HTA Austria
Austrian Institute for
Health Technology Assessment
GmbH

Artificial Intelligence in Health Care with a Focus on Hospitals: Methodological Considerations for Health Technology Assessment

A Scoping Review

Final report

AIHTA Project Report No.: 164| ISSN: 1993-0488 | ISSN-online: 1993-0496

HTA Austria
Austrian Institute for
Health Technology Assessment
GmbH

Artificial Intelligence in Health Care with a Focus on Hospitals: Methodological Considerations for Health Technology Assessment

A Scoping Review

Vienna, November 2024

Project Team

Project leader: Gregor Goetz, MPH MSSc

Authors: Michaela Riegelnegg, MA
Doris Giess, MBBS MPH
Gregor Goetz, MPH MSSc

Project Support

Systematic literature search: Hendrikje Rödiger, MPH
Prof.in. Dr. Cornelia Henschke.
We thank H. Rödiger and C. Henschke (Technische Universität Berlin) for providing related full-text articles (scoping review; ASSESS-DHT).

Hand search: Doris Giess, Michaela Riegelnegg

Internal review: Priv.-Doz. Dr. phil. Claudia Wild

External Review: Mag. Dr. Alexander Degelsegger-Márquez

Correspondence: Michaela Riegelnegg, michaela.riegelnegg@aihta.at

Cover photo: @ TarikVision – stock.adobe.com

This Study was supported by
Generative AI (Claude by Anthropic, Version 3.5 Sonnet, 2024) provided assistance with several work steps.

This report should be referenced as follows:

Riegelnegg M, Giess D, Goetz G. Artificial Intelligence in Health Care with a Focus on Hospitals: Methodological Considerations for Health Technology Assessment. A scoping review. AIHTA Project Report No.: 142; Jahr. Vienna: HTA Austria – Austrian Institute for Health Technology Assessment GmbH.

Conflict of interest

All authors and the reviewers involved in the production of this report have declared they have no conflicts of interest in relation to the technology assessed according to the Uniform Requirements of Manuscripts Statement of Medical Journal Editors (www.icmje.org).

Disclaimer

The external reviewers did not co-author the scientific report and do not necessarily all agree with its content. Only the AIHTA is responsible for errors or omissions that could persist. The final version and the policy recommendations are under the full responsibility of the AIHTA.

IMPRINT

Publisher:
HTA Austria – Austrian Institute for Health Technology Assessment GmbH
Garnisongasse 7/Top20 | 1090 Vienna – Austria
https://www.aihta.at/

Responsible for content:
Priv.-Doz. Dr. phil. Claudia Wild, managing director

AIHTA Project Reports do not appear on a regular basis and serve to publicize the research results of the Austrian Institute for Health Technology Assessment.

AIHTA Project Reports are only available to the public via the Internet at http://eprints.aihta.at/view/types/hta_report.html.

AIHTA Project Report No.: 164

ISSN 1993-0488

ISSN online 1993-0496

Content

Executive Summary ... 10

Zusammenfassung ... 12

1 Introduction ... 16
1.1 Definitions ... 16
1.2 Artificial Intelligence in Health Care: Advancements, Challenges and regulation ... 17
1.3 Regulations ... 18
1.3.1 The EU AI Act ... 18
1.3.2 Medical Device Regulation ... 19
1.4 Artificial Intelligence in Austria ... 20
1.5 HTA Methodology: EUnetHTA Core Model ... 21
1.6 Objectives and Scope ... 25

2 Methods ... 27
2.1 Step 1: Identification of HTA methods guidance and assessments ... 27
2.1.1 Flow Diagram ... 28
2.2 Step 2: Analysis of HTA methods documents ... 28
2.3 Step 3: Analysis of HTA assessments: application areas and methodological approaches ... 29
2.4 Step 4: Practice recommendations for Austria ... 30

3 Results ... 31
3.1 HTA-Methods ... 31
3.1.1 NICE – Evidence standards framework ... 34
3.1.2 Framework from AQuAS ... 37
3.1.3 HAS guidance documents ... 39
3.1.4 Framework from FinCCHTA ... 41
3.1.5 HTW AI checklist ... 42
3.2 Assessments on AI-enabled digital health technologies ... 44
3.2.1 Diagnosis and Screening ... 44
3.2.2 Treatment ... 67
3.2.3 Prediction ... 69
3.3 Methodological considerations for the implementation of AI in Austria ... 74
3.3.1 Thematic analysis of AI-specific themes ... 74
3.3.2 Guide for procurement of AI-enabled DHTs ... 80

4 Discussion ... 85

5 Conclusions and recommendations ... 91

6 References ... 92

Appendix ... 99
Overview of HTA Institutions ... 99
Overview of methods documents and current use ... 100
Extraction Tables Assessments ... 102
Extraction tables methods ... 128
Overview Domains and questions to Themes ... 133
Thematic analysis of methods guidance documents ... 135
Search Strategy ... 136

List of figures

Figure 1-1: Connection between AI/machine learning ... 17
Figure 2-1: Flow chart of study selection (PRISMA Flow Diagram) ... 28
Figure 3-1: Evidence Standards Framework – Tiers ... 35
Figure 3-2: AI-specific themes ... 76
Figure 3-3: Guide for procurement decisions ... 84

List of tables

Table 1-1: Terminologies in the context of artificial intelligence ... 16
Table 1-2: PIC-Problem, Interest, Context for Inclusion of Information ... 26
Table 2-1: Definitions of Application Areas ... 29
Table 3-1: Overview of frameworks and checklists ... 33
Table 3-2: Overview of assessments by primary function of the assessed AI-enabled digital health technologies ... 44
Table 3-3: AI in Diagnosis – Radiology ... 48
Table 3-4: AI in Diagnosis – Internal Medicine ... 53
Table 3-5: AI in Diagnosis – Dermatology ... 56
Table 3-6: AI in Diagnosis – Ophthalmology ... 58
Table 3-7: AI in Diagnosis – Pathology ... 60
Table 3-8: AI in Diagnosis – Patient-Clinician-Interaction ... 63
Table 3-9: AI in Diagnosis – General Medicine ... 65
Table 3-10: AI in Diagnosis – Neurology ... 66
Table 3-11: AI in Treatment – Radiology ... 68
Table 3-12: AI in Prediction – Palliative Care ... 71
Table 3-13: AI in Prediction – Patient management ... 73
Table 3-14: Overview of AI specific HTA methods charted against standard HTA methods ... 79
Table 3-15: Checklist for decision-makers ... 82
Table A-1: HTA Institutions ... 99
Table A-2: Overview of institutions with published AI specific methods or utilising DHT frameworks with guidance on AI-enabled DHTs ... 100
Table A-3: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 1 ... 102
Table A-4: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 2 ... 106
Table A-5: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 3 ... 111
Table A-6: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 4 ... 114
Table A-7: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 5 ... 118
Table A-8: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 6 ... 121
Table A-9: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 7 ... 125
Table A-10: Overview of HTA Methodologies ... 128
Table A-11: Overview Domains and questions to Themes ... 133
Table A-12: Thematic analysis of methods guidance documents ... 135

List of abbreviations

AGI.....Artificial General Intelligence
AI.....Artificial Intelligence
AIHTA.....Austrian Institute for Health Technology Assessment
AIS.....Acute Ischemic Stroke
AMA.....American Medical Association
AMSTAR.....Assessment of Multiple Systematic Reviews
AQUAS.....Agency for Health Quality and Assessment of Catalonia
AUROC.....Area Under the Receiver Operating Characteristic curve
CADTH.....Canadian Agency for Drugs and Technologies in Health
CADe.....Computer Aided Detection
CADx.....Computer Aided Diagnosis
CASP.....Critical Appraisal Skills Programme
CAST.....Computer-Assisted Triage
CCT.....Controlled Clinical Trial
CEA.....Cost-Effectiveness Analysis
CMD.....Connected Medical Device
CNEDiMTS.....Medical Device and Health Technology Evaluation Committee (French)
CNN.....Convolutional Neural Network
CONSORT-AI....Consolidated Standards of Reporting Trials Artificial Intelligence
COPD.....Chronic Obstructive Pulmonary Disease
CORE-MD.....Coordinating Research and Evidence for Medical Devices
CRD.....Centre for Reviews and Dissemination
CT.....Computed Tomography
CTCA.....CT Coronary Angiography
CTP.....CT Perfusion
CUR.....Current Use
CXR.....Chest X-Ray
DALYs.....Disability-Adjusted Life Years
DBT.....Digital Breast Tomography
DGT.....Digital Health Technology
DHTC.....Danish Health Technology Council
DICOM.....Digital Imaging and Communications in Medicine
DL.....Deep Learning
DHT.....Digital Health Technology
DRTS.....Digital Retinopathy Tele-Screening
DTAC.....Digital Technology Assessment Criteria
DWMS.....Digital Wound Management System
EBM.....Evidence-Based Medicine
ECG.....Electrocardiogram
ECO.....Economic
ED.....Emergency Department
EDIHTA.....European Digital Health Technology Assessment
EEG.....Electroencephalogram
EFF.....Effectiveness
EHDS.....Electronic Health Data Space
EHR.....Electronic Health Record
ESF.....Evidence Standards Framework
ETH.....Ethical
EU.....European Union
FDA.....Food and Drug Administration
FinCCHTA.....Finnish Coordinating Center for Health Technology Assessment
GDPR.....General Data Protection Regulation
GÖG.....Gesundheit Österreich GmbH (Austrian National Public Health Institute)
GP.....General Practice/Practitioner
HAS.....Haute Autorité de Santé
HCP.....Healthcare Professional
HIS.....Healthcare Improvement Scotland
HRQoL.....Health-Related Quality of Life
HTA.....Health Technology Assessment
HTW.....Health Technology Wales
IACS.....Institute for Health Sciences of Aragon
ICER.....Incremental Cost-Effectiveness Ratio

ICU.....Intensive Care Unit
INAHTA.....International Network of Agencies for Health Technology Assessment
INESSS.....Institut National d'Excellence en Santé et en Service Social
LEG.....Legal
LLM.....Large Language Model
MCDA.....Multi-Criteria Decision Analysis
MDD.....Medical Devices Directive
MDR.....Medical Device Regulation
MHRA.....Medicines and Healthcare products Regulatory Agency
MI-CLAIM.....Minimum Information about Clinical Artificial Intelligence Modeling
ML.....Machine Learning
MRI.....Magnetic Resonance Imaging
NECA.....National Evidence-based Healthcare Collaborating Agency
NHS.....National Health Service
NICE.....National Institute for Health and Care Excellence
NIHR.....National Institute for Health Research
NN.....Neural Networks
NLP.....Natural Language Processing
NPV.....Negative Predictive Value
NRSI.....Non-Randomized Studies of Interventions
OAR.....Organs at Risk
ORG.....Organisational
PACS.....Picture Archiving and Communication System
PAF.....Paroxysmal Atrial Fibrillation
PPV.....Positive Predictive Value
PRISMA.....Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST.....Prediction Model Risk of Bias Assessment Tool
QALYs.....Quality-Adjusted Life Years
QUADAS.....Quality Assessment of Diagnostic Accuracy Studies
RCT.....Randomised Controlled Trial
RNN.....Recurrent Neural Network
ROB.....Risk of Bias
RX.....Radiology
SAF.....Safety
SaMD.....Software as a Medical Device
SHTG.....Scottish Health Technologies Group
SiMD.....Software in a Medical Device
SOC.....Social
SR.....Systematic Review
TEC.....Technical
UDI.....Unique Device Identification
UK.....United Kingdom
USA.....United States of America
WCAG.....Web Content Accessibility Guidelines

Glossary

Term	Definition	Reference
Algorithm	is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.	[1]
Algorithmic bias	describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others. It also occurs when an algorithm produces results that are systemically prejudiced due to erroneous assumptions in the machine learning process.	[1]
Data security	is the practice of protecting digital information from unauthorised access, corruption or theft throughout its entire lifecycle.	[2]
Data-driven digital health technology	contains algorithms trained using patient data or datasets. The algorithms could be adaptive or fixed. It could also use decision thresholds or cut-off values created using patient data or datasets.	[3]
Explainability	is the ability to link and explain the elements taken into account by the algorithm, for example the input variables, and their consequences, for example, on the prediction of a score, and thus on the decision. The explanations must be adapted to the comprehension level of the person for whom they are intended.	[4]
Interpretability	is the ability to render the operation of an artificial intelligence system comprehensible. An algorithm is “interpretable” when its operation is accurately understood, for example, when an expert system models a decision tree	[4]
Noisy data	is a data set that contains meaningless data.	[5]
Real-world data	includes information about the health of individuals or the delivery and/or outcomes of health care that is collected outside of traditional clinical trials and thus reflects results within the context of the particular health care system.	[6]
Real-world evidence	is evidence about the use, safety, and effectiveness of a medical product, technology, or drug that is based on or derived from analysis of data generated in a real-world health care setting.	[6]
Resilience	is the ability of the system to maintain its conformity with performance and/or security requirements in the presence of input data outside its range of use (e.g. due to a sensor fault)	[4]
Robustness	refers to the ability of a model to maintain its performance when faced with uncertainties or adversarial conditions. A robust model should be able to generalise well and provide reliable predictions even when dealing with unforeseen inputs or circumstances.	[7]
Synthetic data	is artificial data that is generated from original data and a model that is trained to reproduce the characteristics and structure of the original data.	[8]
Training (algorithm)	is a machine learning process through which the artificial intelligence system builds a model from data.	[4]
Validation	is a process consisting of testing, observing and optimising (hyperparameters) system behaviour during running to ensure in the range of use, that the output data are in line with the expected results	[4]

Executive Summary

Background

Artificial Intelligence (AI) designed to imitate human cognitive abilities

Artificial intelligence (AI) has evolved from its theoretical foundations to become increasingly prominent in healthcare applications. AI represents a machine-based system designed to imitate human cognitive abilities and is claimed to be able to make predictions, recommendations, or decisions with varying levels of autonomy. Current AI systems are primarily narrow or weak AI, specialised in specific tasks.

various applications for AI-enabled digital health technologies (DHTs)

As healthcare continues to digitalise, AI-enabled digital health technologies (DHTs) are assumed to be integrated across various medical applications, from image-based diagnostics to the analysis of electronic health records. While these technologies offer promising opportunities, they also present challenges related to data privacy and transparency. As AI-enabled DHTs become more prevalent in healthcare settings, questions have emerged regarding whether specialised methodological approaches are needed to evaluate their benefits for hospital procurement decisions.

methods to evaluate benefits for hospital procurement decisions unclear

Methods

methods: search in HTA- webpages, analysis of HTA guidance documents and assessments, recommendations

The study employed a four-step approach: 1) A targeted search in 51 health technology assessment (HTA) institutional webpages to identify methods guidance documents and assessments for AI-enabled DHTs; 2) Analysis of identified methods guidance documents to describe how to assess AI-enabled DHTs' benefits and identification of themes specific for AI; 3) Analysis of identified assessments focusing on applied methods and application areas; and 4) Development of recommendations for Austrian hospitals.

Results

13/51 HTA institutes published 5 guidance documents and 30 assessments

In 51 HTA-institutes 13 published in total five guidance documents and 30 assessments. The assessments primarily focused on diagnostic and screening AI-enabled DHTs (27/30), particularly in radiology (10/27) and internal medicine (7/27). Treatment applications (1/30) were linked to radiotherapy, whilst prediction applications (2/30) were associated with palliative care and patient management. Most AI-enabled DHTs were positioned as assistive tools performing sub-tasks with low-level autonomy and requiring healthcare professional oversight rather than fully autonomous systems.

standard HTA methods apply for some domains

The analysis of the guidance documents showed that while standard HTA methods generally apply for evaluating aspects of comparative effectiveness and safety, AI-specific considerations are needed for:

AI-specific considerations needed for technical aspects and post-deployment monitoring

- technical aspects (training data, data quality),
- ethical considerations (algorithmic bias),
- organisational aspects (human oversight),
- post-deployment monitoring and re-evaluation.

mostly retrospective observational studies were used for the evaluation of AI

The analysis of the HTA reports revealed that primarily retrospective observational studies were used to validate the AI algorithms. The evaluations encompassed both standard methodological approaches – examining clinical effectiveness, organisational aspects, and economic aspects – as well as AI-

specific considerations, particularly focusing on the main functions of AI-enabled DHTs and the characteristics of their training data. However, across most assessments, the available evidence was insufficient to draw robust conclusions about the technologies' benefits.

comprehensive checklist for health care decision-makers incorporating purpose, regulatory requirements, HTA evaluation and post-deployment monitoring

Additionally, based on the analysis of methods guidance documents a comprehensive checklist for health care decision-makers was developed to guide AI procurement, encompassing four main areas: purpose definition, regulatory requirements, evidentiary requirements, and monitoring. For regulatory compliance, AI systems in healthcare must meet requirements from multiple frameworks: the EU AI Act (typically classifies AI in healthcare as "high-risk"), medical device regulation (MDR) (generally Class IIa or higher for diagnostic/therapeutic decisions), and the General Data Protection Regulation (GDPR). The checklist emphasises the importance of validating both "standard" considerations (such as clinical benefits, safety, and implementation requirements) and AI-specific aspects (including training dataset quality, bias mitigation strategies, and human oversight). Regular monitoring and re-evaluation procedures were identified as crucial components for maintaining AI-system performance over time.

Discussion

standard HTA methods may be suitable for some domains, AI-specific components are needed for technical aspects

Limited adaptation of HTA methods for AI-enabled DHTs was observed. Few HTA institutes have developed or applied AI-specific methodological guidance. Standard HTA methods appear suitable for evaluating some aspects of comparative effectiveness and safety, whilst AI-specific considerations are particularly relevant for technical performance, ethical implications, and organisational aspects. The need for agile use of evidence-based medicine methods across the lifecycle of AI- enabled DHTs was highlighted, particularly given that AI algorithms can change through updates requiring re-evaluation.

digital infrastructure a prerequisite for successful implementation of AI-enabled DHTs

The successful adoption of AI-enabled DHTs in Austrian hospitals is closely tied to the country's health data infrastructure. Many digital health technologies currently used in Austria operate as isolated systems, with data confined to individual hospitals. A sophisticated digital infrastructure with high-level interoperability is often a prerequisite for AI-enabled DHTs to work as anticipated.

Conclusion

conventional HTA methods as a basis together with AI-specific considerations

For the assessment of AI-enabled DHTs, conventional HTA methods may suffice as a basis for some assessment domains. However, AI-specific considerations are necessary, particularly when evaluating technical characteristics (e.g., quality of training data), ethical aspects (e.g., algorithmic biases), and organisational impacts (e.g., human oversight, post-implementation monitoring).

recommendations for Austria include the usage of existing DHT frameworks and AI-specific supplements

For Austrian decision-makers, it is recommended to use existing frameworks for digital health technologies as a starting point for evaluation, and to supplement them with AI-specific components from available guidance documents. The EUnetHTA Core Model, together with supplementary guidance documents, provides a practical toolkit for evaluating AI-enabled DHTs for procurement decisions.

Zusammenfassung

Hintergrund

Künstliche Intelligenz (KI) zur Imitation menschlicher Denkprozesse

Künstliche Intelligenz (KI) als maschinelles System zur Imitation menschlicher Denkprozesse gewinnt im Gesundheitswesen zunehmend an Bedeutung. Insbesondere im Krankenhaus kann KI bei spezifischen Aufgaben wie der bildbasierten Diagnostik oder der Analyse von Gesundheitsdaten unterstützend eingesetzt werden.

KI-Systeme sind schwacher KI zuzuordnen

KI-Systeme sind in der Regel schwacher KI zuzuordnen, können abhängig von ihren Trainingsdaten eine spezifische Aufgabe lösen und sich limitiert adaptieren.

Der regulatorische Rahmen für KI-Gesundheitstechnologien wird maßgeblich durch zwei Verordnungen bestimmt:

EU AI Act als Regulierungsrahmen

Der *EU AI Act* etabliert erstmals einen risikobasierten Regulierungsrahmen für KI-Systeme. Die Regulierung unterscheidet dabei drei zentrale Risikokategorien

1. Als *inakzeptables Risiko* werden Systeme eingestuft, die gegen fundamentale Grundrechte verstoßen.
2. Der Kategorie *hohes Risiko* werden KI-Anwendungen in kritischer Infrastruktur, im Bildungswesen und Gesundheitssektor zugeordnet
3. In der Kategorie *geringes/minimales Risiko* fallen unter anderem generative KI-Modelle wie ChatGPT.

Risikoeinstufung von KI-Anwendungen je nach Funktion in der Medizinprodukteverordnung

Die seit 2021 gültige *Medizinprodukteverordnung* (medical device regulation – MDR), für die eine Übergangsregelung bis 2027/2028 besteht, definiert Software einschließlich KI-Algorithmen als potenzielles Medizinprodukt. Die Verordnung führt eine risikobasierte Klassifizierung von Klasse I (geringes Risiko) bis Klasse III (hohes Risiko) ein. KI-Gesundheitstechnologien werden dabei je nach ihrer spezifischen Funktion und ihrem Verwendungszweck in die Risikoklassen IIa bis III eingestuft.

KI-Anwendungen als unterstützende Systeme

Unabhängig von der Regulatorik, werden KI-Gesundheitstechnologien zusätzlich nach ihrem Autonomiegrad klassifiziert, wobei derzeit hauptsächlich assistierende oder unterstützende Systeme zum Einsatz kommen und demnach die Entscheidung weiterhin beim medizinischen Fachpersonal bleibt.

Unsicherheiten in der Methodik zur Nutzenbewertung

Es besteht jedoch Unsicherheit darüber, welche methodischen Ansätze für die Nutzenbewertung von KI-Gesundheitstechnologien im Kontext von Investitionsentscheidungen in österreichischen Krankenhäusern am besten geeignet sind.

Zielsetzung

Ziel war die Identifizierung und Analyse von HTA-Methoden und Berichten zur Entscheidungsunterstützung

Ziel dieses Berichts war die Identifizierung und Analyse verfügbarer methodischer Ansätze zur Bewertung des (Zusatz-)Nutzens von KI-Gesundheitstechnologien zur Unterstützung für Entscheidungsprozesse in Krankenhäusern. Im Fokus standen dabei sowohl existierende Health Technology Assessment (HTA) Methodendokumente als auch HTA-Berichte.

Methode

Scoping Review und qualitative Dokumentenanalyse in 4 Schritten

Es wurde ein Scoping Review in Kombination mit einer qualitativen Dokumentenanalyse durchgeführt. Es kamen dabei vier methodische Schritte zum Einsatz: (1) Eine fokussierte Handsuche auf den Webseiten von 51 HTA-Instituten bis Mai 2024 sowie die Einbeziehung der Ergebnisse einer systematischen Literatursuche des ASSESS DHT-Projekts dienten der Identifikation

relevanter Dokumente. (2) Die HTA-Methodendokumente wurden narrativ gegenübergestellt und thematisch analysiert, um KI-spezifische Inhalte zu identifizieren, die über das EUnetHTA Core Model hinausgehen. (3) Die Analyse der HTA-Berichte erfolgte mittels standardisierter Extraktionstabellen und erfasste zentrale Aspekte wie Technologie, Komparator sowie methodische und KI-spezifische Bewertungskriterien. Die Ergebnisse wurden nach primärer Funktion gemäß MDR und medizinischem Fachbereich in Form von Vignetten kategorisiert. (4) Basierend auf den Erkenntnissen wurden Empfehlungen für die Evaluation und Implementierung von KI-Gesundheitstechnologien in österreichischen Krankenhäusern formuliert.

Resultate

5 Methodenhandbücher und 30 HTA-Berichte identifiziert

Es wurden fünf Methodendokumente und 30 HTA-Berichte von 13 HTA-Institutionen identifiziert. Bei den Methodendokumenten handelt es sich um:

Framework für DHTs – z. B. der Evidence Standards Framework des National Institute for Health and Care Excellence

- Das „**Evidence Standards Framework**“ **(ESF)** des *National Institute for Health and Care Excellence (NICE*), ein Framework zur Bewertung digitaler Gesundheitstechnologien mit spezifischen Standards für KI-Komponenten.
- Das **Assessment Framework** der *Agency for Health Quality and Assessment of Catalonia (AQuAS),* ein Framework zur Bewertung digitaler Gesundheitstechnologien einschließlich KI, bestehend aus 13 Domänen und 41 Dimensionen.
- Das **DIGI-HTA Framework** des *Finnish Coordinating Center for Health Technology Assessment (FinCCHTA*), ein Framework für die Schnellbewertung sich entwickelnder Technologien mit einer spezifischen KI-Domäne.
- Ein **HTA-Methodenhandbuch** der französischen *Haute Autorité de Santé (HAS)* mit ergänzenden KI-spezifischen Empfehlungen und einem Fragebogen mit 42 KI-spezifischen Fragen.
- Eine **KI-spezifische Checkliste** von *Health Technology Wales (HTW),* die zusätzlich zu Standard-HTA-Methoden verwendet werden soll, mit vier spezifischen Bewertungsdomänen für KI.

Anforderungen an die Evidenz sind abhängig von der Kategorisierung

Einige Assessment Frameworks kategorisieren digitale Gesundheitstechnologien nach ihrer Funktion, wobei die Anforderungen (z. B. an die Evidenz) von dieser Kategorisierung abhängen. Das NICE Evidence Standards Framework unterscheidet beispielsweise drei Kategorien:

Tier A: (ohne direkten Patient*innenkontakt) bis Tier C (Technologie zur Diagnose und Behandlung)

- *Tier A:* Technologien ohne direkten Patient*innenkontakt und ohne mögliche gesundheitliche Schäden (z. B. administrative Systeme)
- *Tier B:* Technologien mit Patient*innenkontakt aber geringem Schadenspotenzial (z. B. Gesundheitsinformation, Dokumentation von Symptomen, einfache Monitoringfunktionen)
- *Tier C:* Technologien zur Diagnose und Behandlung von Erkrankungen oder zur aktiven Überwachung von Gesundheitsparametern, die einen direkten Einfluss auf Patient*innenentscheidungen haben und damit ein erhöhtes Schadenspotenzial aufweisen.

21 Standards müssen bei Tier C-Technologien erfüllt werden

Für Tier C gelten die striktesten Evidenzanforderungen: Alle 21 Standards des Frameworks müssen erfüllt werden, während für Tier A und B nur ausgewählte Standards relevant sind.

KI-spezifische Aspekte

Die Analyse der Dokumente zeigte mehrere KI-spezifische Aspekte, die bei einer Bewertung berücksichtigt werden sollten:

technische Aspekte: u. a. Informationen zur Funktion und Modell

Für die **technischen Charakteristika (TEC)** wurden insbesondere Aspekte zur KI-Funktion und zu den verwendeten Daten als wichtig erachtet:

- Informationen zur KI-Funktion und zum verwendeten Modell,
- Qualität der Trainingsdaten,
- Aspekte zum implementierten Algorithmus.

Sicherheit und ökonomische Aspekte

Im Bereich der **Sicherheit** wurde das Datenrisikomanagement als spezifischer Aspekt identifiziert. Bei den **ökonomischen Aspekten** wurden die Supportkosten als zusätzlicher KI-spezifischer Kostenfaktor genannt.

Ethik, Recht und Organisation (z. B. algorithmische Verzerrung)

In den Bereichen **Ethik, Recht und Organisation** wurden folgende KI-spezifische Themen hervorgehoben:

- Algorithmische Verzerrungen (ETH),
- Datenschutz (LEG),
- Menschliche Kontrolle (ORG).

30 HTA-Berichte zu KI-Anwendungen in Diagnostik und Screening, Behandlung und Vorhersage

Die 30 identifizierten HTA-Berichte konzentrierten sich hauptsächlich auf die Bereiche Diagnostik und Screening (27/30), insbesondere in der Radiologie (10/27) und Inneren Medizin (7/27). Therapiebezogene Anwendungen (1/30) betrafen die Bestrahlungsplanung, während prädiktive Anwendungen (2/30) in der Palliativmedizin und im Patientenmanagement zum Einsatz kamen.

Erwartungen u. a. gesteigerte Effizienz und verbesserte diagnostische Genauigkeit

Die Erwartungen an KI-Gesundheitstechnologien umfassen:

- gesteigerte Effizienz durch reduzierte Arbeitsbelastung und Wartezeiten,
- verbesserte diagnostische Genauigkeit,
- optimierte Arbeitsabläufe und Prozesse,
- verbesserte Patient*innenergebnisse und Zugang zur Versorgung.

Standardmethoden für Wirksamkeit und Sicherheit, zusätzliche KI-spezifische Überlegungen für die Beschreibung der Technologie und Ethik

In den HTA-Berichten wurden dieselben Methoden wie für herkömmliche Gesundheitstechnologien für die Bewertung der vergleichenden Wirksamkeit und Sicherheit verwendet, wenngleich KI-spezifische Überlegungen insbesondere in der Beschreibung der Technologie und Trainingsdaten (Algorithmus-Validierung) und bei ethischen Überlegungen (algorithmische Verzerrung) erforderlich sind.

Diskussion

wenige HTA-Institutionen haben spezielle Methoden für die Bewertung von KI entwickelt

Die Analyse zeigt, dass bisher nur wenige HTA-Institutionen spezielle Methodendokumente für die Bewertung von KI-Gesundheitstechnologien entwickelt haben. Während die standardmäßigen HTA-Methoden für einige Bewertungsdomänen geeignet erscheinen, erfordern vor allem technische, ethische und organisatorische Aspekte zusätzliche KI-spezifische Überlegungen.

fortlaufende Entwicklung der Algorithmen als Herausforderung

Eine besondere Herausforderung stellt die fortlaufende Entwicklung der KI-Algorithmen dar, die regelmäßige Update-Assessments erforderlich macht. Hierfür ist es wichtig, die Methoden der evidenzbasierten Medizin während des gesamten Produktlebenszyklus einzusetzen. Zudem ist die erfolgreiche Integration von KI-Gesundheitstechnologien stark von der vorhandenen digitalen Infrastruktur abhängig.

43-KI-Anwendungen in Österreich im Einsatz (2022)

Eine Erhebung der Gesundheit Österreich GmbH (2022) zeigt, dass in Österreich bereits 43 KI-Gesundheitstechnologien im Einsatz sind, mehrheitlich in der Diagnostik (56 %). Die Nutzung ist regional unterschiedlich verteilt, mit einer Konzentration in Wien, Oberösterreich, Tirol, Salzburg und der Steiermark.

Schlussfolgerungen

Standard-HTA-Methoden für einige Bewertungsdomänen, KI-spezifische Ergänzung erforderlich

Für die Bewertung von KI-Gesundheitstechnologien können für einige Bewertungsdomänen herkömmliche HTA-Methoden als Basis herangezogen werden. Allerdings sind KI-spezifische Ergänzungen erforderlich, insbesondere bei der Bewertung der technischen Eigenschaften (z. B. Qualität der Trainingsdaten), ethischer Aspekte (z. B. algorithmische Verzerrungen) und organisatorischer Auswirkungen (z. B. menschliche Aufsicht, Überwachung nach Implementierung).

Empfehlung: Etablierte Methodik ergänzt um KI-Komponenten

Für österreichische Krankenhäuser wird empfohlen, das EUnetHTA Core Model und bestehende Frameworks für digitale Gesundheitstechnologien wie das NICE Evidence Standards Framework oder das weiterentwickelte Framework von AQuAS als Ausgangspunkt zu nutzen und diese mit KI-spezifischen Komponenten aus anderen identifizierten Dokumenten zu ergänzen.

1 Introduction

Künstliche Intelligenz (KI) als maschinelles System zur Imitation menschlicher kognitiver Fähigkeiten

Artificial intelligence (AI), a branch of computer science focused on replicating human cognitive abilities, is rapidly gaining prominence across various sectors. In recent years, there has been a significant surge in scientific publications related to AI, especially in the medical context [9, 10].

1.1 Definitions

1955: Grundstein der Forschung durch das Dartmouth Projekt

The concept of AI was first formally introduced in 1955 with the proposal for the Dartmouth Summer Research Project on AI [11]. The project aimed to explore how a machine could precisely describe and simulate human intelligence. While the workshop did not result in immediate breakthroughs in the field of AI, it laid the foundations for research on AI, covering a broad range of topics such as problem-solving, neural networks, machine learning or language processing [12].

Table 1-1: Terminologies in the context of artificial intelligence

Term	Definition
Artificial Intelligence (AI)	is a machine enabled to imitate human intelligence/a range of techniques that allow computers to perform tasks typically thought to require human reasoning and problem-solving skills [13].
Machine Learning (ML)	is a subfield of AI – a machine can learn to perform given tasks [14].
Deep Learning (DL)	is a subset of machine learning with the primary difference in how the algorithm learns and how much data each type of algorithm uses [15].
Neural Networks (NN)	are computing systems that form the foundation of deep learning algorithms. These networks belong to the broader field of machine learning and are designed to process information similar to how brain neurons communicate with each other [15].
Natural Language Processing (NLP)	is a form of machine learning which recognises speech patterns, enabling it to understand and generate texts. [16].
Large Language Model (LLM)	is a form of deep learning based on large amounts of data that enables it to understand and generate natural language and other content to perform tasks [17].
Convolutional Neural Networks (CNN)	is a form of deep learning that is used for image classification and recognition [18].
Recurrent Neural Networks (RNN)	uses sequential data or time series data. It is a deep learning form commonly used for language translation or speech recognition [19].

KI als maschinelles System mit Vorhersage- und Entscheidungsfähigkeit

derzeit spezialisiert auf Einzelaufgaben

KI erkennt Muster (diskriminativ) oder erzeugt Inhalte (generativ)

More recent studies define AI as a machine-based system that can make predictions, recommendations, or decisions operating with varying levels of autonomy [5] or as a range of techniques allowing computers to perform tasks that typically require human reasoning skills [13]. AI can also be differentiated between strong and weak or narrow AI [20]. Strong AI or artificial general intelligence (AGI) refers to the hypothetical ability to match or exceed human-level intelligence across various cognitive tasks. As of today, AGI remains a theoretical concept that has not been realised. By contrast, most currently available AI systems can be considered narrow or weak. This means that these systems excel in a specific task with limited adaptability and no self-awareness, and they are heavily dependent on training data. Examples of narrow AI

include voice assistants, chess-playing programs, natural language processing, and self-driving cars. Hence, weak AI can be important in assisting [21, 22]. Another distinction between AI categories is discriminative and generative [23]. While generative AI produces new content, discriminative AI seeks to identify patterns and categorise them [23]. Within AI, there are several subcategories, such as machine learning and deep learning. Table 1-1 gives an overview of terms related to AI and Figure 1-1 shows the connection between the categories.

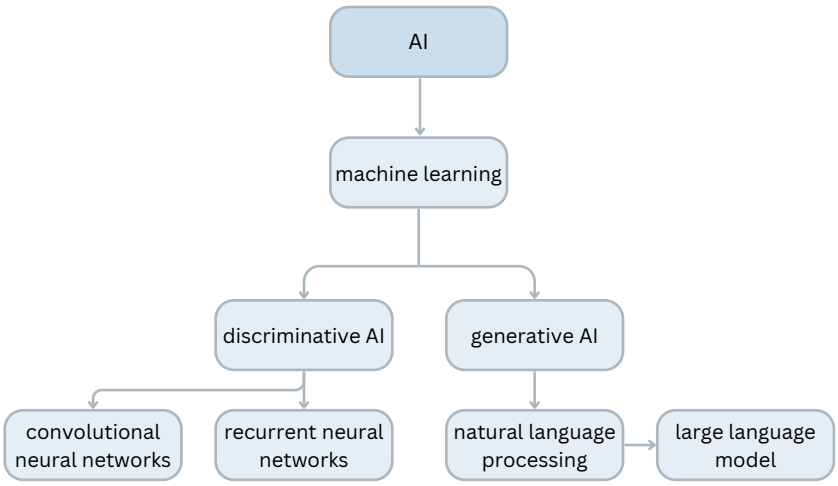


Figure 1-1: Connection between AI/machine learning

1.2 Artificial Intelligence in Health Care: Advancements, Challenges and regulation

Phasen der Digitalisierung – Elektronische Datenakte bis KI-Integration

Healthcare digitalisation has evolved through three significant phases. The first wave introduced electronic medical records and laboratory systems, laying the foundation for digital health infrastructure. The second phase saw the rise of mobile health services, expanding access and patient engagement. Entering the third wave, the integration of AI transforms healthcare by providing enhanced decision support, enabling personalised care, and automating tasks. [24, 25].

KI als Gesundheitstechnologie u. a. bildbasierte Diagnostik, Analyse von Gesundheitsdaten

In the meantime, AI is already utilised in various areas within the healthcare sector [26]. Machine learning and deep learning may be valuable for image-based diagnostics in fields like dermatology and radiology. AI can also analyse electronic health records for predictive modelling, assist in drug development, and power patient-facing tools like chatbots. Natural language processing enhances medical literacy and patient intake processes [27]. In these cases, AI is considered a health technology or digital health technology (DHT)

Klassifizierung nach Autonomie

One crucial aspect to consider is whether AI systems are autonomous in the tasks they perform when used in clinical care. The American Medical Association (AMA) classifies AI applications in medical services into three main categories [28]:

KI kann bei Einzelaufgaben assistieren oder gewisse Arbeitsschritte mit unterschiedlichen Graden der Autonomie unterstützen

- **Assistive**: AI detects relevant data without analysis. Requires physician interpretation.
- **Augmentative**: AI analyses and quantifies data, providing meaningful output. Requires physician interpretation.
- **Autonomous:** AI interprets data and generates conclusions independently. Requires physician conclusion.

Level 1 (Aufzeigen von Handlungsoptionen) bis Level 3 (Schlussfolgerungen, die Kliniker*innen anfechten können)

Autonomous AI is further subdivided into three levels depending on its autonomy:

- **Level I:** AI offers contestable options requiring physician action.
- **Level II:** AI initiates options with alert/override opportunity.
- **Level III:** AI initiates management; the physician must contest if needed.

This classification system, introduced in September 2021, guides the categorisation of AI-enabled DHTs based on the degree of machine involvement and required physician interaction.

Herausforderungen von KI stellen Datenschutz, mangelnde Transparenz der Algorithmen dar

derzeit fehlende Methodik zur Nutzenbewertung

Although AI offers promising opportunities, risks and challenges are associated with it [12, 29]. One of the biggest concerns is the security and privacy of sensitive health data processed by these systems. Another challenge is the lack of transparency and explainability of some AI algorithms, making it difficult to understand how decisions are made. Due to these challenges, AI systems used in healthcare are subject to market approval regulations for high-risk medical devices. Providers of these systems must implement a risk management and quality management system, along with a data governance protocol, and provide technical documentation and user instructions [12]. However, there is currently a lack of methodological guidance for assessing the benefits of AI systems in different fields of application in hospitals.

1.3 Regulations

Regulierung in der Europäischen Union (EU)

This section will describe two regulations defined by the European Union (EU) that are relevant to AI-enabled DHTs. These include the EU AI Act and the Medical Device Regulation (MDR).

1.3.1 The EU AI Act

2024: EU-AI Act Regulierung basierend auf Risikolevel:

inakzeptabel (Verstoß gegen Grundrechte)

Given the rapid development of AI, regulations for its use are crucial. The EU AI Act 2024 established the first framework for regulating AI systems based on their risk levels: unacceptable, high, and low/minimal risk [30].

- **Unacceptable risk:** Systems that violate fundamental rights, such as those capable of cognitive behavioural manipulation, social scoring, and specific biometric identification systems, will be banned. In severe cases, exceptions may apply to law enforcement to achieve a substantial public interest.

- **High-risk AI systems:** These include AI used in critical infrastructure, education, healthcare, employment, and law enforcement. Such systems will be rigorously assessed before and after market release to ensure safety and compliance. Users can file complaints about high-risk systems with national authorities. AI-enabled DHTs are considered high-risk AI systems.

hoch (in u. a. Bildung, Gesundheitswesen)

- **Low/minimal risk:** Generative AI models like ChatGPT, while not classified as high-risk, must adhere to transparency and copyright requirements, such as disclosing AI-generated content and preventing illegal content creation. High-impact general-purpose AI models, which could pose systemic risks, will undergo thorough evaluations and must report serious incidents to the European Commission. Additionally, any AI-modified content, such as deepfakes, must be clearly labelled to inform users.

gering/minimal (generative KI (z. B. ChatGPT))

This layered approach aims to establish AI-enabled DHTs that are safe and transparent, aligning with the EU's commitment to protecting fundamental rights and public safety. Requirements for deployers of high-risk AI systems include the assignment of human oversight to natural persons who have the necessary competence, compliance with the registration obligations, as well as the application for notification by conformity assessment bodies [30].

Ziel: sichere und transparente KI-Gesundheitstechnologien

In addition to the EU AI Act, the European Commission created the European Health Data Space (EHDS) [31]. It is a comprehensive EU initiative that aims to transform how health data is accessed and shared across Europe. Operating on two key levels, it addresses both primary and secondary use of health data. For primary use, the EHDS enables citizens to control and access their electronic health data securely across borders, including medical records, e-prescriptions, and laboratory results. For secondary use, it establishes a framework for utilising health data in research, innovation, policy-making, and regulatory activities under strict security and privacy conditions.

Europäischer Raum für Gesundheitsdaten (EHDS): EU-weiter Zugang und Nutzung von Gesundheitsdaten

1.3.2 Medical Device Regulation

The EU Medical Device Regulation (EU-MDR 2017/745), effective since 2021, is a comprehensive set of regulations governing the production and distribution of medical devices in the EU, aiming to replace the previous Medical Devices Directive (MDD). Due to implementation challenges, the full adoption of the EU-MDR has been postponed several times, resulting in a transitional arrangement that allows some medical devices regulated under the previous directive to be used without proof of compliance with the new regulation until 2027/2028, depending on the specific risk class of the device.

2017: Medizinprodukteverordnung (EU-MDR) – in Kraft seit 2021

Übergangsregelung bis 2027/2028

The EU-MDR introduces fundamental changes in three key areas: clinical evidence requirements, the involvement of independent expert panels, and the role of notified bodies, including enhanced competency requirements. The EU-MDR has particular significance for AI-enabled DHTs, as it explicitly recognises software, including AI algorithms, as potential medical devices [32-34]. This means that many AI-based health solutions may fall under the MDR's scope, requiring compliance with its standards related to, for instance, safety, effectiveness, and transparency.

Einführung neuer Anforderungen an die klinische Evidenz

Relevanz für KI-Gesundheitstechnologien

Under both regulations, a medical device is defined as "any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human

eine Software kann, je nach Funktion, ein Medizinprodukt sein

beings" for diagnosis, prevention, monitoring, prediction, prognosis, treatment, or alleviation of a disease, injury, or disability", According to the MDR, software is also considered an active medical device[1] [33].

Klassifizierung nach potenziellem Risiko von I (geringes Risiko) bis III (hohes Risiko)

The MDR classifies medical devices into classes ranging from I to III based on their potential risk [35]:

- Class I: lowest risk devices, including most non-invasive devices
- Class IIa: low to medium-risk devices
- Class IIb: medium to high-risk devices
- Class III: highest-risk devices

abhängig von Funktion und Zweck: KI als Risikoklasse IIa-III

Software can be allocated to all classes depending on its function and purpose. It is generally classified as IIa if intended to aid in diagnostic or therapeutic decisions [32]. It can also be classified as class III if it could lead to death or irreversible health deterioration [32].

Ziel: genauere Evaluierungsprozesse

This classification system is designed to increase the likelihood that high-risk devices undergo more stringent conformity assessment procedures [35].

Practical Implications

KI-Gesundheitstechnologien unterliegen der EU-Regulierung

An AI-enabled DHT used for diagnostic or therapeutic clinical decision-making in a hospital must comply with the MDR and the EU AI Act. This includes classifying the system as a medical device and adhering to high-risk AI systems' transparency and safety requirements [32], including the CE-marking through a decentralised notified body [36]. For healthcare providers and AI developers, these regulations imply a need for comprehensive risk management and quality management systems, along with data governance protocols. They must also provide technical documentation and user instructions. Meeting these standards is essential for market approval and the continued use of AI systems in healthcare [32].

1.4 Artificial Intelligence in Austria

Expert*inneninterviews (2022)

The Austrian National Public Health Institute (Gesundheit Österreich GmbH, GÖG) conducted expert interviews in 2022 to understand how AI is currently used in Austrian hospitals, experimental or implemented [37].

