Chew and Ngiam *BMC Medicine* (2025) 23:244
https://doi.org/10.1186/s12916-025-04076-0

BMC Medicine

REVIEW

Open Access

Artificial intelligence tool development: what clinicians need to know?

Boon-How Chew[1,2*] and Kee Yuan Ngiam[1,3]

Abstract

Digital medicine and smart healthcare will not be realised without the cognizant participation of clinicians. Artificial intelligence (AI) today primarily involves computers or machines designed to simulate aspects of human intelligence using mathematically designed neural networks, although early AI systems relied on a variety of non-neural network techniques. With the increased complexity of the neural layers, deep machine learning (ML) can self-learn and augment many human tasks that require decision-making on the basis of multiple sources of data. Clinicians are important stakeholders in the use of AI and ML tools. The review questions are as follows: What is the typical process of AI tool development in the full cycle? What are the important concepts and technical aspects of each step? This review synthesises a targeted literature review and reports and summarises online structured materials to present a succinct explanation of the whole development process of AI tools. The development of AI tools in healthcare involves a series of cyclical processes: (1) identifying clinical problems suitable for AI solutions, (2) forming project teams or collaborating with experts, (3) organising and curating relevant data, (4) establishing robust physical and virtual infrastructure, and computer systems' architecture that support subsequent stages, (5) exploring AI neural networks on open access platforms before making a new decision, (6) validating AI/ML models, (7) registration, (8) clinical deployment and continuous performance monitoring and (9) improving the AI ecosystem ensures its adaptability to evolving clinical needs. A sound understanding of this would help clinicians appreciate the development of AI tools and engage in codesigning, evaluating and monitoring the tools. This would facilitate broader use and closer regulation of AI/ML tools in healthcare settings.

Keywords Artificial intelligence, Machine learning, Development, Deployment, Infrastructure, Integration clinical workflow

*Correspondence:
Boon-How Chew
chewboonhow@upm.edu.my
[1] Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore C/O NUHS Tower Block, Level 8, 1E Kent Ridge Road, Singapore 119228, Singapore
[2] Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia
[3] Department of Surgery, Division of General Surgery (Thyroid and Endocrine Surgery), National University of Singapore, University Surgical Cluster, National University Hospital National University Health System Corporate Office, Singapore, Singapore

Background

The transformation and realisation of digital medicine [1] and smart healthcare [2] hinge upon the active and cognizant participation of clinicians in the entire development process and cycle [3, 4]. Clinicians educated in these aspects could provide invaluable insights during the design stage of digital health technologies, including artificial intelligence (AI)-enabled tools and systems [5]. This could ensure that these solutions meet their unique needs and preferences, emphasise patient safety and quality of care and facilitate seamless integration into clinical workflows [3]. Additionally, their endorsement and support foster adoption and acceptance among other healthcare providers, drive innovation and ultimately optimise clinical outcomes [6, 7].

AI is the cognitive ability of machines made possible by mathematically designed neural networks (see the Glossary in the Additional File 1). The electronic neural networks are built to mimic the human neuronal plexus and are programmed to manage a myriad of data according to their categories and to assign different factors of data different weights. The weights are decided from given data on specified outcomes, and this process is continuously being improved with ongoing receival of data. When vast amounts of data are interconnected, it enables new discoveries and creates opportunities that can transform personal experiences and advance science in nearly all aspects of human life [8]. For example, interconnected health data can lead to early detection of diseases through predictive analytics, while personalised education platforms use linked learning data to tailor teaching methods to individual needs. In science, combining datasets across disciplines can uncover patterns such as predicting climate change impacts or accelerating drug development through AI-driven simulations. The learning computer models are machine learning (ML) versions of AI, and deep learning (DL) is a version with multiple layers of neural networks. The predictive ability of such models is evaluated against the annotated or labelled outcome via weights applied to each variable when they are included in the models. The models are self-learning, improving their performance through repeated adjustments (iterations) using a method called backpropagation which optimises the model by minimising errors via a process known as stochastic gradient descent. This process in gauging the best weights for variables in the model as they progress through their different levels of complexity at the different neural layers. Other ML models include natural language processing and computer vision. Transformers are present when DL models differentially weigh the importance of each part of the input data and make natural-language processing possible (ChatGPT stands for the Chat Generative Pretrained Transformer). Augmented intelligence (AugI) is an AI that supplements and enhances humans' ability instead of substituting it. There are FDA-approved software, applications, programmes and devices that use AI to interpret a broad range of imaging modalities and diagnostic and prognostic assistance and help outline possible treatments for clinicians [9, 10]. It is important to discern between programmed computer systems and applications that mimic AI tools and systems but are not considered true AI. Table 1 shows examples of similar tools in these two categories in the healthcare industry.

