database records. However, the filters were developed and validated using approximately 500 AI medical device intervention references and the authors are confident that the filters work effectively to retrieve research literature that has AI terminology of relevance to medical device interventions in the searchable fields of their MEDLINE and Embase records. As discussed above, it is advised that the filters should be used as part of a ‘multi-stranded’ literature search approach to find evidence on AI medical device interventions when details such as brand names or manufacturer names for the interventions are known.

A recognised limitation of using the relative recall method to create Gold Standard sets for search filter development is that identifying a relevant set of references is only as good as the original individual literature searches (8). However, to help mitigate against this, several systematic reviews on AI medical device interventions that used a variety of literature search approaches were used to source references for the Gold Standard set (see Supplementary File 5).

It is acknowledged that subjective decisions were made to select the candidate search terms for the filters and that frequency analysis was not conducted to inform their inclusion and combinations. Furthermore, it is acknowledged that the candidate search terms were not reviewed by people with expertise in the field of AI medical device interventions. However, as the references that were used to identify the search terms for the filter had been selected for inclusion for published NICE AI medical device intervention topics as well as systematic reviews on AI medical device intervention topics, the authors are confident that the terminology used for the filters is correct and relevant.

The NICE AI medical device intervention filters do not contain subject headings and it is possible that the addition of subject headings could have increased their recall. However, both the MEDLINE and Embase filters met

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

the >90percent recall target and so the omission of subject headings did not have a significant impact on the filters' retrieval effectiveness.

It is recognised that the fields of AI medical device intervention technologies and AI technologies in general are rapidly evolving, and it is possible that new AI terminology of relevance for the medical device interventions could come into prominence. The authors of this paper will endeavour to update the filters with new AI technology terminology as required.

Conclusion

The novel NICE AI medical device intervention filters retrieve evidence about the interventions effectively with high recall. It is also likely that the filters are more efficient than previously used untested search strategies for AI medical device interventions that included subject headings plus a wide range of indirect as well as direct terminology for AI technologies. They can be used by information professionals, researchers, and health care professionals for range of AI medical device intervention topics including the effectiveness of the interventions for specific health conditions, evaluation methods, or safety and regulatory issues.

Acknowledgements

None

Sources of Funding

This research received no specific funding from any agency, commercial or not-for-profit sectors.

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Conflicts of Interest: None.

Competing interests: The authors declare none.

Implications for practice

Prior to the development of the NICE AI medical device intervention search filters, it could be challenging to retrieve evidence about the interventions from bibliographic databases. As a result, relevant evidence about AI medical device interventions could have been missed in searches.

The NICE AI medical device intervention search filters improve searching practice for the interventions. The filters retrieve evidence about AI medical devices from MEDLINE and Embase (Ovid) with high recall. They can be applied to literature searches to retrieve evidence on a range of AI medical device topics including the effectiveness of the interventions for specific health conditions, evaluation methods, or safety and regulatory issues.

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

References

1. National Institute for Health and Care Excellence [Internet] Digital health. 2024 [cited 2024 December 03]. Available from: https://www.nice.org.uk/about/what-we-do/digital-health

2. National Institute for Health and Care Excellence [Internet] Corporate document [ECD7]: Evidence standards framework for digital health technologies. 2022 [cited 2024 December 03].. Available from: https://www.nice.org.uk/corporate/ecd7

3. Medicines and Healthcare products Regulatory Agency [Internet]. Software and Artificial Intelligence (AI) as a Medical Device. 2024. [cited 2024 December 03]. Available from: https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device

4. Bélisle-Pipon JC, Couture V, Roy MC, Ganache I, Goetghebeur M, Cohen IG. What Makes Artificial Intelligence Exceptional in Health Technology Assessment? *Front Artif Intell* [Internet]. 2021 Nov [cited 2024 December 03];2;4:736697. Available from: https://www.frontiersin.org/articles/10.3389/frai.2021.736697/full

5. NHSX [Internet] Artificial Intelligence: How to get it right Putting policy into practice for safe data-driven innovation in health and care. 2019 [cited 2024 December 03]. Available from: https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf

6. National Institute for Health and Care Excellence [Internet] PMG20 Developing NICE guidelines: the manual. 2024 [cited 2024 December 03]. Available from: https://www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

7. Ayiku, L., Hudson, T., Glover, S., Walsh, N., Adams, R., Deane, J., & Finnegan, A. The NICE MEDLINE and Embase (Ovid) health apps search filters: Development of validated filters to retrieve evidence about health apps. International Journal of Technology Assessment in Health Care [Internet]. 2021 [cited 2024 December 03];37(1), E16. Available from: https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/abs/nice-medline-and-embase-ovid-health-apps-search-filters-development-of-validated-filters-to-retrieve-evidence-about-health-apps/458C19CD00C034F336DD32AE6627231F

8. Jenkins, M. Evaluation of methodological search filters: A review. Health Info Libr J. [Internet]. 2004 Aug [cited 2024 December 03];21:148–63. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1471-1842.2004.00511.x

9. Glanville, J., Bayliss, S., Booth, A., Dundar, Y., Fernandes, H., Fleeman, ND., Foster, L., Fraser, C., Fry-Smith, A., Golder, S., Lefebvre, C., Miller, C., Paisley, S., Payne, L., Price, A., & Welch, K. So many filters, so little time: the development of a search filter appraisal checklist. J Med Libr Assoc [Internet]. 2008 Oct [cited 2024 December 03];96(4):356-61. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2568852/

10. Damarell, R.A., May, N., Hammond, S., Sladek, RM., & Tieman, JJ. Topic search filters: a systematic scoping review. Health Info Libr J [Internet]. 2019 December [cited 2024 December 03];36: 4-40. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/hir.12244

11. Sampson, M., Zhang, L., Morrison, A., Barrowman, NJ., Clifford, TJ., Platt, RW., Klassen, TP., & Moher, D. An alternative to the hand searching Gold Standard: validating methodological search filters using

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

relative recall. BMC Med Res Methodol [Internet]. 2006 July [cited 2024 December 03];18(6):33. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-6-33

12. Cooper, C., Dawson, S., & Lefebvre, C. Searching for medical devices - Practical guidance. Res Syn Meth [Internet]. 2022 [cited 2024 December 03];13(1):144-154. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1002/jrsm.1524

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Table 1: Recall of the draft MEDLINE filter against the Test Set

Database	No. of Test Set references	No. of Test Set references retrieved by filter	Recall
MEDLINE	66	53	80%

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Table 2: Recall of the revised filters against their Development Sets

Database	No. of Development Set references	No. of Development Set references retrieved by filter	Recall
MEDLINE	155	149	96%
Embase	146	143	98%

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Table 3: Validation of the NICE MEDLINE and Embase AI medical device intervention filters

Database	No. of Validation Set references	No. of Validation Set references retrieved by filter	Recall
MEDLINE	309	297	96%
Embase	300	289	96%

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press

Figure 1: Final NICE MEDLINE and Embase AI medical device intervention filters

1 algorithm*.ti,kf.

2 (algorithm* adj2 (learn* or automate* or detect* or predict* or treatment* or therap* or radiolog* or AI or DL or data or dataset* or base* or classif*)).ab.

3 artificial intelligen*.ti,ab,kf.

4 AI.ti,kf.

5 (machine adj2 learn*).ti,ab,kf.

6 machinelearn*.ti,ab,kf.

7 (deep adj2 learn*).ti,ab,kf.

8 deeplearn*.ti,ab,kf.

9 neural network*.ti,ab,kf.

10 (convolutional adj1 network*).ti,ab,kf.

11 automate*.ti.

12 (automate* adj3 (system* or score* or software* or analysis* or analyse* or risk* or evaluat* or tool* or detect* or process*)).ab,kf.

13 (vector machine* or svm*).ti,ab,kf.

14 radiomic*.ti,ab,kf.

15 ((supervised or unsupervised) adj3 (classifier* or prediction*)).ti,ab,kf.

16 or/1-15

https://doi.org/10.1017/S0266462324004823 Published online by Cambridge University Press