43 KI-Gesundheitstechnologien eingeteilt in Diagnostik, Behandlungsverbesserung und (Risiko-)Vorhersage

The study identified 43 distinct AI-enabled DHTs. These were categorised into three primary application areas: diagnostics (24 products, 55.8%), treatment improvements (12 products, 27.9%), and (risk) prediction (eight products, 18.6%). However, not all these AI-enabled DHTs are operational within the Austrian healthcare system. The implementation status varies: 17 products (39.5%) are in regular operation, 15 (34.9%) are in the pilot project phase, and 11 (25.6%) are either part of ongoing studies or were rapidly introduced in response to the COVID-19 pandemic [37].

[1] Active device means any device, the operation of which depends on a source of energy other than that generated by the human body for that purpose, or by gravity, and which acts by changing the density of or converting that energy [35].

Einsatz von KI vor allem im Zuge von Pilot- oder Forschungsprojekten

unterschiedliche Verteilung in den österreichischen Bundesländern

Adopting AI projects in healthcare varies significantly across Austria's nine federal states, revealing a notable disparity in implementation. AI is utilised most frequently in Vienna and Upper Austria, with 26 and 23 AI projects, respectively. Tyrol, Salzburg, and Styria each have 20 AI initiatives, indicating a robust embrace of this technology in these regions. However, there is a marked decrease in AI project numbers in other states: Lower Austria has implemented six projects, while Burgenland has four. The adoption rate further diminishes in Vorarlberg with three projects, and Carinthia shows only two initiatives for AI-enabled DHT [37]. This increasing adoption of AI systems in Austrian hospitals highlights the need for suitable assessment methods.

1.5 HTA Methodology: EUnetHTA Core Model

EUnetHTA Core Model als Standardwerk für HTA-Bewertungsmethoden

Fokus: Studiendesigns, Endpunkte und Synthesemethoden

The Health Technology Assessment (HTA) methodology encompasses various approaches to evaluate different health technologies. The EUnetHTA Core Model provides a framework for conducting and sharing HTA information, serving as a foundation for international best practices [38]. This chapter describes HTA methods, focusing on appropriate study designs, relevant endpoints, and synthesis methods, and highlights the medical technology's different methodological approaches according to intention (therapeutic, diagnostic or screening) [38].

therapeutische Medizinprodukte

- Therapeutic medical devices are technologies used to treat or manage diseases or conditions. Therapeutic medical devices can be used alone or with other medical devices to support, modify, replace or restore biological functions or structures [39].

Diagnostika können Triage Tests, Ersatztests und Zusatztests zur Optimierung der Diagnostik sein

- Diagnostic technologies confirm and monitor diseases, while screening technologies assess populations to identify potential health risks [38]. Based on pre-test probability, diagnostic tools serve three main roles: triage tests quickly rule out conditions in low-risk cases, replacement tests improve upon existing methods, and add-on tests provide detailed confirmation when risk levels increase. This framework optimises diagnostics by balancing efficiency, accuracy, and cost.

Screeningverfahren haben meist ein hohes Verzerrungsrisiko

- Screening programs assess populations for potential health risks to improve outcomes [38]. Their effectiveness depends on disease factors, test accuracy, and treatment success. Fundamental biases affect screening evaluation: healthy screened bias (participants tend to be healthier), length-time bias (detection favours slower diseases), lead-time bias (earlier diagnosis appears as more prolonged survival), and overdiagnosis (finding harmless abnormalities).

Assessment Domains

EUnetHTA Bewertungsdomänen

The EUnetHTA Core Model consists of a set of domains for which specific research questions, possible outcome measures, study designs, and synthesis methods are defined [38]. Herein, the domains are described as relevant, focusing on areas involving the assessment of AI-enabled therapeutic medical devices and diagnostic/screening technologies.

Clinical Effectiveness (EFF)

Study Design:

Clinical effectiveness assessment prioritises RCTs, distinguishing between efficacy (explanatory) and effectiveness (pragmatic) designs [38]. When RCT data is insufficient or unavailable, inclusion criteria may need to be expanded towards non-randomised studies of interventions (NRSI). For diagnostic technologies, the EUnetHTA Core Model recommends including diagnostic accuracy studies to evaluate a test's ability to correctly identify disease presence or absence compared to a reference standard. Linked evidence may be used when direct evidence is unavailable. For change-in-management assessment, diagnostic before-after studies and time series are suggested. Registry data reflects routine care. Direct comparisons are preferred, but indirect methods may be used for limited head-to-head evidence.

klinische Effektivität (EFF)
Studiendesigns: vorrangig RCTs

wenn keine RCTs verfügbar: andere Studiendesigns; z. B. diagnostische Genauigkeitsstudien und Kohortenstudien

For screening technologies, analysing results from multiple screening intervals helps estimate long-term effectiveness, and modelling studies are valuable for comparing various screening strategies [38]. Time trend studies analysing changes in disease frequency can be helpful but have potential biases. Case-control studies can compare different screening policies but cannot reliably estimate the difference between screening and no screening.

Modellierungen v. a. zur Evaluation von Screeningverfahren sinnvoll

Endpoints:

Primary endpoints include *mortality*, *morbidity*, and *quality of life* [38]. For diagnostic technologies specifically, test accuracy metrics, such as *sensitivity*, *specificity*, *positive predictive value (PPV)*, and *negative predictive value (PPN)*, are additional key measures that influence the overall clinical decision-making process.

Endpunkte: u. a. Mortalität, Morbidität und Lebensqualität

Synthesis Methods:

Meta-analyses are preferred for combining quantitative data from multiple studies, allowing for an aggregated effectiveness assessment [38]. Where pooling of data is not possible due to heterogeneity in study designs or outcomes, narrative synthesis is used to describe overall trends and findings.

vorrangig Meta-Analysen oder narrative Synthesen

Safety (SAF)

Study Design:

The safety of therapeutic medical devices and diagnostic technologies is typically assessed through RCTs, observational studies, registries, and post-marketing surveillance, which is critical for detecting long-term adverse events and device failures [38]. Prospective cohort studies are valuable for assessing safety in real-world settings and clinical trials, especially for diagnostic technologies where false positives or negatives may lead to harm.

Sicherheit (SAF)
Studiendesigns: vorrangig RCTs und prospektive Beobachtungsstudien

Endpoints:

For medical devices, safety endpoints include device-related adverse events, malfunctions, and procedure-related complications [38]. Critical safety concerns for diagnostic technologies include risks from *false positive/negative results*, which could lead to inappropriate treatments or unnecessary follow-up procedures.

Endpunkte: unerwünschte Ereignisse, Risiken durch falsch-positive/negative Ergebnisse

Screening technologies require a shallow tolerance threshold for harm due to their application in healthy populations [38]. It is crucial to consider *indirect harms* such as false positives, false negatives, and overdiagnosis.

indirekte Schäden sind bei Screeningverfahren zu berücksichtigen

Synthesis Methods:

Meta-analyses of safety data are conducted when multiple studies report adverse events or complications [38]. If there is significant heterogeneity or limited data, narrative synthesis helps integrate findings from available studies to provide a clearer picture of the safety profile of the technology.

Meta-Analysen oder narrative Synthesen

Costs and Economic Evaluation (ECO)

Study Design:

Cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) are the most common designs used to assess the economic value of medical devices and diagnostic technologies [38]. These methods compare the costs of the technology with health outcomes, often expressed in terms of quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs).

ökonomische Aspekte (ECO)
Studiendesigns: Kosten-Effektivitätsanalyse und Kosten-Nutzen-Analyse

Endpoints:

Key endpoints include *incremental cost-effectiveness ratios (ICERs)*, *budget impact*, and *resource utilisation* [38]. These measures are critical in determining whether a technology provides sufficient benefit relative to its cost, especially for expensive medical devices or widespread diagnostic technologies.

Endpunkte: Budgetfolgen, Ressourcennutzung

For screening technologies, economic evaluations should include the costs of the entire screening programme, including organisation and follow-up [38]. The impact on healthy, working-age populations (e.g., lost productivity) should be considered. The complexity of screening models and limited empirical data present additional challenges in economic evaluations.

Kosten des gesamten Programms sind bei Screeningverfahren zu berücksichtigen

Synthesis Methods:

Decision-analytic models are often used to simulate the long-term costs and benefits of the technology, mainly when trial data are limited or short-term [38]. Systematic reviews of economic evaluations can also provide aggregated data on the cost-effectiveness of similar technologies across different settings.

Entscheidungsanalytische Modellierung, systematische Übersichtsarbeiten

Ethical Analysis (ETH)

Study Design:

Qualitative research methods, such as stakeholder interviews, focus groups, and surveys, are typically used to explore ethical considerations [38]. These studies often involve input from patients, clinicians, and policymakers.

ethische Aspekte (ETH)
Studiendesign: qualitative Forschung

Endpoints:

Ethical considerations include *fair access to technology*, *patient autonomy*, *equity*, and the potential for *overdiagnosis* or *misuse* of diagnostic technologies [38]. For medical devices, ethical issues may also involve the implications of implantable or life-supporting devices. For diagnostic technologies, it is essential to consider the aim of the diagnostic test and its role in the diagnostic pathway, as well as evaluate unintended implications, including potential harm to healthy individuals.

Endpunkte: gerechter Zugang zur Technologie, Gerechtigkeit; Diagnostik: Ziel des Tests, Rolle in der Versorgung

Screening technologies raise specific ethical issues related to targeting healthy or asymptomatic individuals and require careful consideration of the balance between benefits and harms, including *overdiagnosis* and *overtreatment* [38].

Screening: Balance zwischen Nutzen und Schäden (Überdiagnose und -behandlung)

Synthesis Methods:

Narrative synthesis integrates the ethical perspectives of various stakeholders and balances ethical concerns with clinical and economic outcomes [38]. Multi-criteria decision analysis (MCDA) may also weigh different ethical factors.

narrative Synthese

Organisational Aspects (ORG)

Study Design:

Observational studies, case studies, and surveys are often employed to assess the impact of medical devices and diagnostic technologies on healthcare organisations [38]. These studies examine how introducing new technologies affects workflows, staff training, and resource allocation.

organisatorische Aspekte (ORG) Studiendesigns: Beobachtungsstudien

Endpoints:

Organisational endpoints include *workflow efficiency*, *staff competency*, *resource allocation*, and *integration of new technologies into existing healthcare systems* [38]. The impact on laboratory throughput and reporting accuracy is particularly important for diagnostic technologies.

Endpunkte: Effizienz im Arbeitsablauf, Integration in bestehende Systeme

Assessing screening technologies involves evaluating the entire screening system, from identification to treatment [38]. Various objectives should be considered, including introducing a new test, eligibility population changes, screening interval, or delivery method.

Screening: Einführung des neuen Tests, Screening Intervall

Synthesis Methods:

Narrative synthesis is commonly used to summarise the effects of these technologies on healthcare organisations. In some cases, qualitative synthesis of case studies or observational research can provide insights into how well a technology is integrated into clinical practice [38].

Synthesemethoden: narrative Synthese, qualitative Synthese

Patients and Social Aspects (SOC)

Study Design:

Qualitative studies, such as interviews and focus groups, are often used to capture patient and societal perspectives on medical devices and diagnostic technologies [38]. These studies help understand the broader social implications and the acceptance of technologies among different patient groups.

Patient*innen-bezogene und soziale Aspekte (SOC) qualitative Studiendesigns

Endpoints:

Key endpoints include *patient satisfaction*, *quality of life*, *accessibility*, and *social equity* [38]. For diagnostic technologies, additional considerations include the psychological impact of receiving test results (positive or negative) and how testing impacts the patient's social and daily life.

Endpunkte: Patient*innenzufriedenheit, Zugänglichkeit

It is crucial to focus on access equity and factors affecting participation in screening technologies [38]. Communication strategies and information delivery for informed decision-making are particularly important in screening programs.

Screeningverfahren: Kommunikations-strategien

Synthesis Methods:

Thematic analysis and qualitative synthesis are employed to summarise patient and societal feedback on the technology, offering insights into the social dimensions of its use and the potential barriers to widespread adoption [38].

thematische Analyse und qualitative Synthese

Legal Aspects (LEG)

Study Design:

Legal reviews and case studies are typically used to explore health technologies' regulatory and legal implications [38]. This includes analysing compliance with medical device regulations, intellectual property laws, and data protection standards.

juristische Aspekte (LEG) Gutachten und Fallstudien

Endpoints:

Legal endpoints involve *regulatory compliance*, *intellectual property*, and *liability concerns* related to the performance of the device or technology [38]. For diagnostic technologies, data privacy and patient consent issues are also important.

Endpunkte: Einhaltung gesetzlicher Vorschriften und Haftungsfragen

Synthesis Methods:

Narrative synthesis summarises the legal issues associated with the technology, often combining legal reviews with expert consultations or stakeholder input to ensure compliance and identify potential risks [38].

narrative Synthese

1.6 Objectives and Scope

There is uncertainty regarding whether medical technologies utilising AI require specialised methods to evaluate their utility in assisting procurement decisions in Austrian hospitals. If such methods are needed, it is unclear which ones are most appropriate in specific clinical areas. Hence, it is crucial to examine currently available methodological approaches for assessing these technologies' (added) value.

Ziel: Übersicht zu HTA-Methoden für KI-Gesundheits-technologien

Considering this context, this report aims to address the following research questions (RQ):

Forschungsfragen (FF)

- **RQ1**: How can hospitals assess the potential clinical benefit of AI-enabled digital health technologies? What HTA methods and frameworks can be used for AI procurement and implementation decisions?

  FF1: Identifizierung bestehender Methoden und Frameworks
- **RQ2:** What evaluation methods were used in previous HTA reports to assess the additional clinical benefits of AI applications in specific areas? What are the potential applications of AI systems assessed in these HTA reports, and what are the expectations for their added benefit?

  FF2: verwendete Methoden in HTA-Berichten, Anwendungsgebiete von und Erwartungen an KI
- **RQ3:** Based on the information gathered, what specific recommendations (e.g. requirements of evidence, quality assurance, and clinical applications) can be made for the successful implementation of AI systems in Austrian hospitals?

  FF3: Empfehlungen für österreichische Krankenhäuser

Table 1-2: PIC-Problem, Interest, Context for Inclusion of Information

	Inclusion	Exclusion
Problem	Uncertainty whether existing methods are sufficient/missing methodological guidance to assess the benefits of AI systems in various application areas in hospitals	-
Interests	HTA methods guidance for AI-enabled digital health technologies in hospital settings Application areas, potential benefits and challenges of AI from an HTA perspective Practical recommendations for Austrian decision-makers	-
Context	Austrian health care system/hospitals, AI	-
Language	All languages	-
Publication Type	HTA methods guidance documents, including content relevant to AI, HTA reports concerning AI systems	HTA papers not mentioning AI, other syntheses of evidence

Abbreviations: AI … artificial intelligence, HTA … health technology assessment

2 Methods

This study employed a four-step methodological approach. First, we conducted a targeted search to identify existing HTA methods, guidance documents and assessments for AI-enabled DHTs. Second, identified methods guidance documents were analysed to describe how the benefit of AI-enabled DHTs can be assessed, identifying key aspects to consider. Third, identified assessments were analysed, concentrating on applied methods for evaluating the benefit of AI-enabled DHTs. Fourth, recommendations for Austrian hospitals were formulated.

4 methodische Schritte: (1) Identifizierung von HTA-Methodendokumenten und Assessments, (2) Analyse der Dokumente, (3) Analyse der Assessments und (4) Ableitung von Empfehlungen

This scoping review was pre-registered using the Open Science Foundation [40] platform. No major changes to the protocol applied. In light of numerous articles dealing with potential use cases of AI in the near future, we changed the methods for RQ1. We described application areas based on AI technologies already assessed in HTA instead of identifying application areas based on a focused search.

2.1 Step 1: Identification of HTA methods guidance and assessments

In Step 1, we conducted a focused hand search on 51 HTA institutional webpages to identify both methods guidebooks and assessment reports on AI (a full list of institutions is provided in Table A-2). The search, limited to May 2024, initially included documents in English and German. No language filter was applied. Documents not available in English or German were translated using Google Translate, DeepL or ChatGPT. We contacted the respective INAHTA members directly to verify the translated information if no English summary was available. Experts from HTA organisations participating and presenting research in AI methods sessions were additionally contacted.

Schritt 1: fokussierte Handsuche in 51 HTA-Instituten bis Mai 2024

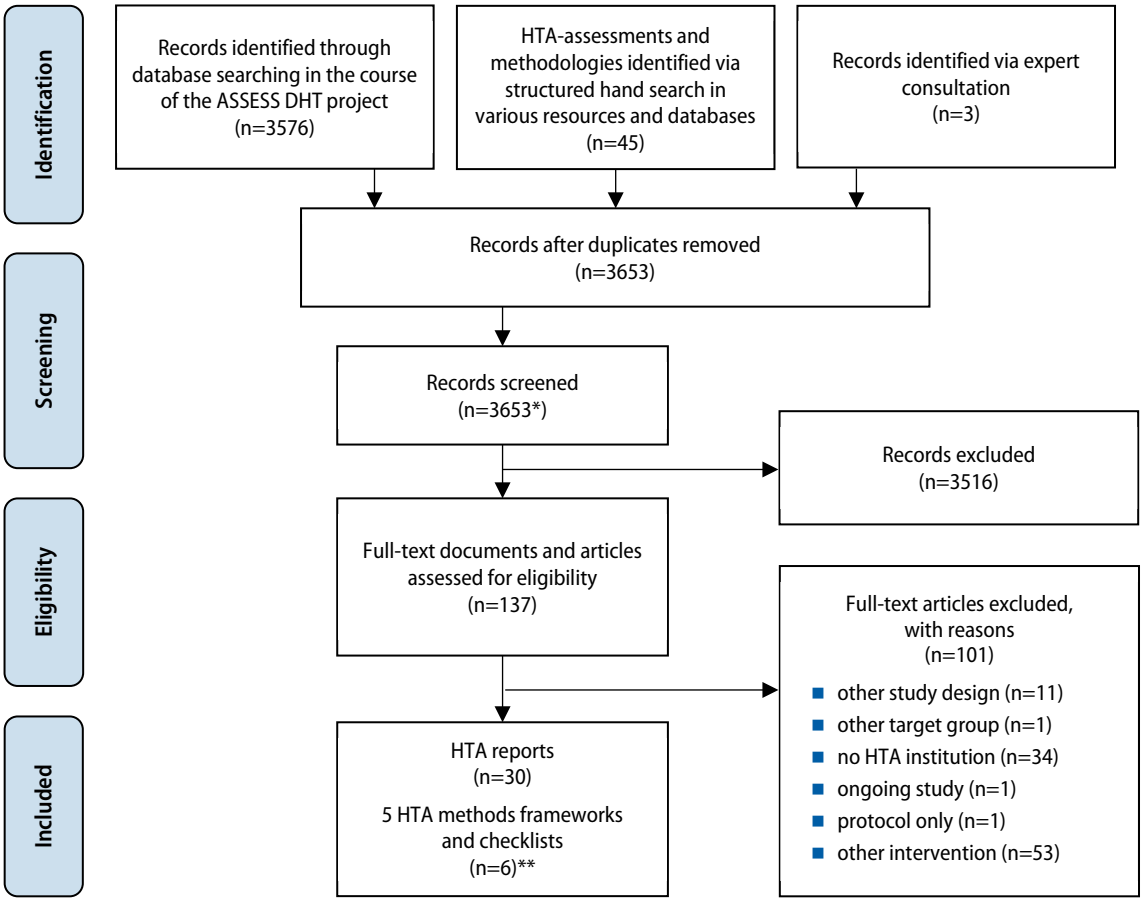
For selecting relevant documents, two independent researchers (MR, DG) followed pre-defined inclusion criteria (depicted in Table 1-2) and were discussed with another researcher (GG). HTA guidance documents were defined as any document guiding HTA practitioners (or other stakeholders involved in HTA) in assessing the benefit of AI-enabled DHTs. AI-enabled DHTs were defined broadly, covering both AI software as a medical device (SaMD) and AI software in a medical device (SiMD) [41]. The selection process is displayed in Figure 2-1.

Auswahl relevanter Dokumente anhand vorab festgelegter Kriterien

Furthermore, we utilised the results of a previous systematic search as part of the ASSESS DHT project (see Appendix Search Strategy) [42]. This search had a broader scope, focusing on assessment frameworks. Two authors from the Technical University of Berlin (HR, CH) screened the abstracts and full texts obtained from this existing search. The full-text articles were provided and subsequently screened for eligibility specific to our current project.

zusätzliche systematische Suche durch die Technische Universität Berlin im Zuge des ASSESS DHT Projekts

2.1.1 Flow Diagram



* *3576 records screened in the course of the ASSESS DHT Project by HR and CH*

** *For one institution, there were supplementary documents, which were treated as a single document during the analysis.*

Figure 2-1: Flow chart of study selection (PRISMA Flow Diagram)

2.2 Step 2: Analysis of HTA methods documents

Schritt 2: qualitative Analyse der HTA-Methodendokumente

Identifizierung KI-relevanter Aspekte durch Stichwortsuche

Using a structured approach, we conducted a qualitative document analysis of HTA method documents. To identify guidance relevant to AI within these documents, we searched for content using the following keywords:

- "Artificial intelligence"
- "Machine learning"
- "Deep learning"
- "Algorithms"
- "Neural networks"
- "Data-driven"
- "Predictive analytics"

Datenextraktion durch eine Wissenschafterin; Kontrolle durch eine/n Zweite/n

Following the keyword search, one researcher (DG) extracted pertinent information into a tabular format (see Table A-10). This extraction was then verified by a second researcher (MR or GG) for accuracy and completeness. We also recorded the general characteristics of each guidance document, including its purpose, intended users, and overall structure.

narrative Beschreibung der Dokumente und thematische Analyse von Themen, die noch nicht im EUnetHTA Core Model enthalten sind (KI-spezifische Themen)

The guidance documents were narratively described. A thematic analysis was then conducted in the context of AI-specific content: to analyse the guidance documents qualitatively, we standardised the text passages identified through keyword search into themes (see Table A-11). These themes were then mapped against the EUnetHTA Core Model, identifying gaps. Finally, we consolidated the identified themes into overarching AI-specific topics and categorised these topics according to the EUnetHTA Core Model (see Table A-12).

2.3 Step 3: Analysis of HTA assessments: application areas and methodological approaches

Schritt 3: Analyse von HTA-Methoden in publizierten Berichten

To summarise methodological approaches used in assessment practice and gain insights into specific application areas, we described the current methodological approaches in previous HTA assessments on AI-enabled DHTs.

Extraktion von u. a. Studiendesign, Endpunkten und Schlussfolgerungen

Relevant data were extracted through piloted extraction tables (see Table A-3 to Table A-9). Central information included, among others, technology and comparator (main function, expectations of AI-enabled DHT), inclusion criteria of HTA (e.g. study design, defined endpoints, synthesis methods), general study methodology, AI-specific assessment criteria, methodological characteristics of included studies in HTA reports and the conclusions on the evidence of the HTA reports.

Erstellung von Vignetten nach Funktion und Anwendungsgebiet

Based on this information, vignettes were created clustered according to functionalities of the assessed AI-enabled DHTs using a classification system of the MDR [33] and medical specialities. These application areas are diagnosis, treatment, prediction, prognosis, prevention, and monitoring. To allocate the assessments to the appropriate application area, we followed the definitions depicted in Table 2-1.

Table 2-1: Definitions of Application Areas

Application Area	Description
Diagnosis	Determining the specific medical condition causing a problem [43].
Treatment	Caring for a patient to combat disease or disorder, excluding diagnostic procedures [43].
Prediction	Analysing patient characteristics to predict risk or determine treatment [44].
Prognosis	Describing the prospect of recovering from injury or disease [45]. While prediction aims to calculate the risk of a disease, prognostics calculate the risk of particular health states occurring [46].
Prevention	Action is taken to decrease the chance of getting a disease or condition [47].
Monitoring	Continuously evaluating a patient's condition over time, either invasively or non-invasively [43, 48].

2.4 Step 4: Practice recommendations for Austria

Schritt 4: Ableitung von Empfehlungen für österreichische Entscheidungsträger*innen

Based on the findings from Steps 1-3, we developed recommendations for Austrian decision-makers relevant to hospital procurement decisions of AI-enabled DHTs. Our recommendation development focused on:

- Potential frameworks and checklists for assessing AI in healthcare
- Potential Requirements for AI-enabled DHTs (e.g., evidentiary)
- Additional considerations for implementing AI in the Austrian healthcare context

Ziel: Anleitung zur Evaluierung und Implementierung

These recommendations aim to provide practical guidance for Austrian decision-makers on evaluating and implementing AI-enabled DHTs, informed by international best practices.

3 Results

In this chapter, we present the results of our analysis, including an overview of the identified methods guidance documents and HTA reports. Of 51 HTA institutions, 13 (26%) have published methods guidance documents and/or conducted assessments on the benefit of AI-enabled DHTs. Six guidance documents were identified, of which two [4, 49] were supplementary documents by the same institution, hence treated as a single document during the analysis. Consequently, five guidance documents and 30 assessments were identified (see Table A-2). These documents and assessments were identified during the focused hand search. No relevant studies were identified in the additional systematic literature search.

Übersicht zu den identifizierten; 6 Methodendokumenten (2 überlappend – folglich als ein Dokument gezählt) und 30 Assessments

3.1 HTA-Methods

The five methods guidance documents identified were developed by HTA bodies in European countries. Among the five guidance documents, the National Institute of Health and Care Excellence (NICE) [3], the Agency for Health Quality and Assessment of Catalonia (AQuAS) [50] and the Finnish Coordinating Center for Health Technology Assessment (FinCCHTA) [24] established three DHT assessment frameworks. One [3] was originally for DHTs only. Then it was complemented with AI. Another document from the French Haute Autorité de Santé (HAS) is a more generic HTA guidance [4, 49]. The final guidance, provided by Health Technology Wales (HTW), stands out as an AI-specific checklist to support HTA practitioners when dealing with AI-enabled health technologies [51].

5 Methodendokumente von europäischen HTA-Institutionen

In terms of scope, NICE [3] and AQuAS [50] focus broadly on DHTs. These frameworks primarily focus on the value assessment of DHTs, and all also address AI-enabled DHTs. The HAS guidance [4, 49] centres on connected medical devices (CMDs), which may also include AI-enabled DHTs, particularly for procedural and reimbursement purposes. FinCCHTA, a DHT framework, extends its scope beyond traditional DHTs to cover emerging technologies, including mHealth, AI, and robotics [24]. The HTW checklist is the most narrowly focused, dealing specifically with AI-enabled DHTs in healthcare.

3 generische DHT-Bewertungsframeworks, ein KI spezifische Checklist und ein Methodenhandbuch mit KI spezifischen Empfehlungen

The deployment status of these frameworks varies. Three of the five documents (NICE, HAS, and FinCCHTA) are already in use, indicating that they are well-established within their respective contexts [3, 4, 24, 49]. However, AQuAS [46] and HTW [47] are still testing their developed guidance documents, suggesting they are being piloted or refined before full implementation.

3 Dokumente werden bereits verwendet, 2 noch pilotiert

In terms of structure, the Evidence Standards Framework (ESF) developed by NICE is organised into five groups and 21 standards explicitly tailored to digital health technologies [3]. Furthermore, it classifies technologies into three tiers, depending on their potential risks. This classification determines the extent of evidence requirements that must be fulfilled. This classification is also used by the AQuAS framework [50]. AQuAS [50] and FinCCHTA [24] structure their frameworks in domains, with 13 domains (AQuAS) and 11 do-

2 Dokumente verwenden eine Risikoklassifizierung auf Basis derer eine gewisse Anzahl an Standards eingehalten werden muss

mains (FinCCHTA) covering areas such as safety, efficacy, technical stability, and data protection. The HAS guidance [4], while not explicitly structured into domains, includes key procedural and methodological elements related to connected medical devices, such as real-life data collection and automatic data processing in one document. The second document [49] consists of 42 questions related to AI. The HTW AI checklist is the most concise in its structure, offering general AI guidance and four specific domains of assessment [51]. HAS [4], HTW [51] and FinCCHTA [24] do not use a classification based on potential risks.

zu KI-spezifischen Themen zählt neben Informationen zu KI-Funktion, Trainingsdaten und menschlicher Aufsicht, die Überwachung nach Implementierung

Themes relevant to AI are embedded across several of the documents. For example, the ESF includes six standards directly related to AI [3]. AQuAS incorporates AI considerations in 12 of its 13 domains, showing its substantial focus on the integration of AI in health technologies [50]. FinCCHTA also touches on AI within its broader scope of emerging technologies, comprising 22 questions [24]. Meanwhile, the HTW checklist and the HAS guidance address concerns such as data sources, retraining capabilities, validation, autonomy, and the need for human oversight in the care process [4, 49, 51]. However, when charting these AI-relevant domains and questions to the EUnetHTA Core Model, topics such as information on AI function and model, training data quality, human oversight, data risk management, re-evaluation, and post-deployment monitoring were considered AI-specific topics.

Table 3-1: Overview of frameworks and checklists

Name/Institution	ESF/NICE [3]	NA/AQuAS [50]	NA/HAS [4, 49]	DIGI-HTA/FinCCHTA [24]	NA/HTW [51]
Country	England	Spain	France	Finland	Wales
Type of guidance	DHT Framework	DHT Framework	HTA Guidance	DHT Framework	AI checklist
Scope of application	DHTs	DHTs, AI	CMDs connected medical devices	mHealth, AI, robotics	AI
Status of deployment	In use	In testing	In use	In use	In testing
Purpose	Value assessment, assist manufacturers	Value assessment	Value assessment, assist manufacturers	Rapid assessment of developing technologies	Assist HTA authors when dealing with AI-enabled DHTs
User	Evaluators, innovation team and manufacturers	HTA agencies, researchers, developers of DHT, decision-makers and regulators	Manufacturers, distributors, and service providers	Representatives of wellbeing service counties (e.g., for procurement), technology companies, healthcare organisations	HTA practitioners
Structure	3 tiers, 5 groups and 21 standards specific for DHT	13 domains	Generic procedural and methodological HTA manual; Descriptive grid (42 questions)	11 domains	General information on AI and 4 domains within checklist
Guidance relevant to AI	6/21 Standards	12/13 Domains	4/4 Areas; 42/42 Questions	1/11 Domains; 22 Questions	4/4 Domains; 31 Questions
AI-specific themes identified in:	6 Standards	2 Domains	2 Areas/40 Questions	19 Questions	11 Questions

Abbreviations: A … area, AI- artificial intelligence, AQuAS … Agency of Health Quality and Assessment of Catalonia, CMD … connected medical device, CUR … current, D … domain, DHT … digital health technology, ECO … economic, EFF … effectiveness, ESF … evidence standards framework, ETH … ethical, FinCCHTA … Finnish Coordinating Center for Health Technology Assessment, HAS … Haute Autorité de Santé, HTW … Health Technology Wales, LEG- legal, NA … not applicable, ORG … organisational, Q … question, SOC … social, TEC-technical

3.1.1 NICE – Evidence standards framework

NICE developed one identified methodological guidance document relevant to AI [3]. The ESF for DHTs was first published in 2018 and updated in 2022. While the previous version could not be used for AI-enabled DHTs, this has been changed for the updated version.

2018: "Evidence Standards Framework" (ESF) von NICE (Update 2022)

General Description

Purpose

The NICE ESF for DHTs outlines the criteria for the evidence required to demonstrate the value of a DHT within the health and social care system of the United Kingdom [3]. A DHT may include smartphone apps, standalone software, online tools for treating or diagnosing conditions, preventing diseases, or improving system efficiencies, and programs that can be used to analyse data from medical devices such as scanners, sensors, or monitors.

Ziel: Anforderungskriterien für Nutzenbewertung von DHT

It includes evidence of technology performance in its intended use and its economic impact in relation to financial risk [3]. The ESF is meant to supplement existing regulatory and technical standards for DHTs. Its primary goal is to evaluate DHTs to ensure their expected function and offer good value for money. The performance evidence standards are set at achievable levels for DHT companies while maintaining a high enough standard to ensure confidence in the technology from the health and social care system. This approach aims to promote the effective and innovative use of DHTs. The value standards are aligned with NICE's evaluation methods to assess the potential economic impact of a DHT. The ESF is created in the context of a health and social care system looking for innovative ways to enhance care while reducing costs.

Systematik zu Evidenzanforderungen zum Nutzenbeleg und zur Qualitätssicherung

Users

The ESF is intended to be used by evaluators and innovation teams, especially in the NHS and care system for purchasing decisions [3]. It can also be used by DHT companies to facilitate commissioning or purchasing decisions.

Zielgruppe: Evaluator*innen, Unternehmen

Structure

The ESF is divided into four sections [3]. Section A describes technologies suitable for evaluation using the ESF, section B explains the classification of DHTs, section C provides an overview of evidence standards tables, and section D is about early deployment standards for evidence-generation programs. In the ESF 21 standards for the evaluation are arranged in five groups. Nine standards are covered in the group "design factors", four standards in "describing value", three standards in "demonstrating performance", two standards in "delivering value" and three standards in "deployment considerations".

Systematik:

21 Standards in 5 Gruppen (Design, Nutzenbeschreibung, Performance, Nutzenbeleg, Implementierung)

DHTs are further classified in tiers based on the potential risk to service users and to the system (see Figure 3-1) [3]. Tier A includes DHTs which are intended to save costs or release staff time, without direct patient, health or care outcomes. When DHTs help communicating about health and care, promote good health and enable health and care diaries, they classify as Tier B. DHTs for treating and diagnosing medical conditions, or guiding care choices are classified as Tier C. While Tier A technologies are excluded from some standards, all standards apply to Tier C DHTs [3].

3 Risikostufen: Stufe A: kein direkter Patient*innenkontakt bis Stufe C: Behandlung und Diagnose

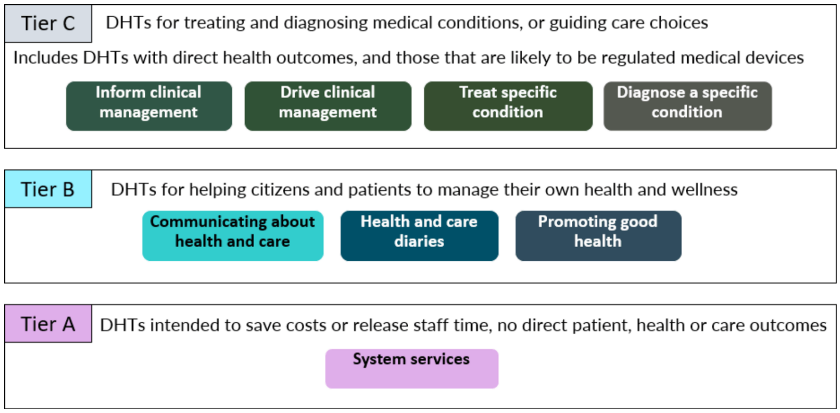


Figure 3-1: Evidence Standards Framework – Tiers [3]

Guidance relevant to AI

The ESF considers certain standards to be relevant to evaluating so-called "data-driven" DHTs that may include AI components [3]. AI components are broadly defined as any DHT with fixed or adaptive machine learning components – AI standards are covered in the topics design features, performance demonstration, and deployment considerations.

KI-relevant sind DHT mit statischen oder adaptiven Komponenten des Maschinellen Lernens (ML)

Design factors

For **design factors**, ESF standards 4, 5, and 6 are highlighted as relevant to AI. The following aspects are as a result of this recommended within the ESF [3]:

Design (Standards 4,5,6 KI-relevant)

Consider health and care inequalities and bias mitigation (Standard 4). For all data-driven DHTs that include AI, all information needs to be available on actions in designing an intervention in order to be able to mitigate against algorithmic bias.

Überlegungen zu Ungleichheit und Verzerrungsreduktion

Embed good data practices in the design of DHT (Standard 5).
The ESF states that information on the following should be available:

Integration guter Datenpraktiken

Information über u. a. Trainingsdaten und Validierungsdaten

- source and size of training and validation data,
- the process of establishing 'ground truth',
- data collection methods,
- information if synthetic data was used,
- diversity of the training and validation data
- if it is representative of the intended target population

For DHTs incorporating machine learning, the ESF refers in this standard to the guiding principles developed by the U.S. Food and Drug Administration, Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA)[52].

spezielle Prinzipien für DHTs mit ML

Define the level of professional oversight (standard 6). Experts should continuously review the output. This should uphold best clinical practice by monitoring the trend of the AI output to ensure alignment with and calibration for best practice.

professionelle Aufsicht und kontinuierliche Kontrolle der Outputs

Demonstrating performance

For **demonstrating performance**, standards 15 and 16 include AI-specific components [3]:

Performance (Standards 15,16 KI-relevant)

Show real-world evidence (RWE) that the claimed benefits can be realised in practice (Standard 15): The ESF recommends the following aspects relevant to DHTs using AI. The performance of these technologies may be affected by local deployment factors. The ESF highlights that the technology may run offline or "in silent mode" for an evaluation. Silent mode allows to evaluate the performance of the DHT using local data inputs before its implementation into clinical or care pathways. This can be used to show whether the expectations of a DHT performance are met. Information on the following aspects should be available:

RWE für den Nachweis der erfolgreichen Anwendung

für Evaluierungsprozesse soll die Technologie offline oder im "silent mode" laufen

- current service provision (in the United Kingdom),
- current best practice,
- user acceptability,
- successful integration without unintended negative consequences for users/services,
- demonstration of improvement in outcomes (clinical/cost-effectiveness).

Information zur Vergütung und ein, Nachweis von Verbesserung von klinischen oder ökonomischen Ergebnissen

Companies or evaluators should agree on a plan for measuring changes in performance over time (Standard 16). To uphold this standard, the ESF recommends discussing the following:

Hersteller brauchen einen Plan zur Performanzmessung bei Veränderungen, sowie

- future plans for updating the DHT (incl. time intervals to retrain, change functionality or when novel versions are available),
- sources of retraining data (incl. how the quality of data will be assessed),
- processes in place for measuring performance over time (detecting impacts on performance),
- processes in place to detect decreasing performance in certain groups of people over time,
- existence of an independent overview process for reviewing changes in performance,
- an agreement on the reporting of changes in performance (evaluators, patients, carers, health and care professionals).

Informationen zu Updates

und Leistungsveränderungen (messen, entdecken, berichten)

If the intended purpose of the DHT is altered, or if new functions are introduced that modify its intended use and ESF classification, a re-evaluation should be conducted. For DHTs classified as medical devices, changes to their intended purpose would necessitate a regulatory reassessment, which could lead to reclassification as a medical device and require additional evidence to support the change.

eine Reevaluierung ist erforderlich bei Veränderungen in Zweck und Funktion

Deployment considerations

As for **deployment considerations**, standards 19 and 20 are relevant for DHTs using AI [3]:

Implementierungsaspekte

Ensure transparency about requirements for deployment (Standard 19): The ESF highlights that the company should provide comprehensible descriptions of any deployment data. The ESF hereby mentions the following:

Sicherstellung der Transparenz soll gesichert sein, sowie: …

- full description of input data (incl. a data dictionary),

- quantification of the level of tolerance for incomplete data (e.g., missing data or data with insufficient quality) and handling outliers,
- data flow map for deployment to increase efficiency in deployment,
- data requirements (e.g., data formats, completeness, quality, and standardisation requirements such as digital imaging and communication in medicine),
- Infrastructure requirements.

… Informationen zum Umgang mit unvollständigen Daten,

und Anforderungen an Daten (Format, Qualität)

Describe strategies for communication, consent and training processes to allow the DHT to be understood (Standard 20). The ESF considers it necessary for the company to uphold communication with various stakeholders, including service users and health care professionals. The goal is to describe the output, features, benefits, and limitations of the DHT. A so-called model card is proposed to help people understand when and how the DHT may be used. The communication and training should also describe the output of the DHT, including information on how the output should be interpreted. The ESF highlights in this context that output may, for instance, be a risk score, a probability of different diagnoses, or a recommendation for other tests. Furthermore, the company should also provide a description of their approach to training of end users (incl. a process for service user consent if needed). This should allow the benefits of a specific DHT to be realised in practice.

Kommunikationsstrategien müssen beschrieben sein, sowie:

Output, Nutzen und Limitationen

möglicher Output einer DHT sind Risikowerte, Diagnosen und Empfehlungen

3.1.2 Framework from AQuAS

The Health Technology Assessment Framework created by AQuAS (Spain) outlines the adaption of health technologies for DHTs [50].

AQuAS: Dokument zu DHTs (inclusive KI)

General Description

For the process of developing a framework, multiple methodologies were piloted to address the assignment and objectives [50]. This resulted in a proposed assessment framework, which was then refined through a workshop with national experts in HTA and digital health. To establish evidence standards for DHTs based on their risk classification, the team used the classification in Tiers A to C described in the updated ESF from NICE. This process involved seminars with the ESF authors and a subsequent consensus workshop with the original group of experts.

Altogether, 13 domains, 41 dimensions, nine subdimensions, and 21 levels of evidence were identified to be considered in the evaluation of DHT.

Risikoklassifizierung nach ESF (NICE)
13 Domänen
41 Dimensionen

Purpose

The aim is to facilitate the evaluation process of DHT, identify the value of digital technology and inform decision-making [50].

Nutzenevaluierung von DHTs

User

The guidelines are intended to be used by HTA agencies, researchers, developers of DHT, decision-makers and regulators [50].

Zielgruppe: HTA-Institutionen, Entscheidungsträger*innen, Hersteller*innen

Structure

The assessment framework consists of 13 domains, 38 dimensions and 9 subdimensions. For each domain, a detailed description is provided, as well as orientation questions, sources of information that can be used to respond to this domain, and observations [50]. A detailed description of the domains is provided in the Appendix.

Systematik: Aufbau in 13 Domänen, 38 Dimensionen und 9 Subdimensionen

Guidance relevant to AI

In the guidance document, twelve out of thirteen domains were considered relevant to AI. These domains include [50]:

KI-relevante Domänen

u. a. die Beschreibung der Technologie, Sicherheit, sowie soziokulturelle Aspekte

- **description of the health problem:** including the prevalence and incidence of the health problem, the target population and information regarding the standard therapeutic approach,
- **description of the technology:** main characteristics of the technology, regulation or licenses required, the requirements of the technology, potential added value, including the dimension adoption and subdimensions use and integration,
- **content:** evaluation of completeness, accuracy, and timeliness of the content (written, visual and auditory information), including the dimension adequacy of the information,
- **safety:** risks, harms and unwanted side effects due to the intervention, including the dimension clinical safety[2],
- **clinical efficacy and effectiveness:** clinical benefits and impact on quality of life under controlled or uncontrolled conditions,
- **economic aspects:** economic costs of acquisition, maintenance, and use – including the dimensions costs and resource use,
- **human and sociocultural aspects:** acceptability, ease of use, perceived benefit, as well as accessibility to the service or health care, changes in workflows and roles,
- **ethical aspects:** assessment of ethical concerns of the technology
- **legal and regulatory aspects:** degree to which the technology complies with the regulations and standards of the country and region, including privacy and transparency,
- **organisational aspects:** changes in workflows or professional functions, human resources and training,
- **technical aspects:** usability and ease of use, adaptability, design, technical stability, interpretability and reproducibility, scalability, technical effectiveness, performance, post-deployment monitoring, generalisability,
- **environmental aspects:** measurement of the environmental impact, e.g. through carbon emissions, use of raw materials, energy consumption, as well es environmental benefits.

AQuAS aligned the domains with the ESF. The domains are similar to the EUnetHTA Core Model domains with additional domains.

Domänen in Anlehnung an das EUnetHTA Core Model

[2] Technical safety was not considered AI-relevant by AQuAS

3.1.3 HAS guidance documents

The Haute Authorité de Santé (HAS) uses their standard guidance to evaluate health technologies by the Medical Device and Health Technology Evaluation Committee (CNEDiMTS). For digital health technologies, the CNEDiMTS published a separate methods guidance document relevant to specific features of the clinical evaluation of a connected medical device [4]. In addition, the general dossier submission was extended to guidance relevant to digital health technologies [49].

HAS verwendet zur Evaluierung ein generisches Dokument zu DHTs, sowie ein Einreichungsformular

General description of guidance

Purpose

The identified HAS guidance documents relevant to digital health technologies, labelled as connected medical devices (CMD), which include AI-enabled DHTs, serve two key purposes [4, 49]:

Ziel: Anforderungskriterien für Nutzenbewertung von connected medical devices (DHT)

- to assist companies manufacturing or operating CMDs in incorporating appropriate clinical trials into their development strategy and
- to outline the evidence requirements considered in Health Technology Assessment (HTA) for CMDs.