The integration of AI into routine clinical care is accelerating, which is now reviewing patient histories, drafting physician notes, offering patient instructions, and not just reading X-rays and histopathological images [11–14]. The appropriate use of AI technology in healthcare defined as ethical, clinically validated and seamlessly integrated applications that enhance patient care

Table 1 A comparison of programmed computer systems and similar AI tools or systems

No	Programmed computer systems	Similar AI tools or systems
1.	**Rule-based Clinical Decision Support Systems** • Drug–drug interaction checkers • Clinical guidelines adherence alerts	**ML-based Clinical Decision Support Systems** • Deep learning models for predicting adverse drug reactions • Reinforcement learning models for personalised treatment recommendations
2.	**Heuristic Diagnostic Systems** • Symptom checkers in telemedicine platforms • Triage systems in emergency departments	**ML-based Diagnostic Systems** • Deep learning models for medical image interpretation (e.g., radiology, pathology) • Natural language processing (NLP) models for clinical note analysis
3.	**Chatbots for Healthcare Consultation** • Chatbots for scheduling appointments • Symptom assessment chatbots for initial patient triage	**NLP-based Healthcare Chatbots** • Chatbots using transformers for conversational AI • Chatbots with sentiment analysis for understanding patient emotions
4.	**Pattern Recognition Systems** • ECG interpretation software • Radiology image analysis tools for detecting abnormalities	**ML-based Pattern Recognition Systems** • Deep learning models for heart disease diagnosis • Convolutional neural networks (CNNs) for medical image classification
5.	**Clinical Documentation Templates** • Electronic health record (EHR) templates for progress notes • Surgical procedure documentation templates	**Natural Language Generation (NLG)-based Clinical Documentation Systems** • NLG models for automatically generating clinical notes • NLG templates for surgical procedure documentation
6.	**Prescription Order Entry Systems** • Computerised physician order entry systems • Prescription order templates in EHRs	**NLP-based Prescription Order Entry Systems** • NLP models for extracting medication information from unstructured text • NLP models for detecting potential medication errors
7.	**Diagnostic Coding Assistance Tools** • ICD- 10 code suggestion tools • CPT code lookup databases	**ML-based Coding Assistance Systems** • ML models for automated diagnostic coding

The content of the table was adapted from that given by ChatGPT3.5

and efficiency which is also potentially cost-effective [15]. This is when improved quality care by reducing variation, being safe and expedient [16, 17], transforming reactive healthcare to a more proactive approach, and focusing on health promotion, disease prevention and health management rather than disease treatment, resulting in fewer hospitalisations, fewer doctor visits and fewer treatments [18]. AI applications are projected to reduce annual healthcare expenditures in the USA by USD 150 billion by 2026, primarily through increased efficiency, improved diagnostics, and optimised care delivery [18]. However, all this AI advancement is not without great challenges from development to deployment [19–21], integration in clinical workflows [22, 23] and influences on doctor–patient consultation [24, 25]. Persistent concerns about integrating AI systems into existing clinical consultations include alert fatigue [26, 27], data quality, data security, transparency and accountability, alignment with standards and guidelines and unintended consequences along with model design requirements, and retention of clinician autonomy [28]. The human factors and AI systems that may affect medical professionals' interactions with technology could be related to perceptions of training data quality, performance of AI systems, explainability, adaptability, medical expertise (young versus experienced clinicians), technological expertise, personality, cognitive biases (proper understanding and use of AI outputs) and trust in the whole ecosystem [29]. Table 2 shows the real present challenges of AI technology in healthcare and its more certain progress in the near future. Globally, the AI healthcare market was valued at USD 20.9 billion in 2024 and is anticipated to grow at a compound annual growth rate of 48.1% reaching an estimated USD 148.4 billion by 2029. This growth reflects expanding investments in AI-driven technologies and services across the healthcare sector [30].

This review explains the usual path for new AI tool development and deployment in healthcare and clinical services. It concurs with other evaluation frameworks [31, 32] (Table 3) and could be extended to include assessments of health economic benefits [33]. When selecting an evaluation framework, users should consider the specific objectives of their study as some frameworks focus on quality evaluation, transparent reporting or risk of bias, while others address specific stages, designs or disciplines. Depending on the purpose, a single framework or a combination can guide study planning, implementation and reporting to ensure robust and impactful outcomes. However, articles with sufficient and clear technical explanations of the AI development process for clinicians are scarce [31, 32, 34]. Clinicians with a sound understanding of the whole development process of AI tools would help them engage in codesign, effective collaboration [35], evaluation and monitoring of the tools, and further facilitate broader use and closer regulation of these tools in healthcare settings [36].

Some reporting guidelines are study design specific (TRIPOD-AI for prognostic and diagnostic studies, STARD-AI for diagnostic test studies, SPIRIT/CONSORT and SPIRIT/CONSORT-AI for clinical trials), stage specific (DECIDE-AI for early clinical studies) or discipline specific (CHEERS-AI for health economy, IDEAL for surgery, and CLAIM and FUTURE-AI for radiology) [39].

How are AI tools and models developed according to clinicians' needs?

This focused integrative review attempts to update and delineate practical knowledge on AI tools or model development throughout the whole process for clinicians. The review questions are as follows: What is the typical process of AI tool development in the full cycle? What are the important concepts and technical aspects of each step? The approach includes a targeted literature review and synthesised summaries from online courses, including but not limited to the AI for healthcare by the National University of Singapore (https://nusmed.emeritus.org/ai-for-healthcare), the No Code AI and Machine Learning: Building Data Science Solutions by the Massachusetts Institute of Technology (https://professionalonline2.mit.edu/no-code-artificial-intelligence-machine-learning-program) and the European Information Technologies Certification Academy (EITCA) Artificial Intelligence Academy (https://eitca.org/certification/eitca-ai-artificial-intelligence-academy/). It strives to provide adequate technical knowledge that is immediately useful for clinicians to appreciate the development of AI tools and is able to engage

Table 2 The present challenges and future uses of AI technology in healthcare

Present challenge	Certain AI technology uses in future
1. Availability of high quality and quantity of data 2. Computing power that has sufficient storage and speed 3. Expertise and talents in data science, software engineering and domain experts under good data governance and policy 4. Robustly tested models or tools 5. Regulation of AI development	1. Mundane, repetitive and tasks require multifactual processing in administrative tasks 2. Images recognition AI models in radiology, pathology and ophthalmology 3. Natural language models that read free text and recognises speeches

Table 3 Evaluation frameworks of AI studies

No	Evaluation/reporting framework	Content
1.	APPRAISE-AI [32]	This evaluates quality of AI studies in the model development across 6 domains: clinical relevance, data quality, methodological conduct, robustness of results, reporting quality and reproducibility. These domains include 24 items with a maximum overall score of 100 points. Higher points indicating stronger methodological or reporting quality
2.	MI-CLAIM checklist [37]	Minimum information about clinical artificial intelligence modelling (MI-CLAIM) is a tool to improve transparent reporting of AI algorithms in medicine. It aims to enable a direct assessment of clinical impact including fairness and bias, and second, to allow rapid replication of the technical design process of any legitimate clinical AI study. The six parts are: (1) study design comprises clinical setting, performance measures, population composition and current baselines to measure performance against, (2) data partitions for model training and testing, (3) optimisation and final model selection, (4) performance evaluation to be reported at the model itself, and the model's clinical performance metrics, (5) model examination as a "sanity check," to uncover biases, to understand model behaviour, (6) reproducible pipeline by complete sharing of the code
3.	CODE-EHR checklist [38]	The CODE-EHR framework aims to improve the design and reporting of research studies using structured electronic health-care data. It requests for clarity on reporting and defines a set of minimum and preferred standards for the processes involved in (1) coding, dataset construction and linkage, (2) details and transparency of the preceding step, (3) disease and outcome definitions, (4) analysis, and (5) research governance which emphasises on patient and public engagement throughout the development process. Researchers are advised to use this checklist in the design phase of their study to ensure that important criteria for successful research and research impact are being used
4.	DECIDE-AI reporting guideline [39]	This comprises key items to be reported in early-stage clinical studies of AI-based decision support systems in healthcare to facilitate the appraisal of these studies and replicability of their findings. It has 17 AI-specific reporting items (with 28 subitems) and 10 generic reporting items with a paragraph for explanation for each of this item
5.	SPIRIT-AI reporting guideline [40]	The SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) extension is a reporting guideline for clinical trial protocols evaluating interventions with an AI component. It includes 15 new items in addition to the core SPIRIT 2013 of 33 items. SPIRIT-AI requires clear descriptions of the AI intervention including instructions and skills needed for use, the setting in which the AI intervention will be integrated, considerations for the handling of input and output data, the human-AI interaction and analysis of error cases
6.	CONSORT-AI extension [41]	This includes 14 new items in addition to the core CONSORT 2010 items. It recommends investigators to provide clear descriptions of the AI intervention, including instructions and skills required for use, the setting in which the AI intervention is integrated, the handling of inputs and outputs of the AI intervention, the human-AI interaction and providing analysis of error cases
7.	The TRIPOD [42] and TRIPOD-AI [43]	TRIPOD + AI contains a 27-item checklist that aims to promote the complete, accurate, and transparent reporting of studies that develop a prediction model or evaluate its performance. Complete reporting will facilitate study appraisal, model evaluation and model implementation. It assists in reporting research in which a multivariable prediction model is being developed (or updated), or validated (tested) using any (supervised) ML technique. The checklists are not a quality appraisal tool
8.	PROBAST [44] and PROBAST-AI [45, 46]	The Prediction model Risk Of Bias ASsessment Tool (PROBAST) comprises four domains (participants, predictors, outcome and analysis) and contains 20 signalling questions to facilitate risk of bias assessment of prediction model studies from the study design, conduct to analysis. PROBAST-AI comprises two components: model development and model evaluation. In model development, users assess quality and applicability with 16 targeted signalling questions, while model evaluation uses 18 targeted questions to assess risk of bias and applicability. Both components share four domains—participants and data sources, predictors, outcome and analysis—with the prediction model's applicability specifically rated in the first three domains
9.	STARD-AI [47]	Standards for Reporting of Diagnostic Accuracy Studies AI Extension (STARD-AI) is used to report diagnostic accuracy/test studies. STARD-AI is underdevelopment
10.	MINIMAR [48]	MINimum Information for Medical AI Reporting (MINIMAR) guides on the minimum information necessary to understand intended predictions, target populations, and hidden biases, and the ability to generalise these emerging technologies in four sections: (1) information on the population providing the training data, (2) training data demographics, (3) detailed information about the model architecture and development and (4) model evaluation, optimisation and validation to clarify how local model optimisation can be achieved and enable replication and resource sharing
11.	CLAIM [49]	Checklist for Artificial Intelligence in Medical Imaging (CLAIM) is modelled after the STARD guideline and has been extended to address applications of AI in medical imaging that include classification, image reconstruction, text analysis, and workflow optimisation to guide complete reporting of research
12.	CHEERS-AI [50, 51]	Consolidated Health Economic Evaluation Reporting Standards-AI assist in describing health economic evaluations to estimate the value for money (cost effectiveness) of AI interventions