User

The guide is designed for manufacturers, distributors, and service providers who intend to submit application dossiers for inclusion in the list of products and services qualifying for reimbursement.

Zielgruppe: Hersteller*innen & Dienstleister*innen

Structure

The structure of the guidance follows the structure of a comprehensive procedural and methodological guidance for HTA conducted by CNEDiMTS. The reader is referred to the website of the HAS for more details [53].

der Aufbau der Leitlinie entspricht einer HTA-Methodenanleitung

Guidance relevant to AI

The HAS guidance documents emphasise that while CMDs are subject to the same overall evaluation criteria for reimbursement decisions as other medical technologies, there are key areas that require special attention due to the unique nature of CMDs [49]:

DHT-Evaluierungskriterien entsprechen jenen für andere Gesundheitstechnologien, aber mit besonderem Fokus auf:

- *Optimised Clinical Development:* This area focuses on demonstrating the device's value through carefully designed clinical trials. The primary outcomes of interest are the clinical benefit provided by the CMD, its acceptability to patients and healthcare providers, and its impact on quality of life. These are considered essential elements in establishing the device's value. Additionally, the guidance suggests evaluating secondary outcomes such as the device's accessibility, how well it aligns with or improves upon the current standard of care, and its effects on healthcare organisation and delivery.

Nachweis klinischer Wirksamkeit und Verbesserung in Zugang und Organisation von Versorgung

- *Prerequisite Requirements:* Before a CMD can be considered for evaluation by the (CNEDiMTS), it must meet certain criteria. These include strict adherence to data processing and hosting regulations, with particular emphasis on compliance with the General Data Protection Regulation (GDPR). The device must also have obtained CE marking,

Voraussetzungen sind Zulassung, Einhaltung von Datenschutz und Qualitätssicherung

indicating conformity with EU health, safety, and environmental protection standards. Finally, manufacturers must have a robust quality assurance process in place to ensure consistent product performance and safety.

- *Algorithm Development Transparency*: Given the central role of algorithms in many CMDs, the guidance requires detailed information about their development. This includes a clear explanation of how variables were chosen and selected, the rationale behind the model selection, details about the learning mechanisms employed, and information about the training data used. Manufacturers must also demonstrate the ongoing relevance of the algorithm, describe processes for regular verification, and provide evidence of efforts to identify and mitigate potential biases.

Anforderungen an Transparenz zur Entwicklung des Algorithmus sind verwendete Daten, Variablen, Verifikation

- *Real-World Data Collection:* The final area emphasises the importance of ongoing data collection after the device has been approved and is in use. This serves two primary purposes: to monitor the device's performance over time in real-world settings and to confirm the medical benefits of the technology as it evolves. Post-registration studies are specifically mentioned to achieve these goals, highlighting the recognition that CMDs may continue to develop and change after initial approval.

Anforderungen an laufende Anwendungsbeobachtungen mittels Datensammlungen

In addition, according to the dossier for submission, a description of the functions built for medical devices, including decision-making systems based on machine learning processes, is required [4]. It is based on 42 questions, including the following topics:

42 Fragen im Einreichungsdossier zur Beschreibung der Technologie, den konkreten Trainingsdaten und Performance-Charakteristiken

- description of the purpose of the technology – what is the target population, and what is the claimed benefit,
- information on the samples and input data used for initial model learning and relearning, as well as a description of input data involved in decision-making,
- description of training, validation, and testing before and after medical device deployment – what is the type of learning, what is the update frequency, and when are humans involved in retraining,
- functional characteristics include performance and qualification, system robustness, system resilience, and explainability and interpretability.

All questions are provided in the supplementary material.

HAS acknowledges that AI knowledge, legislation, standards and good practices are rapidly changing, which is why the guide will be amended over time [4].

HAS aktualisiert das Dokument regelmäßig

3.1.4 Framework from FinCCHTA

General Description

The Digi-HTA framework was developed by FinCCHTA in 2019 [24]. It is also available as an Excel-Sheet and follows a question-answer format.

Veröffentlichung 2019

Purpose

The Digi-HTA is a method developed to assess the suitability of digital products and services for use in social and health care and well-being sectors [54]. It evaluates various aspects of digital solutions to support decision-making in healthcare technology adoption.

Ziel:
(rasche) Nutzenbewertung von DHTs

The Digi-HTA Framework has been developed for the rapid assessment of developing technologies [54]. The Digi-HTA is the first framework that combines the topics of AI and robotics in one framework. The Digi-HTA framework deliberately excludes legal, social and ethical aspects, as these are complex and time-consuming to assess.

rechtliche, soziale und ethische Aspekte werden ausgeschlossen

User

The following users are intended to benefit from using Digi-HTA:

- Representatives of wellbeing services counties (for procurement decisions, market surveys, and piloting)
- Technology companies (to demonstrate product suitability and gain expert evaluation for product development)
- Healthcare organisations (for assessing new technologies before implementation)

Zielgruppe:
HTA-Institutionen, Entscheidungsträger*innen, Hersteller*innen

Structure

The Digi-HTA framework consists of eleven domains and a set of questions for each domain [54]. The first domain, *company information*, *focuses on the business model of the company that offers* the technology, including the use of quality management systems. *Product Information* covers the technical details, including the product's regulatory approvals, technology readiness level (TLR), intended use, and user groups. The *Cost* domain evaluates the financial aspects, such as the initial setup and maintenance costs, as well as any long-term operational expenses or uncertainties. *Effectiveness* assesses the product's clinical benefits, its impact on user behaviour, and any improvements to healthcare processes supported by available evidence from studies or trials.

Systematik:
Aufbau in klassische HTA Module (Fragebogen & technologiespezifische Domänen)

Clinical Safety addresses potential risks, side effects, and adverse events associated with the product, ensuring safety is maintained during its use [54]. *Technical Stability* ensures the product operates reliably, evaluating error handling, downtime, and system performance. The *Usability and Accessibility* domain ensures the product is designed for all users, including those with disabilities, and assesses whether real-world user testing has been performed. *Data Security and Protection* evaluates compliance with data protection laws like GDPR, focusing on the security and privacy of patient information.

zusätzlich werden Anwendbarkeit, Zugänglichkeit und Datenschutz beschrieben

For technologies involving *AI*, specific criteria are used, such as transparency, data sources, retraining processes, and ethical compliance in decision-making systems [54]. *Interoperability* examines the product's ability to integrate with existing healthcare systems, ensuring smooth data exchange. Finally, the

Domänen zu KI und Robotik

Robotics domain focuses on the safety aspects of robotic technologies, such as avoiding collisions, ensuring safe interactions with patients, and evaluating infrastructure requirements like charging and operational times.

Guidance relevant to AI

zusätzliche Kriterien für KI-Gesundheitstechnologien

For AI-enabled DHTs, the framework introduces additional evaluation criteria to ensure safe and transparent use of AI in healthcare [54]:

- **Problem Definition:** The framework evaluates whether the AI technology is designed to address a clearly defined healthcare problem.
- **Transparency:** It assesses whether healthcare professionals can understand the AI's operational logic and decision-making processes.
- **Data Sources**: The relevance and security of the data used for training AI models are examined to ensure that AI systems are using appropriate data.
- **Handling Incomplete/Noisy Data:** The framework considers whether the AI system can operate effectively with incomplete or noisy data.
- **Retraining Capabilities:** It investigates how the AI system can be retrained as new data becomes available, ensuring the system's continuous improvement.
- **Validation:** AI solutions are evaluated to determine whether they have been properly validated in the environment where they will be used.
- **Care Process Changes:** The framework assesses how the introduction of AI may alter existing care processes in healthcare settings.
- **Staff Training**: It considers whether additional training is required for healthcare professionals to work with the AI system.
- **Autonomy vs. Human Oversight:** The criteria evaluate whether the AI system operates autonomously or if human approval is needed for its decisions

All domains and questions of the FinCCHTA framework can be found in the supplementary material.

3.1.5 HTW AI checklist

Checklist derzeit noch nicht veröffentlicht

Health technology Wales (HTW) developed a checklist/"aide memoir" to be used in addition to standard HTA methods for assessing AI in health technologies [51]. The checklist is currently not published.[3]

Purpose

Ziel: strukturierte Nutzenbewertung

To provide a structured approach for assessing AI technologies, ensuring comprehensive evidence collection [51]. The HTW checklist provides general assistance when dealing with AI-enabled DHTs and a checklist for different domains.

[3] As this checklist is exclusively developed for AI in health technologies, no distinction between "general description" and "guidance relevant to AI" is foreseen.

User

HTA practitioners may benefit from using the checklist in various stages of the production of an HTA [51].

Zielgruppe: HTA- Institutionen

Structure

The **general assistance** to review authors covers AI-specific considerations, AI-specific search terms, and generic definitions of AI that may be used in HTA reports [51].

Systematik: Allgemeine und spezifische methodische Anleitung

As for *AI-specific considerations*, review authors should acknowledge potential bias in training datasets, particularly with regard to ethnicity and other risk factors. Hence, it is suggested to select AI-specific outcomes for the evaluation, including, for instance:

Berücksichtigung von potenziellen Verzerrungen

- usability and acceptability,
- area under the curve (AUC),
- rate of reclassification (for diagnostic AI),
- reduction in unnecessary procedures,
- patient outcomes like time to diagnosis and overall survival.

The use of adequate *AI-specific search terms* is further highlighted, entailing different terminology of AI (such as "machine learning", "deep learning", and "AI-assist") [51]. HTW hereby provides a comprehensive list of AI-specific search terms.

Beispiele für adäquate Suchwörter

HTW further provides generic definitions of AI that may be used in the introduction of an HTA report. These definitions focus especially on discriminative AI models.

Bereitstellung generischer Definitionen

The developed **checklist** breaks down the assessment into four domains:

Checklist zu 4 Domänen

1. *Training:* Assess the machine learning model, dataset representativeness, comparator appropriateness, and cross-checking of AI outputs.
2. *Clinical Setting and Use*: Evaluate integration in clinical pathways, user identification, training needs, and potential impacts on health inequalities.
3. *Outputs:* Determine clarity of AI information provided, real-time feedback, and clinical benefits.
4. *Ongoing Support:* Review update plans, pricing models, ongoing support clarity, data monitoring, and evaluation processes.

Each domain contains a set of paired questions, typically ranging from five to ten questions per domain [51]. These questions are designed to be answered using a standard critical appraisal format with the following options: yes, partially yes, partially no, no, and no information available. Additionally, each domain includes a dedicated space for the reviewer to provide supporting information and justification for their judgments.

Methode: KI-spezifische Module in Form eines Fragebogens

The full checklist can be found in the supplementary material.

3.2 Assessments on AI-enabled digital health technologies

30 Assessment von 11 HTA-Institutionen identifiziert

Overall, we identified 30 assessments on AI (assisted) interventions from 11 HTA bodies (Table 3-2). The majority of the assessments focused on areas of radiology and internal medicine. AI functions could be allocated to diagnosis and screening[4], treatment and prediction, with the majority of assessments including AI as diagnostic support. No assessments in the functional category of prevention, prognosis, and monitoring were identified.

Table 3-2: Overview of assessments by primary function of the assessed AI-enabled digital health technologies

HTA body	Country	Function			
		Diagnosis and Screening	Prediction	Treatment	Total
AIHTA	Austria	1			**1**
AQuAS	Spain	1			**1**
CADTH	Canada	3	2		**5**
DHTC	Denmark	1			**1**
HIS	Scotland	1			**1**
HTW	Wales	4			**4**
IACS	Spain	1			**1**
INESSS	Canada	1			**1**
NECA	South Korea	3			**3**
NICE	England	10		1	**11**
NIHR	England	1			**1**
Total		**27**	**2**	**1**	**30**

Abbreviations: AIHTA … Austrian Institute for Health Technology Assessment, AQuAS … Agency of Health Quality and Assessment of Catalonia, CADTH … Canadian Agency for Drugs and Technologies in Health, DHTC … Danish Health Technology Council, HIS … Health Improvement Scotland, HTW … Health Technology Wales, IACS … Institute for Health Sciences of Aragon, INESSS … Institut National d'Excellence en Santé et en Service Social, NECA … National Evidence-based healthcare Collaborating Agency, NICE … National Institute of Health and Care Excellence, NIHR … National Institute of Health Research

3.2.1 Diagnosis and Screening

27 HTA-Berichte zu Diagnostik und Screening

For the functional category of diagnosis and screening, 27 assessments conducted by 11 HTA bodies were identified. Assessed AI-enabled DHTscan be used in diverse medical specialties such as radiology, neurology, oncology, general medicine, ophthalmology and dermatology. The assessments were conducted by the Austrian Institute for Health Technology (AIHTA) in Austria [55], the Danish Health Technology Council (DHTC) in Denmark [56], the National Evidence-based healthcare Collaborating Agency NECA in South Korea [57-59], AQuAS and the Institute for Health Sciences of Aragon (IACS) in Spain [60, 61], the Canadian Agency for Drugs and Technologies in Health (CADTH) and another Canadian HTA institute (Institut National d'Excel-

[4] Diagnosis and screening were combined due to numerous hybrid AI-enabled DHTs performing both functions.

lence en Santé et en Service Social INESSS) [62-64], HTW in Wales [65-68], Health Improvement Scotland (HIS) in Scotland [69] and NICE and the National Institute of Health Research (NIHR) in England [70-81].

134 KI-Produkte

In total, 134 AI products were mentioned in the assessments. The primary function of AI was to assist in the diagnostic pathway, either as support of image review [58, 62, 63, 65, 66, 69-76, 78, 79, 81], during a test procedure (spirometry [68], colonoscopy [56, 60], dermatoscopy [61, 80], electroencephalogram (EEG) [77], electrocardiogram (ECG) [57, 59, 67]), or to facilitate patient-clinician interaction [55, 64]. Some technologies described in the assessments (n=10) have a secondary function as screening technologies The exact AI type was not described in detail in most of the included assessments, however, if described, AI was referred to being a type of machine learning or deep learning.

Radiology

10/27 HTA-Berichte zu KI in der Radiologie, v. a. zur Unterstützung bei der Bildbefundung

Ten identified assessments [58, 62, 69-76] evaluated interventions with AI-health technologies in the field of radiology. Of these, six assessments [70-75] were conducted by NICE (England), while the remaining assessments were conducted by HIS (Scotland, n=1) [69] NIHR (England, n=1)[76], by NECA (South Korea, n=1)[58] and CADTH (Canada, n=1) [62]. The reports by NICE were either full guidances [75], medical innovation briefings [70, 72, 74], or early value assessments/health technology evaluations [71, 73], NIHR published a systematic review and cost-effectiveness analysis [76]. The HIS report was an "innovative medical technology overview" [69]. The report by NECA was an early evaluation of the potential of a health technology [58]. The assessment written by CADTH was a health technology review [62]. An overview of this application area is provided in Table 3-3.

Features of interventions and comparator

2 Hauptfunktionen (56 Produkte): unterstützte Bildbefundung und Priorisierungsfunktion

Two main functions (56 AI products) were identified across all assessments, with the primary purpose of AI being to support a radiologist in various tasks [58, 62, 69-76]. Consequently, AI-supported imaging review (e.g., x-ray, mammography, computed tomography (CT)) was one function in the included assessment, and the second main function of three assessments also was improving prioritisation of radiologist's review acting as a triage alert system [58, 73, 75]. The comparator was non-assisted image review in all assessments.

KI zur Klassifizierung von Notfällen

The CADTH report [62] evaluated the integration of ChatGPT into radiology workflows for clinical decision support and expediting radiology reports. HIS assessed a deep learning model embedded in software that analyses chest X-rays to identify high-risk images [69]. NECA evaluated an AI-based technology for diagnostic support and large vessel occlusive stroke screening. It used non-contrast CT scans to classify emergency cerebral large vessel occlusions, determine their hemisphere, and notify healthcare professionals [58].

KI zur Analyse von Kopf-CTs

NICE evaluated several AI-enabled DHTs across different applications. For brain CTs, AI technologies were assessed for enhancing stroke diagnosis by identifying, quantifying, and notifying clinicians of acute stroke-related brain structures [75]. Another NICE assessment focused on AI analysis of brain CT scans to detect various abnormalities, including stroke, trauma, and dementia, with the AI reporting findings and alerting radiologists to critical cases for prioritisation [70].

In chest CT analysis, NICE evaluated AI-enabled DHTs using fixed algorithms for lung nodule detection and volume measurement [74], as well as deep learning systems for highlighting, labelling, and prioritising abnormal findings [71]. NICE also considered AI in chest X-ray analysis for identifying and prioritising abnormalities [73].

KI zur Analyse von Lungen CTs

Additionally, NICE assessed AI in mammography, focusing on detecting and characterising suspicious features and predicting malignancy likelihood [72]. The NIHR evaluated AI-assisted review of brain CTs for detecting stroke-specific brain abnormalities and notifying medical specialists [76].

KI zur Analyse von Mammographien

The AI systems described in the CADTH report [62] can be classified as "augmentative AI". These tools serve as add-ons for clinical decision support and documentation. However, they do not interpret and prioritise, requiring physicians to do these tasks and thus falling under the augmentative AI category. Similarly, the NICE report on chest CTs for lung nodule detection [74] also discusses augmentative AI technology. In contrast, the AI-enabled DHTs used in chest X-rays [69, 73], brain CTs [58, 70, 75], chest CTs [71, 76], and mammography [72] are classified as autonomous AI, level 1. While these systems analyse images and interpret them for risk management and prioritisation, they still require a physician's action, distinguishing them from autonomous levels 2 and 3.

Autonomie: augmentative bis Level 1 autonome KI (ärztliche Entscheidungen werden benötigt)

AI system training data were reported in five AI products, with a range of 250,000 to 4,400,000 images used to train the algorithm to recognise patterns [69, 72]. In one assessment [70], the company claims for one product that algorithms have been trained on clinical examples and validated against consensus guidelines. In another assessment [75], it is reported that data from scans in clinical practice are not used to further develop algorithms in the software. Still, they are developed using CT scans held by the company or accessed through research studies. Consequently, fixed algorithms are used in clinical practice. Further specifics regarding the precise algorithms used were sparsely reported.

Trainingsdaten für 5 Produkte vorhanden und reichen von 250.000 bis 400.000 Bilder

The AI systems are intended to increase efficiency through, for instance, reducing the radiologists' workload (e.g., review time of images), patients' waiting time, and improving diagnostic overall performance.

Erwartungen an die Technologie sind u. a. verbesserte Effizienz

Inclusion criteria and study methodology of assessments

The identified HTA reports included any studies evaluating the respective AI system. The reviews selected the AI-assisted imaging process or AI-assisted triage system as the intervention of interest. Patients with suspected lung cancer undergoing chest X-rays [69], referred from primary care who are undergoing chest X-rays [73], with suspected brain abnormalities or suspected stroke [58, 70, 75, 76], undergoing mammograms for screening or diagnostic purposes [72], patients who have been referred for chest CT [71] and patients undergoing a chest CT scan because of suspected lung cancer [74] were included in the HTA reports. Usual care consisted of conventional imaging review and was the comparator in all assessments.

HTA-Berichte inkludierten alle Studiendesigns, welche die passende Intervention und Population eingeschlossen haben

The identified assessments predefined the following endpoints for evaluating the benefit of these AI-assisted technologies: Five HTA reports defined clinical effectiveness endpoints consisting of time to review, treatment, referral, or diagnosis [70, 73-76]. Four identified HTA reports defined endpoints related to diagnostic accuracy, including sensitivity and specificity [70, 73-75]. Some further reviews also predefined ethical (e.g. equality considerations)

5/10 Berichte hatten vordefinierte Endpunkte zur klinischen Wirksamkeit; 4/10 Berichte zur diagnostischen Genauigkeit

[70] or social [74, 76] (e.g. clinician acceptance [76]), organisational (e.g. length of stay [70, 76]) and economic (e.g. cost-effectiveness [74, 76]) endpoints. Five assessments did not predefine any endpoints [58, 62, 69, 71, 72] and included any endpoints defined by the identified studies.

2 Institute befolgten die NICE-Standard-Methodik

The identified assessments from NICE [70-75] and NIHR [76] followed the methodology described in the NICE Diagnostics Assessment Program manual [82] or the general methods of the NICE early value assessments [83].

Methodological characteristics of included studies in assessments

The identified studies were predominantly observational. Additionally, studies utilised retrospective data to validate AI algorithms.

eingeschlossene Studien in den HTA-Berichten waren primär retrospektive Beobachtungsstudien zur Validierung des Algorithmus

Thirteen studies evaluated AI-supported review of mammograms (retrospective observational studies using data to develop and validate the algorithm). These studies were partly included in the form of conference proceedings (n=3) or conference abstracts (n=3) [72], two conference abstracts of retrospective studies [71], as well as 24 retrospective studies and three prospective studies for chest CTs in suspected lung cancer [74]. Eleven studies were included for chest X-rays in patients with suspected lung cancer, covering one retrospective observational study, two systematic reviews and two cohort studies [69]. Eleven retrospective studies were included for AI analysis of chest X-rays in patients referred from primary care [73]. Included studies for the evaluation of AI in brain scans covered nine retrospective observational studies using data to develop and validate the algorithm, of which seven were diagnostic accuracy studies, one prospective comparative study and one study with a historical control group [70]. One assessment concerning the detection of stroke in CTs included 22 studies (retrospective observational studies to validate the accuracy of the algorithm, either as a full paper or conference abstract) [76] and another report three studies (manufacturer clinical trial results report, conference abstracts) [58]. One further assessment included 15 retrospective observational studies using data to validate the accuracy of the algorithm [75]. One report (horizon scan) of ChatGPT as clinical decision support in radiology did not specify included studies and study designs [62].

beschriebene Endpunkte beinhalten u. a. Sicherheit, Wirksamkeit, diagnostische Genauigkeit und organisatorische Aspekte

Regardless of predefined endpoints, the final assessment documents described a variety of domains and endpoints covering safety (patient safety [69], technical failure [69, 73], adverse events [69, 76]), diagnostic accuracy (AUROC [58, 72, 75], sensitivity and specificity [58, 70, 72-76], segmentation accuracy [71], accuracy rate [62]), clinical effectiveness (time to diagnosis or treatment [58, 70, 72, 74-76], recall rate [72], generalisability and time to report [73, 74], triage outcomes and time to referral [73], time to decision [69]), equality considerations [69, 70, 72-74], patient and social aspects (patient and user experience [69], people most likely to benefit [73, 74]), organisational aspects (required infrastructure for implementation, training requirement, effect on hospital resource requirements [72], resource consequences [70, 71], practical implications, standardisation of report writing [73], length of stay [70, 75, 76], number of treatments [75], ease of use [76], and turnaround times and documentation time [62]), and cost and economic aspects (technology purchase and implementation cost [70-72], cost effectiveness [73, 74, 76], basic budget analysis [69]).

Challenges

Herausforderungen beinhalten die Einbettung in bestehende Infrastruktur und Unklarheiten zu Trainingsdaten

The assessments identified several key challenges associated with AI in radiology. Training bias and uncertainty of training data were the most frequently mentioned issues noted in three assessments [62, 69, 71]. Compatibility problems between hardware and software and insufficient IT capacity were reported twice [70, 71]. Other challenges included concerns about the generalisability of AI systems [76], data protection issues [62], acceptance of AI technology [62], and equality concerns [72].

Conclusions on the evidence

Overall, assessments concluded on diagnostic performance, clinical effectiveness, cost-effectiveness, and ethical aspects.

unzureichende Evidenz zu den meisten KI-Anwendungen deutet auf eine gute diagnostische Genauigkeit hin

HTA assessors' conclusions were somewhat positive regarding the diagnostic performance of certain technologies used as an add-on to clinical review. However, NICE judged the evidence on diagnostic performance to be limited for AI-assisted review of mammography, brain CT, and chest CT scans [70-73], for which it was highlighted that the sensitivity increased, whilst specificity decreased when using AI next to conventional clinical review [74]. In one assessment [73] evaluating an AI-supported review of chest X-rays, no evidence was found. There is insufficient evidence for AI-assisted CT reviewing in suspected stroke [75], but the NECA report indicated high diagnostic accuracy [58]. CADTH concluded that there is a potential for generating inaccurate information and diagnoses [62].

unzureichende Evidenz zum klinischen Nutzen

Four HTA reports (HIS, NICE, NIHR) concluded on clinical effectiveness, of which three highlight that the evidence on the assessed technology (chest-X-ray, brain CT) is insufficient [69, 75, 76]. The other one (NICE) could not identify evidence for the population of interest [73].

Großteils unzureichende Evidenz zur Kosteneffektivität

Regarding cost-effectiveness, the author's conclusion was somewhat positive in one assessment by NICE [74], indicating cost-effectiveness when used alongside clinician review for people having chest CT scans as part of a targeted lung cancer screening program. Most assessments, however, concluded that there is insufficient evidence of cost-effectiveness [69, 71, 72]. In brain CTs, limited evidence suggests a greater resource use than standard care [70]. No evidence was found for the population of interest in chest x-rays [73] as well as for the detection of stroke in CT scans [76].

No conclusions were derived for other domains.

Table 3-3: AI in Diagnosis – Radiology

Vignette 1: AI in Diagnosis – Radiology	
AI main functions	AI-assisted image review
AI type (e.g. machine learning, large language model, CNN, unspecified)	Discriminative AI (ML/Deep Learning/CNN)
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=9) AI-specific HTA methods: dice coefficient: (n=1)
N of assessments	10
HTA institutions/country	NICE/England (n=6), HIS/Scotland (n=1), NIHR/England (n=1), NECA/South Korea (n=1), CADTH/Canada (n=1)
Reported outcomes	Diagnostic accuracy (e.g. sensitivity, specificity) Clinical effectiveness (e.g. time to review/treatment/decision) Ethical (e.g. equity considerations)

Vignette 1: AI in Diagnosis – Radiology	
Reported outcomes *(continuation)*	Patient & Social (e.g. patient experience) Organisational (e.g. required infrastructure) Cost & Economic (e.g. implementation cost, cost-effectiveness)
AI products in assessments	n=50 Screening and diagnosis: Mammography (n=5) Diagnosis: Brain CT (n=17), Chest X-rays (n=14), Chest CT (n=13), Radiology (n=1)
Expected benefits	Efficiency (reduced workload and waiting times) Diagnostic Accuracy (increased detection of suspicious findings) Workflow and Process Improvement (enhanced triage, prioritisation) Patient Outcomes (HRQoL)
Conclusion of evidence syntheses	**Mammography:** *NICE:* Limited evidence suggests that the technology might improve performance and reduce workload in the review of mammograms. No clinical validation studies or evidence for cost-effectiveness were identified. **Brain CT/Stroke diagnosis:** *NIHR:* The available evidence is not suitable to determine the clinical effectiveness of AI-assisted CT scan review systems, and there is currently no evidence for superiority in cost-effectiveness *NICE:* Uncertainty about comparative diagnostic accuracy in AI-assisted CT scan review in suspected stroke. Inconclusive evidence on clinical efficacy (time to treatment, patient outcomes). Some evidence indicating cost-effectiveness (NICE STROKE) Need for more quality evidence on accuracy, clinical and cost-effectiveness *NECA:* Evidence suggests the potential for high diagnostic accuracy and potential improvement in patient care through early detection and intervention/treatment **Chest X-Rays:** *HIS:* No published evidence on clinical effectiveness, cost-effectiveness or safety identified *NICE:* The available evidence is not suitable to assess accuracy and clinical effectiveness in the population of interest. Unknown cost-impact and safety results. Limited evidence suggests reduced time to reporting and treatment and increased identification of treatable cancers. Further evidence is needed on diagnostic accuracy in addition to clinician review, on risk of false positives/negatives and impact on review time. **Chest CT:** *NICE:* Evidence suggests comparable accuracy with clinician review and faster review times depending on the experience level of the clinician reviewing. Increased sensitivity at reduced specificity For screening purposes, evidence suggests cost-effectiveness when used alongside clinician review, but costs vary depending on the technology. Not enough evidence on use outside targeted screening **CRX reporting:** Limited evidence raises concerns about the validity of training data and the potential for generating fictitious or inaccurate information. No evidence of cost-effectiveness.
Challenges	Training bias/uncertainty of training data (n=3) Compatibility of hardware and software/lack of IT capacity (n=2) Generalisability (n=1) Data protection (n=1) Acceptance (n=1) Equality concerns (n=1)

Abbreviations: AI … artificial intelligence, CADTH … Canadian Agency for Drugs and Technologies in Health, CNN … convolutional neural networks, CRX … chest x-ray, CT … computed tomography, HIS … Health Improvement Scotland, HRQoL … health-related quality of life, HTA … health technology assessment, IT … information technology, ML … machine learning, n … number, NECA … National Evidence-based healthcare Collaborating Agency, NICE … National Institute of Health and Care Excellence, NIHR … National Institute of Health Research.

Internal Medicine

7/30 HTA-Berichte in der inneren Medizin

Seven assessments evaluated AI- enabled DHTs in the field of internal medicine. One was conducted by DHTC (Denmark) [56], one by AQuAS (Spain) [60], two by NECA (South Korea) [57, 59], two by HTW (Wales) [67, 68], and one by NICE (England) [81]. One report was considered a systematic review [60], the reports by NECA were evaluations of the potential [57, 59], the two assessments by HTW were topic exploration reports [67, 68], and the NICE report was an early value assessment/health technology evaluation. An overview of this application area is provided in Table 3-4.

Features of interventions and comparator

7 Produkte (Koloskopie, EKG-Interpretation, unterstützende Befundung in CTCA-Scans und Spirometrie)

Overall, seven AI products were included in the studies. Two assessments evaluated AI-assisted colonoscopy [56, 60] compared to standard colonoscopy. Further, three assessments compared AI-assisted ECG diagnosis and analysis [57] or interpretation [59, 67] to the standard of care or interpretation without AI assistance, and the remaining reports assessed AI-assisted analysis of computed tomography coronary angiography (CTCA) scans in comparison with standard CTCA scan combined with a clinical assessment of risk factors [81] and AI-assisted diagnosis of chronic obstructive pulmonary disease (COPD) compared with standard spirometry [68].

Identifizierung von präkanzerösen Läsionen

AQuAS and DHTCs evaluated AI in colonoscopy as a supporting tool for detecting and characterising precancerous lesions. This technology utilises deep learning or convolutional neural networks (CNN – see Table 1-1) algorithms to analyse real-time images during colonoscopy procedures, alerting clinicians to suspicious areas [56, 60]. The AI complements the clinician's review to enhance the detection process [60].

KI-Bewertung bei COPD und EKG-Diagnostik

HTW assessed two distinct AI applications in healthcare. First [68], they analysed the effectiveness of a high-resolution CO_2 sensor combined with an AI platform for COPD detection. This system measures changes in respiratory gas movement indicative of COPD diagnosis. Second [67], HTW evaluated an AI-assisted ECG pattern interpretation system. This mobile digital application digitises ECG tracing images. AI models interpret the raw signal to classify and diagnose cardiac rhythm abnormalities and cardiovascular conditions, supporting clinicians in their diagnostic process.

KI-Bewertung bei EKG-Diagnostik und zur Analyse von Entzündungsgraden in Arterien

NECA evaluated two AI applications for ECG analysis. The first [59] uses recurrent neural networks (RNN – see Table 1-1) algorithms to analyse 12-lead ECGs with normal rhythm, presenting the probability of paroxysmal atrial fibrillation to assist healthcare professionals in diagnostic decision-making. The second NECA report [57] describes an AI-based algorithm that analyses 12-lead ECGs to detect the possibility of acute myocardial infarction, displaying a score and risk level for diagnostic support. NICE assessed an AI algorithm that assists in analysing inflammation levels in coronary arteries from CTCA scans [81].

Autonomie: augmentative bis Level-1 autonome KI

One technology (diagnosis of COPD) [68] may be classified as "augmentative AI", which solely measures changes in gas movement without interpreting the results. In contrast, the other technologies fall under Level 1 autonomy. These more advanced systems analyse and interpret results from colonoscopies [56, 60], ECGs [57, 59, 67], and CTCA scans [81]. However, despite their analytical capabilities, they still require a physician's oversight and final decision-making.

Training data was not reported in any of the respective assessments.

keine Informationen zu Trainingsdaten

In colonoscopy, AI is expected to improve the detection of neoplastic changes [56, 60]. In ECG, AI is claimed to improve [57, 81], speed up diagnosis and treatment, improve patient outcomes, and act as a triage function [57]. It is also intended to improve patient flow [67], accuracy and patient management [59], as well as risk prediction, increase the motivation of patients to moderate risks and optimise prevention and treatment strategies [81]. AI is expected to enhance COPD detection [68].

Erwartungen an KI sind u. a. Verbesserungen der Effizienz und Genauigkeit im Patientenmanagement

Inclusion criteria and study methodology of assessments

One report [60] included RCTs, diagnostic test accuracy studies, systematic reviews, economic evaluations and observational studies. Another one included primary studies (observational studies) as well as systematic reviews [67]. The report by NICE included RCTs, CCTs, comparative or non-comparative observational studies, before and after studies, model development and retrospective studies, validating the algorithm, and cost-effectiveness analyses [81]. The other reports did not define eligibility criteria for the study design [56, 57, 59, 68].

Einschlusskriterien bezüglich Studiendesigns variierten in den HTA-Berichten

The population covered adults undergoing colonoscopy [56], patients with suspected colorectal precancerous lesions [60], patients presenting with symptoms of myocardial infarction [57], patients with suspected cardiovascular disease in primary or emergency care [67], patients with a history of atrial fibrillation [59], patients with stable cardiac symptoms [81] as well as patients with suspected COPD [68].

Population in Abhängigkeit von der Intervention

One assessment [56] predefined all variables from included studies on safety, efficacy, effectiveness, and efficiency for the evaluation. Five assessments did not specify predefined outcomes [56, 57, 59, 67, 68]. One assessment [81] determined accuracy, clinical effectiveness (e.g. time to results, quality of life), as well as cost and economic aspects (e.g. technology purchase) in advance.

7 Berichte definierten Endpunkte vor

Methodological characteristics of included studies in assessments

Overall, assessments included RCTs, qualitative studies, observational studies, diagnostic performance studies, and systematic reviews.

In detail, for colonoscopy, two RCTs, seven qualitative studies and one interview were included in one report [56], and the other included two RCTs [60]. In the context of ECG, three observational diagnostic test accuracy studies were included [57], as well as three observational studies and eight systematic reviews [67]. In atrial fibrillation, one retrospective cohort and one diagnostic performance study were included [59]. For the assessment of AI in CTCA scans, one retrospective observational study to develop the algorithm was included [81]. Four observational studies and systematic reviews were included to evaluate AI in the detection of COPD [68].

überwiegend retrospektive Beobachtungsstudien zur Validierung des Algorithmus

Described endpoints in the assessments were diagnostic accuracy (sensitivity, specificity [57, 67, 68], AUROC [57, 67], NPV, PPV [57], diagnostic odds ratio, comparative diagnostic accuracy [67]), clinical effectiveness (differences in detection of adenoma [56], adenoma detection rate [60], time to diagnosis [59, 67, 68, 81], time to treatment [67], referrals to spirometry [68], quality of life [67, 68, 81], patient satisfaction [68]), organisational aspects (clinician's acceptance, incorporation in clinical routine, risk of overtreatment [56], number of referrals [67], resource use [67, 81], and infrastructure requirements [81]), ethical aspects (equity considerations [59, 81]), patient and social as-

berichtete Endpunkte inkludieren u. a. diagnostische Genauigkeit, klinische Wirksamkeit, und organisatorische Aspekte

pects (attitude, preferences, experiences [56], acceptance [56, 59], and practicality [59]), and cost and economic aspects (cost effectiveness [56, 59, 60, 68], cost of technology [67, 81]).

Challenges

Herausforderungen stellen die Qualität der Trainingsdaten, sowie Verzerrungen im Algorithmus und mangelndes Vertrauen in die Technologie dar

The integration of AI in internal medicine presents several challenges, according to the HTA reports. A primary concern is the potential increase in false positives and overtreatment, which could lead to unnecessary interventions [56, 60]. The quality and representativeness of training data used to develop AI systems are crucial, as biased datasets may result in poorly performing algorithms across diverse patient populations [68]. This relates closely to the issue of generalisability, where AI models developed in specific contexts may not translate correctly to other clinical settings or patient groups [67]. Additionally, health care professionals might distrust AI technologies [60].

Conclusions on the evidence

Overall, the authors concluded on diagnostic accuracy, clinical effectiveness, organisational aspects and cost-effectiveness.

mögliche hohe diagnostische Genauigkeit bei 5 Indikationen

The authors concluded somewhat positively on diagnostic performance. NECA indicated a potential high accuracy in detecting myocardial infarction [54] and a potential high accuracy in detecting atrial fibrillation [59]. The topic exploration report by HTW also indicated diagnostic accuracy for detecting myocardial infarction when using the AI-assisted ECG interpretation.[67]. Another topic exploration report indicated diagnostic accuracy in the detection of COPD [68]. The NICE report concluded that AI accurately identified high-risk patients in CTCA scans based on one study [81].

mögliches Risiko von Überbehandlungen bei Darmkrebs; mögliche Wirksamkeit bei der Identifizierung von Myokardinfarkten; keine Evidenz zu CTCA-Scans

Considering clinical effectiveness, the evidence indicates a risk of overtreatment in colon cancer, as reported by DHTC [56], and AQuAS [60] concludes that there was increased frequency in the detection of small adenomas (<10mm), but no statistically significant difference in the detection of adenomas above 10mm in size. One NECA evaluation indicates clinical effectiveness due to time savings in detection and intervention concerning myocardial infarction [57]. No evidence concerning patient outcomes was found in the NICE report comparing AI-assisted CTCA scan review with the standard of care [81].

mögliche Ressourcenersparnis bei Myokardinfarkt

Relating to organisational aspects, the NECA report evaluating the potential of AI in the detection of myocardial infarction indicates resource savings [57]. There was no evidence for this endpoint in detecting COPD [68].

keine Evidenz zur Kosteneffektivität in 4 Indikationen, jedoch möglich bei Vorhofflimmern

No evidence was found to conclude on the cost-effectiveness of AI in colonoscopy [60], ECG interpretation [67], detection of COPD [68] and AI-assisted CTCA scan analysis [81]. NECA reports that there is a potential for AI to be cost-effective in the detection of atrial fibrillation [59].

The authors did not conclude on outcomes concerning other domains.

Table 3-4: AI in Diagnosis – Internal Medicine

Vignette 2: AI in Diagnosis – Internal Medicine	
AI main functions	AI-assisted colonoscopy, AI-assisted ECG diagnosis and interpretation, image analysis in CTCA scans, AI-assisted spirometry
AI type (e.g. machine learning, large language model, CNN, unspecified)	Discriminative AI (ML/DL/CNN)
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=5) AI-specific HTA methods: NICE ESF for DHT (n=2)
N of assessments	7
HTA institutions/country	AQuAS/Spain (n=1), DHTC/Denmark (n=1), NECA/South Korea (n=2), HTW/Wales (n=2), NICE/England (n=1)
Reported outcomes	Diagnostic accuracy (e.g. sensitivity, specificity) Clinical effectiveness (e.g. time to diagnosis, quality of life) Safety (e.g. adverse events) Organisational (e.g. resource use) Patient & Social (e.g. user experience) Cost & Economic (e.g. cost-effectiveness)
AI products in assessments	n=7 Colonoscopy (n=2), Spirometry (n=1), ECG (n=3), CATCA (n=1),
Expected benefits	Efficiency (enhance detection) Workflow and Process Improvement (patient flow, patient management) Patient outcomes (improvement of outcomes, enhance motivation)
Conclusion of evidence syntheses	**Colonoscopy:** *DHTC:* CADe-assisted colonoscopy should not be implemented as a decision support tool as evidence suggests risk of overtreatment *AQuAS*: ADR was increased with AI- assistance in colonoscopy. More small adenomas (<10mm) were detected, but no statistically significant difference in detection of adenomas >10mm. No evidence of economic impact found. **Spirometry:** *HTW:* Evidence suggests diagnostic accuracy in AI-assisted spirometry but no evidence of clinical effectiveness or cost-effectiveness **ECG:** *HTW:* Evidence suggests high accuracy/non inferiority, impact on resource savings and faster time to intervention in high-risk patients. No evidence on clinical outcomes and cost-effectiveness identified *NECA:* Evidence for the potential of high diagnostic accuracy. Evidence for the potential of resource savings and clinical effectiveness with faster diagnosis and intervention/treatment for high-risk patients. Insufficient evidence on time savings. Further evidence will be appraised for safety and effectiveness in a new HTA evaluation **CTCA:** The technology might increase accuracy to identify people at risk of heart attack or cardiac death compared to the standard risk assessment alone, but the comparator in the study did not reflect UK standard practice. No evidence on patient outcomes or cost-effectiveness. Not recommended until further evidence is generated.
Challenges	Possible increase of false positives/overtreatment (n=2) Representativeness of training data (n=1) Generalisability (n=1) Distrust in health care professionals (n=1)

Abbreviations: ADR … adenoma detection rate, AI … artificial intelligence, AQUAS … Agency for Health Quality and Technology Assessment, CADe … computer aided detection, CNN … convolutional neural networks, CTCA … computed tomography coronary angiography, DHT … digital health technology, DHTC … Danish Health Technology Council, DL … deep learning, ECG … electrocardiogram, ESF … Evidence Standards Framework, HTA … Health Technology Assessment, HTW … Health Technology Wales, ML … machine learning, mm … millimetres, n … number, NECA: National Evidence based healthcare Collaborating Agency, NICE … National Institute of Health and Care Excellence, UK … United Kingdom

Dermatology

2 HTA-Berichte in der Dermatologie

Two assessments were identified in the field of dermatology. One was a health technology assessment conducted by IACS (Spain) [61], and one further report was a medical innovation briefing conducted by NICE (England) [80]. An overview of this application area is provided in Table 3-5.

Features of interventions and comparator

2 Hauptfunktionen (12 Produkte) – Identifizierung und Priorisierung von Melanomen

Two functions (twelve products) were identified throughout the assessments [61, 80]. The primary function of the AI-enabled DHTs was the assistance in dermoscopy to detect melanoma. A triage function was further described in one assessment [80]. The AI systems described in both assessments can be classified as Level 1 autonomous, indicating a physician is still required.

KI-unterstützte Dermoskopie oder Muttermalanalyse – Bildanalyse und Klassifizierung

The intervention is either AI-assisted digital dermoscopy [61] or AI-assisted mole analysis [80]. A dermoscopy is a non-invasive technique to magnify skin structures not visible to the naked eye. The digital form takes images of a bigger surface to diagnose and monitor numerous skin lesions. The integrated algorithms classify the images and support triage. In one assessment [80], AI systems are not necessarily a dermoscopy add-on but are also used with smartphones or tablets. These technologies are compared to manual dermoscopy and photography or standard of care using a dermoscope or clinical assessment only.

Trainingsdaten für ein Produkt vorhanden

Training data of one AI system (DERM, [80]) were described as proportion of historical and prospectively collected images from UK population. There was no information on training data for other products.

Erwartungen an KI sind unter anderem erhöhte Melanomerkennung und erhöhte diagnostische Genauigkeit

AI-enabled DHTs are expected to increase detection of melanoma in early treatable stages, increase accuracy in diagnosing malignant melanoma. Decrease in unnecessary excision of benign lesions and improve monitoring of patients at risk for malignant melanoma [61], as well as reducing waiting lists and streamlining workflows in dermatology referrals, and earlier diagnosis and treatment of skin cancer and earlier reassurance for people with benign lesions [80].

Inclusion criteria and study methodology of assessments

alle Studiendesigns ausgenommen Fall-Kontrollstudien wurden inkludiert

Regarding eligibility criteria, one study included any study evaluating digital dermascopy and manual photography apart from case-control studies [61]. No further details were described in the other assessment.

One assessment included patients undergoing skin checks [80], and the other one included patients with lesions suspicious of malignant melanoma, including patients with risk factors for cutaneous melanoma [61].

1/2 HTA-Berichten definiert die Endpunkte vor

One assessment [61] predefined the following outcomes: diagnostic accuracy (e.g. sensitivity, specificity, PPV, NPV), clinical effectiveness (e.g. number of excised lesions, reduction in excised lesions, number of lesions needed to excise to diagnose malign melanoma), organisational aspects (e.g. differences in the use of resources), patient and social aspects (e.g. patient acceptance and compliance), and cost and economic aspects (e.g. budget impact, health and social costs). The other assessment [80] did not specify predefined outcomes.