Table 3 (continued)

No	Evaluation/reporting framework	Content
13.	IDEAL checklists [52, 53]	The Innovation, Development, Exploration, Assessment, and Long-term (IDEAL) Framework describes the five stages through which surgical therapy innovation normally passes. Each IDEAL stage is defined by key research questions These are intended to provide a minimum list of concepts authors should include in a report of surgical and device innovation. It can also be used both prospectively to help plan a study and retrospectively to assist in appraisal. IDEAL-D Framework for Device Innovation is a consensus statement on the preclinical stage of development (Stage 0) [54]
14.	FUTURE-AI checklist [55]	The guiding principles of FUTURE-AI are 1) Fairness, 2) Universality, 3) Traceability, 4) Usability, 5) Robustness and 6) Explainability. They aim to guide developers, evaluators and other stakeholders in delivering medical AI tools in health imaging that are trustworthy and optimised for real-world practice
15.	OPTICA [56]	Organisational PerspecTIve Checklist for AI solutions adoption (OPTICA) is a comprehensive and practical checklist tool to assess an adoption of AI solutions in health care organisations. It was developed through a consensus process involving multiple subject-matter domain experts and decision-makers across the authors' organisation. It comprises 13 chapters, each containing 3 to 12 checklist items, totalled 77. No scoring but checklist items that require a qualitative, case-specific evaluation process
16.	ALTAI [21, 57]	Assessment List for Trustworthy AI (ALTAI) provided by the European Commission's High-Level Expert group for Artificial Intelligence. It comprises seven requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, nondiscrimination and fairness, (6) societal and environmental well-being and (7) accountability, and 60 questions in total

with developers, vendors and researchers when considering clinical adoption and codesigning a new tool. ChatGPT 3.5, 4o and o1 (OpenAI, San Francisco, CA, USA) were used to assist in drafting and language editing of portions of this review. The authors have reviewed and edited the content produced by ChatGPT for accuracy and integrity, and accept full responsibility for the final version of the manuscript.

Overview of the AI development process

The development of AI in healthcare involves a series of cyclical processes (Fig. 1). It begins by identifying clinical problems suitable for AI solutions, forming project teams or collaborating with experts, and organising and curating relevant data. The establishment of robust infrastructure and architecture supports subsequent stages, including the exploration of AI neural networks on open access platforms and the validation of AI/ML models.