Methodological characteristics of included studies in assessments

Overall, the identified assessments included seventeen [61] and fourteen [80] studies. The institutions primarily identified observational studies.

unterschiedliche Studiendesigns wurden inkludiert

The assessment conducted by IACS included one sequential clinical trial, eight prospective cohorts, one retrospective cohort, two qualitative studies, one budget model, one HTA assessment and one systematic review [61]. The assessment conducted by NICE included one non-randomised comparative study, four prospective observational studies, five cross-sectional studies, three retrospective observational studies to develop and validate the algorithm, and one prospective study, including a retrospective analysis in the second phase of the sequential trial. [80].

evaluierte Endpunkte waren u. a. klinische Wirksamkeit, organisatorische und ökonomische Aspekte

For one assessment [61], all predefined outcomes were evaluated. The other assessment evaluated the following outcomes without predefining them: diagnostic accuracy (sensitivity, specificity), clinical effectiveness (time to diagnosis), safety (issues with diagnostic inaccuracies), organisational aspects (infrastructure needs for implementation, training, effect on hospital resources), ethical aspects (equality considerations), and cost and economic aspects (technology purchase and implementation cost) [80]).

Challenges

falsch-positive und negative Ergebnisse, sowie unterschiedliche Hauttöne stellen Herausforderungen dar

The application of AI in dermatology faces several challenges. False positives and negatives remain a concern, potentially leading to misdiagnosis and inappropriate treatment decisions [80]. AI systems may struggle to identify rare forms of skin cancer, limiting their effectiveness in comprehensive skin health assessments. A critical issue is the difficulty in accurately analysing certain skin tones, which could exacerbate healthcare disparities [80]. Patient trust and compliance pose another hurdle.

Ärzt*innen brauchen Zeit zur Einarbeitung

Additionally, clinicians require adequate training to effectively utilise AI technologies, which demands time and resources [61]. The generalisability of AI studies in dermatology is often questioned. Lastly, the field suffers from a lack of RCTs [80].

Conclusions on the evidence

mögliche Kosteneffektivität von KI in 1 HTA-Bericht

The authors concluded on cost and economic aspects, clinical effectiveness and diagnostic accuracy. IACS [61] suggests that in people with high or very high-risk factors for malignant melanoma, initial screening with manual dermoscopy or digital dermoscopy followed up with digital dermoscopy periodically may be more cost-effective than the standard of care with manual dermoscopy alone. Evidence also suggests a potential benefit for persons with a high-risk factor.

mögliche diagnostische Genauigkeit in 1 HTA-Bericht

The NICE report [80] concludes that the diagnostic accuracy of the technologies is comparable or superior to the standard of care. However, there is insufficient evidence to conclude on clinical and cost-effectiveness.

Table 3-5: AI in Diagnosis – Dermatology

Vignette 3: AI in Diagnosis – Dermatology	
AI main functions	AI-assisted digital dermoscopy AI algorithms analyse the digital images to review and inform the detection of possible melanoma by presenting a risk score.
AI type (e.g. machine learning, large language model, CNN, unspecified)	ML/DL
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA Methods
N of assessments	n=2
HTA institutions/country	NICE (n=1), IACS (n=1)
Reported outcomes	Diagnostic accuracy (sensitivity, specificity, PPV, NPV, AUROC) Clinical effectiveness (time to diagnosis, N of excised lesions, N of lesions identified at in situ stage, reduction in excised lesions) Ethical (equality considerations) Organisational (infrastructure needs for implementation, position in care pathway) Patient & social: (patient acceptance and compliance) Cost & Economic (e.g. health and social costs, budget impact, technology purchase and implementation cost)
AI products in assessments	n=10
Expected benefits	Increased diagnostic accuracy Earlier diagnosis of melanoma in treatable stages Improved monitoring of patients at risk Reduce waiting lists and improve efficiency
Conclusion of evidence syntheses	NICE: Evidence suggests comparable or superior diagnostic accuracy Lack of evidence for clinical and cost-effectiveness in the intended population and settings IACS: Use of AI technology for initial screening and follow-up in patients with high or very high-risk factors might be more cost-effective
Challenges	False positives/negatives (n=1) Inability to diagnose rare forms of skin cancer (n-1) Difficulties in certain skin tones(n=1) Trust/Compliance of patients with instructions (n=1) Adequate training to use technology by clinicians n=1) Study generalisability (n=1) Lack of RCTs (n=1)

Abbreviations: AI … artificial intelligence, AUROC … area under the ROC curve, CNN … convolutional neural networks, DL … deep learning, HTA … health technology assessment, IACS … Institute for Health Sciences of Aragon, ML … machine learning, N … number, NICE … National Institute of Health and Care Excellence, NPV … negative predictive value, PPV … positive predictive value

Ophthalmology

In the field of ophthalmology, two assessments were identified. One assessment was conducted by INESSS (Canada) [63], and the other one was a medical innovation briefing by NICE (England) [79]. An overview of this application area is provided in Table 3-6.

2 HTA-Berichte in der Ophthalmologie

Features of interventions and comparator

Two functions (eight products) have been identified. The main functions of the AI system include assisting in diabetic retinopathy telescreening [63] or retinal scan [79]. One further function is the triage support. These functions were compared to the standard screening conducted by a health care professional. AI-enabled DHTs described in both assessments can be classified as Level 1 autonomous.

2 Funktionen (8 Produkte) – KI zur Diagnostik der diabetischen Retinopathie, KI zur Priorisierung

In detail, AI is used, after the retinal image is taken by a health care professional. It assesses and grades the image immediately and provides support to aid in the decision on triage and further patient management [63]. NICE evaluates AI-assisted detection and grading of diabetic retinopathy. Some technologies can incorporate patient history data and assess disease activity and image quality [79].

KI analysiert das Netzhautbild und klassifiziert es

One assessment [79] reported training data from 0.5 million patients and two million retinal images.

Trainingsdaten in einem HTA-Bericht

The AI technologies are expected to be integrated into screening programs of symptomatic patients [79] and to improve the screening service [63]. Additionally, they aim to enable earlier diagnosis and management of diabetes mellitus-associated retinopathy [63], as well as lead to cost savings [63, 79]. It's important to note that AI is not intended to diagnose diabetic retinopathy alone.

Erwartungen an KI ist u. a. eine frühere Diagnose der diabetischen Retinopathie

Inclusion criteria and study methodology of assessments

Eligibility criteria concerning study types were not reported. The identified HTA reports included patients with diabetes mellitus.

The intervention is the AI-assisted screening and diagnosis of diabetic retinopathy. It can either work by directly taking colour images with a retinal camera, uploading them on a cloud server and producing a report [63, 79], or it analyses images and automatically removes images that show no signs of diabetic retinopathy [63, 79].

keine vordefinierten Einschlusskriterien zu Studiendesigns und Endpunkten

Neither of the studies specified predefined outcomes.

Methodological characteristics of included studies in assessments

One assessment [63] included 23 studies (real-world observational studies, experimental studies, study type not described), and the other one [79] included seven studies (prospective and retrospective observational studies).

RWE, prospektive und retrospektive Studien inkludiert

Although not predefined, the following endpoints were selected to evaluate the AI-enabled DHTs: both reports evaluated sensitivity and specificity for assessing clinical effectiveness. Additionally, one assessment measured PPV, NPV and AUC. Concerning organisational aspects, infrastructure for implementation, impact on resources [63, 79], training and the position in the care pathways were considered. Other selected outcomes covered ethical aspects (equality [79] or equity considerations [63], cost and economic aspects (technology purchase and implementation cost, as well as cost-effectiveness) and legal aspects (data security).

berichtete Endpunkte waren u. a. klinische Wirksamkeit und organisatorische Aspekte

Challenges

The integration of AI in ophthalmology presents several challenges, as presented in the NICE report [79]. Data safety is a paramount concern, as the sensitive nature of patient information requires robust security measures to prevent unauthorised access. Integrating AI-enabled DHT into existing healthcare infrastructures poses technical and operational hurdles that need to be overcome for efficient implementation. A critical issue is the lack of trust in AI-enabled DHTs, which is compounded by uncertainties surrounding the liability for errors when AI is involved in diagnosis or treatment decisions. There is also a risk of AI-enabled DHTs missing important pathologies, potentially leading to delayed or missed diagnoses.

Herausforderungen stellen Datensicherheit, die Einbettung in bestehende Infrastrukturen und mangelndes Vertrauen in KI dar

Conclusions on the evidence

unzureichende Evidenz für klinische und ökonomische Wirksamkeit in 2 HTA-Berichten

INESSS concluded on organisational aspects, indicating that AI-assisted diabetic retinopathy screening could be used to reduce the staff needed to identify and grade diabetic retinopathy. There was insufficient evidence for the clinical and cost-effectiveness of the respective technology [63].

The NICE assessment concluded that there is insufficient evidence for clinical effectiveness, reporting similar to lower clinical efficacy than the standard of care for retinal screening [79].

The authors did not conclude on other outcomes.

Table 3-6: AI in Diagnosis – Ophthalmology

Vignette 1: AI in Diagnosis – Ophthalmology	
AI main functions	AI-assisted detection of diabetic retinopathy
AI type (e.g. machine learning, large language model, CNN, unspecified)	ML
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA Methods
N of assessments	2
HTA institutions/country	NICE/England (n=1), INESSS/Canada (n=1)
Reported outcomes	Diagnostic accuracy (e.g. sensitivity, specificity, PPV, NPV) Ethical: Equality considerations Organisational (infrastructure needs for implementation, position in care pathway) Cost & Economic (e.g. cost-effectiveness, technology purchase and implementation costs)
AI products in assessments	n=8
Expected benefits	Improved screening and triage of patients that need specialist follow-up by integration into current pathway as first/second/arbitrary screener Earlier diagnosis Workload reduction Cost savings
Conclusion of evidence syntheses	NICE: Evidence suggests potential to reduce staff requirements in screening. Need for more high-quality evidence for clinical and cost-effectiveness as well as comparison of different products INESS: Evidence supporting benefits is uncertain with comparable to lower clinical efficacy reported Potential to be integrated into clinical pathway Need for more research on clinical and cost-effectiveness
Challenges	Data safety (n=1) Integration in the healthcare system (n=1) Ownership of errors, lack of trust in the systems (n=1) Lack of trust (n=1) Missing of pathologies (n=2)

Abbreviations: AI … artificial intelligence, CNN … convolutional neural networks, HTA … health technology assessment, INESSS … Institut National d'Excellence en Santé et en Service Social, ML … machine learning, n … number, NICE … National Institute for Health and Care Excellence, NPV … negative predictive value, PPV … positive predictive value

Pathology

2 HTA-Berichte in der Pathologie

Two assessments were identified in the field of pathology. One assessment was conducted in Wales by HTW (topic exploration report) [65], and the other one in England by NICE (medical innovation briefing) [78]. An overview of this application area is provided in Table 3-7.

Features of interventions and comparator

KI zur Diagnostik von Prostatakrebs (3 Produkte)

The intervention of interest covers the AI-assisted diagnosis of prostate cancer (three products). The AI algorithm-based software is expected to identify an area of interest on whole slide prostate biopsy images with the highest likelihood of harbouring cancer while automatically grading and measuring according to the Gleason scale to assist pathologists in decision-making. The AI technology is compared to the diagnosis without AI assistance. The AI-enabled DHTs in the assessments can be classified as Level 1 autonomous.

1 HTA-Bericht beschreibt Trainingsdaten

Training data was reported in one assessment [78] using the digital archive from the Memorial Sloan Kettering Cancer Centre in the United States. No further information was given on the AI algorithm. The authors explained that while alternative AI-based systems exist, the company claims that the algorithm has been developed to be highly robust to variations in slide preparation from different institutions and does not need a per-site calibration using pixel annotations or other forms of calibration data.

effizientere Biopsie erwartet

The technology of interest is expected to either increase accuracy [65, 78] (both) and the speed of prostate biopsy results [65], improve productivity and patient outcomes, and save costs [78].

Inclusion criteria and study methodology of assessments

One assessment [65] planned on including RCTs, SRs, economic evaluations and relevant case studies.

Einschlusskriterien für einen Bericht vordefiniert

The population of interest were men undergoing prostate biopsy. Predefined outcomes covered diagnostic accuracy (e.g. sensitivity and specificity) [65, 78], clinical effectiveness (e.g. time to diagnosis) [65, 78], organizational aspects (e.g. procedure time [65], effect on resources), ethical aspects (e.g. equality considerations) [78], and cost and economic aspects (e.g. cost effectiveness [65], technology purchase [78]). The other assessment [78] did not specify outcomes in advance.

Methodological characteristics of included studies in assessments

u. a. wurden Beobachtungsstudien und systematische Übersichtsarbeiten eingeschlossen

One assessment [65] included eleven studies, and one further assessment [78] five to evaluate the benefit of the technology of interest. The included studies covered two prospective and five retrospective observational studies, as well as one cross-sectional study, one HTA assessment and two systematic reviews [65]. The second assessment included five retrospective observational studies [78].

beschriebene Endpunkte waren u. a. diagnostische Genauigkeit, klinische Wirksamkeit und Organisatorisches

Described outcomes in the included studies were diagnostic accuracy (sensitivity, specificity, PPV and NPV [65, 78]), as well as test concordance [78]), clinical effectiveness (time to diagnosis [65, 78], and number of repeat biopsies), organisational aspects (procedure time [65], infrastructure needs for implementation, training, effect on hospital resources [78]), ethical aspects (equality considerations [78]), and cost and economic aspects (cost-effectiveness [65], technology purchase and implementation cost [78])

Challenges

Einbettung in bestehende Infrastruktur als Herausforderung

The integration of AI in pathology faces several challenges. Hardware-software compatibility is a hurdle, as AI systems often demand computational resources and infrastructure that may not be readily available or easily integrated. The issue of training bias and uncertainty in training data is another critical concern, as AI algorithms are only as good as the data they are trained on [78]. Lastly, the impact of AI on decision-making in pathology is uncertain. While AI can potentially enhance diagnoses, there might be concerns about over-reliance [65].

Conclusions on the evidence

The authors concluded on diagnostic performance, clinical effectiveness, cost-effectiveness and organisational aspects.

limitierte Evidenz deutet auf vergleichbare oder höhere diagnostische Genauigkeit hin

Regarding diagnostic performance, HTW [65] concluded that evidence indicates comparable accuracy (AI-assisted diagnosis of prostate biopsies vs. without AI). NICE [78] states that limited evidence suggests that the device may increase diagnostic performance compared to the biopsy without AI.

ungenügend Evidenz zur klinischen Wirksamkeit

Clinical effectiveness is increased by a shorter time to diagnosis, as indicated by HTW [65], and NICE [78] concludes that there is insufficient evidence for the outcome of interest.

ungenügende Evidenz zur Kosteneffektivität

Considering cost-effectiveness, there is insufficient evidence [65, 78] to conclude the utility of AI-assisted diagnosis of prostate biopsies.

mögliche Reduktion in der Ressourcennutzung

In the context of organisational aspects, evidence identified by HTW suggests reduced resource use in AI-assisted diagnosis of prostate biopsies [65]. The authors did not conclude on other outcomes.

No conclusions were derived on other outcomes.

Table 3-7: AI in Diagnosis – Pathology

Vignette 5: AI in Diagnosis – Pathology	
AI main functions	AI algorithm identifies areas of interest/suspicious for cancer on whole slide biopsy images and grades and measures according to the Gleason scale to assist clinician diagnosis
AI type (e.g. machine learning, large language model, CNN, unspecified)	ML/DL
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=1) AI-specific HTA methods: NICE ESF for DHT (n=1)
N of assessments	2
HTA institutions/country	NICE (n=1); HTW (n=1)
Selected outcomes	Diagnostic accuracy (e.g. sensitivity, specificity, PPV, NPV) Clinical effectiveness (e.g. time to diagnosis, number of repeat biopsies) Ethical: Equality considerations Safety (e.g. adverse events) Organisational (procedure time, infrastructure needs for implementation) Legal (privacy issues) Cost & Economic (e.g. cost-effectiveness, technology purchase and implementation costs)
AI products in assessments	n=3
Expected benefits	Efficiency: Increased diagnostic accuracy and productivity, cost savings Patient outcomes: Improved patient outcomes

Vignette 5: AI in Diagnosis – Pathology	
Conclusion of evidence syntheses	HTW: Evidence suggests comparable accuracy, reduced resource use and shorter time to diagnosis, but unclear evidence on patient outcomes and cost-saving NICE: Limited evidence suggests increased diagnostic performance and productivity
Challenges	Hardware-Software compatibility (n=1) Training bias/uncertainty of training data (n=1) Impact on decision making (n=1)

Abbreviations: AI … artificial intelligence, DHT … Digital Health Technology, DL … deep learning, ESF … Evidence Standards Framework, HTA … health technology assessment, HTW … Health Technology Wales, ML … machine learning, N … number, NICE … National Institute for Health and Care Excellence, NPV … negative predictive value, PPV … positive predictive value

Patient-Clinician Interaction

2 HTA-Berichte zur Patient*innen-Kliniker*innen-Interaktion

In the field of patient-clinician interaction, two assessments were identified. One was an update of a NIHR systematic review conducted in Austria by AIHTA [55], and one further assessment (horizon scan) was conducted in Canada by CADTH [64]. An overview of this application area is provided in Table 3-8.

Features of interventions and comparator

Hauptfunktion: Selbstdiagnose/Chatbot zur Fragenbeantwortung (44 Produkte)

The main function was a system for self-diagnosis [55] and a chatbot created to answer patient's questions [64] (44 products). The technology was compared to standard of care, such as face-to-face appointments and phone hotlines [55]. A comparator in the other assessment was not applicable [64]. The AI-enabled DHTs included in both assessments are related to Level 1-autonomy.

Unterstützung für Patient*innen und Kliniker*innen

In detail, the AIHTA report described AI-enabled DHT designed to aid symptomatic patients with self-diagnosis or support healthcare staff with triage through evaluation of the conditions, suggesting potential diagnoses and options for management [55]. CADTH analyses chatbots trained to generate responses to questions or participate in human-like conversations [64].

Trainingsdaten in einem HTA-Bericht beschrieben

Training data was not reported in one assessment [55]. The other assessment reported that "the AI models are trained on large sets of text-based closed data sets" and that this information is used to generate responses to questions [64].

eine Erwartung ist u. a. der anonyme Zugang zu Informationen

Expectations of the technology are the relief of staff workload, constant availability, anonymous access to information, symptom assessment, and cost savings [64].

Inclusion criteria and study methodology of assessments

verschiedene Studiendesigns wurden eingeschlossen

One report [55] defined RCTs, non-randomised controlled trials (NRCTs), observational studies, register studies, reviews and evaluation reports as studies of interest, while the other report [64] included any publication on AI Chatbots or conversational agents in healthcare settings. Conference abstracts and grey literature would be only included if additional information was provided.

symptomatische Patient*innen wurden eingeschlossen

Either symptomatic patients [55] or patients seeking healthcare information were included in the assessments [64]. Digital symptom-checker applications were compared to usual care [55]. In the other assessment, a specific comparator was not applicable [64].

One assessment did not specify outcomes in advance [64], and the other assessment [55] defined diagnostic accuracy (e.g. sensitivity, specificity), clinical effectiveness (e.g. time to diagnosis and quality of life), organisational aspects (e.g. number of visits to physician), and patient and social aspects (e.g. acceptability).

Endpunkte wurden in 1 HTA-Bericht vordefiniert

Methodological characteristics of included studies in assessments

Overall, one assessment included nine [55], and the other 15 [64] studies.

The AIHTA report included one systematic review, five case vignette studies, two prospective observational studies and one case-control study. [55]. The CADTH report included one systematic review and three scoping reviews [64].

insgesamt wurden 24 Studien eingeschlossen

Regardless of the predefined outcomes, the assessments evaluated the following: one assessment evaluated diagnostic accuracy (sensitivity, specificity [55]). Both reports assessed clinical effectiveness (quality of life and patient satisfaction, time to diagnosis, length of illness, severity of illness [55], behavioural change, mental health symptom improvement, physical activity and health promotion [64]), patient and social aspects (acceptability [55] and user experience [64]). Additional outcomes covered safety (adverse events, patient harm [64]), ethical aspects (transparency, algorithm bias, accessibility and equity [64], organisational aspects (number of physician visits [55]), legal aspects (privacy [64])and cost and economic aspects (costs of products [64]).

beschriebene Endpunkte waren u. a. klinische Wirksamkeit, Sicherheit und organisatorische Aspekte

Challenges

There are some challenges to the integration of AI-enabled DHTs in the patient-clinician interaction. Generalisability and algorithm bias represent a major concern, as AI systems trained on specific datasets may not perform equally well across diverse patient populations, potentially exacerbating existing healthcare disparities [64]. Accessibility is another crucial issue, as the adoption of AI-enabled DHTs may be limited by cost, technological infrastructure, and the digital divide, potentially creating or widening gaps in healthcare quality between different regions or socioeconomic groups [64]. Data protection also emerges as a critical challenge, given the sensitive nature of medical information [55].

Verzerrungen im Algorithmus, Zugänglichkeit und die Einbettung in vorhandene Infrastruktur stellen Herausforderungen dar

Conclusions on the evidence

The authors concluded on diagnostic accuracy, clinical effectiveness, safety, cost-effectiveness and organisational aspects.

Regarding clinical effectiveness, AIHTA [55] concluded that there is currently insufficient evidence to show a medical benefit when it comes to the utilisation of symptom-checker applications. The CADTH assessment suggests the effectiveness of providing information to support behavioural changes, improve mental health symptoms, promote health, and support physical activity when using chatbots [64].

1 HTA- Bericht beschreibt unzureichende Evidenz für die klinische Wirksamkeit, 1 HTA-Bericht suggeriert Wirksamkeit

Concerning diagnostic accuracy, evidence was considered insufficient for the symptom-checker, according to the AIHTA [55].

unzureichende Evidenz für diagnostische Genauigkeit

Safety concerns have been identified in the CADTH assessment regarding some chatbot solutions' lack of real-time updates [64].

zu wenig Updates als Risiko

Concluding on cost-effectiveness and organisational aspects, evidence was inconsistent [55], or not reported [64]. The authors did not conclude on other outcomes.

inkonsistente Evidenz zur Kosteneffektivität

Table 3-8: AI in Diagnosis – Patient-Clinician-Interaction

Vignette 1: AI in Diagnosis – Patient-Clinician-Interaction	
AI main functions	Systems for self-diagnosis and to engage in simulated conversations using human-like language
AI type (e.g. machine learning, large language model, CNN, unspecified)	Large language model
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=2)
N of assessments	2
HTA institutions/country	AIHTA/Austria (n=1), CADTH/Canada (n=1)
Reported outcomes	Diagnostic accuracy (e.g. sensitivity, specificity) Clinical effectiveness (e.g. quality of life, time to diagnosis, length of illness) Safety (e.g. adverse events) Ethical (e.g. transparency) Organisational (e.g. adjusted treatment plans) Patient & Social (e.g. acceptability, user experience) Cost & Economic (e.g. costs of products)
AI products in assessments	n=48 Digital health applications (n=38) Chatbots (n=10)
Expected benefits	Efficiency (reduce workload) Workflow and Process Improvement (reduce hospital stays) Patient Outcomes (anonymous access to information)
Conclusion of evidence syntheses	CADTH: Potential effectiveness identified for providing information to inform behavioural changes, improve mental health symptoms and support health and physical activity promotion but evidence remains unclear on clinical effectiveness, particularly without human interference. Safety concerns regarding lack of real-time updates. No evidence of cost-effectiveness was identified. AIHTA: The analysis of the evidence showed that for symptom-checkers there is currently insufficient evidence to show a medical or organisational benefit, as well as diagnostic accuracy.
Challenges	Generalisability and algorithm bias (n=1) Accessibility (n=1) Data protection (n=1)

Abbreviations: AI … artificial intelligence, AIHTA … Austrian Institute for Health Technology Assessment, CADTH … Canadian Agency for Drugs and Technologies in Health, CNN … convolutional neural network, HTA … Health Technology Assessment, LLM … large language model, n … number

General Medicine

In the field of general medicine one assessment (full guidance) was identified, conducted in the Wales by HTW [66]. An overview of this application area is provided in Table 3-9.

1 HTA-Bericht in der Allgemeinmedizin

Features of interventions and comparator

The main function (six products) is wound management [66]. The technology was compared to usual care without AI assistance. In detail, HTW analysed AI-assisted 3D-imaging of wounds, whereby AI analyses and monitors wounds and associated wound care. Digital images can be taken by a health care professional or the patients themselves. They are securely uploaded and can be used for initial assessment, including measurement of the wound and tissue analyses (identification of infections). The AI-enabled DHTs included in the assessment can be classified as Level 1-autonomous.

Hauptfunktion (6 Produkte): Management von Wunden – Bildaufnahme von Wunden und automatische Messung

Training data was not reported.

In wound management, AI is expected to measure wounds more accurately, support the transfer of care to other health care professionals, monitor wound healing or changes in wounds, and assist remote care [66].

Erwartungen sind u. a. genauere Wundmessungen

Inclusion criteria and study methodology of assessments

The included report [66] defined systematic reviews, RCTs, single-arm trials, and evidence from lower priority sources in their inclusion criteria.

unterschiedliche Studiendesigns wurden eingeschlossen

Patients receiving wound care in any setting were included in the assessments [66]. Integrated digital wound care management systems were compared to usual care [66].

The predefined endpoints to evaluate the benefit of the AI-enabled DHT covered diagnostic accuracy (e.g. test-retest reliability), clinical effectiveness (e.g. wound healing outcomes), safety (e.g. adverse events), organisational (e.g. resource use), and patient and social aspects (e.g. patient adherence to treatment) [66].

u. a. klinische Wirksamkeit und Sicherheit wurden als Endpunkte vordefiniert

The report was a rapid systematic review using standard HTA methods adapted from the Cochrane Rapid Reviews Methods Group [84] and the NICE guidelines manual [85].

Methoden von Cochrane und NICE wurden verwendet

Methodological characteristics of included studies in assessments

The assessment included 18 studies [66], of which nine were cross-sectional studies and nine were feasibility studies.

insgesamt 18 Studien eingeschlossen

Considering review outcomes, the assessment evaluated diagnostic accuracy (reproducibility, accuracy of wound measurement, test-retest or interrater reliability, concurrent validity [66]), clinical effectiveness (quality of life and patient satisfaction, wound healing outcomes, time for wound healing, resolution of infection and number of amputations [66]), safety (adverse events [66]), organisational aspects (resource use, length of hospital stay, completion and accuracy of documentation [66]), patient and social aspects (patient adherence to treatment [66]).

beschriebene Endpunkte waren u. a. klinische Wirksamkeit, diagnostische Genauigkeit, und organisatorische Aspekte

Challenges

The adoption of AI in general medicine faces several practical and technical challenges [66]. Connectivity issues, particularly unreliable Wi-Fi networks, can hinder the real-time use of AI tools in clinical settings. Communication between hospitals and general practitioners presents another significant hurdle, as seamless data sharing and integration of AI insights is crucial for continuity of care but often hampered by incompatible systems. Furthermore, AI-enabled DHTs can struggle with accurately assessing certain types of wounds and analysing diverse skin tones.

Internetprobleme und die Kommunikation zwischen den Krankenhäusern als Herausforderungen

Conclusions on the evidence

Regarding clinical effectiveness, the assessment claims that there is insufficient evidence for the routine adoption of digital wound management instead of the standard of care [66]. In the context on organisational and patient and social aspects, the assessment concludes that these outcomes cannot be evaluated [66]. The authors did not conclude on other outcomes.

unzureichende Evidenz für klinische Wirksamkeit keine Evaluierung organisatorischer Aspekte möglich

Table 3-9: AI in Diagnosis – General Medicine

Vignette 7: AI in Diagnosis – General Medicine	
AI main functions	Automatic wound measurements (digital imaging, automatic assessment, centralised digital dashboard)
AI type (e.g. machine learning, large language model, CNN, unspecified)	Discriminative AI/ML
HTA methods (e.g. applied methodology, evidentiary criteria)	AI-specific HTA methods: NICE ESF for DHT (n=1)
N of assessments	1
HTA institutions/country	HTW/Wales (n=1)
Reported outcomes	Diagnostic accuracy (e.g. reproducibility) Clinical effectiveness (e.g. wound healing outcomes) Safety (e.g. adverse events) Organisational (e.g. resource use, length of stay) Patient & Social (e.g. patient adherence) Cost & Economic (e.g. costs of products)
AI products in assessments	n=6
Expected benefits	Efficiency (reduce workload, accurate measurement) Workflow and Process Improvement (facilitate transfer of care, assist remote care)
Conclusion of evidence syntheses	There is insufficient evidence to support routine adoption, as the impact on clinical management, healthcare resource use and patient outcomes cannot be evaluated.
Challenges	Connectivity (Wi-Fi) Communication between hospitals and GPs Difficulties in certain wounds and skin tones

Abbreviations: AI … artificial intelligence, CNN … convolutional neural networks, DHT … Digital Health Technologies, ESF … Evidence Standards Framework, GP … general practitioner, HTW … Health Technology Wales, ML … machine learning, n … number, NICE … National Institute for Health and Care Excellence

Neurology

One assessment, a horizon scan [77] conducted by CADTH (Canada), was identified in the field of neurology. An overview of this application area is provided in Table 3-10.

1 HTA-Bericht in der Neurologie

Features of interventions and comparator

The AI system's main function (one product) is to assess seizure burden and treatment effects in patients with suspected nonconvulsive seizures by analysing portable EEGs. The technology is compared to conventional EEG. The AI technology can be classified as Level 1-autonomous.

Hauptfunktion: Evaluierung der Anfallslast bei nicht-konvulsiven epileptischen Anfällen

In detail, the algorithm evaluates the EEG signal over five minutes. It assesses seizure burden and treatment effects in patients with suspected non-convulsive seizures to support diagnosis, treatment decisions and patient management [77].

EEG-Analyse für 5 Minuten

Training data is not described in detail. However, CADTH reported that the thresholds used in the algorithm are based on those defined by the American Clinical Neurophysiology Society.

Trainingsdaten wurden nicht beschrieben

The technology is expected to identify patients with nonconvulsive seizures more quickly, particularly in centres without consistent access to EEG or a specialist, enhance initiation of treatment, reduce under or overtreatment, reduce the length of hospital stay, and improve patient outcomes.

erwartete schnellere Identifizierung eines nicht-konvulsiven epileptischen Anfalls

Inclusion criteria and study methodology of assessments

unterschiedliche Studiendesigns wurden eingeschlossen

The authors prioritised any publications on the technology of interest and point-of-care EEG. Conference abstracts and grey literature were included only if they provided additional information.

Patient*innen in der Notaufnahme und Intensivstation

The population consisted of patients in emergency departments and intensive care units. AI-supported portable point-of-care EEG device was compared to conventional EEG.

Endpoints of interest were not predefined.

Methodological characteristics of included studies in assessments

9 Studien inkludiert, Großteils retrospektive Beobachtungsstudien

The identified assessment included nine studies, of which five were retrospective studies, three were non-randomised prospective studies, and one was a cost-effectiveness analysis.

beschriebene Endpunkte waren u. a. klinische Wirksamkeit und organisatorische Aspekte

The endpoints defined in the included studies were diagnostic accuracy (sensitivity, specificity, PPV and NPV), clinical effectiveness (time to correct diagnosis, adjusted treatment plans, reduced treatment escalation, faster discharge), safety (adverse events), organisational aspects (adjusted treatment plans, faster discharge), and cost and economic aspects (cost-effectiveness).

Challenges

Akzeptanz und Unterrepräsentation gewisser Gruppen

The implementation of AI in the field of neurology faces some challenges [77]. These include the acceptability of the technology by patients and health care providers, as well as the potential underrepresentation of certain groups within the training data, leading to the potential to perform poorly across diverse patient populations.

Conclusions on the evidence

Evidenz deutet auf klinische Wirksamkeit hin

mögliche Verkürzung des Krankenhausaufenthalts

Regarding clinical effectiveness, limited evidence suggests that the technology could avoid delayed treatment for patients with suspected nonconvulsive seizures by accessing conventional EEG systems. The technology was associated with shorter stays, changes in treatment plans, fewer escalations in antiseizure medication, and decreased patient transfers to tertiary care. The authors did not conclude on further outcomes.

Table 3-10: AI in Diagnosis – Neurology

Vignette 8: AI in Diagnosis – Neurology	
AI main functions	AI-assisted image review
AI type (e.g. machine learning, large language model, CNN, unspecified)	Discriminative AI (ML/Deep Learning/CNN)
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=9) AI-specific HTA methods: dice coefficient: (n=1)
N of assessments	1
HTA institutions/country	CADTH/Canada (n=1)
Selected outcomes	Diagnostic accuracy (e.g. sensitivity, specificity) Clinical effectiveness (e.g. time to diagnosis, reduced treatment escalation, faster discharge) Safety (e.g. adverse events) Organizational (e.g. adjusted treatment plans) Cost & Economic (e.g. cost-effectiveness)
AI products in assessments	n=1

Vignette 8: AI in Diagnosis – Neurology	
Expected benefits	Efficiency (efficient patient identification, reduction of overtreatment) Workflow and Process Improvement (reduce hospital stays) Patient Outcomes (HRQoL)
Conclusion of evidence syntheses	**EEG:** *CADTH:* Limited evidence from small retrospective studies suggests that the AI system could avoid delayed treatment for patients with suspected nonconvulsive seizures. Use of the AI system was associated with shorter hospital stays, changes in treatment plans, fewer escalations in antiseizure medication and a decrease in patient transfers to tertiary care. Larger prospective trials are required to demonstrate a benefit in real-world settings.
Challenges	Acceptability Underrepresented groups in training data

Abbreviations: AI … artificial intelligence, CADTH … Canadian Agency for Drugs and Technologies in Health, CNN … convolutional neural networks, DL … deep learning, EEG … electroencephalogram, HRQoL … health-related quality of life, HTA … health technology assessment, ML … machine learning, n … number

3.2.2 Treatment

1 HTA-Bericht in der Radiotherapie

One early value assessment [86] conducted by NICE (England) was identified, which assessed AI-enabled DHTs potentially used in radiotherapy. An overview of this application area is provided in Table 3-11.

Features of interventions and comparator

Hauptfunktion (11 Produkte) radiotherapeutische Konturierung von Organen

AI-assisted technologies (11 products) were evaluated for their role in radiotherapy contouring. This AI-enabled DHT supports outlining radiation target areas and identifying organs at risk (OARs), both crucial steps in radiotherapy treatment planning. The primary function of the AI system is to generate initial contours, which are reviewed and refined by healthcare professionals before being used for patient treatment. Manual contouring, atlas-based contouring, and model-based segmentation were considered comparators in the assessment [86]. The AI-enabled DHTs are related to Level 1 autonomy, requiring a physician.

keine Details zu den Trainingsdaten vorhanden

The NICE report does not provide detailed information about the type of AI and specific training data used for the AI-enabled DHTs in radiotherapy contouring. It acknowledges that these technologies have been trained on medical images (e.g., CT, magnetic resonance imaging) but does not specify the datasets, their sources, or the diversity of the data used for training [86].

reduziertes Arbeitspensum und verringerte Wartezeiten als Erwartungen

Based on the NICE report, the technology aims to improve efficiency and accuracy in the treatment process. Specifically, it is expected to reduce workload, increase time for patient-facing tasks, decrease patient waiting times, and potentially lower costs. Moreover, the technology may improve contouring consistency, better align with clinical guidelines, and aid in managing complex contouring cases [86].

Inclusion criteria and study methodology of assessments

verschiedene Studiendesigns wurden eingeschlossen

The authors prioritised quantitative and qualitative study types, including RCTs, real-world evidence, and systematic reviews [86].

Komparator war manuelle Konturierung

Patients undergoing radiotherapy were the population of interest. They received AI-based treatment contouring compared to standard of care, which is manual contouring, atlas-based contouring and model-based segmentation [86].

Predefined endpoints of interest were accuracy (e.g. dice coefficient), clinical effectiveness (e.g. alignment with guidelines, patient satisfaction), organisational (e.g. impact on resource use), ethical, patient and social (e.g. user experience), as well as cost and economic aspects (cost-consequence analysis) [86].

vordefinierte Endpunkte waren u. a. klinische Wirksamkeit und organisatorische Aspekte

Methodological characteristics of included studies in assessments

The identified assessment found 79 relevant studies, of which 15 were prioritised, including eight prospective studies, four retrospective studies, a retrospective study with a prospective part, and two conference abstracts, one from a blinded prospective, the other one from a retrospective evaluation of the algorithm [86].

15 Studien wurden priorisiert, darunter prospektive und retrospektive Studien

The endpoints included accuracy (dice coefficient[5] and qualitative measures, consistency), clinical effectiveness (alignment with guidelines, impact on radiology treatment planning quality, patient satisfaction), organisational aspects (usability, impact on resource use, staff and training performance), patient and social aspects (acceptability, user experience), ethical aspects (equality considerations, algorithmic bias), and cost and economic aspects (cost-consequence analysis) [86].

beschriebene Endpunkte waren klinische Wirksamkeit, organisatorische und soziale Aspekte

Challenges

The integration of AI in treatment faces some challenges, particularly in algorithm bias and generalisability. These issues are especially evident when dealing with anatomical variations, obesity, and scar tissue, which can affect accuracy. Additionally, estimating true resource use in AI-assisted treatments presents a complex challenge since the number of healthcare professionals involved in treatment can vary [86].

Verzerrung im Algorithmus und anatomische Variabilität als Herausforderungen

Conclusions on the evidence

NICE concluded that nine technologies can be used [86]. Concerning clinical effectiveness, evidence indicates that AI contouring performs similarly to the comparators but may have difficulties with specific anatomic sites or difficult positions. There was strong evidence for the potential usefulness of the technology. Considering organisational evidence, it suggests time saving compared to manual contouring. Concluding on cost and economic aspects, a cost-consequence analysis suggests a potential cost saving, depending on individual technology costs. The authors did not conclude on other outcomes.

Evidenz indiziert ähnliche Konturierung wie manuell starke Evidenz für Brauchbarkeit mögliche Zeitersparnis

Table 3-11: AI in Treatment – Radiology

Vignette 9: AI in Treatment – Radiology	
(Medical) specialty	Radiology
AI main functions	AI-assisted radiotherapy treatment contouring
AI type (e.g. machine learning, large language model, CNN, unspecified)	NR
HTA methods (e.g. applied methodology, evidentiary criteria)	Not AI specific
N of assessments	NICE/England (n=1)

[5] A detailed description of the dice coefficient can be found in Radiology

Vignette 9: AI in Treatment – Radiology	
Selected outcomes	Accuracy: Dice segmentation coefficient and qualitative measurement Clinical effectiveness: Acceptability of contours, alignment with guidelines, impact on RX treatment and planning Patient & Social: user experience and satisfaction Ethical: equality considerations Organisational impact: usability, clinician experience Cost & economic: cost-effectiveness
AI products in assessments	n=11
Expected benefits	Efficiency: Reduced workload, reduce waiting lists, cost savings Accuracy: Improved accuracy and consistency and compliance with guidelines, manage complex contouring problems
Authors conclusion:	Sufficient evidence for potential benefits from the technologies (can be used once DTAC approved but must be used with HCP review) AI-assisted contouring performs similarly to standard contouring but potential difficulties with specific anatomic sites, atypical anatomy or difficult positions
Challenges	Algorithm bias and generalisability (anatomical variations, obesity, scar tissue) Estimation of true resource use as the number of involved HCPs may vary as comparator.

Abbreviations: AI … artificial intelligence, CNN … convolutional neural network, DTAC … Digital Technology Assessment Criteria, HCT … healthcare professional, NICE … National Institute for Health and Care Excellence, NR … not reported, RX … radiotherapy,

3.2.3 Prediction

Two HTA reports were identified in the field of prediction. All of the reports used methodologies typically seen in the early phase of the life cycle of a health technology, namely horizon scans [87, 88].

2 HTA-Berichte in der Vorhersage

Palliative care

One HTA report (horizon scan) from CADTH (Canada) investigated an AI-based nudging tool to assist conversations on end-of-life care planning [88]. An overview of this application area is provided in Table 3-12.

1 HTA-Bericht in der Palliativversorgung (Gespräche in der letzten Lebensphase)

Features of interventions and comparator

The intervention of interest defined in the CADTH report [88] was an AI-based nudging tool (two products). The technology is a decision support tool using prompts and alerts to aid clinicians in whether and when a discussion on end-of-life planning is appropriate for their patients. It consists of two components: First, a machine learning mortality prediction algorithm incorporated into an electronic health record (EHR) aims to identify patients for whom palliative care would be appropriate to be discussed. Second, nudges and prompts are then sent to clinicians. The CADTH report [88] identified two AI-nudging technologies designed for clinicians to better identify individuals among patients with cancer with whom an end-of-life conversation should be considered. The nudges are related to Level 1-autonomy, requiring a physician.

KI-basiertes Nudging-Tool (2 Produkte) zur Benachrichtigung von Kliniker*innen, um Palliativgespräche durchzuführen

KI analysiert die Gesundheitsakte zur Sterblichkeits-Vorhersage

Of the two identified nudges, one is commercially available in the United States, whilst the University of Pennsylvania developed the other AI-based nudging system [88]. The commercially available nudge uses an *N-dimensional eigenspace* algorithm and can be used for all cancer types. The algorithm uses various clinical data and billing information (e.g., diagnostic codes, cancer staging). Socioeconomic and behavioural data (such as purchasing chan-

für die Entwicklung des Algorithmus wurden klinische Daten, sowie sozioökonomische und Verhaltensdaten verwendet

nels and life stage) were used for model development. However, it is unclear whether variables in those domains were used to predict individual mortality risk. The AI-based nudge not currently commercially available can also be used in all cancer types and uses a *gradient-boosted tree*, using structured EHR data (e.g., demographic, laboratory and clinicopathologic) as input data.

Entscheidungsfindung ohne KI (z. B. mit der Palliative Performance Scale) als Komparator

The selected comparator of the CADTH report [88] was decision-making on palliative care without AI assistance. In the Canadian care setting, this consisted of using specific tools such as the Palliative Performance Scale. In the CADTH report [88], it is noted that end-of-life conversations are often considered too late in Canada, often caused by prognostic uncertainty and optimism bias.

Erwartungen an KI waren unter anderem eine leichtere Identifizierung von Patient*innen

The expectations of the AI-based nudging tools are to increase the number of end-of-life care planning conversations between clinicians and patients and the number of referrals to palliative care [88]. The technology could also help clinicians to easily identify patients with palliative care needs and consequently improve care for these patients.

Inclusion criteria and study methodology of assessments

alle Studiendesigns wurden eingeschlossen

As for eligible study designs, the CADTH report [88] included any publication on combining AI-based mortality prediction models and behavioural interventions.

Endpunkte wurden nicht vordefiniert

The CADTH report [88] did not specify any predefined outcomes.

The CADTH report [88] utilised a horizon scan methodology typically used in the early phase of the lifecycle of a health technology.

Methodological characteristics of included studies in assessments

3 Studien wurden eingeschlossen

The CADTH report [88] identified two prognostic cohort (algorithm validation) studies, a stepped-wedged cluster randomised trial and a real-world implementation study meeting their predefined eligibility criterion.

interne Validität wurde im Zuge der Validierungsstudien (prognostische Kohortenstudien) getestet

The validation studies tested the internal validity of the predictive performance of N-dimensional eigenspace and the gradient-boosted tree algorithms in 3,671 and 24,582 patients and a threshold for high-risk patients of 5% and 40%, respectively. The settings were community practice and tertiary practice in the United States, using the area under the receiver operating characteristics curve (AUROC), sensitivity, specificity, and positive and negative predictive values as outcomes. These studies did not evaluate the external validation [88].

eine Interventionsstudie untersuchte die Wirksamkeit von gezielten Gesprächsanreizen auf die Palliativversorgung am Lebensende

The stepped-wedged cluster randomised trial compared the nudge developed by the non-commercially available AI algorithm of the University of Pennsylvania with usual care in patients with cancer. The study involved 20,506 patients and 41,021 patient encounters, including 5,520 (13.5%) high-risk patient encounters, using endpoints such as the number of serious illness conversations, end-of-life systemic therapy, and the impact on hospice enrolment, length of stay, inpatient death and end-of-life intensive care unit use [88].

Studie zu möglichen Veränderungen in Palliativ-Konsultationen

The CADTH report [88] identified one further before-after study measuring a potential change in palliative care consults and hospice referrals in patients with cancer.