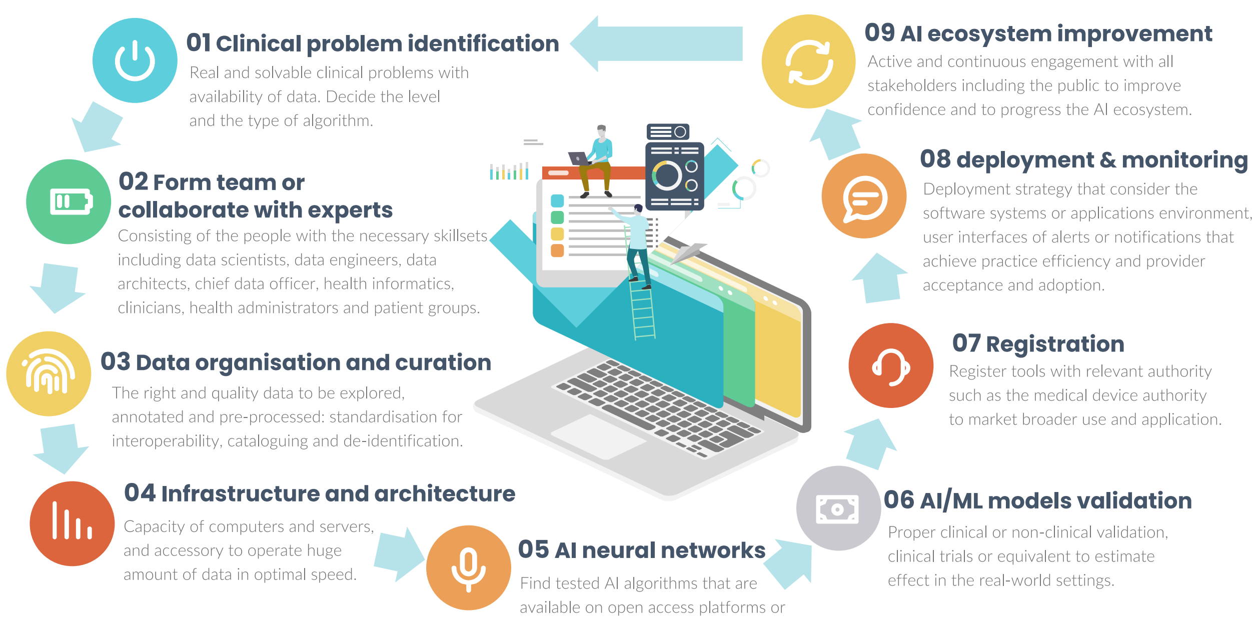

Fig. 1 AL/ML tool development process

Following registration procedures, clinical deployment and continuous performance monitoring occur. Finally, a commitment to improving the AI ecosystem ensures its adaptability to evolving clinical needs.

Clinical problem identification

This first step is the most important starting point for the rest of the development process (see Fig. 2). Some clinical and biomedical problems in healthcare services could be best resolved with the help of an automated solution. These are problems or challenges that are technically factual, mechanical, repetitive and complex in nature owing to the need to process multiple aspects of healthcare services, people in the health system or patient characteristics (see tips and examples in Table 4). DL/AugI/ML does not help address personal values, health beliefs or emotions that change until these constructs are measured in certain ways. Identifying the problem includes deciding on the level of the problem for the AI technology to solve. This method is descriptive, diagnostic, predictive or prescriptive and uses either assistive or autonomous AI algorithms (Table 4). Descriptive AI models are about estimating the quantity of a certain condition, diagnostic models are about the probability of occurrence of certain conditions, prognostic models are predicting certain outcomes, and prescriptive models suggest the most likely efficacious treatment. The order denotes an incremental level of value and complexity to be expected in the development of the tool. Be as specific and clearly defined as possible with all the variables, especially the outcome variable (supervised learning models).

Forming a project team or collaborating with experts

A successful team must consist of individuals with the right skillsets. This includes data scientists for data validation, transformation, curation, and visualisation for AI/ML models; data engineers to implement data workflows, such as storage; data architects to design the system architecture for data repositories; and a chief data officer to establish the data governance structure and policies. Clinicians, healthcare administrators and relevant stakeholders, including patients and public groups, are essential for the planning, development, deployment and sustained use of the AI/ML tool. Additionally, health informatics professionals and business or industry partners should be considered. To secure funding, the project

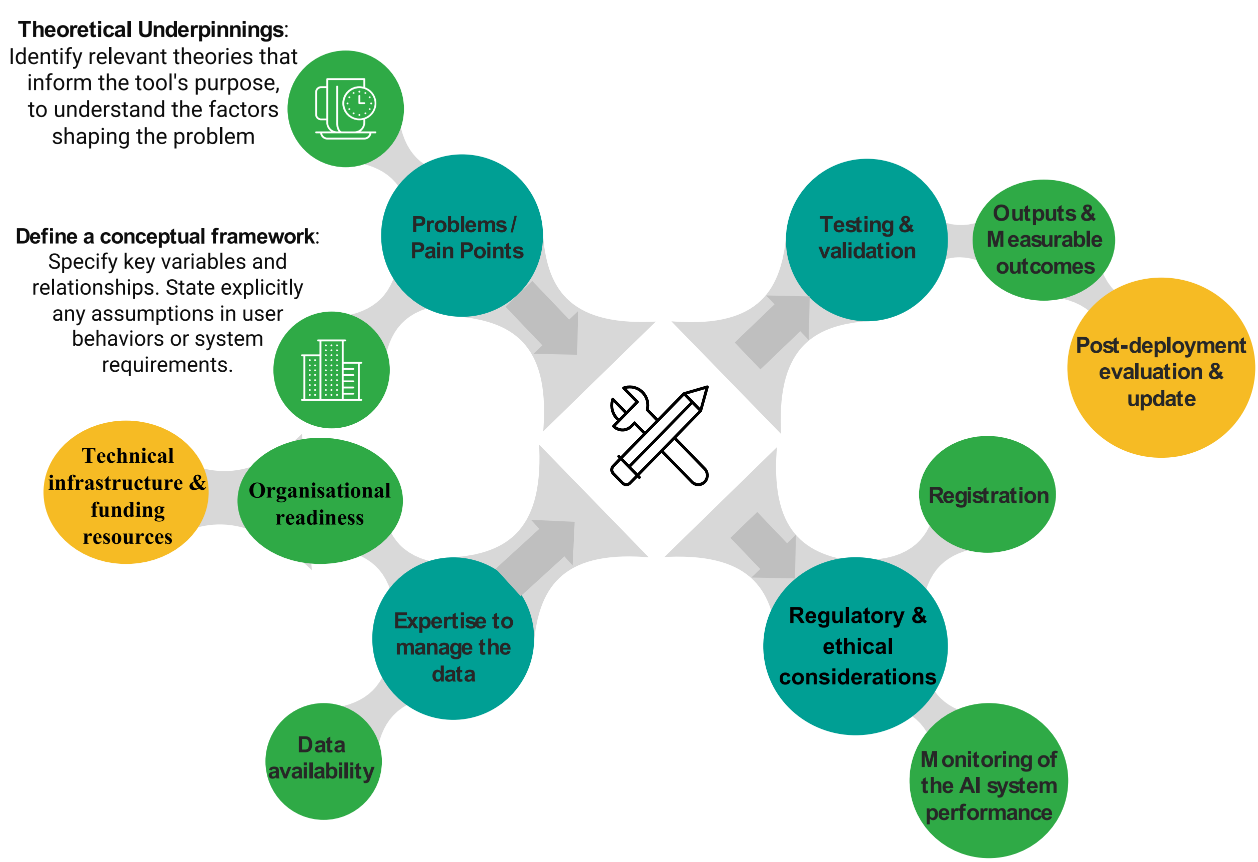

Fig. 2 The process of identifying AI tools for clinical problems

Table 4 Tips for identifying clinical problems for AI solutions

Tips in identifying a clinical problem for AI solutions	Examples
1. **Understand Healthcare Challenges** a. Can begin with your organisation's value statements, then ponder on the current challenges and pain points within healthcare services. These could include issues related to i. diagnosis accuracy, ii. treatment effectiveness, iii. patient outcomes, iv. workflow inefficiencies, v. resource allocation, or vi. healthcare accessibility b. May conduct medical audits to ascertain the relevant aspects or potential areas, or engage colleagues in services or those in the administration to explore and prioritise issues to be resolved c. Identify potential use cases where AI can have a significant impact on improving patient outcomes, enhancing clinical decision-making, optimising healthcare delivery or reducing healthcare costs d. Prioritise use cases based on factors such as clinical relevance, feasibility, scalability, and potential for positive impact e. Literature review could be conducted to further understand related or prevalent problems, emerging trends, and possible interventions to improve the issues in healthcare delivery 2. **Analyse Healthcare Data** a. Analyse available healthcare data, including electronic health records, medical imaging data, genomic data, wearable device data, and healthcare claims data. Look for patterns, trends, and anomalies that may indicate areas of concern or opportunities for intervention 3. **Evaluate Feasibility and Resources** a. Assess the feasibility of implementing AI solutions for identified problems within the current healthcare ecosystem. Consider factors such as data availability, technical infrastructure, expertise needed, funding resources, and organisational readiness 4. **Regulatory and Ethical Considerations** a. Consider regulatory requirements, ethical considerations, privacy concerns, and data security issues when identifying AI-driven solutions for healthcare problems. Ensure compliance with relevant regulations such as data protection regulations	**Descriptive (Assistive AI)**: An assistive AI algorithm that provides descriptive insights into patient demographics and healthcare utilisation patterns within a hospital system. This algorithm analyses historical data from electronic health records to generate reports and visualisations depicting patient demographics, admission rates, length of stay, and common diagnoses. Clinicians and hospital administrators can use these insights to better understand patient populations, allocate resources effectively, and optimise healthcare delivery processes **Diagnostic (Assistive AI)**: An assistive AI algorithm for medical image analysis that assists radiologists in diagnosing breast cancer from mammography images. This algorithm uses DL techniques to analyse mammography images and detect suspicious lesions or abnormalities indicative of breast cancer. Radiologists can review the algorithm's findings alongside their own assessments to improve diagnostic accuracy and reduce the risk of false positives or false negatives **Predictive (Autonomous AI)**: An autonomous AI algorithm for predicting patient readmissions to the hospital within 30 days of discharge. This algorithm leverages ML algorithms trained on historical patient data, including demographics, medical history, diagnosis codes, and previous hospitalisation records. By analysing these data points, the algorithm generates personalised risk scores for each patient, indicating the likelihood of readmission. Healthcare providers can use these predictive insights to intervene proactively and provide targeted interventions to high-risk patients such as care coordination, medication management or follow-up appointments to prevent readmissions and improve patient outcomes **Prescriptive (Autonomous AI)**: An autonomous AI algorithm for personalised treatment recommendation in oncology. This algorithm analyses genomic data, tumour characteristics, treatment history, and clinical outcomes from a large database of cancer patients. Based on this analysis, the algorithm generates personalised treatment plans tailored to each patient's unique profile, including chemotherapy regimens, targeted therapies or/and immunotherapies. Oncologists can use these prescriptive recommendations to make informed treatment decisions and optimise patient outcomes while minimising the risk of adverse effects or treatment resistance