Challenges

algorithmische Verzerrung stellt eine Herausforderung dar

Concerning the integration of AI-enabled DHTs in palliative care, some challenges were noticed, especially in generalisability and algorithmic bias. Another challenge is the possible alert fatigue in health care professionals, especially in the case of constant alerts [88].

Conclusions on the evidence

limitierte Evidenz zur prädiktiven Leistung

nicht schlüssige Evidenz zur Akzeptanz

In the CADTH report [88], it is concluded that there is currently no commercially available AI-based nudging tool in Canada. Although acknowledging the need for tools to identify and improve care for palliative care patients, the evidence on predictive performance and generalisability due to lack of external validation was interpreted to be limited. CADTH [88] further found inconclusive evidence on patient and user acceptance. CADTH [88] raises concerns about the equity and the generalisability of the AI models. CADTH [88] further stated that there is no evidence of cost-effectiveness. The authors did not conclude on other outcomes.

Table 3-12: AI in Prediction – Palliative Care

Vignette 10: AI in Prediction – Palliative Care	
(Medical) speciality	Palliative Care
AI main functions	predict mortality/predict death risk to identify patients for end-of-life conversations
AI type (e.g. machine learning, large language model, CNN, unspecified)	NR
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=1)
N of assessments	CADTH/Canada (n=1)
Reported outcomes	Safety: Patient safety, Data security Diagnostic/Predictive accuracy: Sensitivity, Specificity Clinical effectiveness: improvement of palliative care planning and delivery Organisational: clinician attitude and response
AI products in assessments	Palliative care prediction tool: n=2 [6]
Expected benefits	Efficiency: increased end-of-life planning conversations between patients and clinicians, increased referrals to end-of-life care, improved end-of-life care
Authors conclusion:	CADTH: Limited evidence on predictive performance and generalisability due to lack of external validation. Inconclusive evidence on patient and user acceptance and equality considerations. No evidence of cost-effectiveness.
Challenges	Lack of generalisability and algorithm bias Alert fatigue as a barrier to implementation

Abbreviations: AI … artificial intelligence, CADTH … Canadian Agency for Drug and Technologies in Health, CNN … convolutional neural networks, HTA … health technology assessment, ML … machine learning, N … number, NR … not reported

[6] Assessment refers to 2 nudge applications being developed but none have been approved for use in Canada.

Triage/patient management

One HTA report (horizon scan) from CADTH (Canada) investigated an AI-based patient flow application [87]. An overview of this application area is provided in Table 3-13.

1 HTA-Bericht im Patient*innenmanagement

Features of interventions and comparator

The intervention of interest defined in the CADTH report [87] was an AI-based patient flow application (seven products). The technology consists of AI algorithms using data from EHR to predict and monitor patient's movement through different stages of treatment/care over time. The CADTH report [87] did not evaluate specific products but exemplified diverse AI-based patient flow systems that are either in development (n=3) or in use (n=1) in Canada. Furthermore, three AI-based appointment scheduling systems are already in use in Canada. Patient flow models used mostly machine learning algorithms. The AI-enabled DHTs is related to Level 1-autonomy, requiring a physician.

KI-basierte Applikation für den Patient*innenfluss (7 Produkte); Vorhersage und Überwachung von Patient*innenbewegungen durch die Datenanalyse aus Gesundheitsakten

Information on AI algorithms and training data was sparsely reported.

The comparator to the AI-based patient flow application was defined as human workflow without AI assistance (standard patient management).

Standard-Patient*innen-management als Komparator

The technology of interest is expected to improve patient flow, support volume forecasting to match the demand of needs with supply resources, and improve the use and allocation of available resources [87].

Erwartungen ist u. a. verbesserte Ressourcenplanung

Inclusion criteria and study methodology of assessments

The CADTH report [87] included any studies using an AI or ML intervention to manage patient flow or for appointment scheduling. No further inclusion criteria, e.g., outcomes or study designs, were applied.

alle Studiendesigns wurden inkludiert

Predefined outcomes were not specified.

Endpunkte wurden nicht vordefiniert

The CADTH report [87] utilised a horizon scan methodology typically used in the early phase of the lifecycle of a health technology.

Methodological characteristics of included studies in assessments

The CADTH report [87] found two narrative reviews and three retrospective studies for AI patient flow management tools. In the reviews and primary studies, the setting was mostly inpatient with diverse patient populations. The comparators to ML prediction models were mostly (multivariate) human workflow prediction models without ML integration. As for outcomes, the reviews and studies included by CADTH mostly defined predictive performance outcomes such as accuracy, sensitivity, positive-predictive value and negative predictive value. For AI patient appointment scheduling tools, CADTH [87] found one systematic review with 11 studies and patients mostly in the outpatient setting. CADTH [87] stated that the identified SR neither reported on the AI model nor the selected comparator. As for outcomes were, missed appointment outcomes (e.g., volume of double booking), resource allocation outcomes (e.g., waiting time), and other outcomes such as visit requests and prediction length.

2 narrative Reviews und 3 retrospektive Studien wurden inkludiert

KI-Vorhersagemodelle wurden mit Vorhersagemodellen ohne KI verglichen

Challenges

Herausforderungen stellen Datensicherheit und -schutz dar

The integration of AI in patient management faces several changes [87]. Privacy and security issues are paramount. Real-world healthcare settings present numerous obstacles to AI implementation, including infrastructure limitations, resistance to change from healthcare professionals, and difficulties in seamlessly integrating AI tools with existing workflows and electronic health record systems. The lack of transparency in AI decision-making processes creates issues of trust and accountability. Furthermore, algorithmic bias remains a critical concern.

Conclusions on the evidence

Evidenz zur Wirksamkeit nicht eindeutig, mögliche Verbesserung des Arbeitsablaufs

The authors of the CADTH report stated that the evidence of the effectiveness of the technology on patient flow to change clinical outcomes and patient experience is unclear. CADTH found some evidence for the ability to forecast the volume of patients as well as to improve workflow and efficiency. CADTH found no evidence of cost-effectiveness [87]. The authors did not conclude on other outcomes.

Table 3-13: AI in Prediction – Patient management

Vignette 11: AI in Prediction – Triage/Patient management	
(Medical) specialty	Triage/Patient management
AI main functions	Predict disease probability/predict and monitor patient's movement
AI type (e.g. machine learning, large language model, CNN, unspecified)	ML
HTA methods (e.g. applied methodology, evidentiary criteria)	General HTA methods (n=1)
N of assessments	CADTH/Canada (n=1)
Reported outcomes	Safety: Patient safety, Data security Diagnostic/Predictive accuracy Clinical effectiveness: length of stay, predicting improvement Ethical: diversity, accessibility Organisational: user acceptance, implementation requirements Patient/social: patient perspective
AI products in assessments	Triage/patient management: n= 7[7]
Expected benefits	Efficiency: Improved use and allocation of resources, improved care, support volume forecasting and match demand of needs with supply resources Accuracy: Improved risk prediction
Authors conclusion	CADTH: Some evidence for effectiveness in forecasting patient volume and improving workflow and efficiency but impact on clinical outcomes, patient experience and cost-effectiveness unclear
Challenges	Privacy and security issues Obstacles to implementation in real-world health care Lack of transparency Algorithmic bias

Abbreviations: AI … artificial intelligence, CADTH … Canadian Agency for Drug and Technologies in Health, CNN … convolutional neural networks, HTA … health technology assessment, ML … machine learning, N … number, NR-not reported

[7] Unspecified examples of AI applications in development or in use in various health centres in Canada, one was named as Discharge Predictor (Signal 1)

3.3 Methodological considerations for the implementation of AI in Austria

Übersicht über die thematische Analyse und einen strukturierten Implementierungsansatz

This chapter provides an overview of the thematic analysis of AI-specific considerations and provides healthcare decision-makers with a structured approach to procuring AI-enabled DHTs. It serves as a practical roadmap for healthcare organisations considering AI implementation.

3.3.1 Thematic analysis of AI-specific themes

Analyse KI-spezifischer Themen

In chapter 3.1, guidance documents with AI-relevant content are described. As some guidance relevant to AI is also addressed within conventional HTA methods, we focus on newly identified methodological aspects currently not described within the EUnetHTA Core Model in this chapter (AI-specific themes, see Figure 3-2). Themes not covered within the EUnetHTA Core Model were found for the technology's safety, technical, ethical, economic, legal and organisational aspects. One further aspect was identified: monitoring of performance throughout the life cycle.

Technical aspects

technische Aspekte (TEC)

To assess the technical performance of an AI-enabled DHT, all documents highlight the importance of information on training data and data requirements. Additionally, various further aspects are related to the deployed algorithm.

zentrale Themen in den Dokumenten: Trainings-Methodik, Umgang mit inkompletten Datensätzen,

NICE's evidence standards framework [3] emphasises the importance of information on training data and data collection methods, including validation data, synthetic data, and the diversity and representativeness of the datasets (Standard 5,19). The framework highlights transparency: companies should clearly describe their management for incomplete data and specify their data requirements, including formats and standardisation needs.

Transparenz,

In the framework by *AQuAS* [50] an AI-specific theme within the domain TEC covered transparency. It should be clear how data is used and who has access to data.

Repräsentativität, Erklärbarkeit und Interpretierbarkeit,

The guidance documents by *HAS* [4, 49] consider it to be important to have extensive requirements regarding training data, data representativeness, and retraining processes. It covers specific questions (Q) like information on AI models and their explainability and interpretability, as well as data origin, pre-processing, missing data and outliers (Q5-12,21,22,29,30,40,41) [4], and the additional area (A) "data requirements" (A2) [49].

KI-Model-Funktion,

The methods document by *FinCCHTA* (Digi-HTA) [54] focuses on providing information about training data (Q8,9,16,17) and how to handle incomplete and noisy data (Q12,13), specifying the AI model, function and relevance (Q2,3,4,18) and retraining processes (Q14,15).

Anforderungen an Datenqualität und Training, sowie Strategien zu erneutem Training

In the *HTW* checklist [51], especially (input) data quality is highlighted by mentioning training, incomplete data, data quality and training requirements (D1.2,1.3,1.4,2.7,2.9). Furthermore, information on deployed AI models (D1.1) and retraining is covered within the checklist. Also, real-time feedback and the handling of data output (D3.3;4.7-4.8) are AI-specific themes that need to be considered.

Safety

To assess safety, one document [4] includes several AI-specific themes:

SAF

One guidance document by *HAS* [4] underlines the importance of data risk management. Companies or manufacturers should provide information on how to handle risks in training data (Q28), how to measure errors (Q39), how input data anomaly can be detected and what impact it could have (Q37,38).

Informationen zu Datenrisikomanagement

Economic Aspects

One document [51] considers AI-specific economic aspects.

ECO

HTW [51] adds a theme for the domain of economic aspects. The theme support costs (D4.4) is considered to be AI-specific.

Informationen zu laufenden (Support-)Kosten

Ethical, Legal and Organisational Aspects

All documents highlight important ethical, legal and organisational issues when evaluating AI-enabled DHTs.

ETH/LEG/ORG

Considering ethical aspects, *NICE* [3] highlights that companies should describe any actions taken in the design of the DHT to mitigate against bias that could lead to inequity (Standard 4).

Beschreibung, wie gegen Verzerrungen im Algorithmus vorgegangen wird und wie Datenschutz eingehalten werden kann

Concerning legal aspects, *AQuAS* [50] underlines the importance of evaluating the degree of compliance with current privacy and data protection legislation (D9.1).

AI-specific themes within the domain of organisational aspects comprise human oversight in the documents by *HAS, FinCCHTA*, *NICE* and *HTW* [3, 4, 51, 54]. A natural person should oversee AI-enabled DHTs. It should be clear how humans are involved in the development (HAS, Q24)), use (HAS, Q34, HTW, D1.7; FinCCHTA, Q22), and retraining (HAS, Q25)).

menschliche Kontrolle wird in 4/5 Methodendokumenten als wichtig eingestuft

Further aspects in the post-deployment phase

Post-deployment monitoring is a crucial aspect of the assessment of AI-enabled DHTs. It does not represent a new domain but needs to be considered for all other domains. AI-enabled DHTs can change their performance, which needs re-evaluation (see Figure 3-2).

Überwachung nach Implementierung bei KI erforderlich

NICE [3] mentions in Standard 6 the monitoring of the output. In addition, Standard 15 considers the performance evaluation, and Standard 15 highlights information on re-evaluation and performance monitoring, especially post-deployment.

Monitoring des Outputs, Performance Evaluation und …

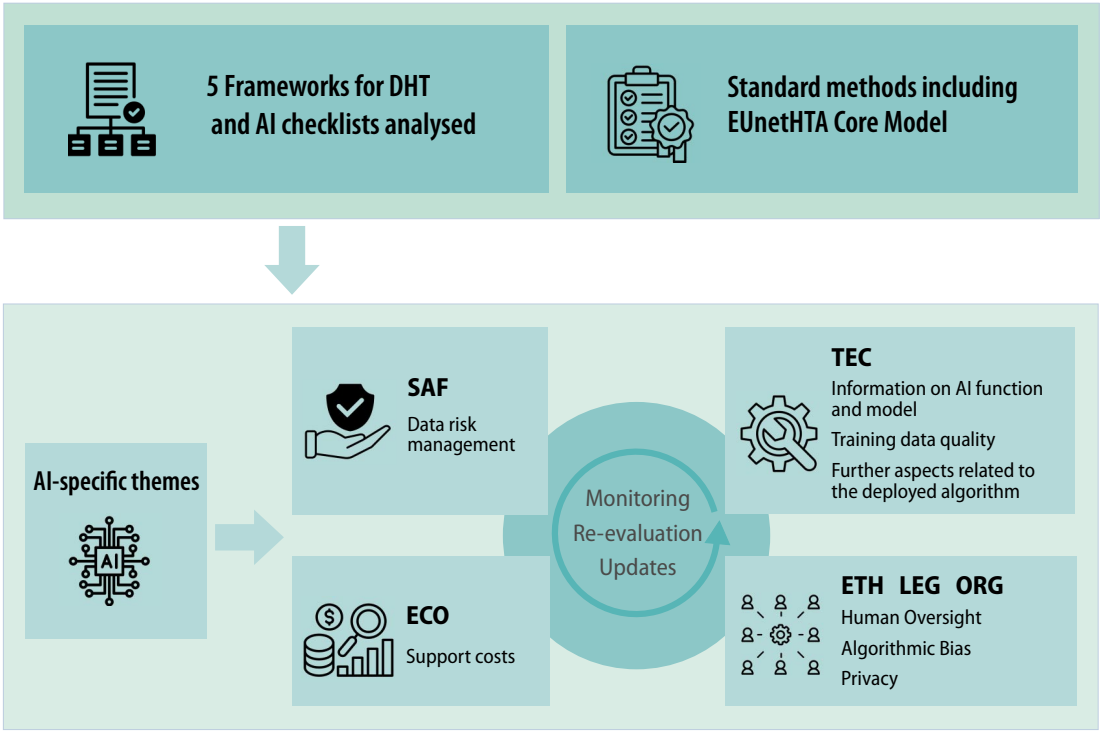
In the guidance document by *AQuAS* [50], post-deployment monitoring (D11.3) was also identified as relevant for AI and not yet covered in the EUnetHTA Core Model.

Several AI-specific questions are reported in one of the guidance documents by *HAS* [4]. They comprise update strategies (Q20,23), performance monitoring (Q34), performance evaluation (Q33), changes in performance (Q27,36), and performance thresholds (Q32,35).

… Update-Strategien bzw. Re-Evaluierung

FinCCHTA [54] includes two questions that are especially related to post-deployment strategies of AI-enabled DHTs concerning information on re-training (Q12,13).

How to handle post-deployment is included in the *HTW* checklist [51] with the domain support clarity (D4.3), information on updates, and monitoring (D4.1,4.4,4.5)



Abbreviations: AI- artificial intelligence, DHT- digital health technology, ECO- economic, ETH- ethical, LEG- legal, ORG- organisational, SAF- safety, TEC- technical

Figure 3-2: AI-specific themes

Considerations beyond EUnetHTA Core Model in HTA-Assessments

Methodische Aspekte außerhalb des EUnetHTA Core Models

In this chapter, endpoints and methods which are not covered in the EUnetHTA Core Model but were applied in the included HTA assessments are described. These methods and endpoints were considered AI-specific considerations. An overview is provided in Table 3-14.

Diagnosis

Radiologie: Trainingsdaten und KI-Hauptfunktion

Some AI-specific endpoints were detected in *radiology*. For the domain TEC, AI-specific themes comprised training data (reported for AI-assisted review of mammograms and chest x-rays. [69, 72]), and further information on the AI algorithm (reported for AI-assisted review of brain CTs and mammograms [70, 72, 75]). The main AI functions were described in all assessments. Concerning diagnostic performance, AI-specific endpoints comprise concordance [74], test failure rate [73] and the dice coefficient[8] [71].

[8] The dice coefficient represents a statistical metric to judge the similarity of two samples used for the measure of segmentation accuracy [71]. The concordance is the agreement between two variables [89]. Test failure rate is a metric used in software engineering to measure the percentage of failed test cases over a particular period [90].

In *internal medicine*, two reports [67, 68] used somewhat AI-specific methods, referring to the NICE ESF [3]. Consequently, additional outcomes were taken into account to evaluate the effectiveness of Tier C digital health technologies. Concerning the domain TEC, the main function was reported in all assessments, while further information on the algorithm was reported for AI-assisted colonoscopy in one assessment [60] and for the AI-assisted review of CTCA scans [81]. Considering the diagnostic performance, AI-specific endpoints were the F1 score, which refers to predictive accuracy [59][9].

innere Medizin: 2 Berichte verwenden den ESF von NICE

zusätzliche Beschreibung der KI-Hauptfunktion und weitere Informationen zum Algorithmus

In *dermatology*, one assessment [80] described AI-specific themes. Concerning the domain TEC, additional information on the AI algorithm was reported.

Dermatologie: zusätzliche Informationen zum Algorithmus

In *ophthalmology*, both assessments included AI-specific themes. NICE reported additional information on the AI-algorithm (domain TEC) [79]. INESSS addresses data security as part of the domain SAF [63]. In this context, data security is related to risk management, thus trying to avoid adverse events.

Ophthalmologie: Datensicherheit berichtet

In *pathology*, HTW [65] incorporated some AI-specific methodology in its assessment by following the NICE ESF [3]. However, HTW's AI-specific content was limited to describing the main function of the AI system within the domain TEC. The NICE assessment [78] was more comprehensive in its AI coverage. Beyond describing the main AI function, it provided detailed information about the AI algorithm, diagnostic performance metrics, and concordance measures.

Pathologie: 1 Bericht verwendet den ESF; 1 Bericht genauere Informationen zum Algorithmus

Concerning *patient-clinician interaction*, the CADTH report [64] tackled AI-specific topics in describing the domain *TEC* the main function, and the training data of the respective AI-enabled DHT. Furthermore, the domain ETH was considered in more detail, considering the algorithmic bias. Also, the domain LEG, including privacy issues, was identified.

Patient*innen-Kliniker*innen Interaktion: algorithmische Verzerrung

In *general medicine*, the report [66] used somewhat AI-specific methods, referring to the NICE ESF. Consequently, to evaluate the effectiveness of Tier C digital health technologies, some additional outcomes should be considered. Described were, however, AI-specific themes, such as domain TEC (AI main function) and diagnostic performance (interrater reliability). The interrater reliability refers to the degree of agreement between two AI-based tests that rate the same condition [89].

Allgemeinmedizin: ein Bericht verwendet ESF von NICE

In *neurology*, AI-specific themes were in light of the domain TEC, concerning the main function as well as additional information on the AI algorithm [77].

Neurologie: Fokus auf KI-Algorithmen

Treatment

In *radiotherapy planning*, AI-specific themes included the description of the main function within the domain TEC. Furthermore, the dice coefficient was used to assess diagnostic accuracy [86].

Radiotherapie: KI-Hauptfunktion

[9] This assessment was still allocated to the application area "diagnosis", since the main function was to assist health care professionals in diagnostic decisions.

Prediction

Vorhersage: in der Palliativmedizin wurde prädiktive Genauigkeit als KI-spezifisch betrachtet

In *palliative care*, AI-specific themes concerning the domain TEC were the description of the main function as well as additional information on the AI algorithm. Predictive accuracy was considered an AI-specific performance measure. The report included a novel study design for the effectiveness domain (EFF), a stepped-wedge cluster randomised controlled trial, currently not depicted in the EUnetHTA Core Model. In the domain EFF, a new study design was identified: the stepped-wedge cluster randomised trial.

auch im Patient*innenmanagement wurde prädiktive Genauigkeit berichtet

Concerning patient flow and management the following AI-specific themes were considered: Concerning the domain TEC, the main AI function was described. Performance was measured through predictive accuracy. [87].

Table 3-14: Overview of AI specific HTA methods charted against standard HTA methods

Function	Number of documents	Diagnosis and Screening								Treatment	Prediction	
		Radiology	Internal Medicine	Dermatology	Ophthal-mology	Pathology	Patient Clinician-Interaction	General Medicine	Neurology	Radiology	Triage and patient man-agement	Palliative care
AI-specific guidance	5	TEC (Information on AI function and model, training data quality, further aspects related to the deployed algorithm) SAF (data risk management) ECO (support costs), ETH, LEG, ORG: (human oversight, algorithmic bias)										
N of assessments	30	10	7	2	2	2	2	1	1	1	1	1
Domain Technical Characteristics	30/30	X	X	X	X	X	X	X	X	X	X	X
Domain Effectiveness & Test performance metrics	27/30	X	X	X	X	X	X	X	X	X	X	X
	25/30	X	X	X	X	X	X	X	X	X	X	X
Domain Safety	7/30	X	X	X	X		X	X	X			
Domain Economic	22/30	X	X	X	X	X					X	
Domain Ethical	17/30	X	X	X	X	X	X			X	X	X
Domain Social	12/30	X	X	X			X	X		X	X	
Domain Legal	1/30						X					
Domain Organisational	21/30	X	X	X	X	X	X	X	X		X	X

Note: X = described domain in at least one assessment of the respective application area.
Grey: themes beyond the EUnetHTA Core Model were described.
Blank: not described.

3.3.2 Guide for procurement of AI-enabled DHTs

4-schrittige Anleitung für die Beschaffung von KI-Anwendungen

The guide for procurements of AI-enabled DHTs includes four steps. First, they are listed, and then each step is further described:

- determining purpose,
- assessing regulatory requirements,
- HTA-evaluation,
  - AI-relevant considerations,
  - AI-specific considerations.
- Monitoring across the life cycle

Purpose

Zweckbestimmung der KI-Anwendung

The first step in implementing an AI-enabled DHT is determining its fundamental purpose. Furthermore, decision-makers should consider the effects on the existing healthcare processes. A comprehensive understanding of who will interact with the AI-system is also essential. The primary user groups may include clinical staff, administrative personnel, and potentially patients.

Regulatory Requirements

regulatorische Anforderungen (EU AI Act, MDR und Datenschutzverordnung)

AI systems in healthcare must comply with multiple regulations: the EU AI Act, MDR, and GDPR. Healthcare AI is typically classified as "high-risk" under the EU AI Act, requiring rigorous pre- and post-market safety assessments. Under MDR, AI software used for diagnostic or therapeutic decisions is generally classified as Class IIa or higher, with potential Class III classification if it could cause death or irreversible health deterioration [33].

Prinzipien zum Datenschutz müssen mit der KI-Anwendung kompatibel sein

Regarding data protection, GDPR principles must be carefully balanced with AI capabilities [91]. While AI often requires extensive data processing, organisations can achieve compliance through:

- purpose limitation flexibility that allows data reuse when compatible with original collection purposes,
- data minimisation through pseudonymisation rather than reducing data quantity,
- clear information to patients about AI-based processing purposes and limitations,
- implementation of privacy by design principles,
- appropriate safeguards for profiling and automated decision-making,
- strict controls on data re-identification [91].

Risiko- und Datenmanagement verpflichtend

Healthcare providers and AI developers must implement comprehensive risk management, quality management systems, and data governance protocols [30]. In addition, the EHDS, once fully implemented needs to be considered in the context of data security, management and training [31].

HTA-Evaluation

Frameworks für DHTs können verwendet und mit KI-spezifischen Komponenten ergänzt werden

Alongside the EUnetHTA Core Model, framework for DHTs such as NICE's evidence standard framework or the adapted version from AQuAS can be used as a starting point also for an evaluation of AI-enabled DHTs and should be supplemented with components of other guidance documents such as the AI-specific checklists developed from HTW or Digi-HTA. These frameworks already highlight the evidentiary requirements depending on the risk-classification of the medical device.

agile EBM-Methoden für die Bewertung von KI

However, the need for agile use of evidence-based medicine methods across the lifecycle of an AI-enabled DHT was underpinned by both HTA methods guidance documents and published HTA reports.

AI-relevant considerations

Wirksamkeit und Sicherheit sind KI-relevant

Our thematic analysis and the analysis of currently conducted HTA reports suggest that standard HTA methods may currently apply for evaluating some domains such as aspects of comparative effectiveness and safety of AI-enabled DHTs. Also, considerations of population, setting applicability, technological characteristics, user experience, implementation requirements, equity impacts, and cost implications are crucial.

AI-specific considerations

technische, ethische und organisatorische Aspekte sind KI-spezifisch

AI-specific themes in guidance were found specifically for the evaluation of technical performance (such as validating the algorithm), ethical considerations (such as algorithmic bias) and organisational aspects (human oversight). Key considerations include the quality and representativeness of training datasets, strategies for performance monitoring and bias mitigation, level of human oversight required, strategies for monitoring, and protocols for re-evaluation as the AI-enabled DHT evolves in use.

Monitoring across lifecycle

kontinuierliche Überwachung über den Lebenszyklus

Continuous monitoring of AI-enabled DHTs is particularly important due to their learning capabilities and potential performance changes. Comprehensive monitoring throughout the entire lifecycle includes surveillance strategies, re-evaluation protocols, update strategies, and clear support structures. These considerations are relevant across all assessment domains and help maintain the quality of the technology. If performance changes significantly alter the technology's intended purpose, the process should be restarted from the initial purpose definition of the checklist (see Table 3-15).

Table 3-15: Checklist for decision-makers

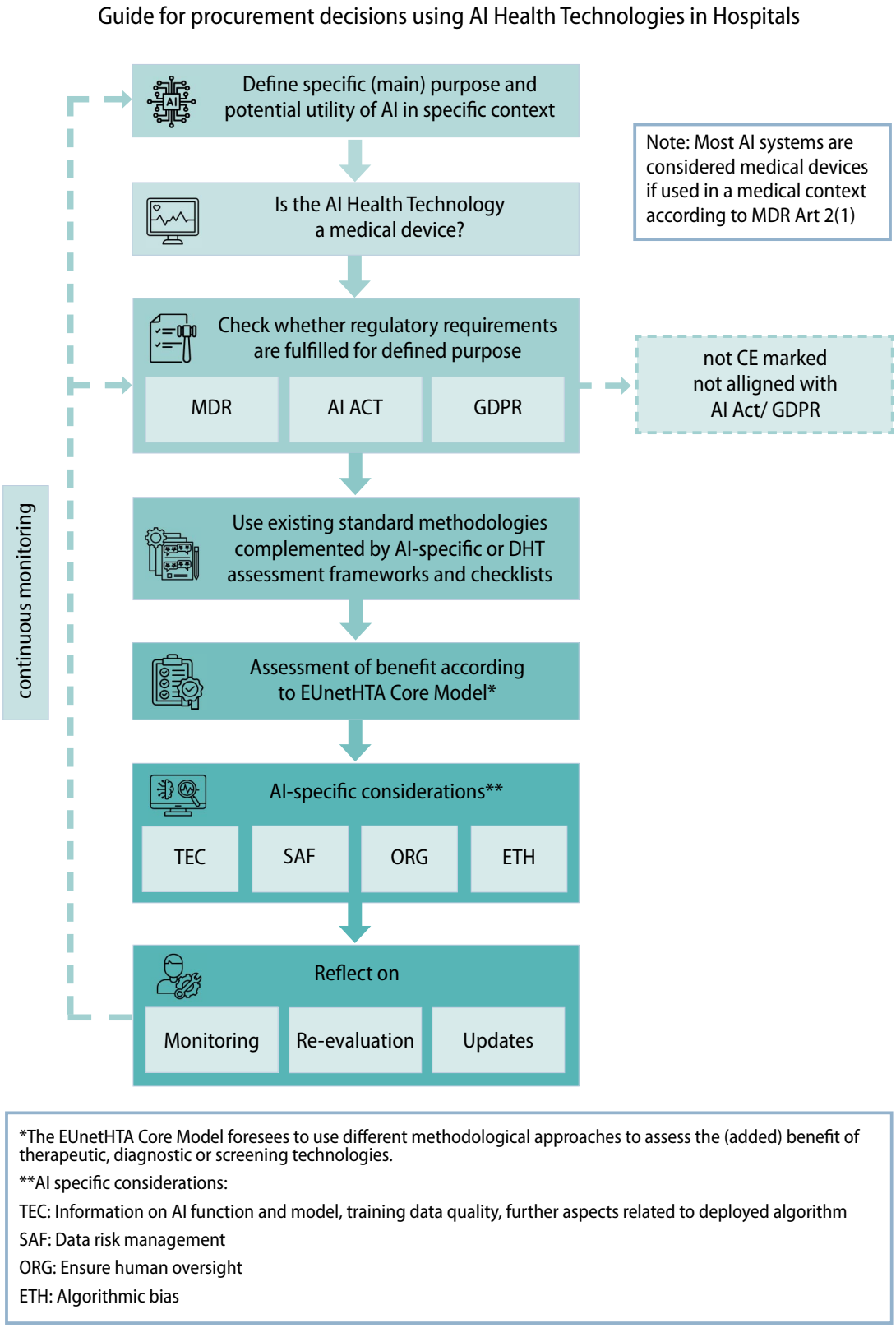
Checklist	
Purpose	
	What is the main purpose of the AI and what is the main utility?
	Which specific healthcare processes will be affected?
	Who are the intended users (healthcare professionals, patients, administrators)?
Regulatory Requirements	
Medical Device Classification	
	Is it considered a medical device under MDR?
	What is its risk classification under MDR (Class I, IIa, IIb, or III)
	What is its risk classification under EU AI Act (high-risk, low-risk)?
	Does the AI-system adhere to high-risk AI systems transparency and safety requirements? (see MDR, EU AI Act)
	Is a valid CE marking present?
Data Protection and Privacy	
	Does the AI-enabled DHT comply with GDPR requirements?
	Are there procedures for patient consent and data rights?
	Consider the EHDS once fully implemented.
HTA Evaluation	
	Reflect on who will conduct the assessment, if HTA-reports are not yet available
AI relevant considerations (covered in standard methodology[10])	
CUR	What are the main characteristics of the health problem, including the proposed AI solution, and the specific patient populations and clinical settings where it can be implemented?
TEC	What are the main characteristics of the AI-enabled DHT?
EFF	What are the clinical benefits and quality of life impact of the AI-enabled DHT, and are the benefits superior to those of existing alternatives?
SAF	Are there risks or possible undesirable effects caused by the AI-enabled DHT that could lead to physical or psychological harm to patients or professionals?
ETH	Does the AI-enabled DHT have an impact on inequalities?
SOC	What is the user experience of the AI-enabled DHT?
ORG	Does the implementation of the AI-enabled DHT involve the training of the professional team?
ECO	What are the costs of acquiring, maintaining and using the AI-enabled technology at the patient and health system level?
AI-specific considerations (not covered in standard methodology)	
TEC	Which data sets were used for training and validating the DHT? Is there a strategy how to handle incomplete data? What is the type of machine learning? How will the performance be measured?
SAF	Are there strategies on data risk management foreseen? How can anomalies of the AI-enabled DHT in operational use be detected?
ETH	Are there strategies to mitigate algorithmic bias in the AI-enabled DHT?
ORG	What is the level of professional oversight? Is staff's approval needed for action, proposed by the AI-enabled DHT? Has the output been cross-checked by a qualified human?
ECO	Is it clear what ongoing support is available for adopters and what it would cost?
Monitoring of performance	
	Define strategies on post-deployment for the AI-enabled DHT.
	How often will the AI-enabled DHT be monitored and by whom?
	How will changes in performance be detected and measured?
	When should a re-assessment of the AI-enabled DHT be conducted?
Check again in case of changes in performance and purpose	

Abbreviations: AI … Artificial Intelligence, CUR … Current Use, DHT … Digital Health Technology, ECO … Economic, EFF … Effectiveness, EHDS … Electronic Health Data Space, ETH … Ethical, EU … European Union, GDPR … General Data Protection Regulation, HTA … Health Technology Assessment, MDR … Medical Device Regulation, ORG … Organisational, SAF … Safety, SOC … Social; TEC … Technical.

[10] E.g. the EUnetHTA Core Model

der Ablauf für den Beschaffungsprozess kann in einem Flowchart dargestellt werden und unterstreicht die Relevanz für kontinuierliches Monitoring

Figure 3-3 illustrates a structured decision-making guide for the procurement of AI-enabled DHTs in hospitals. The guide follows a sequential process, beginning with defining the AI system's specific purpose and utility within the healthcare context. Following the initial assessment of whether the technology qualifies as a medical device (noting that most AI systems in medical contexts are classified under MDR Article 2 [1]), the guide progresses through regulatory compliance verification encompassing MDR, AI ACT, and GDPR requirements. The process incorporates standard methodological approaches to an assessment, with assessment domains selected depending on the function and expectations of the AI-enabled DHT. Domains should be complemented by AI-specific/general DHT assessment frameworks. AI-specific considerations are essential for technical (TEC), safety (SAF), organisational (ORG), and ethical (ETH) assessment domains. Notably, the guide emphasises continuous monitoring throughout the life cycle, comprising re-evaluation and update strategies, as indicated by the dashed line connecting back to earlier stages. This cyclical approach ensures ongoing assessment of the implemented AI-enabled DHTs.



Abbreviations: AI- Artificial Intelligence, DHT- Digital Health Technology, ETH – Ethical, EU- European Union, GDPR- General Data Protection Regulation, MDR- Medical Device Regulation, ORG- Organisational, SAF- Safety, TEC- Technical

Figure 3-3: Guide for procurement decisions

4 Discussion

Summary of Findings

Übersicht über Methoden und Assessments von KI Anwendungen

This review examined methodological approaches to evaluate artificial intelligence (AI) health technologies and provided an overview of internationally evaluated technologies structured according to their primary function. We found five relevant methods guidance documents and 30 HTA reports on AI-enabled digital health technologies (DHTs). Application areas were concentrated in medical fields such as radiology, internal medicine, dermatology, pathology, ophthalmology, general medicine, patient-clinician interaction, radiotherapy planning, palliative care, and patient management.

Methods and Frameworks

Methodendokumente: 3 Assessment-Frameworks, 1 KI-spezifische Checkliste und 1 HTA-Methodenhandbuch

The identified guidance documents describe the evaluation of digital health technologies, including those incorporating AI (NICE, AQuAS, FinCCHTA, HAS [3, 4, 24, 49, 50]) or are designed specifically for AI-enabled DHTs (HTW) [51]. These guidance documents were designed to facilitate the evaluation of health technologies. Furthermore, the documents are primarily intended for HTA agencies, researchers, developers/manufacturers, decision-makers and regulators. Some frameworks are also designed to assist companies in preparing submission dossiers for reimbursement applications. The guidance documents typically describe the technology, clinical effectiveness, economic aspects, safety, and ethical considerations. Some guidance documents are primarily based on AI-specific questions, including checklists. These typically include technology description, clinical effectiveness, economic aspects, safety, and ethical considerations.

KI-Aspekte in Bewertungsdomänen: u. a., technisch, ethisch

The guidance documents [3, 4, 24, 49-51] highlight several aspects relevant to AI-enabled DHTs. Some of these aspects do not exclusively address AI-enabled DHTs but health technologies in general. Consequently, themes were charted against the EUnetHTA Core Model. If they were not depicted, they were considered-AI specific. Across the included guidance documents, AI-specific topics (aggregated themes) were information on AI function and model, training data quality, further aspects related to the deployed algorithm, data risk management, support costs, algorithm bias, privacy and liability, human oversight, implementation in daily routine, monitoring, re-evaluation and updates, which should be considered for all domains.

Assessments on AI-enabled digital health technologies

KI in Assessments vorrangig im Bereich Diagnose und Screening

Most AI-enabled DHTs described in identified assessments (n=30) focused on diagnostics and screening (27/30), particularly in radiology (10/27), followed by internal medicine (7/27). The main function in radiology was AI-assisted imaging review (e.g. in CTs and X-rays). In other application areas, AI-enabled DHTs are, for example, used in electrocardiogram (ECG) interpretations, colonoscopies, and dermoscopies. The "treatment" function (1/30) was linked to a radiotherapy application (contouring areas). At the same time, "prediction" was associated with palliative care (1/30, nudging tool to initiate end-of-life conversations) and patient management (1/30, predict and monitor patient flow). Notably, across all categories, AI was consistently positioned as a tool to help healthcare professionals in specific tasks rather than perform a given task fully autonomously. In most of the subtasks, an AI was involved. The AI-enabled DHT concludes and offers contestable diagnosis/

management options that require action from a physician or health care professionals to implement. This level of autonomy could be classified as a low level of autonomy according to the AMA AI taxonomy [28].

erwartete KI-Vorteile: u. a. Effizienz, diagnostische Genauigkeit, Verbesserung der Versorgungsabläufe

Throughout the included assessments, AI was expected to enhance efficiency by reducing workload and waiting times while increasing diagnostic accuracy through improved detection of suspicious findings. It aims to improve workflows and processes by enhancing triage, prioritisation, patient flow, and management. AI is anticipated to positively impact patient outcomes, including health-related quality of life, by facilitating earlier diagnosis and improving the identification of at-risk patients. AI may also enhance patient access to information, assist in remote care, and facilitate care transfers.

erwähnte Herausforderungen: u. a., Datenqualität, Generalisierbarkeit und Integrierung in bestehende Systeme

Despite these expectations, HTA reports mentioned several challenges across application areas. These include concerns about data quality, algorithmic bias, generalisability, and integration with existing systems. In radiology and internal medicine, AI demonstrates the potential for improved diagnostic accuracy and efficiency, but evidence was considered to be currently limited. Dermatology and ophthalmology applications show promise in screening and diagnosis but struggle with diverse skin tones, rare conditions and overdiagnosis. The identified HTA reports considered the evidence currently insufficient for AI in pathology [65, 78], and general medicine [66]. AI shows potential for time-saving and improved resource allocation for treatment planning [86], advance care planning in palliative care [88], and patient management [87], but more robust evidence on clinical and cost-effectiveness is needed.

HTA-Berichte in unterschiedlichen Zeitpunkten des Produktzyklus

While some HTA reports [55, 60, 61, 66, 74-76, 86] were fully investigating the clinical benefits of AI systems informing investment decisions (full guidances, full HTAs, systematic review), numerous identified reports ([58, 59, 64, 67-73, 77-81, 87, 88]) were HTAs that are typically performed in the early phase of the life cycle of a health technology such as horizon scans and early value assessments, at a point in time when little sound evidence is available.

Standardmethoden häufig für Wirksamkeit und Sicherheit verwendet

Some institutions [65-68] altered their standard methodological approach when evaluating AI-DHTs, using the evidence standards framework from NICE [3]. Of note is that the majority of the HTA reports adhered to standard HTA methods for assessing the *effectiveness* and *safety*, without substantial methodological deviations (i.e., regarding requirements to endpoints, study designs and synthesis methods) from their institutional methods guidance or highlighting methodological AI-specific-themes. Also, the assessment of the accuracy of the algorithm and related *performance* metrics were considered in most HTA reports (25/30) without clearly mentioning whether these can be regarded as linked evidence for effectiveness. Data security is of special interest in AI-enabled DHTs [63].

alle HTA-Berichte erwähnen KI-spezifische Themen

All HTA reports addressed AI-specific themes, primarily focusing on describing the main function of each AI-enabled DHT. Some HTA reports (13/30) went further by including information about the training data used to develop the AI system, as well as additional information on the algorithms [60, 64, 69-73, 75, 77-81, 88].

1 HTA-Bericht beschreibt ethische und gesetzliche Aspekte

Most HTA reports addressed ethical aspects, primarily focusing on equality and equity. One horizon scan [64] highlighted the importance of evaluating algorithmic bias in AI-enabled DHTs. Only one horizon scan [64] briefly examined legal considerations, focusing on privacy compliance and liability with current legislation.

HTA-Berichte verwendeten vorwiegend Standardmethoden bei organisatorischen Aspekten

Assessments followed conventional approaches by using methods like the EUnetHTA Core Model, evaluating standard endpoints such as resource utilisation and length of hospital stays. Similarly, employed traditional methods assessing endpoints like user acceptance and communication strategies. The *economic* analysis also adhered to standard methodological approaches, with no AI-specific adaptations identified.

KI-Evaluation erfordert agile Verwendung von EBM-Methoden

Our analysis also indicates the need for agile use of EBM methods and consideration of HTA across the whole life cycle of AI-enabled DHTs [92, 93]. For instance, the fact that AI algorithms can change through updates requiring re-evaluation was highlighted in multiple methods guidance documents [3, 4, 49-51, 54]. The European Health Data Space (EHDS) could play a crucial role in this context. Through its dual framework of primary and secondary use, the EHDS could enable both secure cross-border patient data sharing and facilitated access to health data for research and innovation under strict security protocols. This infrastructure could significantly enhance algorithm development and validation by providing access to larger, standardised datasets across Member States [31].

EHDS könnte die KI-Entwicklung fördern

und ist stark Kontextabhängig

As some of the assessed AI-enabled DHTs are highly context-dependent, and questions related to implementation on specific levels (what works in which context) is evidently important [94]. For instance, an AI-enabled DHT that was identified in one assessment [88] aimed to nudge clinicians when end-of-life conversations should be considered. Most of the available studies included in the assessment were conducted in the United States and Canada.

digitale Infrastruktur als Voraussetzung für erfolgreiche Implementierung

The successful adoption of digital health technologies more broadly in Austrian hospitals is closely tied to the country's health data infrastructure. Many digital health technologies presently used in Austria operate as isolated systems, with data confined to individual hospitals [55]. A digital infrastructure with systema and data interoperability is often a prerequisite for AI-enabled DHTs to work as anticipated [95]. Addressing this need is one of several objectives within two broader initiatives: the Artificial Intelligence Mission Austria 2030 [96] and the 2023 healthcare reform.

Embedding our study into existing knowledge

Einbettung in bestehendes Wissen: AI-MIND mit ähnlichem Fokus und Ergebnissen

To our knowledge, this is the first study to analyse AI content in HTA methods guidance and HTA reports. However, similar projects have already been conducted in the context of AI and DHT. The AI-Mind project conducted a Delphi survey to evaluate the relevance of topics from the EUnetHTA Core Model along with 20 additional literature-derived topics, categorising them as critical, important, or not important [97, 98]. Key findings identified AI model accuracy, data bias, and human oversight as critical elements, ethical analysis was mentioned most often, followed by clinical effectiveness. In a parallel effort, AQuAS [50] undertook a broader survey examining domains relevant to DHTs, AI, and mHealth, analysing 26 references encompassing 102 frameworks. Their work, which included surveys, a thematic analysis to address terminology variations and a consensus workshop, resulted in additional domains. However, for several domains (CUR, SAF, EFF, ECO, ETH, and LEG) the EUnetHTA Core Model was considered to be sufficient for the evaluation. While AQuAS considered the ethical and legal aspects adequate [50], this contrasts with both the Delphi study from the AI-MIND project and the findings of this report. While the AI-MIND Project [97] viewed the EUnetHTA Core Model as insufficient for AI evaluation, the fact that numerous HTA reports identified in our analysis utilised standard HTA meth-

Ethische Aspekte haben bei AI-MIND Priorität

odology aligned with EUnetHTA Core suggests that it may serve as a valuable foundation, though it needs to be complemented with AI-specific considerations. This approach acknowledges both the established utility of the EUnetHTA Core Model and the need for its adaptation to address unique AI challenges.

Ongoing research on guidance for evaluating AI

laufende Forschungsprojekte in Europa

The **CORE-MD** (Coordinating Research and Evidence for Medical Devices) project provides complementary research focused on the regulation of AI in medical device software. Recently, a comprehensive review [12] captured several important recommendations on medical AI that were not covered in our analysis. For clinical studies, they highlighted CONSORT-AI, which offers reporting guidelines for clinical trials involving AI interventions. They also noted reporting guidelines like MI-CLAIM by Norgeot et al. [99], which proposes minimum information standards for clinical AI modelling. Next to the need for guidance in the context of HTA, the review authors also highlight that regulatory guidance needs to keep pace with the rapid deployment of AI-enabled medical devices. As evidence of this rapid growth, they note that as of October 5, 2022, the FDA had already approved 521 AI-enabled medical devices since 1997. This underscores the pressing need for timely and appropriate regulatory frameworks to ensure the safe and effective implementation of AI in healthcare. The authors suggest focusing regulatory efforts on addressing gaps in current guidance and on challenges unique to AI devices, such as ensuring appropriate use cases, managing iterative software changes, and conducting effective post-market surveillance.

HTA-Methodologie zur Bewertung von DHTs

The **ASSESS DHT** [42] and **EDIHTA** projects [100] plan to provide guidance for HTA of DHTs, under which terms also AI-enabled DHTs are covered. Currently, taxonomies and assessment frameworks are in development. In this context, numerous scoping reviews are currently ongoing with a broader scope including multiple HTA frameworks and checklists that may also have relevant domains for HTA. These European projects will most likely also produce HTA guidance highly relevant for AI-enabled DHTs.

Leitlinien zur Berichterstattung und Erfassung des Verzerrungsrisikos

Furthermore, there are ongoing projects developing reporting guidelines and risk of bias assessments, of which one may be highly relevant for HTA [12]: The PROBAST-AI checklist. It is designed to be a standardised tool that may help evaluate AI prediction models for potential biases. It will enable various professionals – including researchers, doctors, reviewers, and policymakers – to thoroughly assess how these AI models are designed, implemented, and analysed [101].

Diskussionen zur Regulatorik von KI-Anwendungen

There is ongoing debate about how to effectively regulate AI-enabled medical devices that can learn and adapt over time. According to an expert review of regulatory frameworks for AI medical devices, the FDA's 2021 action plan offers a structured framework with predetermined plans on how to handle changes in performance change and real-world monitoring requirements [102]. In contrast, it is argued that the EU's Medical Device Regulation and AI Act [30] lack specific guidance, with minimal detail on implementing algorithm change protocols.

regulatorische Herausforderungen bezüglich Datenverarbeitung, …

Recent research [36] highlights several key regulatory challenges in AI-enabled DHTs that need to be addressed. First, the definition of "public interest" in health data processing requires clarification to ensure appropriate data use. Second, clear frameworks are needed for managing intellectual prop-

erty rights of AI-generated outputs. Third, specific guidelines must be developed for monitoring AI systems and managing algorithm updates. Of particular importance is the interaction between the EU AI Act and the MDR. This interaction affects the CE-marking process, as AI-enabled DHTs must comply with both frameworks to enter the European market [36].

… Updates und Re-Zertifizierung

Under EU MDR (part C in [33]), software medical devices need recertification (new unique device identification [UDI] – device identifier) for significant changes affecting original performance, safety, intended use, or data interpretation capabilities. This includes modifications to algorithms, database structures, platforms, architecture, interfaces, or interoperability. Data interpretation capabilities are particularly crucial, as changes here can directly impact diagnostic or therapeutic decisions. Minor revisions only require an updated version identifier (new UDI-production identifier), covering bug fixes, non-safety usability improvements, security patches, and efficiency updates. Manufacturers must carefully assess changes to data interpretation features to determine if recertification is necessary, ensuring compliance and patient safety while balancing innovation in medical software development [33]. Some critique suggests the EU approach could burden manufacturers and hospitals, potentially affecting procurement decisions and patient access to new technologies [102]. To address this, developing specific EU guidelines for algorithm changes and monitoring is recommended, which would tackle regional concerns around data protection and public AI perceptions [102]. Further work at the EU level may be needed to refine these regulations.

EU-MDR definiert Kriterien für Rezertifizierung von Software als Medizinprodukt, jedoch sind spezifische Leitlinien für die Bewertung des Ausmaßes der Algorithmus-Änderungen ausstehend

Implementation in Austria

A study conducted in 2022 by Austria's National Public Health Institute (GÖG) [37] examined AI usage in Austrian hospitals. The research identified 43 AI products, primarily in diagnostics (56%), treatment improvements (28%), and prediction (19%). Implementation varies, with 40% in regular operation, 35% in pilot phases, and 26% in studies or pandemic responses.

zumindest 43 KI-Anwendungen in Österreich

Among the AI-enabled DHTs that were assessed in the 30 HTA reports, three are currently in use in Austria: Veolity, assessed in two reports [71, 74] assists in reviewing chest CT scans and is implemented in Tyrol. In the same region, two additional chest CT review systems, Contextflow Search and AI-Rad Companion Chest CT [74] are in use. Related versions of the AI-Rad Companion system exist, specifically for chest x-ray (AI-Rad Companion Chest X-Ray) [73] and for contouring organs (AI- Rad Companion Organs R) [86]. These tools are, however, not yet implemented in Austria. In Upper Austria the GI Genius™, evaluated in two HTA reports [56, 60] is being used in clinical practice. This tool enhances standard colonoscopies by detecting and alerting clinicians to potential abnormalities.

3 KI-Anwendungen, die in HTA-Berichten evaluiert wurden, sind in Österreich in Pilotierung

Strengths and Limitations

This HTA report provides a comprehensive overview of the current HTA methodologies and assessments in the context of AI-enabled DHTs. A key strength of this report is its systematic mapping of methods guidance documents and conducted HTA assessments. A further strength may be the categorisation by function and medical specialty, helping to gain an understanding of where AI-enabled DHTs may be currently predominantly used and how international HTA bodies currently assess the benefit of such health technologies.

systematische Kartierung der Methodendokumente als Stärke

Fokus auf HTA-Berichte und Methodendokumente

The focus of our report was mostly AI-enabled DHTs for hospitals. Focusing on HTA methods documents and assessments is a strength for this purpose. Yet, we have not captured AI-enabled DHTs that could potentially influence the health care system's level, since all included AI-enabled DHTs could be classified as Tier C, according to NICE [3].

unveröffentlichte Leitlinien und Methoden können übersehen worden sein

Limitation durch Einschlussrestriktion zu HTA-Berichten und Methoden

Moreover, while the report benefitted from inputs via the INAHTA network and the ASSESS DHT survey, it is possible that unpublished guidance documents or emerging methodologies not yet formalised into official guidelines were missed. This is particularly pertinent given the relatively early stage of AI technology integration into healthcare, where many frameworks are still in development. Furthermore, the exclusion of assessment frameworks and guidance documents developed by non-HTA institutes, although available through the ASSESS DHT project [42], limits the scope of this report to a specific subset of existing resources.

manche KI-Systeme mit mehr als einer Funktion, Fokus auf „primäre Funktion"

The assessments provide an overview on HTA methods and application areas of AI. Yet, it is worth noting that in some cases, a clear allocation to a single functional category was challenging due to AI applications having multiple functions and a lack of precise definition of the AI function in the assessments.

keine Qualitätsbewertung der eingeschlossenen HTA-Berichte

Furthermore, no formal risk of bias assessment was conducted on the identified documents, as the aim was not to synthesize evidence for specific interventions. While this approach aligns with the report's objective, it may limit the interpretability of the methodologies in the context of evidence quality.

5 Conclusions and recommendations

HTA-Methoden benötigen (begrenzte) Anpassungen

One can see HTA methods need (a limited) adaption when applying them to artificial intelligence (AI) enabled digital health technologies (DHT). However, few HTA institutes have developed or applied AI-specific methodological guidance to evaluate AI-enabled DHTs yet.

es besteht Bedarf an der Entwicklung von konkreten Leitlinien für KI-Anwendungen

The EUnetHTA Core Model and available supplementary guidance documents such as digital health technology (DHT) assessment frameworks and AI checklists provide a toolkit for evaluating the benefits of AI-enabled DHTs for procurement decisions. Yet, there is a need for thorough scientific discussion and development of more concrete guidance for assessing AI-enabled DHTs. This currently takes place, in numerous ongoing European research projects on assessment frameworks that will provide more concretised guidance also for AI.

Empfehlungen:

The following recommendations are derived for Austrian decision makers:

neben dem EUnetHTA Core Model, sollten Frameworks für DHTs und KI-spezifische Checklists verwendet werden

- Alongside with the EUnetHTA Core Model, we recommend using an existing framework for DHTs such as NICE's evidence standard framework or the adapted version from AQuAS as a starting point for an evaluation of AI-enabled DHTs and adding components of other guidance documents such as the AI-specific checklists developed from HTW or Digi-HTA.

kontinuierliche Überwachung ist erforderlich

- AI algorithms may change over time and require continuous monitoring and interference by clinical experts (users) in case they potentially affect the quality of delivered care. Regular auditing of the AI-application is recommended.

Endpunkte und Studiendesigns sind abhängig von der Hauptfunktion

- For the evaluation of several domains, standard HTA methodology guided by the EUnetHTA Core Model can be applied and supplemented with AI-specific components. In doing so, study designs and outcomes should be defined depending on the primary function of the AI-enabled DHT.
- As AI-enabled DHTs are sensitive to input data and the underlying data infrastructure, efforts should be undertaken to ensure AI-enabled DHTs work as anticipated and continuously perform as demonstrated in (clinical) studies.

6 References

[1] Florida State University Libraries. Algorithm. 2021 [cited 21.10.2024]. Available from: https://guides.lib.fsu.edu/algorithm.

[2] Makady A. d. B., A.; Hillege, H.; Klungel, O.; Goettsch, W. What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. Value in Health. 2017;20(7):858-865. DOI: https://doi.org/10.1016/j.jval.2017.03.008.

[3] National Institute for Health and Clinical Excellence. Evidence standards framework for digital health technologies. 2022 [cited 04.04.2024]. Available from: https://www.nice.org.uk/corporate/ecd7/resources/evidence-standards-framework-for-digital-health-technologies-pdf-1124017457605.

[4] Haute Autorité de Santé. Dossier submission to the Medical Device and Health Technology Evaluation Committee. 2020 [cited 06.05.2024]. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2020-10/guide_dm_vf_english_publi.pdf.

[5] Wright G. Noisy Data. n.D. [cited 21.20.2024]. Available from: https://www.techtarget.com/searchbusinessanalytics/definition/noisy-data.

[6] Canadian Agency for Drugs and Technologies in Health. Real-World Evidence: A Primer. 2024 [cited 21.10.2024]. Available from: https://www.cda-amc.ca/real-world-evidence-primer#:~:text=Real%2Dworld%20evidence%20(RWE),real%2Dworld%20health%20care%20setting.

[7] Dubrov V. Understanding Machine Learning Robustness: Why It Matters and How It Affects Your Models. 2023 [cited 12.11.2024]. Available from: https://medium.com/@slavadubrov/understanding-machine-learning-robustness-why-it-matters-and-how-it-affects-your-models-5e2cb5838dab.

[8] Riemann R. Synthetic Data. 2024 [cited 21.10.2024]. Available from: https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en.

[9] Boverhof B. J. R., W.K.; Visser, J.J.; Uyl-de Groot, C.A.; Rutten-van Mölken, M.P. Broadening the HTA of medical AI: A review of the literature to inform a tailored approach. Health Policy and Technology. 2024;13.

[10] Farah L. D.-S., J.; Martin, T.; Nguyen, P.; Borget, I.; Martelli, N. Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A systematic review. Artificial Intelligence in Medicine. 2023;140:13. DOI: https://doi.org/10.1016/j.artmed.2023.102547.

[11] McCarthy J. M., M.; Rochester, N.; Shannon, C. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. AI Magazine. 2006;27(4):3. DOI: https://doi.org/10.1609/aimag.v27i4.1904.

[12] Fraser A. G., Biasin E., Bijnens B., Bruining N., Caiani E. G., Cobbaert K., et al. Artificial intelligence in medical device software and high-risk medical devices – a review of definitions, expert recommendations and regulatory initiatives. Expert Rev Med Devices. 2023;20(6):467-491. Epub 20230508. DOI: 10.1080/17434440.2023.2184685.

[13] Academy of Medical Royal Colleges. Artificial Intelligence in Healthcare. 2019 [cited 08.05.2024]. Available from: https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf.

[14] Lekadir K. Q., G.; Garmendia, A.; Gallin, C. Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. Brussels: 2022 [cited 08.05.2024]. Available from: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf.

[15] IBM. AI vs. machine learning vs. deep learning vs. neural networks: What's the difference? 2023 [cited 12.11.2024]. Available from: https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks.

[16] Khurana D., Koli A., Khatter K. and Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2023;82(3):3713-3744. Epub 20220714. DOI: 10.1007/s11042-022-13428-4.

[17] IBM. What are large language models (LLMs)? n.D. [cited 06.05.2024]. Available from: https://www.ibm.com/topics/large-language-models.

[18] Tian Y. Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. IEEE Access. 2020;8:14. DOI: 10.1109/ACCESS.2020.3006097.

[19] IBM. What is a recurrent neural network (RNN)? n.D. [cited 06.05.2024]. Available from: https://www.ibm.com/topics/recurrent-neural-networks.

[20] A Guide To Artificial Intelligence In Healthcare: The Medical Futurist; 2019.

[21] Ng G. W. L., W.C. Strong Artificial Intelligence and Consciousness. J Artif Intell Conscious. 2020;7(1):63-72. DOI: 10.1142/S2705078520300042.

[22] Sheikh H. P., C.; Schrijvers, E. Artificial intelligence: Definition and Background. Mission AI. 2023:15-41. DOI: 10.1007/978-3-031-21448-6_2.

[23] Canorea E. Discriminative AI vs Generative AI: Keys to understanding them. 2024 [cited 19.06.2024]. Available from: https://www.plainconcepts.com/discriminative-ai-vs-generative-ai/#:~:text=Traditional%20or%20discriminative%20AI%20is,images%2C%20music%2C%20videos.

[24] Haverinen J. K., N.; Falkenbach, P.; Maijala, A.; Kolehmainen, T.; Reponen, J. Digi-HTA: Health technology assessment framework for digital healthcare services. FinJeHeW. 2019;11(4):326-341.

[25] Reponen J., Winblad I. and Hamalainen P. Current status of national eHealth and telemedicine development in Finland. Stud Health Technol Inform. 2008;134:199-208.

[26] Bures D., Hosters B., Reibel T., Jovy-Klein F., Schramm J., Brendt-Müller J., et al. Die transformative Wirkung von künstlicher Intelligenz im Krankenhaus. Die Innere Medizin. 2023(64):1025-1032. DOI: https://doi.org/10.1007/s00108-023-01597-9.

[27] Gilvary C., Madhukar N., Elkhader J. and Elemento O. The Missing Pieces of Artificial Intelligence in Medicine. Trends Pharmacol Sci. 2019;40(8):555-564. Epub 20190702. DOI: 10.1016/j.tips.2019.06.001.

[28] American Medical Association. CPT Appendix S: AI taxonomy for medical services & procedures. 2024 [cited 14.10.2024]. Available from: https://www.ama-assn.org/system/files/cpt-appendix-s.pdf.

[29] Prasser F., Riedel N., Wolter S., Corr D. and Ludwig M. Künstliche Intelligenz und sichere Gesundheitsdatennutzung im Projekt KI-FDZ: Anonymisierung, Synthetisierung und sichere Verarbeitung von Real-World-Daten. Bundesgesundheitsbl. 2024(67):171-179.

[30] European Commission. Regulation of European Parliament and of the Council. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021 [cited 05.04.2024]. Available from: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.

[31] European Health Data Space. The European Health Data Space (EHDS). n.D. [cited 13.11.2024]. Available from: https://www.european-health-data-space.com/#:~:text=The%20AI%20compares%20the%20patient's,effectiveness%20before%20seeking%20market%20approval.

[32] Bretthauer M. G., S:; Hassan, C.; Ahmad, O.; Mori, Y. . The New European Medical Device Regulation: Balancing Innovation and Patient Safety. Annals of Internal Medicine. 2023;176(6):844-848. DOI: 10.7326/m23-0454 %m 37068279.

[33] European Commission. Regulations on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. 2017 [cited 30.07.2024]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745.

[34] European Commission. Richtlinie 93/42/EWG des Rates. 1993 [cited 15.10.2024]. Available from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:DE:HTML.

[35] Medical Device Coordination Group. Guidance on classification of medical devices. 2021 [cited 30.07.2024]. Available from: https://health.ec.europa.eu/system/files/2021-10/mdcg_2021-24_en_0.pdf.

[36] Schmidt J., Schutte N. M., Buttigieg S., Novillo-Ortiz D., Sutherland E., Anderson M., et al. Mapping the regulatory landscape for artificial intelligence in health within the European Union. NPJ Digit Med. 2024;7(1):229. Epub 20240827. DOI: 10.1038/s41746-024-01221-6.

[37] Degelsegger-Márquez A., Dick D. and Trunner K. Telemedizin und Künstliche Intelligenz im intramuralen Bereich Österreichs. Ereignisbericht.: 2022 [cited 3.4.2024]. Available from: https://jasmin.goeg.at/id/eprint/2443/1/Telemedizin_und_KI_in_Krankenanstalten_bf.pdf.

[38] EUnetHTA Joint Action 2. Work Package 8. HTA Core Model® version 3.0. 2016 [cited 03.06.2024]. Available from: https://www.eunethta.eu/wp-content/uploads/2018/03/HTACoreModel3.0-1.pdf.

[39] EUnetHTA Joint Action 2. Therapeutic medical devices. 2015 [cited 21.10.2024]. Available from: https://www.eunethta.eu/wp-content/uploads/2018/01/Therapeutic-medical-devices_Guideline_Final-Nov-2015.pdf.

[40] Open Science Foundation. OSF. n.D. [cited 21.10.2024]. Available from: https://osf.io/.

[41] European Commission. Annexes to the Proposal for a Regulation of the European Parliament and of the Council. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.: 2021 [cited 05.04.2024]. Available from: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF.

[42] ASSESS DHT. Development & harmonisation of methodologies for assessing digital health technologies in Europe. 2024 [cited 21.10.2024]. Available from: https://assess-dht.eu/.

[43] Albahri A. S., Alwan, J.K., Taha, Z.K., Ismail, S.F., Hamid, R.A., Zaidan, A.A., Albahri, O.S., Zaidan, B.B., Alamoodi, A.H., Alsalem, M.A. IoT-based telemedicine for disease prevention and health promotion: State-of-the-Art. Journal of Network & Computer Applications. 2020;173.

[44] Jen M. Y. S., M.; Varacallo, M. Predictive Medicine. Treasure Island: 2022 [cited 06.06.2024]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441941/.

[45] Hansebout R. R., Cornacchi S. D., Haines T. and Goldsmith C. H. How to use an article about prognosis. Can J Surg. 2009;52(4):328-336.

[46] van Smeden M., Reitsma J. B., Riley R. D., Collins G. S. and Moons K. G. Clinical prediction models: diagnosis versus prognosis. J Clin Epidemiol. 2021;132:142-145. DOI: 10.1016/j.jclinepi.2021.01.009.

[47] National Cancer Institute. Prevention. n.D. [cited 21.20.2024]. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/prevention.

[48] Chishti A. W., I.A. Patient monitoring techniques. Surgery (Oxford). 2019;37(8):450-459. DOI: https://doi.org/10.1016/j.mpsur.2019.05.002.

[49] Haute Autorité de Santé. Medical device evaluation by the CNEDiMTS (Medical Device and Health Technology Evaluation Committee). Guide to the specific featrues of clinical evaluation of a connected medical device (CMD) in view of its application fo reimbursement. 2019 [cited 06.05.2024]. Available from: https://www.has-sante.fr/upload/docs/application/pdf/2021-08/guide_to_the_specific_features_of_clinical_evaluation_of_cmd_in_view_of_its_application_for_reimbursement.pdf.

[50] Segur-Ferrer J., Moltó-Puigmarti C., Pastells-Peiró R. and Vivanco-Hidalgo R. Health Technology Assessment Framework: Adaption for Digital Health Technology Assessment. Madrid: Agència de Qualitat i Avaluació Sanitàries de Catalunya, 2023.

[51] Batten L. N.-T., A.; England, C.; Jarrom, D. Heath Technology Assessments of Artificial Intelligence: Special Considerations And Development Of A Checklist. [Checklist]. In press 2024.

[52] U.S. Food and Drug Administration, Health Canada and Medicines & Healthcare products Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles. 2021 [cited 05.04.2024]. Available from: https://www.fda.gov/media/153486/download.

[53] Haute Autorité de Santé. Haute Autorité de Santé. n.D. [cited 21.10.2024]. Available from: https://www.has-sante.fr/jcms/pprd_2986129/en/home.

[54] Haverinen J., Turpeinen M., Falkenbach P. and Reponen J. Implementation of a new Digi-HTA process for digital health technologies in Finland. Int J Technol Assess Health Care. 2022;38(1):e68. Epub 20220819. DOI: 10.1017/S0266462322000502.

[55] Jeindl R. G., G. Prozess und Bewertung digitaler Gesundheitsanwendungen – am Beispiel der „Symptom-Checker". Wien: HTA Austria – Austrian Institute for Health Technology Assessment GmbH, 2021 [cited 12.06.2024]. Available from: https://eprints.aihta.at/1348/1/HTA-Projektbericht_Nr.141.pdf.

[56] Behandlungsradet. Use of artificial intelligence as clinical decision-support in colonoscopy for the diagnosis of neoplastic disease. 2023 [cited 06.06.2024]. Available from: https://behandlingsraadet-classic.azureedge.net/media/bbjjjro3/use-of-artificial-intelligence-as-clinical-decision-support-in-colonoscopy.pdf.

[57] National Evidence-based Healthcare Collaborating Agency. Artificial intelligence-based diagnostic assistant test for acute myocardial infarction using 12-lead electrocardiogram data. 2024 [cited 06.06.2024]. Available from: https://nhta.neca.re.kr/nhta/publication/nhtaU0601V.ecg?pub_seq=1033.

[58] National Evidence-based Healthcare Collaborating Agency. Artificial intelligence-based screening test for emergent large vessel occlusion using non-contrast brain CT images. 2024 [cited 06.06.2024]. Available from: https://nhta.neca.re.kr/nhta/publication/nhtaU0601V.ecg?pub_seq=1047.

[59] National Evidence-based Healthcare Collaborating Agency. Artificial intelligence-based diagnostic assistant test for paroxysmal atrial fibrillation using 12-lead electrocardiogram data. 2024 [cited 06.06.2024]. Available from: https://nhta.neca.re.kr/nhta/publication/nhtaU0601V.ecg?pub_seq=1046.

[60] Gallastegui E. P.-V., C.; Estrada, M.; Priego, B.; Vivanco-Hidalog, R. Inteligencia artificial para la detección y caracterización de lesiones precancerosas colorrectales en la colonoscopia. Barcelona: Agència de Qualitat i Avaluació Sanitàries de Catalunya, 2023.

[61] Del Cura Bilbao A. L. M., H.; Marín Sánchez, J.; Blas Diez, M. Dermatoscopia Digital. Instituto Aragonés de Ciencias de la Salud, 2022.

[62] Canadian Agency for Drugs and Technologies in Health. Implications of ChatGPT on Radiology Workflow. Canadian Journal of Heatlh Technologies. 2023;3(11):16. DOI: https://doi.org/10.51731/cjht.2023.789.

[63] N'Diaye Mombo N. A., S.; Brabant, J.; Alami, H. Artificial intelligence-assisted diabetic retinopathy tele-screening. Québec: Institut national d'excellence en santé et en services sociaux, 2021 [cited07.06.2024]. Available from: https://www.inesss.qc.ca/fileadmin/doc/INESSS/Rapports/Technologies/INESSS_IA_TDRD_EC.pdf.

[64] Clark M. B., S. Chatbots in Health Care: Connecting Patients to Information. Canadian Journal of Heatlh Technologies. 2024;4(1):22.

[65] Health Technology Wales. Artificial intelligence assisted tools for diagnosis of prostate cancer from whole slide digital biopsy images. Health Technology Wales, 2023 [cited 07.06.2024]. Available from: https://healthtechnology.wales/wp-content/uploads/2023/09/TER480-Galen-Prostate-Solution-WEB.pdf.

[66] Evans A. E., C.; Jarrom, D.; Hasler, E.; Boyce, R.; Shepherd, R.; Hughes, S. Integrated Digital Wound Care Management Systems to assess and manage people receiving wound care. Health Technology Wales, 2023 [cited 06.06.2024]. Available from: https://healthtechnology.wales/wp-content/uploads/2023/03/EAR051-DWMS-WEB.pdf.

[67] Health Technology Wales. Artificial intelligence-assisted ECG intepretation for people with suspected cariodvascular disease in primary/emergency care. Health Technology Wales, 2024 [cited 07.06.2024]. Available from: https://healthtechnology.wales/wp-content/uploads/TER491-PMcardio.pdf.

[68] Health Technology Wales. Artificial intelligence for the diagnosis of chronic obstructive pulmonary disease at point-of-care. Health Technology Wales, 2024 [cited 07.06.2024]. Available from: https://healthtechnology.wales/wp-content/uploads/TER511-POC-AI-to-diagnose-COPD.pdf.

[69] Health Improvement Scotland. Artificial intelligence supported clinician review of chest x-rays from patients with suspected lung cancer. Health Improvmenet Scotland, 2024 [cited 06.06.2024]. Available from: https://shtg.scot/our-advice/artificial-intelligence-supported-clinician-review-of-chest-x-rays-from-patients-with-suspected-lung-cancer/.

[70] National Institute for Health and Care Excellence. Artificial intelligence for analysing CT brain scans. 2020 [cited 09.06.2024]. Available from: https://www.nice.org.uk/advice/mib207/resources/artificial-intelligence-for-analysing-ct-brain-scans-pdf-2285965396121029.

[71] National Institute for Health and Care Excellence. Artificial intelligence for analysing chest CT images. 2021 [cited 09.06.2024]. Available from: https://www.nice.org.uk/advice/mib243/resources/artificial-intelligence-for-analysing-chest-ct-images-pdf-2285965631267269.

[72] National Institute for Health and Care Excellence. Artificial intelligence in mammography. 2021 [cited 09.06.2024]. Available from: https://www.nice.org.uk/advice/mib242/resources/artificial-intelligence-in-mammography-pdf-2285965629587653.

[73] National Institute for Health and Care Excellence. Artificial intelligence – derived software to analyse chest X-rays for suspected lung cancer in primary care referrals: early value assessment. 2023 [cited 09.06.2024]. Available from: https://www.nice.org.uk/guidance/hte12/resources/artificial-intelligencederived-software-to-analyse-chest-xrays-for-suspected-lung-cancer-in-primary-care-referrals-early-value-assessment-pdf-50261967918277.

[74] National Institute for Health and Care Excellence. AI-derived computer-aided detection (CAD) software for detecting and measuring lung nodules in CT scan images. 2023 [cited 09.06.2024]. Available from: https://www.nice.org.uk/guidance/dg55/resources/aiderived-computeraided-detection-cad-software-for-detecting-and-measuring-lung-nodules-in-ct-scan-images-pdf-1053873334213.

[75] National Institute for Health and Care Excellence. Artificial intelligence (AI)-derived software to help clinical decision making in stroke. 2024 [cited 09.06.2024]. Available from: https://www.nice.org.uk/guidance/dg57/resources/artificial-intelligence-aiderived-software-to-help-clinical-decision-making-in-stroke-pdf-1053876693445.

[76] Westwood M. R., B.; Grimm, S.; Armstrong, N.; Wijnen, B.; Ahmadu, C.; et al. Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2024;28(11):234. DOI: 10.3310/RDPA1487.

[77] Madakadze C. M., S. Artificial Intelligence – Enhanced Rapid Response Electroencephalography for the Identification of Nonconvulsive Seizure. Canadian Journal of Heatlh Technologies. 2023;3(12):15.

[78] National Institute for Health and Care Excellence. Paige Prostate for prostate cancer. 2021 [cited 09.06.2024]. Available from: https://www.nice.org.uk/advice/mib280/resources/paige-prostate-for-prostate-cancer-pdf-2285965868093125.

[79] National Institute for Health and Care Excellence. AI technologies for detecting diabetic retinopathy. 2021 [cited 09.06.2024]. Available from: https://www.nice.org.uk/advice/mib265/resources/ai-technologies-for-detecting-diabetic-retinopathy-pdf-2285965755558853.

[80] National Institute for Health and Care Excellence. Digital technologies for the detection of melanoma. 2022 [cited 09.06.2024]. Available from: https://www.nice.org.uk/advice/mib311/resources/digital-technologies-for-the-detection-of-melanoma-pdf-2285967623291845#:~:text=DERM%20(Skin%20Analytics),except%20where%20specific%20exclusions%20apply.

[81] National Institute for Health and Care Excellence. CaRi-Heart for predicting cardiac risk in suspected coronary artery disease: early value assessment. 2023 [cited 09.06.2024]. Available from: https://www.nice.org.uk/guidance/hte4/resources/cariheart-for-predicting-cardiac-risk-in-suspected-coronary-artery-disease-early-value-assessment-pdf-1396170676933.

[82] National Institute for Health and Clinical Excellence. Diagnostics Assessment Programme manual. 2011 [cited 12.08.2024]. Available from: https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-diagnostics-guidance/Diagnostics-assessment-programme-manual.pdf.

[83] National Institute for Health and Clinical Excellence. Early value assessment interim statement. 2022 [cited 12.08.2024]. Available from: https://www.nice.org.uk/process/pmg39/resources/early-value-assessment-interim-statement-pdf-72286784283589.

[84] Garritty C. G., G.; Nussbaumer-Streit, B.; King, V.; Hamel, C.; Kamel, C.; Affengruber, L.; Stevens, A. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol. 2021;130:13-22. DOI: 10.1016/j.jclinepi.2020.10.007.

[85] National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. 2014 [cited 15.10.2024]. Available from: https://www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869.

[86] National Institute for Health and Clinical Excellence. Artificial intelligence technologies to aid contouring for radiotherapy treatment planning: early value assessment. 2023 [cited 09.06.2024]. Available from: https://www.nice.org.uk/guidance/hte11/resources/artificial-intelligence-technologies-to-aid-contouring-for-radiotherapy-treatment-planning-early-value-assessment-pdf-50261966238661.

[87] Canadian Agency for Drugs and Technologies in Health. Artificial Intelligence for Patient Flow. Canadian Journal of Heatlh Technologies. 2024;4(4):20.

[88] Xie W. B., R. Artificial Intelligence Decision Support Tools for End-of-Life Care Planning Conversations. Canadian Journal of Heatlh Technologies. 2023;3(12):16.

[89] Joos C., Albrink K., Hummers E., Muller F., Antweiler K., Schroder D., et al. Concordance of data collected by an app for medical history taking and in-person interviews from patients in primary care. JAMIA Open. 2024;7(4):ooae102. Epub 20241007. DOI: 10.1093/jamiaopen/ooae102.

[90] Software.com. Test failure rate. n.D. [cited 21.10.2024]. Available from: https://www.software.com/engineering-metrics/test-failure-rate.

[91] Sartor G. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. Brussels: 2020 [cited 29.10.2024]. Available from: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf.

[92] Pfaff H. and Schmitt J. Shifting from Theoretical Best Evidence to Practical Best Evidence: an Approach to Overcome Structural Conservatism of Evidence-Based Medicine and Health Policy. Gesundheitswesen. 2024;86(S 04):S239-S250. Epub 20240815. Von der theoretisch besten Evidenz zur praktisch besten Evidenz: ein Ansatz zur Uberwindung des Strukturkonservatismus in der evidenzbasierten Medizin und Gesundheitspolitik. DOI: 10.1055/a-2350-6435.

[93] Trowman R. M., A.: Ollendorf, DA. Health technology assessment 2025 and beyond: lifecycle approaches to promote engagement and efficiency in health technology assessment. International Journal of Technology Assessment in Health Care. 2023;39(1). DOI: 10.1017/S0266462323000090.

[94] Nielsen K. M., M. What works for whom in which circumstances? On the need to move beyond the 'what works?' question in organizational intervention research. Human Relations. 2017;70(1):40-62. DOI: https://doi.org/10.1177/0018726716670226.

[95] Lehne M., Sass J., Essenwanger A., Schepers J. and Thun S. Why digital medicine depends on interoperability. NPJ Digit Med. 2019;2:79. Epub 20190820. DOI: 10.1038/s41746-019-0158-1.

[96] Bundesministerium für Verkehr I. u. T. and Bundesministerium für Digitalisierung und Wirtschaftsstandort. Aritifical Intelligence Mission Austria 2023 – Die Zukunft der Künstlichen Intelligenz in Österreich gestalten. 2018 [cited 21.10.2024]. Available from: https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/aimat.html.

[97] Di Bidino R., Daugbjerg, S.B., Papavero, S.C., Hebold Haraldsen, I.R.J., Cicchretti, A. Health Technology Assessment Framework for Artificial Intelligence-Based Technologies: Results of a Delphi Expert Survey. AI-Mind Project. [Framework]. In press 2024.

[98] Daugbjerg S D. B. R., Cicchetti A. OP103 What To Include In A Health Technology Assessment Of Artificial Intelligence-Based Technologies: Results Of A Delphi Expert Survey. International Journal of Technology Assessment in Health Care. 2023;29(S1):29-29. DOI: doi:10.1017/S0266462323001162.

[99] Norgeot B. Q., G.; Beaulieu-Jones, B.; Torkamani, A.; Dias, R.; Gianfrancesco, M.; Arnaout, R.; Kohane, I.; Saria, S.; Topol, E.; Obermeyer, Z.; Yu, B.; Butte, A. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med. 2020;26(9):1320-1324. DOI: 10.1038/s41591-020-1041-y.

[100] European Digital Health Technology Assessment. EDiHTA. 2024 [cited 21.10.2024]. Available from: https://edihta-project.eu/?playlist=03c706c&video=4c173c7.

[101] Collins G. S., Dhiman P., Andaur Navarro C. L., Ma J., Hooft L., Reitsma J. B., et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021;11(7):e048008. Epub 20210709. DOI: 10.1136/bmjopen-2020-048008.

[102] Gilbert S., Fenech M., Hirsch M., Upadhyay S., Biasiucci A. and Starlinger J. Algorithm Change Protocols in the Regulation of Adaptive Machine Learning-Based Medical Devices. J Med Internet Res. 2021;23(10):e30545. Epub 20211026. DOI: 10.2196/30545.

Appendix

Overview of HTA Institutions

Table A-1: HTA Institutions

Institution	Country	Institution	Country
Agency for Care Effectiveness	Singapore	Health Technology Wales	United Kingdom
Agencia de Evaluación de Tecnolo	Spain	Health Sciences Institute in Aragon	Spain
Andalusian Agency for Health Technology Assessment	Spain	Institute of Health Economics	Canada
The Agency for Regional Healthcare	Italy	National Authority for Assessment and Accreditation in Healthcare	Tunisia
Agency for Healthcare Research and Quality	USA	Institut national d'excellence en santé et en services sociaux	Canada
Adelaide Health Technology Assessment	Australia	National Authority for Assessment and Accreditation in Healthcare	Tunisia
Austrian Institute for Health Technology Assessment	Austria	Institut national d'excellence en santé et en services sociaux	Canada
Agência Nacional de Saúde Suplementar	Brazil	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen	Germany
Agency for Health Technology Assessment and Tariff System	Poland	Belgian Health Care Knowledge Centre	Belgium
Assistance publique- Hopitaux de Paris	France	Health Technology Assessment Section	Malaysia
Agency of Health Quality and Assessment of Catalonia	Spain	National Evidence-based healthcare Collaborating Agency	Korea
Australian Safety and Efficacy Register of New Interventional Procedures-Surgical	Australia	National Institute for Health and Care Excellence	United Kingdom
Galician Agency for Health Technology Assessment	Spain	National Institute for Value and Technologies in Healthcare	Slovak Republic
Center For Outcomes Research And Economic Evaluation For Health	Japan	National Institute for Health Research	United Kingdom
Canada's Drug Agency	Canada	Norwegian Institute of Public Health	Norway
Central Administration of Health Technology Assessment	Egypt	Ontario Health	Canada
Center for Drug Evaluation	China	Basque Office for Health Technology Assessment	Spain
National Committee for Technology Incorporation	Brazil	Pharmac	New Zealand
The Danish Health Technology Council	Denmark	Regione Emilia-Romagna	Italy
Finnish Coordinating Center for Health Technology Assessment	Finland	Salidat Kairbekova National Research Center for Health Development	Kazakhstan
The Federal Joint Committee	Germany	Swedish Agency for Health Technology Assessment and Assessment of Social Services	Sweden
Gesundheit Österreich GmbH	Austria	Department of HTA at the State Expert Centre of the Ministry of Health	Ukraine
Uruguay – Health Assessment Division	Uruguay	Swiss Federal Office of Public Health	Switzerland
Haute Autorité de Santé	France	HTA Unit in A.Gemelli Teaching Hospital	Italy
Health Information and Quality Authority	Ireland	Zorginstituut Nederland	The Netherlands
Healthcare Improvement Scotland	Scotland	The Netherlands Organisation for Health Research and Development	The Netherlands

Overview of methods documents and current use

Table A-2: Overview of institutions with published AI specific methods or utilising DHT frameworks with guidance on AI-enabled DHTs

Institution	Developed Framework/ Methodology with AI component	Framework with AI component in use	Identified Frameworks with AI component					
			AQuAS	DIGI HTA	ESF	HAS	HIS	HTW Checklist
ACE	☐	☐						
AETSA	☐	☐						
AGENAS	☐	☐						
AHRQ	☐	☐						
AHTA	☐	☐						
AIHTA	☐	☐						
ANS	☐	☐						
AOTM	☐	☐						
APHP	☐	☐						
AQUAS	☑	☐	✓					
ASERNIP	☐	☐						
AVALIA T	☐	☐						
CADTH	☐	☐						
C2H	☐	☐						
CA-HTA	☐	☐						
CDE	☐	☐						
CONITEC	☐	☐						
DHTC	☐	☐						
DIGEMID	☐	☐						
EUnetHTA	☐	☐						
FinCCHTA	☐	☑		✓				
GBA	☐	☐						
GOEG	☐	☐						
HAD	☐	☐						
HAS	☐	☑				✓		

Institution	Developed Framework/ Methodology with AI component	Framework with AI component in use	Identified Frameworks with AI component					
			AQuAS	DIGI HTA	ESF	HAS	HIS	HTW Checklist
HIQA	☐	☐						
HIS	☐	☑			✓	✓		
HTW	☑	☑			✓			✓
IACS	☐	☐						
IECS	☐	☐						
IETS	☐	☐						
IHE	☐	☐						
INEAS	☐	☐						
INESSS	☐	☐						
IQWiG	☐	☐						
KCE	☐	☐						
MaTHAS	☐	☐						
NECA	☐	☐						
NICE	☐	☑			✓			
NIHO	☐	☐						
NIHR	☐	☑						
NIPH	☐	☐						
OH	☐	☐						
OSTEBA	☐	☐						
PHARMAC	☐	☐						
RER	☐	☐						
SK NRCHD	☐	☐						
SBU	☐	☐						
SEC	☐	☐						
SFOPH	☐	☐						
UVT	☐	☐						
ZIN	☐	☐						
ZonMw	☐	☐						

Note: ☑ *AI specific methods/guidelines ,* ☐ *AI or DHT mentioned;* ☐ *AI not mentioned*

Extraction Tables Assessments

Table A-3: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 1

Author/Institution, year	CADTH, 2023 Health Technology Review [62]	HIS/SHTG, 2024 MIB [69]	NECA, 24 Assessment of potential [58]	NICE, 2024 Full Guidance [75]	NICE, 2020 MIB [70]
Study Characteristics					
Country	Canada	Scotland, UK	South Korea	UK	UK
Functional Category[11]	Diagnosis	Diagnostics	Diagnosis	Diagnosis, clinical decision making	Diagnosis
(medical) specialty	Radiology	Radiology	Radiology	Radiology	Radiology
Population(s)	Radiologists/Radiology departments	Patients with suspected lung cancer	Patients presenting with symptoms of suspected stroke	Patients with acute suspected stroke	Patients with suspected brain abnormalities
Intervention	Chat GPT supported radiology reporting	AI supported chest x-ray review	AI-based emergency cerebral vessel occlusion screening test using non-contrast CT	AI- derived software assisted CT scan review in patients with acute suspected stroke	AI-supported brain CT analysis
Comparator	Standard Workflow	Standard of care/Chest Xray reviewed by radiologist and diagnosis of cancer by MDT	Standard of care/general CT diagnostics/review by a neurosurgeon	Standard of care/CT scan review by an HCP without AI assistance. Non enhanced CT scans can be reviewed by a variety of trained HCP, whilst CTA and CTP scans need to be reviewed by a specialist.	Standard of care/ CT scan reviewed by radiologists
Eligibility criteria	NR Predefined review outcomes: not specified The main objective of this report are to summarize potential applications of the technology for improving radiology workflow efficiency and its strengths and limitations in this use.	Studies on the use of AI to analyse CXR with UK context Predefined review outcomes: not specified	Predefined review outcomes not specified	All comparative study designs: study designs will be included in a hierarchical manner (RCTs, CCTs, observational studies), i.e. CCTs and observational studies will only be considered for inclusion where no RCTs are identified, or where there are concerns about the applicability (e.g. non-UK settings) or risk of bias for identified RCTs Predefined review outcomes: Clinical effectiveness Cost effectiveness	Most relevant or best available evidence relating to the clinical effectiveness of the technology *Diagnostic accuracy:* sensitivity, specificity *Clinical effectiveness*: time to treatment *Organisational:* Resource consequences, length of stay *Ethical:* Equality considerations *Cost & Economic:* technology purchase and implementation cost
Number of included studies	Not recorded in detail	11	1	15	11

[11] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

Author/Institution, year	CADTH, 2023 Health Technology Review [62]	HIS/SHTG, 2024 MIB [69]	NECA, 24 Assessment of potential [58]	NICE, 2024 Full Guidance [75]	NICE, 2020 MIB [70]
AI Characteristics					
AI Product name | AI Type (CNN, LLM, unspecified, others)	Chat GPT	Annalise Enterprise CXR qXR	ELVO, HEURON StroCare Suite, HEURON	Accipio (MaxQ AI) Aidoc (Aidoc) BioMind (BioMind.ai) BrainScan CT (brainscan.ai) Cercare Perfusion (Cercare Medical) CINA Head (Avicenna) CT Perfusion 4D (GE Healthcare) e-Stroke (Brainomix) icobrain ct (icometrix) Neuro Solutaion (Nanox.AI) qER (Qure.ai) RapidAI (Ischemaview) Viz (Viz.ai)	head (Aidoc) e-CTA/e-ASPECTS (Brainomix) icobrain (Icometrix) qER (Qure) Zebra triage (Zebra Medical Vision) DLCExpert (Mirada Medical)
Training data	NR	Annalise: >250.000 CXR images qXR: over 4.4 million CXR images	NR	NR	DLCExpert: The company claims the algorithms have been trained on clinical examples and validated against consensus guidelines.
Information on AI algorithm/model specifications	No	No	No	Data from scans in clinical practice is not used to further develop algorithms in the software. They are developed using CT scans held by the company or accessed through research studies. In clinical practice fixed algorithms are used.	No
AI main functions	Diverse roles in radiology workflow support: ■ Expediting radiology reports ■ Clinical decision support ■ Supporting writing intensive tasks	Using a machine/deep learning model, the software analyses CXRs to identify high-risk images and labels them for urgent clinician review. Annalise.ai scans for 124 potential issues with 34 considered priority findings and acts as a triage system alerting clinicians to urgent review. qXR analyses the CXR image with a processing time of 20 seconds and also alerts clinicians to urgent review. (deep learning AI technology)	AI-based technology to provide diagnostic support and screening for large vessel occlusive stroke. Non-contrast CT scans are transmitted to an AI-based software to classify an emergency cerebral large vessel occlusion, to clarify in which hemisphere it occurred and to notify HCPs of the classification results.	AI Analysis of Images enhances stroke diagnosis by identifying, quantifying, and in some cases notifying clinicians of brain structures related to acute stroke. The software included in this assessment cannot update itself.	AI software for analysis of CT brain scans (including NCCT, CTA and CTP) to detect abnormalities (stroke, trauma, dementia) The systems report preliminary findings, alert radiologists to critical cases and prioritise cases that need urgent review Most software integrates into the current imaging systems and the results can be viewed as visual results on DICOM output images. email notifications or a web browser

Author/Institution, year	CADTH, 2023 Health Technology Review [62]	HIS/SHTG, 2024 MIB [69]	NECA, 24 Assessment of potential [58]	NICE, 2024 Full Guidance [75]	NICE, 2020 MIB [70]
Expectations for the technology	■ Relieving pressure on radiology resources and staff ■ Reduce waiting times ■ Support final diagnosis ■ Cancer screening ■ Administration functions	■ Identification of high risk CXR early during clinical pathway leads to prioritization of patients in need for urgent further assessment and possibly treatment.	■ Improved diagnosis of cerebral vessel occlusion on non-contrast CT ■ Notifying clinicians of priority patients ■ Faster access to treatment ■ Reduce volume burden	■ Enhanced triage, prioritisation, transfer and treatment ■ Support of evaluation for time-sensitive treatments ■ Rapid report turnaround and multisite scan reviews	■ Faster review of images and treatment of patients if required ■ Improved patient outcomes ■ Prevention of missing subtle changes ■ Reduction of clinician workload
Study Methodology					
Types of Included studies	NR	*Published evidence:* Retrospective Observational (n-1) SR (n=2), cohort study with prospective and retrospective phase (n=2)	Manufacturer clinical trial results report (n=1) Conference Abstracts (study design unclear) (n=2)	Retrospective observational studies using data to validate the accuracy of the algorithm (n=15)	Retrospective data analysis to validate algorithm (n=9) Prospective comparative observational study (n=1) Observational study with historic control group (n=1)
Described outcomes and endpoints	*Clinical effectiveness: Workflow improvement*	*Accuracy (sensitivity, specificity, PPV< NPV)* *Clinical Effectiveness:* time to decision CT/no Ct, number of treatable cancers identified *Patient &Social: patient experience* *Cost and economic: basic budget analysis*	*Diagnostic accuracy: sensitivity, specificity, accuracy, AUROC* *Clinical effectiveness:* Time to diagnosis, time to treatment	*Diagnostic accuracy:* sensitivity specificity, PPV, NPV *Clinical effectiveness*: mortality, mRS, time to treatment, reliability to aid decision making number of treatments, mortality, mRS, HRQoL *Organisational:* number of treatments, length of stay *Cost & Economic:* Cost-effectiveness Ethical: equality considerations	*Diagnostic accuracy:* sensitivity, specificity *Clinical effectiveness:* length of stay, time to treatment *Organisational:* Resource consequences *Ethical:* Equality considerations (no equality issues related to the use of AI *Cost & Economic:* technology purchase and implementation cost
Overall methodology	Health Technology Review	Innovative Medical Technology review examining potential impact on health and social care in Scotland. Highlights strengths/weaknesses of the evidence base.	The potential of this technology was determined by a committee of innovative health technology experts based on data submitted by the applicant and their expert opinions on its potential.	Committee discussion based on a literature search for published evidence Systematic review methods will follow the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care, NICE Diagnostics Assessment Programme manual and the Cochrane Handbook for Diagnostic Test Accuracy Reviews.	Med innovation briefing Literature research was carried out in accordance with the interim process and methods statement of NICE Expert opinions on the technology were invited but are not considered to represent NICE's views.
Assessment framework	NR	NR	NA	NR	NR

Author/Institution, year	CADTH, 2023 Health Technology Review [62]	HIS/SHTG, 2024 MIB [69]	NECA, 24 Assessment of potential [58]	NICE, 2024 Full Guidance [75]	NICE, 2020 MIB [70]
Evidence requirements	NR	NR	NA	NR	NR
AI-specific checklist	No	No	No	No	No
Tools for Risk of Bias Assessment	NR	NR	NA	NR *Strengths and limitations discussed but not formally assessed with a tool*	NR *Strengths and limitations discussed but not formally assessed with a tool*
Author's conclusion	The included limited evidence raises concerns on training data, the potential for generating inaccurate or fictitious information and the lack of domain expertise. No evidence on cost-effectiveness was included.	No published evidence on clinical effectiveness, cost-effectiveness or safety of the technology was identified. An interim analysis of an ongoing service evaluation in an NHS trust suggests reduced time to reporting, reduced time to treatment and increased identification of treatable lung cancers/.	Evidence suggests high diagnostic accuracy and a potential improvement in patient care through early detection and intervention. Evidence is insufficient to prove time savings. The technology will be appraised for safety and effectiveness within a new medical technology evaluation.	No published evidence for nine of the thirteen assessed technologies. Currently, three technologies can be used with evidence generation (with DTAC approval), only with professional review and under maintenance of standard scan protocols Uncertainty about comparative diagnostic accuracy Inconclusive and limited evidence for time to treatment and clinical outcomes Some evidence for cost-effectiveness Need for quality evidence on test accuracy, clinical effectiveness (impact on time to treatment, patient outcomes) and cost-effectiveness	Limited evidence suggests comparable diagnostic accuracy. The resource impact would be greater than standard care, but this might be offset by the added clinical benefit. Need for comparative studies to evaluate clinical effectiveness (including patient outcomes/time to treatment) and cost-effectiveness.