The content of the table was adapted from that given by ChatGPT3.5

must address the tool's ethical aspects, ensuring professional integrity, a clear balance of benefits over harm, justice and trustworthiness, with designated accountability for its implementation.

Data availability, organisation and curation

Relevant real-world data sources must be explored, annotated and preprocessed (Table 5). The availability of high-quality and sufficient variables in the target population is crucial for AI solutions to address clinical problems. These data must be diverse and representative [58], properly labelled and curated to minimise bias and errors. Data will need to go through several stages before becoming useful for AI algorithmic models. These stages include standardisation (coding structured and unstructured data) for interoperability, cataloguing, deidentification (pseudo or anonymisation), cleaning/transformation (validation), and linking and combining different sources into a single dataset. Managing a large amount of quality data within credible data governance structures remains a significant challenge [59].

Infrastructure and architecture for data repository and AI technology

An adequate capacity of computers and servers and accessories for operating large amounts of data at the optimal speed are needed (Table 6). Intelligence processing units (IPUs) are rarer, especially on certain clouds, and are best used in graph-based AI algorithms. In scenarios where high-performance, energy-efficient hardware acceleration is required to handle demanding AI workloads, enabling faster training, inference and deployment of AI models across various applications and industries. Another specialised hardware accelerator developed by Google is the tensor processing unit (TPU). It is specific to ML tasks, particularly those involving TensorFlow and Google's open-source ML framework. Compared with traditional CPUs and GPUs, they offer

Table 5 Data curation for AI tools

No	Data curation step	Explanation
1.	Data Exploration	This involves getting familiar with the dataset by examining its structure, size and basic statistics. This step helps identify any missing values, outliers or inconsistencies in the data. It provides insights into the distribution of features and potential patterns that may exist within the datasets
2.	Linking and Combining Different Sources into One Dataset	This involves integrating data from multiple datasets or sources into a single cohesive dataset. This step allows for a comprehensive analysis of data from various sources, enabling insights and patterns that may not be apparent when analysing individual datasets separately
3.	Deidentification (Pseudo or Anonymisation)	Deidentification involves removing or obfuscating personally identifiable information from the dataset to protect individual privacy. This step is crucial for handling sensitive data and ensuring compliance with data protection regulations. Deidentified data can still be used for analysis and modelling while preserving the anonymity of individuals
4.	Data Annotation	This involves labelling or tagging the data with relevant information such as class labels or categories, to prepare it for supervised learning tasks. This step is essential for training ML models as it provides ground truth labels for the algorithm to learn from. Data annotation can be done manually by human annotators or using automated tools, depending on the complexity and scale of the dataset
5.	Data Preprocessing	This critical step includes several substeps to clean and prepare the data for analysis or modelling. This may involve removing noise or irrelevant information, handling missing values, addressing class imbalances, encoding categorical variables and scaling or normalising features. Data preprocessing aims to ensure that the data is in a standardised format and suitable for ML algorithms. This ensures that all features have a similar scale and distribution, which can improve the performance of certain ML algorithms such as gradient descent-based methods. Common techniques include scaling features to have zero mean and unit variance (standardisation) or scaling features to a specified range (min–max scaling)
6.	Data Quality Assurance	This involves ensuring the integrity, accuracy and reliability of the dataset throughout the curation process. This includes conducting thorough checks for errors, inconsistencies or biases in the data, as well as validating the annotations and preprocessing steps. Data quality assurance aims to identify and rectify any issues that could impact the performance or validity of the ML models trained on the dataset
7.	Data Splitting	This involves dividing the dataset into training, validation, and test sets to evaluate the performance of ML models. A common practice is to split the data into a 70–30 or 80–20 ratio, with the larger portion allocated for training. The training set is used to train the model, the validation set is used to tune hyperparameters and optimise model performance, and the test set is used to evaluate the final performance of the model on unseen data. Proper data splitting ensures that the model's performance estimates are unbiased and generalisable to new data
The quantity of unique inputs for features/variables in the datasets for AI/ML algorithms is in the thousands. When faced with a lack of data quantity or poor data quality even after data cleaning and preprocessing as mentioned above, there are several strategies that can be employed to overcome these challenges:		
1.	Data Augmentation	This involves generating additional training samples by applying various transformations to the existing data. This includes random rotations, flips, crops or colour adjustments for image data, adding noise or perturbations for other types of data. Data augmentation helps increase the diversity and variability of the training data improving the model's ability to generalise to unseen examples and reducing the risk of overfitting
2.	Feature/variable Engineering	This involves creating new features or transforming existing ones to improve the performance of ML models. This may include extracting relevant information from raw data, combining or aggregating features, or applying mathematical transformations to make the data more informative or discriminative. Feature engineering may improve the model's ability to capture underlying patterns in the data
3.	Ensemble Methods	Combine predictions from multiple weak models to create a stronger and more robust model. Ensemble methods such as bagging, boosting or stacking can help mitigate the effects of noisy or low-quality data by leveraging diverse models and averaging their predictions
4.	Semisupervised Learning	Incorporate unlabelled data along with the limited labelled data to train the model may leverage the abundant unlabelled data to improve model performance and generalisation even when labelled data is scarce
5.	Active Learning	Strategically select which samples to label by iteratively training the model on a small labelled dataset, then using the model to select the most informative samples for annotation. This approach maximises the utility of limited labelling resources
6.	Transfer Learning	Utilise pretrained models on large, relevant datasets and fine-tune them on your smaller or lower-quality dataset. Transfer learning leverages the knowledge learned from the pretrained model to boost performance on the target task with limited data
7.	Domain Knowledge Integration	Incorporate domain knowledge and expertise into the modelling process to guide feature selection, model architecture design and interpretation of results. Domain knowledge can help compensate for data limitations and improve the relevance and accuracy of the model's predictions