Abbreviations: AI … artificial intelligence, CADe … computer aided detection, CADTH … Canadian Agency for Drugs and Technologies in Health, CEA … cost effectiveness analysis, CT … computed tomography, CXR … chest Xray, DHTC … Danish Health Technology Council, DICOM … Digital Imaging and Communications in Medicine, ICER … incremental cost-effectiveness analysis, EHR(S) … electronic health record (system), HIS … Healthcare Improvement Scotland, HL7 … health level 7, MDT … multidisciplinary team, ML … machine learning, PACS … picture archiving and communication system, SHTG … Scottish Health Technologies Group, SR … systematic review

Table A-4: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 2

Author/Institution, year	NICE, 2023 Full Guidance [70]	NICE, 2021 EVA [70]	NICE, 2021 EVA [73]	NICE,2021 MIB [72]	NIHR, 2024 HTA [76]
Study characteristics					
Country	UK	UK	UK	UK	UK
Functional Category[12]	Diagnosis	Diagnosis	Diagnosis	Diagnosis	Diagnosis
(medical) specialty	Radiology	Radiology	Radiology	Radiology	Radiology
Population(s)	Patients undergoing a chest CT scan who have no confirmed chest nodules or lung cancer, or people who are having CT surveillance for a previously discovered chest nodule	Patients referred to chest CT	Patients referred from primary care who are undergoing chest XR	Patients undergoing mammograms for screening or diagnostic purposes	Patients (>18y) presenting in secondary care with suspected stroke or AIS
Intervention	AI-supported detection and measurement of nodules in chest CT scans	AI-supported Chest CT interpretation	AI Analysis of chest X-rays for suspected lung cancer in patients referred from primary care	AI-supported mammography review	AI derived software assisted CT scan review systems assessing non enhanced CT brain scans, CTA and CTP scans reviewed by an HCP other than a neuroradiologist
Comparator	Standard of care/CT scan review by a radiologist or less qualified HCP without AI assistance	Standard of care/Review by radiologist	Standard of care Chest XR review by radiologist/ radiographer with varying levels of experience and without AI assistance	Standard of care/2 readers and arbitration if necessary	Unassisted or AI-assisted review of plain CT brain scan, CTA brain scan, CTA and CTP brain scan by an HCP other than a neuroradiologist
Eligibility criteria	Relevant published evidence on the included technologies and populations Predefined outcomes: *Diagnostic accuracy: e.g.* sensitivity, specificity, concordance *Clinical effectiveness:* e.g. Morbidity, Mortality, HRQoL, time to report/diagnosis/stage of cancer detected *Patient &Social:* Populations most likely to benefit *Cost & Economic:* Cost effectiveness	Most relevant or best available published evidence relating to the clinical effectiveness of the technology. Predefined outcomes: not specified	Peer reviewed comparative studies No studies were identified in the population of interest, so the criteria were widened to include studies that compared X-ray review by AI software with a review by radiologists alone, but not from the population specified. Predefined outcomes: *Diagnostic accuracy:* sensitivity, specificity, test failure rate *Clinical Effectiveness:* time to review, triage outcomes, time to referral, Mortality, Morbidity, HRQoL	Most relevant or best available evidence relating to the clinical effectiveness of the technology Predefined outcomes: not specified	Clinical Effectiveness: Diagnostic accuracy studies, comparative study designs (RCTs, controlled clinical trials, observational studies) Cost Effectiveness: CEA studies examining QUALYs with ≤ AI derived software assisted review technology Predefined outcomes: *Diagnostic accuracy:* sensitivity, specificity *Clinical Effectiveness:* mortality, mRS, time/rate to thrombolysis, HRQoL *Safety* (adverse events) *Organisational:* Length of stay, time in AE (pre-admission/discharge)

[12] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

Author/Institution, year	NICE, 2023 Full Guidance [70]	NICE, 2021 EVA [70]	NICE, 2021 EVA [73]	NICE,2021 MIB [72]	NIHR, 2024 HTA [76]
Eligibility criteria *(continuation)*					*Patient & Social*: clinician acceptability/ease of use *Cost & Economic:* cost effectiveness
Number of included studies	27	2	11	6	22
AI characteristics					
AI Product name | AI Type (CNN, LLM, unspecified, others)	AI-Rad Companion Chest CT (Siemens) AVIEW LCS+ (Coreline Soft) ClearRead CT (Riverain) contextflow SEARCH (contextflow) InferRead CT lung (Infervision) LD-01K (JLK) Lung AI (Arterys) Lung Nodule AI (Fujifilm) qCT (Qure.ai) SenseCare-Lung Pro (Sense Time) Veolity (MeVis) Veye Lung Nodules (Aidence) VUNO Med-Lung CT AI (VUNO)	Veye Chest (Aidence) Icolung (icometrix) Veolity (MeVis)	AI-Rad Companion Chest X-ray (Siemens Healthineers, CADx) Annalise CXR (Analise.ai, CADe/CAST) Auto Lung Nodule Detection (Samsung, CADe) ChestLink (Oxipit, CADe/CAST) ChestView (Gleamer, CADe) Chest X-ray (Rayscape, CADe) ClearRead Xray (Riveraintech, CADe) InferRead DR Chest (Infervision, CADe) Lunit INSIGHT CXR (Lunit, CADe) Milvue Suite (Milvue, CADe/CAST) qXR (qure.ai, CADe) Red dot (Behold.ai, cADe/CADx/CAST) SenseCare-Chest DR PRO (SenseTime, CADe) VUNO Med-Chest X-Ray (VUNO, CADe)	Transpara Mammography (ScreenPoint Medical) | DL -CNN Transpara DBT (ScreenPoint Medical) HealthMammo | DL -CNN (Zebra Medical Vision) ProFound AI for 2D mammography ProFound AI for DBT (iCAD) | DL-CNN	Avicenna CINA LVO Brainomix e-CTA/e-ASPECTS RapidAI®/CTA/LVO/CTP ischemaView Viz LVO, Viz ICH
Training data	NR	Icolung: NR Veye Chest: training data from people aged 50 to 74 in a registry of people who smoke	NR	Transpara: >1,000,000 images from US and EU sites HealthMammo: >500,000 cases from 150 facilities across 3 continents ProFound AI (iCAD): 2,000,000 images	NR
Information on AI algorithm/model specifications	No	No	No	Transpara: deep learning CNN (feature classifiers and image analysis algorithms) HealthMammo Software: deep learning CNN ProFound AI: deep learning CNN (feature classifiers and image analysis algorithms)	No

Author/Institution, year	NICE, 2023 Full Guidance [70]	NICE, 2021 EVA [70]	NICE, 2021 EVA [73]	NICE,2021 MIB [72]	NIHR, 2024 HTA [76]
AI main functions	AI analysis of chest CT images to detect the presence and growth of lung nodules All software in this assessment can detect nodules and measure their volume but the software uses a fixed algorithm that cannot update itself All assessed software integrates with PACS.	Using a machine/deep learning, the software analyses chest CT images, highlights abnormal findings, characterises and labels findings and prioritises images that need urgent review Some technologies allow the images to be transferred from the hospital to the software platform, and when output is returned, it can be used/viewed with hospital systems such as DICOM, PACS or HL7 can be used.	Computer-assisted triage (CAST): Prioritisation and triage of images Computer-aided detection (CADe) and Computer-aided diagnosis (CADx): Identification/diagnosis of abnormalities on Chest Xrays	Machine learning AI models are trained to detect and characterise suspicious mammography features and predict the likelihood of malignancy The AI systems are provided with training images specifying 'ground truth' and instead of learning to classify new cases on predefined rules, the system learns from examples provided and recognises patterns that predict the outcome Systems provide risk scores and triage notifications to aid clinical decision making	AI algorithm to analyse CT Brain images for findings suggestive of stroke to notify an appropriate medical specialist of these findings in parallel to standard of care image interpretation.
Expectations for the technology	■ Reduced time to review and diagnosis ■ Increased detection of lung nodules that need further assessment and that would have been missed by human readers ■ Improved reporting of characteristics to improve decision-making ■ Monitor nodule growth	■ support radiologists in the review process of chest CT images ■ improve diagnostic accuracy ■ faster review time ■ reduce workload	■ Faster review time ■ Prioritisation of relevant patients ■ Faster time to referral for further tests ■ Faster time to diagnosis/treatment ■ Increased accuracy of reporting	■ Reduce workload and waiting times ■ Improve diagnosis especially of very small lesions difficult to interpret by the human eye ■ Reduce unnecessary recalls	■ Facilitates scan review by non-neuroradiologists ■ Enhances triage, prioritisation, and transfer ■ Supports evaluation for time-sensitive treatments ■ Enables rapid report turnaround and multisite scan reviews through integration with radiology CT stations and cloud hosting
Study Methodology					
Types of Included studies	Retrospective studies (n=24) Prospectives studies (n=3)	Non-peer-reviewed conference abstracts of 2 retrospective studies	Retrospective studies (n=11)	Retrospective observational (validation) (n=6) Diagnostic accuracy dstudy (n=1) Conference proceedings (n=3) Conference abstracts (n=3)	Diagnostic accuracy studies/cross sectional studies (n=15), Observational studies (n=7)
Described outcomes and considerations	*Diagnostic accuracy:* sensitivity, specificity, concordance *Clinical effectiveness:* Morbidity, Mortality, HRQoL,time to report/ diagnosis/stage of cancer detected *Ethical:* Equality considerations *Patient &Social:* Morbidity, Mortality, HRQoL, acceptability, Populations most likely to benefit	*Diagnostic accuracy:* segmentation accuracy (dice coefficient) *Ethical:* Equality considerations *Organisational:* resource consequences *Cost & Economic:* technology purchase and implementation cost	*Diagnostic accuracy:* sensitivity, specificity, test failure rate *Clinical Effectiveness:* time to review, triage outcomes, time to referral, Mortality, Morbidity, HRQoL *Organisational:* practical implications (e.g. image review to radiology report), standardisation of report writing, technical failure rate	*Diagnostic accuracy:* AUROC, sensitivity, specificity *Clinical effectiveness:* time to diagnosis, recall rate *Organisational:* required infrastructure for implementation, training requirement, effect on hospital resource requirements *Ethical:* Equality considerations	*Diagnostic accuracy:* sensitivity, specificity *Clinical Effectiveness:* mortality, mRS, time/rate to thrombolysis, HRQoL *Safety* (adverse events) *Organisational:* Length of stay, time in AE (pre-admission/discharge)

Author/Institution, year	NICE, 2023 Full Guidance [70]	NICE, 2021 EVA [70]	NICE, 2021 EVA [73]	NICE,2021 MIB [72]	NIHR, 2024 HTA [76]
Described outcomes and considerations *(continuation)*	*Cost & Economic:* Cost effectiveness		*Patient & Social:* Populations most likely to benefit *Ethical:* Equality considerations *Cost & Economic:* Cost effectiveness	*Cost & Economic:* technology purchase and implementation cost	*Patient & Social*: clinician acceptability/ease of use *Cost & Economic:* cost-effectiveness
Overall methodology	Committee discussion based on a literature search for published evidence Systematic review methods follow the principles of the NICE Diagnostics Assessment Programme manual and the Cochrane Handbook for Diagnostic Test Accuracy Reviews. CEA	Early value assessment/rapid review: Literature Search and multi-stakeholder committee discussions Comments received are individual opinions and do not represent NICE's views Literature research was carried out in accordance with the interim process and methods statement of NICE and includes the most relevant or best available evidence relating to the clinical effectiveness of the technology	Early value assessment/rapid review: Literature Search and multi-stakeholder committee discussions Comments received are individual opinions and do not represent NICE's views The purpose of this early assessment was to assess the evidence on adjunct AI software for analysing chest x-rays for suspected lung cancer and to identify evidence gaps.	Med innovation briefing Literature research was carried out in accordance with the interim process and methods statement of NICE Expert opinions on the technology were invited but are not considered to represent NICE's views.	Systematic Review with a summary of results according to research question, AI-derived technology and study type, following the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care, NICE Diagnostics Assessment Programme Manual and the Cochrane Handbook for Diagnostic Test Accuracy Reviews Cost Effectiveness Analysis was performed using a de novo model based on decision tree model and a state transition model to calculate the mean adjusted costs and quality adjusted life years.
Assessment framework	NR	NR	NR	NR	NR
Evidence requirements	NR	NR	NR	Evidence was evaluated on the background of the UK National Screening Committee (NSC) interim guidance	Clinical Effectiveness: NR Cost Effectiveness: NR
AI-specific checklist	No	No	No	No	No
Tools for Risk of Bias Assessment	ROB2 for RCT ROBINS I for non-RCT, before/after studies, historical control studies and cohort studies CASP checklist for qualitative studies NICE, preferred appraisal tools for other designs	NR *Strengths and limitations discussed but not formally assessed with a tool*	*Final protocol describes that study design appropriate tools will be use, citing the example of tools by the Joanna Briggs Institute (JBI). Deprioritised studies will not be assessed for risk of bias.*	NR *Strengths and limitations discussed but not formally assessed with a tool*	QUADAS-2 for diagnostic test accuracy (n=15) Author's own quality checklist on observational studies (n=7)

Author/Institution, year	NICE, 2023 Full Guidance [70]	NICE, 2021 EVA [70]	NICE, 2021 EVA [73]	NICE,2021 MIB [72]	NIHR, 2024 HTA [76]
Results/Conclusion	For people having chest CT scans as part of a targeted lung cancer screening program, evidence suggests cost-effectiveness, when used alongside clinician review, however there is not enough evidence to determine which of them are most clinically and cost- effective. Evidence suggests increased sensitivity, but reduced specificity, and potential faster time to reporting, depending on the experience level of the clinician reviewing. Further research is needed to compare technologies and evaluate clinical impact. There is not enough evidence to evaluate the use of the technology outside of targeted screening programmes.	Limited evidence from only two retrospective studies suggests comparable accuracy to clinician review but there are concerns regarding the generalisability of training data. The use of the technology has the potential to reduce resource use by reducing workload of staff. There is insufficient evidence to derive conclusions on cost-effectiveness, however, costs may vary due to different technologies. Need for more research on safety, accuracy, clinical efficacy and cost-effectiveness.	The current available evidence does not allow to assess accuracy in the population of interest, clinical efficacy and cost-effectiveness Further evidence is needed on the diagnostic accuracy in parallel to clinician review, on the risk and consequences of false positives/ negatives and impact on review time.	Limited evidence suggests the AI technology may improve performance and save time in mammography. Datasets for clinical validity may not be representative as the studied datasets were from patients with breast cancer which is not representative of the screening scenario. There is insufficient evidence to derive conclusions on cost-effectiveness, however, costs may vary due to different technologies. Authors point out the need for prospective randomised accuracy studies with multiple arms to incorporate comparison of different AI technologies, concordance studies and the need for real-world evidence.	The available evidence is not suitable to determine the clinical effectiveness, and the economic analyses did not provide evidence to prefer the AI assisted strategy over current clinical practice.

Abbreviations: AE … accident and emergency, AI...artificial intelligence, CEA … cost effectiveness analysis, CT … computed tomography, CTA … computed tomographic angiography, CTP … computed tomography perfusion, CO_2 … carbon dioxide, COPD … chronic obstructive pulmonary disease, DHT … digital health technologies, DHTC … Danish Health Technology council, ESF … evidence standards framework … healthcare professional, GP … general practice, HCF … healthcare facility, HRQoL … health related quality of life, HTA … health technology assessment, HTW … Health technology Wales, mRS … modified ranking score, NR … not recorded, NIHR … National Institute of Health Research, NPV … negative predictive value, RCT … randomized controlled trial, SCHARR … Sheffield Centre for Health and Related Research, TIEDieR … Template for Intervention Description and Replication, QUADAS … Quality assessment of diagnostic accuracy studies, QUALY … quality adjusted life year, PPV … positive predictive value, SR … systematic review

Table A-5: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 3

Author, year	AQUAS, 2023 HTA [60]	DHTC, 2023 [56]	HTW 2023, Topic exploration report [68]	HTW 2024, Topic exploration report [67]
Study characteristics				
Country	Spain	Denmark	UK	Wales, UK
Functional Category[13]	Diagnosis	Diagnosis	Diagnosis	Diagnosis
(medical) specialty	Internal Medicine	Internal Medicine	Internal Medicine	Internal Medicine
Population(s)	Adult patients with suspected colorectal precancerous lesions	Adults undergoing colonoscopy	Patients with suspected COPD	Patients with suspected cardiovascular disease in primary/emergency care
Intervention	AI-assisted colonoscopy	CADe Colonoscopy	AI assisted COPD diagnosis including CO_2 sensor devices	AI-assisted ECG interpretation
Comparator	Standard colonoscopy without AI-assistance	Standard colonoscopy	Spirometry/Standard of care	Standard of care/ECG interpretation by clinician
Eligibility criteria	RCTs, diagnostic test studies, SRs, economic evaluations, observational studies Predefined outcomes: all variables from included studies on safety efficacy/effectiveness and efficiency	NR Predefined outcomes: not specified	NR Predefined outcomes: not specified	Primary studies with confirmed regulatory approval of the technology under investigation as well as systematic reviews Predefined outcomes: not specified
Number of included studies	2	9	5	11
AI characteristics				
AI Product name | AI Type (CNN, LLM, unspecified, others)	GI Genius (Medtronic)	NR	N-Tidal Diagnose (TidalSense)	PMCardio (Powerful Medical) KardiaMobile (MTG64)
Training data	NR	NR	NR	NR
Information on AI algorithm/model specifications	The CADx system consists of two algo-rithms, one that classifies each detected polyp as an "adenoma" or "non-adenoma" polyp and provides a description of the appearance, and another algorithm which provides an image quality score for each detected polyp that expresses how clearly the polyp characteristics are displayed. The first algorithm is designed to activate automatically when a new poly is detected, and overlay images to compare and make a decision in real time.	No	No	No

[13] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

Author, year	AQUAS, 2023 HTA [60]	DHTC, 2023 [56]	HTW 2023, Topic exploration report [68]	HTW 2024, Topic exploration report [67]
AI main functions	AI-assisted colonoscopy device supports detection and characterisation of precancerous lesions as a complement to the clinician's review CADe systems use deep learning/CNN algorithms to analyse real-time images during colonoscopy procedures and alert clinicians to suspicious areas in order to improve polyp detection	CADe systems use deep learning/CNN algorithms to analyse real-time images during colonoscopy procedures and alert clinicians to suspicious areas to improve polyp detection	A high-resolution CO_2 sensor is combined with an AI platform to measure changes in the movement of respiratory gases indicative of a diagnosis of COPD to support clinical diagnosis	AI assisted ECG pattern interpretation A mobile digital application digitises an image file of ECG tracing, and the raw signal is interpreted by AI models to classify/ diagnose cardiac rhythm abnormalities and cardiovascular conditions to support clinicians with diagnosis
Expectations for the technology	■ Improved detection of neoplastic changes during colonoscopy	■ Improved detection of neoplastic changes during colonoscopy	■ Point-of-care rapid detection of COPD	■ Aid diagnosis of cardiovascular disease associated with ECG changes ■ Improved triage/patient flow
Study Methodology				
Types of Included studies	RCT (n=2)	RCT (n=2), Qualitative studies (n=7), Interview (n=1)	*Primary evidence:* Observational studies (n=4) *Secondary evidence:* SR (n=1)	*Primary Evidence:* Observational studies (n=3) *Secondary Evidence:* Systematic Reviews (n=8)
Described outcomes and considerations	*Clinical effectiveness:* Adenoma detection rate (ADR), morphology and size, adenomas detected by colonoscopy (APC), proximal vs distal adenomas, n of polyps detected, n of sessile serrated lesions *Safety: non neoplastic resection rate, duration of withdrawal process* Organisational: Implementation considerations, impact on health system Ethical: ethical, political, social impact of the technology (accessibility) *Cost & Economic:* cost-effectiveness	*Clinical Effectiveness:* difference in detection of histologically confirmed Adenomas *Organisational:* clinician's acceptance, incorporation into clinical setting, risk of overtreatment *Patient & Social:* attitude and acceptance, preferences, experiences *Cost & Economic: cost effectiveness* analyses (CEA, ICER)	*Diagnostic Accuracy:* sensitivity, specificity *Clinical Effectiveness:* time to diagnosis, referrals to spirometry, patient satisfaction and quality of life *Cost & Economic:* Cost effectiveness	*Diagnostic accuracy:* AUROC, sensitivity, specificity, diagnostic odds ratio, comparative diagnostic accuracy *Clinical Effectiveness:* time to diagnosis, time to treatment, HRQoL *Organisational:* number of referrals to secondary care, resource use *Cost & Economic:* cost of technology
Overall methodology	Systematic review	NR	Topic explorations are designed to provide a high-level briefing on new topics submitted for consideration by Health Technology Wales. The main objectives of this report are to: ■ Determine the quantity of evidence available for a technology of interest. ■ Identify any gaps in the evidence. ■ Inform decisions on topics that warrant fuller assessment by Health Technology Wales	Topic explorations are designed to provide a high-level briefing on new topics submitted for consideration by Health Technology Wales. The main objectives of this report are to: ■ Determine the quantity of evidence available for a technology of interest. ■ Identify any gaps in the evidence. ■ Inform decisions on topics that warrant fuller assessment by Health Technology Wales

Author, year	AQUAS, 2023 HTA [60]	DHTC, 2023 [56]	HTW 2023, Topic exploration report [68]	HTW 2024, Topic exploration report [67]
Assessment framework	NR	DHTC process	NICE ESF (categorised as Tier C digital health technology)	NICE ESF (categorised as Tier C digital health technology)
Evidence requirements	NR	NR	Tier C DHT (ESF) requires satisfactory evidence to prove its claimed benefits: ■ prospective peer reviewed studies in a similar HC setting ■ real world evaluation of clinical utility with high quality studies confirming improvements in relevant outcomes ■ Economic analysis	Tier C DHT (ESF) requires satisfactory high-quality evidence to prove its claimed benefits, specifically test accuracy studies and concordance studies, to evaluate relevant outcomes: ■ Test accuracy ■ Time to diagnosis ■ Clinical utility
AI-specific checklist	No	No	No	No
Tools for Risk of Bias Assessment	ROB-2 tool for RCT	NR	NR	NR
Author's conclusions	Evidence suggests no safety concerns Overall, the ADR was increased when the AI-technology was used Adenomas < 10 mm were more frequently detected, but there was no statistically significant difference in the detection of adenomas >10 mm. No evidence on economic impact was found	CADe assisted Colonoscopy should not be implemented as a decision support tool as evidence suggests a risk of overtreatment	Evidence suggests diagnostic accuracy. No evidence for impact on time to diagnosis, referral number, resource use and cost-effectiveness was found.	Evidence suggests high accuracy and superiority/non-inferiority. No evidence on clinical outcomes or cost-effectiveness was found.

Abbreviations: AI … artificial intelligence, AIDR … artificial intelligence assisted diabetic retinopathy screening, AIHTA … Austrian Institute of Health Technology Assessment, AMSTAR … assessment of multiple systematic reviews, DWMS … digital wound care management systems, DRTS … digital retinopathy tele-screening, ECG … electrocardiogram, ED … emergency department, EHRS … electronic health record system, ICER … incremental cost effectiveness ratio, NICE ESF … National Institute of Care Excellence Evidence Standard Frameworks, NIHR … National Institute of Health Research, DHT … digital health technology, DRTS … digital retinal screening training systems, HCP … healthcare professional, HTA … health technology assessment, MMA … mobile medical applications, MS … multiple sclerosis, NR … not recorded, QUADAS … quality assessment of diagnostic accuracy studies, RCT … randomized controlled trial, Rob … risk of bias, SR … systematic review, SAT … single arm trial

Table A-6: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 4

Author, year	NECA, 2024, Assessment of potential [59]	NECA, 24 Assessment of potential [57]	NICE,2023 EVA [81]	IACS, 2022 HTA [61]	NICE, 2022 MIB [80]
Study characteristics					
Country	South Korea	South Korea	UK	Spain	UK
Functional Category[14]	Diagnosis	Diagnosis	Diagnosis	Diagnosis	Diagnosis
(medical) specialty	Internal Medicine	Internal Medicine	Internal Medicine	Dermatology	Dermatology
Population(s)	Patients aged 30 and over who currently have a normal sinus rhythm but have a history of atrial fibrillation	Patients presenting with symptoms of MI	Patients with stable cardiac symptoms/suspected CAD undergoing CTCA scan	Persons with lesions suspicious for malignant melanoma including patients with risk factors for CM	Patients undergoing skin checks
Intervention	AI-based algorithm predicting the likelihood of paroxysmal atrial fibrillation from standard ECG	AI-based algorithm to assist diagnosing MI from 12-lead ECG	AI-assisted analysis of CTCA scans to assess inflammation level	(AI-assisted) digital dermatoscopy [15]	AI-assisted mole analysis
Comparator	portable Holter ECG	Standard of care/ECG without AI-assistance and cardiac enzyme assessment (for NSTEMI)	Standard of care/CTCA plus clinical assessment of risk factors for CVD	Manual dermatoscopy + photography	Standard of care/dermoscopy without AI-assistance
Eligibility criteria	NA Predefined Outcomes: not specified	NA	RCT, CCT, comparative or non-comparative observational studies, before and after studies, development and validation studies, cost effectiveness analyses Predefined outcomes: *Predictive accuracy* *Clinical effectiveness: patient outcomes (*cardiac events), HRQoL proportion of patients requiring lifestyle changes/drug treatment, changes to clinical management, time to test results *Cost & Economic:* technology purchase and implementation cost, cost of treatment, additional testing	Any study evaluating DD and DM + photography apart from case-control studies Predefined outcomes: *Diagnostic accuracy: sensitivity, specificity, PPV, NPV* *Clinical effectiveness*: number of excised lesions, number of lesions identified at in situ stage, number of lesions needed to excise to diagnose MM, reduction in excised lesions *Organisational:* efficiency/differences in the use of resources *Patient & Social:* Patient acceptance and compliance *Cost & Economic:* Health and social costs, budget impact	NR Predefined outcomes: not specified

[14] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

[15] Only some DD devices utilize artificial intelligence

Author, year	NECA, 2024, Assessment of potential [59]	NECA, 24 Assessment of potential [57]	NICE,2023 EVA [81]	IACS, 2022 HTA [61]	NICE, 2022 MIB [80]
Number of included studies	2	n=3	1	17	14
AI characteristics					
AI Product name | AI Type (CNN, LLM, unspecified, others)	ECG analysis software, SmartECG-AF | RNN	AiTiAMi	CaRi-Heart (Caristo diagnostics)	DB-MIPS (DM-Dermo MIPS) DermoGenious (Rodenstock) DermoGenius Basic II (Linos Photonics) FotoFinder bodyscan ATBM (FotoFinder Systems GmbH) MicroDERM (Visiomed AG) Mole Max II (Derma Instruments) Mole Expert (DermoScan Gmbh) SolarScan (Polartechnics)	nomela (Moletest Scotland) DERM (Skin Analytics) Molenanalyzer pro (FotoFinder Systems) Skin Vision
Training data	NR	NR	NR	NR	DERM: proportion of historical and prospectively collected images from UK population Others: NR
Information on AI algorithm/model specifications	No	No	All software technologies in clinical settings use fixed algorithms. They cannot adapt in real time using data from the clinical practice setting in which they are used.	No	No
AI main functions	The software analysis 12-lead electrocardiograms with normal rhythm using AI algorithms (RNN) rand presents the probability of PAF to assist HCP in making diagnostic decisions. The ECG software receives data from a 12 lead ECG machine and utilises data from 8 leads to determine the normal rhythm of the patient.	AI-based algorithm analyses the 12-lead ECG and detects the possibility of acute MI Score and risk level are displayed for diagnostic support	AI algorithm-assisted analysis of inflammation level in the coronary arteries predicts 8-year cardiac death risk, atherosclerotic plaque burden and clinical risk factors	(AI-assisted) digital dermoscopy In some digital dermoscopy devices, AI algorithms review and inform the detection of possible melanoma in skin lesions. Assessed with a dermatoscope or tablet/smartphone camera, the algorithm assesses the image and presents a risk score Some devices use deep learning and can continuously update themselves whilst others have to be retrained with new images to update the signal processing algorithm.	AI algorithms review and inform the detection of possible melanoma in skin lesions. Assessed with a dermatoscope or tablet/smartphone camera, the algorithm assesses the image and presents a risk score Some devices use deep learning and can continuously update themselves whilst others have to be retrained with new images to update the signal processing algorithm. All require access to the internet

Author, year	NECA, 2024, Assessment of potential [59]	NECA, 24 Assessment of potential [57]	NICE,2023 EVA [81]	IACS, 2022 HTA [61]	NICE, 2022 MIB [80]
Expectations for the technology	■ Improved diagnosis rate and accuracy ■ Improved patient management through early detection and intervention	■ Improved and faster diagnosis ■ faster treatment ■ Improved patient outcome ■ Triage function (cardiac enzymes only rise after 3-6 hours whilst this technology can detect NSTEMI faster) ■ resource savings	■ Improved risk prediction ■ Increased motivation of patients to moderate risk by adhering to medication and lifestyle changes ■ Optimisation of prevention and treatment strategies	■ Increased accuracy in diagnosing malignant melanoma ■ Increased detection of melanoma in early treatable stages ■ Decrease in unnecessary excision of benign lesions ■ Improved monitoring of patients at risk for MM or previously diagnosed MM	■ Reducing waiting lists and streamline workflows in dermatology referrals ■ Earlier diagnosis and treatment of skin cancer and earlier reassurance for people with benign lesions.
Study Methodology					
Types of Included studies	Retrospective cohort (n=1) retrospective single-centre study (n=1)	Observational diagnostic test accuracy studies(n=3)	Retrospective observational study (n=1)	*Primary Evidence* Sequential clinical trial (n=1) Prospective cohort (n=8) Retrospective cohort (n=3) Qualitative studies (n=2) Budget Model (n=1) *Secondary Evidence* HTA assessment (n=1) SR (n=1)	Non-randomised comparative study (n=1) Prospective observational study (n=4) Cross sectional studies (n=5) Retrospective observational study (n=3) Prospective study with retrospective phase 2 (n=1)
Described outcomes and considerations	*Diagnostic accuracy: sensitivity, specificity, F1-Score* *Clinical effectiveness:* time to diagnosis *Ethical*: Equity considerations *Patient & Social:* Practicality of use, acceptability, HRQoL *Cost-effectiveness:* CEA	*Diagnostic accuracy:* AUROC, Sensitivity, Specificity, NPV, PPV,	*Predictive accuracy* *Clinical effectiveness:* Proportion of patients requiring lifestyle changes/drug treatment, changes to clinical management, time to test results, patient outcomes (mortality, cardiac events), HRQoL *Organisational:* infrastructure for implementation, training, effect on hospital resources *Ethical:* equity considerations *Cost & Economic:* technology purchase and implementation cost, cost of treatment, additional testing	*Diagnostic accuracy:* sensitivity, specificity *Clinical effectiveness:* number of excised lesions, number of lesions identified at in situ stage, number of lesions needed to excise to diagnosis, reduction in excised lesions *Organisational:* efficiency and differences in the use of resources *Patient & Social:* patient acceptance and compliance *Cost & Economic:* health and social costs, budget impact	*Diagnostic accuracy:* sensitivity, specificity, AUROC, *Clinical effectiveness:* time to diagnosis *Safety:* issues with diagnostic inaccuracies *Organisational:* infrastructure needs for implementation, training, effect on hospital resources *Ethical:* Equality considerations *Cost & Economic:* technology purchase and implementation cost

Author, year	NECA, 2024, Assessment of potential [59]	NECA, 24 Assessment of potential [57]	NICE,2023 EVA [81]	IACS, 2022 HTA [61]	NICE, 2022 MIB [80]
Overall methodology	The potential of this technology was determined by a committee of innovative health technology experts based on data submitted by the applicant and their expert opinions on its potential.	The potential of this technology was determined by a committee of innovative health technology experts based on data submitted by the applicant and their expert opinions on its potential.	Early value assessment/rapid review: Literature Search and multi-stakeholder committee discussions	Systematic review	Med innovation briefing Literature research was carried out in accordance with the interim process and methods statement of NICE Expert opinions on the technology were invited but are not considered to represent NICE's views.
Assessment framework	NA	NA	NR	NR	NR
Evidence requirements	NA	NA	NR	NR	NR
AI-specific checklist	No	No	No	No	No
Tools for Risk of Bias Assessment	NA	NA	PROBAST	Osteba Critical Appraisal tool (FLC 3.0)	NR *Strengths and limitations discussed but not formally assessed with a tool*
Outcomes/ Conclusions	Evidence suggests comparable to higher diagnostic accuracy and cost-effectiveness. The technology can be referenced for diagnostic assistance purposes but further research on the influence of the technology on treatment decisions are required. It was decided that the Holter record is similar to the new technology, therefore suggesting a potential and eligibility for a new medical technology assessment.	Evidence for high diagnostic accuracy Evidence for resource savings and faster time to detection and intervention for high-risk patients Evidence suggests a potential, however, further research needed for improvement of medical outcomes in practice.	This technology might more accurately identify people at risk of heart attack or cardiac death than the standard risk assessment alone. The comparator in the study informing the assessment did not reflect standard UK practise of risk assessment, no evidence concerning patient outcomes or cost effectiveness was found. The technology is not recommended for use in the NHS while further evidence is generated.	Evidence suggests that in people with high or very high risk factors for MM, initial screening with manual dermoscopy or digital dermoscopy followed up with digital dermoscopy periodically may be more cost-effective than standard of care with just manual dermoscopy. Evidence suggests a potential benefits for persons with a high-risk factor.	Evidence suggests that diagnostic accuracy of the technologies is comparable or superior to standard of care. Need for high-quality evidence in the intended use, setting and population to assess clinical and cost-effectiveness and impact on the healthcare system.

Abbreviations: AI … artificial intelligence, CASP … Critical Appraisal Skills Programme, CCT … controlled clinical trial, CT … computed tomography, DBT … digital breast tomosynthesis, DTAC … digital technology assessment criteria, EU … European Union, HRQoL … health related quality of life, NICE … National Institute of Care Excellence, NPV … negative predictive value, MDT … multi-disciplinary team, mRS … modified ranking score, NR … not recorded, PACS … picture archiving and communication system, PPV … positive predictive value, RCT … randomised controlled trial, ROB2 … Risk of Bias tool for randomised trials, ROBINS 1 … Risk of Bias in non-randomised studies, US … United States

Table A-7: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 5

Author, year	INESSS, 2021 [63]	NICE, 2021 MIB [79]	HTW 2023, Topic exploration report [65]	NICE, 2021 MIB [78]
Study characteristics				
Country	Canada	UK	UK	UK
Functional Category[16]	Screening, Diagnosis	Diagnosis	Diagnosis	Diagnosis
(medical) specialty	Ophthalmology	Ophthalmology	Pathology	Pathology
Population(s)	Diabetic patients	Patients with Diabetes	Men undergoing prostate biopsy	Men undergoing prostate biopsy
Intervention	AI assisted diabetic retinopathy tele- screening	AI-assisted retina scan	AI assisted diagnosis of prostate biopsy	AI-assisted diagnosis of prostate biopsy
Comparator	Standard of care (telescreening without AI)	Standard of care DR screening by an HCP without AI assistance	Prostate biopsy diagnosed without AI assistance	Standard of care/prostate biopsy slides assessed without AI assistance
Eligibility criteria	NR Predefined outcomes: not specified	NR Predefined outcomes: not specified	Priority to RCTs and SRs Economic evaluations and relevant case studies Predefined outcomes: Diagnostic Accuracy: Sensitivity, Specificity, Accuracy Clinical Effectiveness: Procedure time, time to diagnosis, number of repeat biopsies Organisational aspects:Use of immunohistochemistry	NR Predefined outcomes: not specified
Number of included studies	23	7	11	5
AI characteristics				
AI Product name | AI Type (CNN, LLM, unspecified, others)	CARA/Neoretina EyeArt, IDx-DR V2.0 Retmarker OpthtAI SELANA+	EyeArt (Eyenuk) RetinaLyze (RetinaLyze System A/S) Retmarker (Retmarker)	Galen Prostate Solution (Ibex Medical Analytics) Paige Prostate (Paige AI, Inc) DeepDx (Deep Bio)	Paige Prostate (Paige AI sync), Deep Learning
Training data	NR	EyeArt: Data from 0.5 million patients and 2 million retinal images Others: NR	NR	2012- 2017 Digital archive from Memorial Sloan Kettering Cancer Centre, US

[16] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

Author, year	INESSS, 2021 [63]	NICE, 2021 MIB [79]	HTW 2023, Topic exploration report [65]	NICE, 2021 MIB [78]
Information on AI algorithm/model specifications	No	The algorithms in the technologies were trained using a database of existing human graded images. The algorithms were then tested against another set of images to fine tune them for real-world use.	No	While alternative artificial intelligence (AI)-based systems exist, the company claim that the Paige Prostate algorithm has been developed to be highly robust to variations in slide preparation from different institutions and does not need a per-site calibration using per-pixel annotations or other forms of calibration data.
AI main functions	Artificial intelligence-assisted detection of diabetic retinopathy and other sight-threatening eye conditions A retinal image is taken by an HCP and uploaded to the AI-platform which assesses and grades the image immediately and provides a report to aid decision on triage and further management of patient.	Software algorithm detects and grades diabetic retinopathy Artificial intelligence-assisted detection of diabetic retinopathy and other sight-threatening eye conditions A retinal image is taken by an HCP and uploaded to the AI-platform which assesses and grades the image immediately and provides a report to aid decision on triage and further management of patient. Most technologies can be integrated into the healthcare centre's software system Some technologies can incorporate patient history data and assess disease activity and also assess if the image is of adequate quality for evaluation.	AI algorithm-based software that can identify an area of interest on whole slide prostate biopsy images with the highest likelihood of harbouring cancer while automatically grading/measuring according to the Gleason scale to assist pathologists making a diagnosis.	AI algorithm-based software that can identify an area of interest on whole slide prostate biopsy images with the highest likelihood of harbouring cancer while automatically grading/measuring according to the Gleason scale in order to assist pathologists making a diagnosis.
Expectations for the technology	■ Improved screening service ■ Improved assessment and triage of patients to distinguish which need specialist follow-up ■ Earlier diagnosis and management of DM ■ Resource and cost savings	■ Integration into screening programme ■ Integration into current pathway (first/second/arbitration screener) ■ Cost savings and workload reduction	■ Increased accuracy and speed of prostate biopsy results	■ Improved productivity ■ Cost saving ■ Increased diagnostic accuracy ■ Improved patient outcomes
Study Methodology				
Types of Included studies	2/3 observational studies based on real-world data 1/3 experimental studies *No further details*	Prospective observational studies (n=3) Retrospective observational studies (n=4)	*Primary evidence:* Prospective observational studies (n=2) Retrospective observational studies (n=5) Cross-sectional study (ongoing, n=1) *Secondary evidence:* HTA assessment (n=1), SR (n=2)	Retrospective observational studies (n=5)
Described outcomes and considerations	*Clinical effectiveness:* sensitivity, specificity, pathologies recognised *Organisational Aspects:* practicalities and requirements for integration, position in care pathway, impact on resource	*Diagnostic accuracy:* sensitivity, specificity, PPV, NPV. AUC *Organisational:* infrastructure for implementation, training, effect on hospital resources	*Diagnostic accuracy: sensitivity, specificity, PPV, NPV* *Clinical Effectiveness:* time to diagnosis, number of repeat biopsies *Organisational:* procedure time	*Diagnostic accuracy:* sensitivity, specificity, PPV, NPV, test concordance *Clinical effectiveness:* time to diagnosis *Organisational:* infrastructure needs for implementation, training, effect on hospital resources

Author, year	INESSS, 2021 [63]	NICE, 2021 MIB [79]	HTW 2023, Topic exploration report [65]	NICE, 2021 MIB [78]
Described outcomes and considerations *(continuation)*	Safety: data security *Ethical:* Equity considerations, *Cost & Economic:* Cost effectiveness	*Ethical:* Equality considerations *Cost & Economic:* technology purchase and implementation cost	*Cost & Economic:* Cost effectiveness	*Ethical:* Equality considerations *Cost & Economic:* technology purchase and implementation cost
Overall methodology	Rapid review of clinical performance and efficiency and a reflection on various issues related to the use of the technology	Med innovation briefing Literature research was carried out in accordance with the interim process and methods statement of NICE Expert opinions on the technology were invited but are not considered to represent NICE's views.	Topic explorations are designed to provide a high-level briefing on new topics submitted for consideration by Health Technology Wales. The main objectives of this report are to: ■ Determine the quantity of evidence available for a technology of interest. ■ Identify any gaps in the evidence. ■ Inform decisions on topics that warrant fuller assessment by Health Technology Wales	Med innovation briefing Literature research was carried out in accordance with the interim process and methods statement of NICE Expert opinions on the technology were invited but are not considered to represent NICE's views.
Assessment framework	NR	NR	NICE ESF (categorised as Tier C digital health technology)	NR
Evidence requirements	NR	NR	Tier C DHT (ESF) requires satisfactory evidence to prove its claimed benefits: ■ rospective peer reviewed studies in a similar HC setting ■ real world evaluation of clinical utility with high quality studies confirming improvements in relevant outcomes ■ Economic analysis	NR
AI-specific checklist	No	No	No	No
Tools for Risk of Bias Assessment	NR	NR *Strengths and limitations discussed but not formally assessed with a tool*	NR	NR *Strengths and limitations discussed but not formally assessed with a tool*
Author's conclusion	Evidence suggests AIDR to be an efficient technology, however, the evidence supporting added benefits is uncertain with some studies reporting similar to lower clinical efficacy than standard of care retinal screening. Authors point out the necessity of a well-established DRTS program with a strong organisational structure is essential for the realisation of benefits.	Evidence suggests that the technology could be used to reduce staff needed to identify and grade diabetic retinopathy. There is a need for more high-quality evidence for clinical and cost- effectiveness in general as well as comparing the different technologies on the market.	Evidence suggests comparable diagnostic accuracy, reduced resource use and shorter time to diagnosis Overall evidence is unclear for impact on prostate cancer outcomes or cost-effectiveness	Limited evidence suggests that the device may increase diagnostic performance and productivity. Need for high-quality evidence to assess clinical efficacy, cost-effectiveness and impact on the healthcare system.

Abbreviations: AI … artificial intelligence, CAD … coronary artery disease, CADTH … Canadian Agency for Drugs and Technologies in Health, CTCA … CT coronary angiography, NICE … National Institute of Care Excellence, NR … not recorded, PROBAST … Prediction Model Risk of Bias Assessment Tool, RCT … randomised control trial, SAT … single arm trial

Table A-8: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 6

Author, year	AIHTA, 2021 SR [55]	CADTH, 2024 Horizon Scan [64]	HTW, 2023 FULL GUIDANCE [66]	CADTH, 2023 Horizon Scan [77]
Study characteristics				
Country	Austria	Canada	Wales, UK	Canada
Functional Category[17]	Diagnosis	Diagnosis	Diagnosis	Diagnosis
(medical) specialty	Triage/Patient management	Patient-Clinician Interaction	General Medicine	Neurology
Population(s)	Symptomatic individuals and healthcare professionals	All patients seeking healthcare information	Patients receiving wound care in any setting (in-patient and community)	Patients in ED/ICU
Intervention	Digital Symptom Checker Applications	AI software designed to simulate conversations with patients using humanlike language	Integrated digital wound care management systems (DWMS):	AI supported portable point-of-care EEG device
Comparator	Standard of Care (Face to Face appointments, phone hotlines)	NA	Usual care without AI assistance/Best practice requires regular patient and wound assessments with accurate documentation of wound and treatment plans by a range of different HCP	Conventional EEG
Eligibility criteria	RCTs, NRCTs, Observational studies, Register-studies, Reviews and Evaluation Reports Predefined outcomes: *Diagnostic accuracy:* sensitivity, specificity, PPV, NPV, concordance *Clinical effectiveness:* C time to diagnosis for diagnosis and triage functions, HRQoL, patient satisfaction, length of illness, severity of illness *Organisational:* number of visits to physician or ED *Patients & Social:* acceptability	Any publication on AI Chatbots or conversational agents in healthcare settings. Conference abstracts and grey literature were included only if they provided additional information	SRs, RCTs, SAT, If necessary: evidence from "lower priority" sources[18] Predefined outcomes: *Diagnostic accuracy:* reproducibility, accuracy of wound measurement, test-retest or interrater reliability, concurrent validity *Clinical effectiveness* (wound healing outcomes, time wound healing, resolution of infection, number of amputations) *Safety:* adverse events *Organisational:* (resource use, length of hospital stay, completion and accuracy of documentation) *Patient & Social* (patient adherence to treatment, patient satisfaction and quality of life)	Publications on Ceribell and point-of-care EEG. Conference abstracts and grey literature were included only if they provided additional information
Number of included studies	41 (8 new, 27 from SR that the report is based on)	15	18	9

[17] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

[18] Studies validating AI algorithms with historical images or using non-human models were excluded.

Author, year	AIHTA, 2021 SR [55]	CADTH, 2024 Horizon Scan [64]	HTW, 2023 FULL GUIDANCE [66]	CADTH, 2023 Horizon Scan [77]
AI characteristics				
AI Product name | AI Type (CNN, LLM, unspecified, others)	132 DiGAs (not clear, where AI is integrated), 38 relevant for Austria Various, listed in the Meta-Directory	Various 10 Examples in the report:(Ada, Babylon (eMed), Buoy Health, Florence, Healthily, OneRemission, Senseley, Symptomate, Youper, Woebot Health)	Minuteful for Wounds (Healthy.io) insight (eKare) Cares4Wounds (Tetsuyu HC) Tissue Analytics – Net Health Swift Wound – Swift Medical Wound Viewer – Omnidermal ImageJ software	Ceribell (device)/Clarity (algorithm)
Training data	NR	"AI models are trained on large sets of text-based closed data sets and use that information to generate responses to questions or participate in conversations"	NR	NR
Information on AI algorithm/model specifications	No further information identified in primary studies.	No	No	The clarity algorithm evaluates EEG signal over a 5-minute period. The thresholds for seizure activity that are used by Clarity are based on those described by the American Clinical Neurophysiology Society
AI main functions	AI- software designed to aid symptomatic patients with self-diagnosis or assessing healthcare staff with triage through evaluation of the issue, suggesting potential diagnoses and options for management	Generative pre-trained transformer (Trained to generate responses to questions or participate in human-like conversations.	AI-assisted 3D imaging of wounds to analyse and monitor wound and associated wound care. Digital images can be taken by HCPs or directly by patients. They are securely uploaded and can be used for initial assessment. The technology measures the size and depth of the wound and the type of tissue in the wound bed (assessing if there is an infection or not) Integration with EHRS and use of a centralised dashboard aids planning and resource management.	Fed through a portable headband EEG recorder the AI algorithm assesses seizure burden and effects of treatment within five-minute intervals in patients with suspected nonconvulsive seizures to support diagnosis, treatment decisions and patient management. Seizure burden, not individual events are recorded. Alarm alerts to status epilepticus, Online portal allows remote access to data for review
Expectations for the technology	NR	■ Relieving staff workload ■ 24/7 availability ■ Anonymous access to information for patients ■ Automation of repetitive tasks ■ Symptom Assessment and Triage ■ Assistance of chronic conditions including mental health ■ Appointment scheduling and other admin work ■ Cost saving	■ More accurate measurement of the wound ■ Supporting transfer of care to other health care professionals ■ monitoring wound healing or changes in wounds ■ for multi-disciplinary team wound review ■ assisting with remote care and supporting patient engagement with self-care	■ Faster identification of patients with nonconvulsive seizures, particularly in centres without consistent access to EEG/specialist on site ■ Faster initiation of treatment ■ Reducing under/overtreatment ■ Reduced length of hospital stays ■ improved patient outcome

Author, year	AIHTA, 2021 SR [55]	CADTH, 2024 Horizon Scan [64]	HTW, 2023 FULL GUIDANCE [66]	CADTH, 2023 Horizon Scan [77]
Study Methodology				
Types of Included studies	*Primary Evidence: (n=8)* Case vignettes (n=5) Prospective Observational (n=2) Case-control (1) *Secondary Evidence:* Systematic Reviews (n=1)	Systematic review (n=12) Scoping review (n=3)	Cross-sectional studies (n=9) and feasibility studies (n=9)	retrospective studies (n=5) non-randomised prospective studies (n=3) CEA (n=1)
Described outcomes and endpoints	*Diagnostic accuracy:* sensitivity, specificity, PPV, NPV, concordance *Clinical effectiveness:* C time to diagnosis for diagnosis and triage functions, HRQoL, patient satisfaction, length of illness, severity of illness *Organisational:* number of visits to physician or ED *Patients & Social:* acceptability	*Clinical Effectiveness: e.g.* behavioural change, mental health symptom improvement, HRQoL *Safety:* adverse events, patient harm *Ethical:* transparency, algorithm bias, equality, accessibility Legal: privacy *Patient &Social:* user experience, acceptability, usability	*Diagnostic accuracy:* reproducibility, accuracy of wound measurement, test-retest or interrater reliability, concurrent validity *Clinical effectiveness* (wound healing outcomes, time wound healing, resolution of infection, number of amputations) *Safety:* adverse events *Organisational:* (resource use, length of hospital stay, completion and accuracy of documentation) *Patient & Social* (patient adherence to treatment, patient satisfaction and quality of life)	*Diagnostic accuracy: sensitivity, specificity, PPV, NPV* *Clinical Effectiveness:* time to (correct) diagnosis, adjusted treatment plans, reduced treatment escalation, faster discharge *Safety:* adverse events *Organisational:* adjusted treatment plans, faster discharge *Cost & Economic:* cost-effectiveness
Overall methodology	Update of a NIHR Systematic review on Symptom checker applications	Horizon Scan	Rapid systematic review using standard HTA methods adapted from the Cochrane Rapid Reviews methods Group and the NICE guidelines manual For economic aspects, an economic literature review and a HTW internal cost analysis was performed	Horizon Scan
Assessment framework	modified NICE ESF	NR	NICE ESF (categorised as Tier C digital health technology)	NR
Evidence requirements	NR	NR	Effectiveness (Tier C DHT) considering specific endpoints: ■ Improvement in *organisational endpoints* (documentation, reporting, response time, specialist services, resource use) ■ *Patient-reported outcomes* including quality of life ■ Variation in effectiveness depending on wound type (small/discrete vs. complex/large)	NR

Author, year	AIHTA, 2021 SR [55]	CADTH, 2024 Horizon Scan [64]	HTW, 2023 FULL GUIDANCE [66]	CADTH, 2023 Horizon Scan [77]
Evidence requirements *(continuation)*			■ *Cost-effectiveness* (DWMS vs. usual care) ■ *Change in care pathways* ■ *Management of digital exclusion, connectivity and integration* ■ *Accuracy and reliability* of DWMS in routine practice ■ *Practice implications of unequal DWMS performance across skin tones/lighting levels* *Retention rate and adherence over time for patient-uploaded photos*	
AI-specific checklist	No	No	No	No
Tools for Risk of Bias Assessment	AMSTAR checklist for SR QUADAS-2 for diagnostic studies	NR	No formal tool to assess the RoB of included studies applied	NR
Outcome/Conclusion	The analysis of the evidence showed that for symptom-checkers there is currently insufficient evidence to show a medical or organisational benefit, as well as diagnostic accuracy. Some qualitative evidence for high usability Inconsistent evidence for cost-effectiveness and organisational impact	Safety concerns have been identified regarding lack of real time updates. Evidence suggests that the technology is effective for providing information to support behavioural changes, improving mental health symptoms, health promotion and supporting physical activity. No evidence on cost effectiveness was reported.	There is insufficient evidence to support routine adoption, as the impact on clinical management, healthcare resource use and patient outcomes cannot be evaluated.	Limited evidence suggests that Ceribell could avoid delayed treatment for patients with suspected nonconvulsive seizures with accessing conventional EEG systems. Use of Ceribell was associated with shorter hospital stays, changes in treatment plans, fewer escalations in antiseizure medication and a decrease in patient transfers to tertiary care. Currently the evidence stems from small retrospective studies and to accurately assess the added benefit of this technology, independent randomised trials with large sample size are required.

Abbreviations: AE … accident and emergency, ADR … adenoma detection rate, AQUAS … Agency for Health Quality and Assessment of Catalonia, CADTH … Canadian Agency for Drugs and Technologies in Health, CEA … cost-effectiveness-analysis, CT … computer tomography, DWMS … digital wound management system, ECG … electrocardiogram, HCP … healthcare professional, IACS … Institute for Health Sciences of Aragon, NECA … National Evidence-based healthcare Collaborative Agency (South Korea), PAF … paroxysmal atrial fibrillation, RNN … recurrent neural network

Table A-9: Overview of identified HTA reports: characteristics and utilised methods to evaluate AI systems 7

Author, year	CADTH, 2024 Horizon scan [87]	CADTH, 2023 Horizon scan [88]	NICE, 2023 Full Guidance [86]
Study characteristics			
Country	Canada	Canada	UK
Functional Category[19]	Prediction, Prognosis	Prediction, Prognosis, Organisational Aspects	Treatment
(medical) specialty	Triage/patient management	Palliative Care	Radiology
Population(s)	Patients in healthcare settings	Terminally ill patients	Patients undergoing radiotherapy
Intervention	AI based patient flow applications	AI based nudge tool	AI-based RX treatment contouring
Comparator	NA/prediction models without AI	Standard clinician decision without AI assistance In Canada, decisions to initiate palliative care are made my assessing the patient's condition using specific tools like the Palliative Performance Scale or the HOMR tool.	Standard of care, manual contouring, atlas-based contouring and model-based segmentation.
Eligibility criteria: Study types	Any study on AI or ML interventions to manage patient flow or appointment scheduling. Predefined outcomes: not specified	Any publication on combination of AI based mortality prediction models and behavioural interventions. Conference abstracts and grey literature were included only if they provided additional information Predefined outcomes: not specified	Range of quantitative and qualitative study types including RCTs and real-world evidence, and systematic reviews Predefined outcomes: *Accuracy:* Dice coefficient and qualitative measures, consistency *Clinical Effectiveness:* Acceptability of contours, alignment with guidelines, impact on RX treatment planning quality *Organisational Impact:* usability, Impact on resource use, staff and training performance, *Patient & Social:* user experience and satisfaction *Cost &Economic:* cost- consequence analysis *Ethical:* equality considerations
Number of included studies	6	2	15[20]
AI characteristics			
AI Product name | AI Type (CNN, LLM, unspecified, others)	Names NR: 7 technologies described	Names NR, 2 technologies described	AI- Rad Companion Organs RT (Siemens Healthineers) ART- Plan (TheraPanacea, Oncology Systems; Brainlab DLCExpert (MiradaMedical) INTContour (Carina Medical) Limbus Contour (Limbus AI, AMG Medtech)

[19] Prediction, prognosis, diagnosis, treatment, monitoring, organisational aspects

[20] 79 relevant studies were found but 15 were prioritised.

Author, year	CADTH, 2024 Horizon scan [87]	CADTH, 2023 Horizon scan [88]	NICE, 2023 Full Guidance [86]
AI Product name | AI Type (CNN, LLM, unspecified, others) *(continuation)*			MIM Contour ProtegeAI (MIM Software) MRCAT prostate plus Autocontouring (Philips) MVision Segmentation Service (MVision AI Oy, Xiel) RayStation (RaySearch) AutoContour (Radformation) OSAIRIS (Cambridge University NHS Trust)
Training data	NR	NR	NR
Information on AI algorithm/model specifications	No	■ N-dimensional eigenspace ■ Gradient-boosted tree	No
AI main functions	AI algorithms use data from patient's EHR to predict and monitor their movement through stages of treatment and care over time Some tools can also create automated appointment scheduling	Patients with a high risk of short- term mortality are identified by a ML prediction tool incorporated into the EHRS and clinicians are encouraged through notifications to have end of life conversations with those patients.	Contouring is an important part of Radiotherapy treatment planning in order to make treatment effective and minimise toxicity. AI-guided technologies outline the target radiation volumes and contouring of the organs at risk. They will produce an initial contour to be revised by trained HCP.
Expectations for the technology	■ Improved patient flow ■ Support volume forecasting and match demand of needs with supply resources ■ Improved use and allocation of resources	■ Increase end-of-life planning conversations between patients and clinicians ■ Increase referrals to end-of-life care ■ Improved end-of-life care	■ More efficient workflow, faster contouring preparation ■ Similar quality contours as manual contouring- improve consistency ■ Cost savings
Review Methodology			
Types of Included studies	*Primary evidence:* Retrospective studies (n=3) *Secondary evidence:* SR (n=3)	1 stepped-wedge cluster RCT comparing AI nudging with usual care 2 prognostic cohort study on prognostic performance 1 real word before and after implementation study	Clinical Effectiveness (n=15): Prospective studies (n=8) Retrospective studies (n=4) Retrospective study with prospective part (n=1) Blinded prospective evaluation of the algorithm (n=1, abstract only) Retrospective evaluation of the algorithm (n=1, abstract only)
Described outcomes and endpoints	*Prediction Accuracy (Sensitivity, Specificity, NPV, PPV)* *Clinical Effectiveness:* prediction and improvement of resource use, e.g. number of missed appointments, double bookings wait times *Organisational: user acceptance, implementation requirements* Ethical: inclusion, diversity, equity and accessibility *Patient &Social: patient perspective* *Cost and economic*	*Accuracy: predictive accuracy of mortality:* sensitivity, specificity *Clinical Effectiveness:* improvement of palliative care planning and delivery, generalisability *Ethical:* equity *Organisational:* clinician attitude and response	*Accuracy:* Dice coefficient, H, dosimetric analysis and qualitative measures *Clinical Effectiveness:* Acceptability of contours, alignment with guidelines, *Ethical:* equality considerations (algorithmic bias) *Patient & Social:* user experience and satisfaction

Author, year	CADTH, 2024 Horizon scan [87]	CADTH, 2023 Horizon scan [88]	NICE, 2023 Full Guidance [86]
Overall methodology	Horizon Scan	Horizon Scan	Early value assessment/Rapid review: Literature Search and multi-stakeholder discussions Evidence for this early value assessment was considered from several sources, including a previous early value assessment by the external assessment group. 15 studies were prioritised for analysis, and then a committee of experts deliberated the findings
Assessment framework	NR	NR	NR
Evidence requirements	NR	NR	NR
AI-specific checklist	No	No	No
Tools for Risk of Bias Assessment	No	No	*Strengths and limitations discussed but not formally assessed with a tool*
Author's conclusions	The evidence on the effectiveness of the technology on patient flow to change clinical outcomes and patient experience is unclear There is some evidence for effectiveness in forecasting the volume of patients as well as for improving workflow and efficiency No evidence on cost-effectiveness was found,	There is limited evidence for predictive performance and generalisability due to lack of external validation The evidence on patient and user acceptance and equity considerations is limited and inconclusive. No evidence on cost-effectiveness was found.	Nine technologies can be used more evidence is generated but must be used with an HCP review of the generated contours. There was strong evidence for the potential usefulness Evidence indicates that AI contouring performs similarly to the comparators but may have difficulties with specific anatomic sites, atypical anatomy or difficult positions. Evidence suggests time saving compared to manual contouring and a cost-consequence analysis suggests a potential cost saving, depending on individual technology costs.

Abbreviations: AI … artificial intelligence, CAD … coronary artery disease, CTCA … computed tomography coronary angiogram, CCT … controlled clinical trial, CRD … Centre for Research and Dissemination, DM … diabetes mellitus, HCP … healthcare professional, HRQoL … Health Related Quality of Life, HOMR … hospital one year mortality risk, NR … not recorded, PRISMA … Preferred Reporting Items for Systematic Reviews, QUADAS … Quality Assessment of Diagnostic Accuracy studies, RCT … randomised control trial, Rx … radiology, TBC … to be confirmed, XR … Xray

Extraction tables methods

Table A-10: Overview of HTA Methodologies

HTA Methodologies			
Institution, Year	**Tool (handbook, checklist, e.g.)**	**General Description**	**AI specific**
NICE, 2023 [3]	ESF – Evidence standards framework for digital health technologies	Purpose: outline criteria for the evidence required to demonstrate value of a DHT, including machine learning algorithms. Users: evaluators, innovation teams, DHT companies Structure: ■ 4 Sections (A: description of technologies suitable for evaluation using the ESF; B: classification of DHTs; C: overview of evidence standards tables; D: early deployment standards for evidence-generation programs) ■ 5 Groups: (design factors, describing value, demonstrating performance, delivering value, deployment considerations) ■ 3 Tiers: (A: DHTs without direct patient, health or care outcomes; B: communication about health care; C: treating and diagnosing medical conditions) ■ 21 Standards: ▪ The DHT should comply with relevant safety and quality standards ▪ Incorporate intended user group acceptability in the design of the DHT ▪ Consider environmental sustainability ▪ Consider health and care inequalities and bias mitigation ▪ Embed good data practices in the design of the DHT ▪ Define the level of professional oversight ▪ Show processes for creating reliable health information ▪ Show that the DHT is credible with UK professionals ▪ Provide safeguarding assurances for DHTs where users are considered to be in vulnerable groups, or where peer-topeer interaction is enabled ▪ Describe the intended purpose and target population ▪ Describe the current pathway or system process ▪ Describe the proposed pathway or system process using the DHT ▪ Describe the expected health, cost and resource impacts compared with standard or current care or system processes ▪ Provide evidence of the DHT's effectiveness to support its claimed benefits ▪ Show real-world evidence that the claimed benefits can be realised in practice ▪ The company and evaluator should agree a plan for measuring usage and changes in the DHT's performance over time	For data-driven DHTs (including those with artificial intelligence with fixed or adaptive machine learning algorithms) certain standards in the groups design factors, demonstrating performance and deployment considerations have been marked as more relevant: Design factors: Standard 4,5, and 6 ■ Consider health and care inequalities and bias mitigation ▪ For data-driven DHTs (including those with artificial intelligence), the company should describe any actions taken in the design of the DHT to mitigate against algorithmic bias that could lead to unequal impacts between different groups of service users or people. ▪ The ESF recommends using the Open Data Institute's Data Ethics Canvas to manage ethical issues in data projects. ■ Embed good data practices in the design of DHT ▪ For data driven DHTs, information should be available on source and size of training and validation data, the process of establishing 'ground truth', data collection methods, information if synthetic data was used, diversity of the training and validation data and if it is representative of the intended target population ▪ For DHTs that incorporate machine learning, companies should follow the MHRA guiding principles on good machine learning practice for medical device development (REF -see Literature Methods). ■ Define the level of professional oversight ▪ Expert review of output, monitoring the trend of the outputs to ensure alignment with, and calibration for, best clinical practice, monitoring if output was overridden by professionals Demonstrating performance: Standards 15 and 16 ■ Show real-world evidence that the claimed benefits can be realised in practice Current service provision or best practice, information from pilot sites on acceptability, performance, successful integration without unintended negative consequences for users/services and demonstration of improvement in outcomes (clinical/cost effectiveness)

HTA Methodologies			
Institution, Year	**Tool (handbook, checklist, e.g.)**	**General Description**	**AI specific**
NICE, 2023 [3] *(continuation)*		▪ Provide a budget impact analysis ▪ For DHTs with higher financial risk, provide a cost-effectiveness analysis ▪ Ensure transparency about requirements for deployment ▪ Describe strategies for communication, consent and training processes to allow the DHT to be understood by end users ▪ Ensure appropriate scalability ■ Tool to assess if a digital health technology (DHT) is relevant to the health and social system, what tier(s) of DHT it falls into and what evidence requirements relate to the highest relevant tier category. ▪ The evidence standards are presented in groups related to phases of a DHT life cycle: ▪ Design factors, describing value, demonstrating performance, delivering value and deployment considerations ▪ For each group, evidence standards are described that apply some or all DHT tiers	▪ he performance of technologies may be affected by local deployment factors, it is highlighted that the technology may run offline or "in silent mode" for an evaluation, because this mode allows to evaluate the performance of the DHT using local data inputs, before its implementation into clinical or care pathways. ■ Companies or evaluators should agree on a plan for measuring changes in the performance over time (in DHTs incorporating AI or machine-learning algorithms ▪ Plans on updates to algorithms/versions, sources and quality control of retraining data, processes for performance monitoring (detecting impacts on performance), detect changes in performance, overview process for reviewing changes in performance, and a plan on how and when and where changes should be reported. ▪ Reevaluation is needed if new functions are introduced that modify its intended use and ESF classification Deployment consideration: Standards 19 and 20 ■ Ensure transparency about requirements for deployment ▪ Information on input data, level of tolerance for incomplete data, data requirements (formats, standardisation, completeness, quality), minimum infrastructure for deployment ■ Describe strategies for communication, consent and training processes to allow the DHT to be understood by end users Strategies to communicate to service users/HCP outputs and interpretation, benefits and limitations of the technology, and to provide training.
HTA Wales, 2024 [51]	Health Technology Assessments of Artificial Intelligence Checklist [21]	Purpose: provide a structured approach for assessing AI technologies, ensuring comprehensive evidence collection. HTW highlights that HTA assessments for topics including AI should follow the same processes as usual, using the same questions, but incorporate the checklist with four domains. Users: HTA practitioners Structure: ■ general assistance ■ AI specific considerations and search terms ■ Checklist – 4 domains (Training, Clinical Setting and Use, Outputs, Ongoing Support)	Domain 1: Training ■ Model used ■ Training dataset/representativeness of local population ■ Reporting of outcomes (sensitivity, specificity, NPV, PPV) ■ Comparator reflects standard practice/best care ■ Domain 2: Clinical Setting and use ■ Position in the clinical care pathway, extra step or seamless integration ■ Ideal patient groups/clinical setting ■ Acceptability ■ Training requirements

[21] Currently still in development

HTA Methodologies			
Institution, Year	**Tool (handbook, checklist, e.g.)**	**General Description**	**AI specific**
HTA Wales, 2024 [51] *(continuation)*			■ Data requirements ■ Equality considerations Domain 3: Outputs ■ What information does the user of the technology get, is it real time and how should it be used? ■ Clinical benefit Domain 4: Ongoing support ■ Pricing model ■ Provision/process for ongoing monitoring, support and updates/retraining ■ Data collection and use- privacy considerations ■ Processes for monitoring and evaluating outputs ■ Additional technology or support requirement
HAS, 2019 [4, 49]	Medical device evaluation by the CNEDiMTS: Guide to the specific features of clinical evaluation of a connected medical device in view of its reimbursement	Purpose: To assist companies manufacturing or operating CMDs in incorporating appropriate clinical trials into their development strategy and to outline the evidence requirements considered in Health Technology Assessment for CMDs User: The guide is designed for manufacturers, distributors, and service providers, who intend to submit application dossiers for inclusion in the list of products and services qualifying for reimbursement Structure: generic description., descriptive grid	For medical devices embedding decision-making systems based on machine learning processes, it is required to provide a description of the functions built or subject to change using these technologies. Therefore, the section "descriptive grid" needs to be followed: (detailed questions, see Appendix) 1. Optimised clinical development focusing on the following outcomes: ■ The first challenge for the company in question is to create a clinical development programme that is compatible with the CMD's intended ultimate purpose. For all CMDs for individual use, the evaluation of their impact in terms of clinical benefit, acceptability or improvement of quality of life for users s necessary. Other impacts can also be looked for, especially in terms of access to treatment, standard of care and organization of care ■ The evaluation must in theory cover the technological solution as a whole, that is to say all elements collecting, processing and transmitting information from a remote site, taking treatment organization into account. In some cases, especially where certain components are self-operating, evaluation of the effect specific to the CMD can be a challenge for developers. 2. Prerequisites independent of any evaluation by CNEDiMTS ■ Observance of requirements in terms of processing and hosting of data covered by applicable legislation, especially the GDPR ■ Being granted CE marking, which aims to ensure general safety and performance requirements are met during the device's life cycle. ■ Elements set up by the company for ensuring the quality of the results is managed throughout the period of availability of the CMD to the patient. ■ Observance of data processing and hosting legal requirements, especially general data protection regulation

HTA Methodologies			
Institution, Year	**Tool (handbook, checklist, e.g.)**	**General Description**	**AI specific**
HAS, 2019 [4, 49] *(continuation)*			■ CE marking ■ Quality assurance process in place 3. Detailed information on the development of the algorithm[22] ■ In the event of automatic data processing, the CNEDiMET ■ Choice and selection of variables, model selection, learning mechanism, training data ■ Relevance of algorithm, regular verification, absence of bias 4. Real life data collection ■ to monitor performance over time to confirm medical benefit of an evolving technology with post registration studies In addition HAS has a document comprising 42 questions on data, purpose, model and functional characteristics. These questions are provided in the supplementary material.
FinCCHTA [24]	Digi-HTA	Purpose: The Digi-HTA is a method developed to assess the suitability of digital products and services for use in social and health care and well-being sectors. It evaluates various aspects of digital solutions to support decision-making in healthcare technology adoption. User: representatives of wellbeing services counties (for procurement decisions, market survey, and piloting), technology companies (to demonstrate product suitability and gain expert evaluation for product development), healthcare organisations (for assessing new technologies before implementation) Structure: 11 domains and subquestions The framework considers the following domains: 1. Company information 2. Product information 3. Technical stability 4. Usability and accessibility 5. Interoperability 6. Artificial Intelligence 7. Robotics	AI Domain: 1. Exactly what defined problem is going to be solved by the AI? 2. What is the classification of AI? Visualization only, AI–assisted (e.g., diagnosis/classification/decision), or solely AI–controlled? 3. Could the problem be solved without the AI solution? 4. Is the solution based on machine learning or a neural network? 5. Do the staff have sufficient capacity to understand the operational logic of AI (e.g., do they need additional training)? 6. Are the conclusions and decisions of the AI solution transparent, i.e., can medical staff understand what the decisions are based on? 7. Is the AI solution validated in the environment in which it will be used? 8. What are the data sources for the AI solution? 9. Are the data sources used in the training of AI solutions relevant to a final use case (e.g. are the age and gender composition of training groups comparable to that of real user groups)? 10. Are the access rights required for the use of the data in order, and have data protection (e.g., GDPR) and security issues been taken into account? 11. When it comes to classifier teaching, are there enough data relative to the size of the smallest class? 12. Can the AI solution use incomplete data?

[22] For automated data processing, de CNEDiMTS is not responsible for evaluation the mathematical functioning of the model

HTA Methodologies			
Institution, Year	**Tool (handbook, checklist, e.g.)**	**General Description**	**AI specific**
FinCCHTA [24] *(continuation)*			13. Can the AI solution use noisy data? 14. Is retraining possible for the AI solution? 15. What are the data sources for retraining? 16. How is it ensured that the system is not taught with irrelevant data? 17. How many tests or results are needed for the AI model? 18. Is the algorithm purchased software as a service (SaaS) or its own design? 19. What performance criteria are used? 20. Does the AI solution change care processes? How? 21. When does the AI solution propose an action? How, and who will actually implement it? 22. Is staff's approval needed for action proposed by the AI?
AQUAS, 2024 [50]	Adaptation of HTA assessment framework for Digital Health Technology Assessment	A methodological framework was developed utilising literature review, thematic analysis, consensus workshops and the adaptation of the updated NICE ESF for DHTs and resulted in the description of 13 domains, 41 dimensions, 9 subdimensions and 21 levels of evidence required. Dimensions and subdimensions especially relevant for the assessment process of AI-based technologies are as follows: Domain 8: Ethical aspects: ■ Control, user autonomy, accountability ■ Responsibility ■ Transparency, explainability and interpretability Standard Domain 9: Legal and regulatory aspects: ■ Privacy ■ Transparency Domain 11: Technical aspects: ■ Adaptability (Interoperability, Scalability, Data integration, Transferability) ■ Technical effectiveness (Reliability, Accuracy, Validity, Sensitivity) ■ Generalisability and reproducibility ■ Interpretability and explainability	■ Description of the health problem: ■ Description of the technology (adoption, use, integration) ■ Content (adequacy of the information) ■ Safety (clinical safety) ■ Clinical efficacy and effectiveness ■ Economic aspects (costs, use of resources) ■ Human and sociocultural aspects (acceptability) ■ Ethical aspects ■ Legal and regulatory aspects (privacy, transparency) ■ Technical aspects (scalability, technical effectiveness and performance, post deployment monitoring, generalizability and reproducibility, interoperability) ■ Environmental aspects

Abbreviations: AI … artificial intelligence, AIPA … artificial intelligence prediction algorithm, AQUAS … Agency for Quality and Assessment of Catalonia, AUC … area under the curve, CAIR … clinical artificial intelligence research, CNEDiMTS … National committee for the evaluation of medical devices, CNN … convolutional neural network, DHT … digital health technology, DTAC … digital technology assessment criteria ESF … evidence standard framework , FDA … Food and Drug Administration, HAS … Haute Autorite de Sante, HIS … Health Innovation Scotland, HTA … health technology assessment, LLM … large language model, MI … CLAIM-minimum information about clinical artificial intelligence modeling, MDR … medical drug regulation, MDD … Medical device directive, MHRA … Medicines and Healthcare products Regulatory Agency, ML … machine learning, NHS … national health service, NNT … number to treat, NPV … negative predictive value, PPV … positive predictive value, SHTG … Scottish Health Technologies Group

Overview Domains and questions to Themes

Table A-11: Overview Domains and questions to Themes

ESF Standard [3]	Themes	AQuAS Domain [50]	Themes	HAS Question [4, 49]	Themes	FinCCHTA Question [24]	Themes	HTW Domain [51]	Themes
4	Algorithmic bias	1	Health problem	1	Claimed Use	1	Health Problem	1.1	AI Model identification
5	Training data	2	Technology	2	Specifics on benefit	2	Info AI function	1.2	Training data
5	Target population	2.1	Adoption	3	Target population	3	AI relevance	1.3	Data representativeness
6	Human oversight	2.1.1	Use	4	Setting	4	AI model	1.4	Training data
6	Monitoring	2.1.2	Integration	5	Training data (representativeness)	5	Staff training	1.5	Target population
15	Acceptability	3	Content	6	Training data quality (representativeness)	6	Transparency	1.6	Comparator
15	Performance	3.1	Adequacy of information	7	Data organisation	7	Setting	1.7	Human oversight in use
15	Pathway	4	Safety	8	Data organisation	8	Data sources	1.8	Performance
15	Evaluation	4.1	Clinical safety	9	Input data origin	9	Data sources	1.9	Performance
16	Post deployment monitoring	5	Efficacy and effectiveness	10	Input data pre-processing	10	Accessibility	2.1	Clinical Pathway
19	Training data	6	Economic aspects	11	Missing data	11	Data representativeness	2.2	Target population
19	Incomplete data	6.1	Costs	12	Training data (Outliers)	12	Incomplete data	2.3	Support
19	Data requirements	6.2	Use of resources	13	Data representativeness	13	Noisy data	2.4	Target Population
20	User training	7	Human and social aspects	14	Data organisation	14	Retraining	2.5	Setting
		7.1	Acceptability	15	Input data origin	15	Retraining	2.6	User acceptance
		8	Ethical aspects	16	Data pre-processing	16	Training data (relevant data)	2.7	Training requirement
		9	Legal and regulatory	17	Information on output data	17	Training data (how much data)	2.8	Patient impact
		9.1	Privacy	18	AI model	18	AI model (software)	2.9	Input data quality
		9.2	Transparency	19	Function	19	Performance	2.10	Performance
		10	Organisational	20	Update strategies	20	Impact on current use	3.1	Clarity of output data
		11	Technical aspects	21	Explainability	21	Usability	3.2	Usability
		11.1	Scalability	22	Training data	22	Human oversight	3.3	Real time feedback
		11.2	Technical effectiveness	23	Update strategies			3.4	Clinical pathway fit
		11.3	Post deployment monitoring	24	Human oversight in development			3.5	Cliinical benefit

ESF Standard [3]	Themes	AQuAS Domain [50]	Themes	HAS Question [4, 49]	Themes	FinCCHTA Question [24]	Themes	HTW Domain [51]	Themes
		11.4	Generalisability	25	Human oversight in retraining			3.6	Uncertainties
		11.5	Interoperability	26	Performance measurement			4.1	AI updates
		12	Environmental	27	Changes in performance			4.2	Pricing model
				28	Risks in data training			4.3	Support clarity
				29	Transparency			4.4	Support costs
				30	Transparency (Performance)			4.5	Monitoring
				31	Performance validation			4.6	Data collection
				32	Performance thresholds			4.7	Data output handling
				33	Performance evaluation			4.8	Data output handling
				34	Monitor performance			4.9	Output evaluation
				35	Performance thresholds			4.10	Additional requirements
				36	Changes in performance				
				37	Input data anomaly detection				
				38	Impact of Input data anomaly				
				39	Error measurement (algorithm or user)				
				40	Explainability				
				41	Interpretability				
				42	Quality guarantee				
				HAS AREA					
				1	Acceptance				
				1	Clinical benefit				
				1	Accessibility				
				1	Training data				
				2	Data requirement				
				2	Safety guarantee				

Abbreviations: AQuAS … Agència de Qualitat i Avaluació Sanitàries de Catalunya, FINNCHTA … Finnish Coordinating Center for Health Technology Assessment, HAS, HTW … Health Technology Wales, NICE … National Institute for Health and Care Excellence

Note: Grey – questions/domains/standards from the respective document. Blue – topics not addressed within the EUnetHTA Core Model. All questions/domains were allocated to topics. These topics were charted against the EUnetHTA Core Model and highlighted in blue, of not addressed.

Thematic analysis of methods guidance documents

Table A-12: Thematic analysis of methods guidance documents

	Themes	NICE [3]	AQuAS [50]	HAS [4, 49]	FINCCHTA [24]	HTW [51]
CUR						
TEC	Information on AI function and model Training data quality Further aspects reltated to the deployed algorithm	Standard 5 (information on training data, validation data, data collection methods, synthetic data, diversity and representativeness) Standard 19 (information on training data, incomplete data and data requirements, transparency)	D9.2: Transparency	Q5-17, 22: information on data (representativeness, organisation, data source, pre-processing, missing data, outliers, output data, training data) Q18: Information on AI model Q21,40,41: explainability and interpretability Q29, 30 Transparency A1: training data A2: data requirements	Q8,9,10,11,12,13,14,16,17: Information in training data, data representativeness incomplete and noisy data Q2,3,4,18 Information on AI function, relevance and model	D1.2;1.3,1.4;2.7;2.9: Information on training data, input data quality, training requirements, data representativeness D1.1 Information on AI model D3.1;4.7;4.8 clarity on data output, data output handling
SAF	Data risk management			Q28: Risks concerning training data Q39: Error measurement Q37,38: Input data anomaly detection and impact		
EFF						
ECO						D4.4: Support costs
ETH **SOC** **LEG** **ORG**	Human Oversight Algorithmic Bias Privacy	ETH Standard 4 (mitigate against algorithmic bias; Open Data Institute's Data Ethics Canvas to manage ethical issues)	LEG D9.1: privacy	ORG Q24,25,34: human oversight (in development, use and retraining	ORG: Q22: human oversight	ORG: D1.7: human oversight
Other aspects	Monitoring Re-evaluation Updates	Standard 6 (monitoring of the output) Standard 16 (measuring changes in performance → reevaluation and performance monitoring/post-deployment monitoring) Standard 15 Performance evaluation	D11.3: post-deployment monitoring	Q20,23: Update strategies Q34: performance monitoring Q33: Performance evaluation Q27, 36: Changes in performance Q32,35: performance thresholds	Q12,13: Information on retraining Q20: changes in performance Q14,15: retraining	D4.1 Updates D4.3: support clarity D4.5 Monitoring D4.9 Output evaluation

Abbreviations: A … area, AQUAS … Agència de Qualitat i Avaluació Sanitàries de Catalunya, CUR … current, D … domain, ECO … economic, EFF … effectiveness, ETH … ethical, FINNCHTA … Finnish Coordinating Center for Health Technology Assessment, HAS, HTW … Health Technology Wales, LEG … legal, NICE … National Institute for Health and Care Excellence, ORG … organisational, Q … question, SOC … social, TEC … technical

Note: These is an aggregated version of topics, allocated in Table A-11. These topics were further summarised into 1 to 3 aggregated topics.

Search Strategy

PubMed

#		Results 25 Apr 2024
1	"telemedicine"[MeSH Terms] OR "digital technology"[MeSH Terms] OR "mobile applications"[MeSH Terms] OR "monitoring, ambulatory" [MeSH Terms] OR "digital health"[tiab] OR "digital therapeutic"[tiab] OR "digital health application*"[tiab] OR "DiHA"[tiab] OR "mobile health" [tiab] OR "mHealth"[tiab] OR "telehealth"[tiab] OR "telecare"[tiab] OR "web based intervention*"[tiab] OR "internet based intervention*"[tiab] OR "artificial intelligence"[tiab] OR "medical artificial intelligence"[tiab] OR "medical AI"[tiab]	160,953
2	(technology assessment, biomedical [MeSH Terms] OR "evaluat*"[tiab] OR "apprais*"[tiab] OR "appraisal"[tiab] OR "health technology assessment"[tiab] OR HTA [tiab] OR "technology assessment*"[tiab] OR "technology evaluation*"[tiab])	4,706,876
3	("framework*"[tiab] OR "guideline*" [tiab] OR "guidance*" [tiab])	1,081,521
4	("health" [MeSH Terms] OR "medicine" [MeSH Terms] OR "therapeutics" [MeSH Terms])	6,549,808
((1 AND (2 AND 3)) AND 4) and 2015-2024		**1,976**

Embase

No.	Query	Results	Results Date
#35.	#33 NOT #34	2,931	25 Apr 2024
#34.	#33 AND 'conference abstract'/it	1,283	25 Apr 2024
#33.	#32 AND [2015-2024]/py	4,214	25 Apr 2024
#32.	#17 AND #23 AND #27 AND #31	4,872	25 Apr 2024
#31.	#28 OR #29 OR #30	14,623,066	25 Apr 2024
#30.	'therapy'/exp	11,118,470	25 Apr 2024
#29.	'medicine'/exp	4,344,613	25 Apr 2024
#28.	'health'/exp	945,962	25 Apr 2024
#27.	#24 OR #25 OR #26	1,441,475	25 Apr 2024
#26.	'guidance*':ti,ab	243,995	25 Apr 2024
#25.	'guideline*':ti,ab	783,688	25 Apr 2024
#24.	'framework*':ti,ab	473,657	25 Apr 2024
#23.	#18 OR #19 OR #20 OR #21 OR #22	6,580,276	25 Apr 2024
#22.	'technology assessment*':ti,ab	12,407	25 Apr 2024
#21.	hta:ti,ab	9,870	25 Apr 2024
#20.	'apprais*':ti,ab	91,878	25 Apr 2024
#19.	'evaluat*':ti,ab	6,498,514	25 Apr 2024
#18.	'biomedical technology assessment'/exp	18,176	25 Apr 2024
#17.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16	184,752	25 Apr 2024
#16.	'medical ai':ti,ab	233	25 Apr 2024
#15.	'artificial intelligence':ti,ab	49,173	25 Apr 2024
#14.	'internet based intervention*':ti,ab	980	25 Apr 2024
#13.	'web based intervention*':ti,ab	1,671	25 Apr 2024
#12.	'telecare':ti,ab	1,024	25 Apr 2024
#11.	'telehealth':ti,ab	16,865	25 Apr 2024
#10.	'mhealth':ti,ab	6,990	25 Apr 2024
#9.	'mobile health':ti,ab	7,465	25 Apr 2024
#8.	diha:ti,ab	64	25 Apr 2024

#7.	'digital health application*':ti,ab	327	25 Apr 2024
#6.	'digital therapeutic':ti,ab	347	25 Apr 2024
#5.	'digital health':ti,ab	7,142	25 Apr 2024
#4.	'ambulatory monitoring'/exp	12,359	25 Apr 2024
#3.	'mobile application'/exp	28,033	25 Apr 2024
#2.	'digital technology'/exp	5,754	25 Apr 2024
#1.	'telemedicine'/exp	76,005	25 Apr 2024

HTA Austria
Austrian Institute for
Health Technology Assessment
GmbH