The content of the table was adapted from that given by ChatGPT3.5

Table 6 Different levels of computing power and AI technologies

No	Specification of computer	Level of AI technology
1.	Entry-level CPU (e.g., Intel Core i3, AMD Ryzen 3)	Basic machine learning algorithms, such as linear regression or decision trees, for small-scale data analysis and prediction tasks
2.	Mid-range CPU (e.g., Intel Core i5, AMD Ryzen 5)	More advanced machine learning algorithms including neural networks for tasks such as image recognition, natural language processing and recommendation systems
3.	High-end CPU (e.g., Intel Core i7/i9, AMD Ryzen 7/9, Apple M3 Pro)	High-performance computing (HPC) for training complex deep learning models on large datasets such as those used in medical imaging, autonomous vehicles and financial modelling
4.	Entry-level GPU (e.g., NVIDIA GeForce GTX 1650, AMD Radeon RX 550, Intel Arc A380)	Accelerated computing for training and inference of machine learning models particularly for tasks involving parallel processing such as computer vision, speech recognition and gaming. In general, GPUs are not optimised for AI/ML tools, consume lots of energy and may become outdated in short time
5.	Mid-range GPU (e.g., NVIDIA GeForce RTX 2060, NVIDIA RTX 3060/4060, AMD Radeon RX 5600 XT, AMD RX 6700 XT)	Deep learning training and inference for applications requiring higher computational power and memory bandwidth such as real-time video analytics, virtual reality and autonomous drones
6.	High-end GPU (e.g., NVIDIA GeForce RTX 3080, AMD Radeon RX 6900 XT, AMD RX 7900 XTX, Apple M3 Ultra)	State-of-the-art deep learning research, training of large-scale models (e.g. GPT, BERT), and deployment of AI applications in industries like healthcare, finance and cybersecurity
7.	FPGA (e.g., Xilinx, Intel Stratix 10)	Customised hardware acceleration for specific AI tasks such as real-time inferencing in edge devices, network optimisation and hardware emulation of neural networks
8.	ASIC (e.g., Google TPU v5e, Graphcore IPU, Tesla Dojo)	Specialised microchips optimised for AI workloads, offering unparalleled performance and energy efficiency for tasks like neural network training and inference in data centres and edge devices

The content of the table was adapted from that given by ChatGPT3.5

ASIC Application-Specific Integrated Circuit, BERT Bidirectional Encoder Representations from Transformers, *CPU* central processing unit, *FPGA* Field-Programmable Gate Array, *GPU* graphics processing unit, *GPT* Generative Pre-trained Transformer. The content of the table was adapted from that given by ChatGPT3.5, and 4o

significant speedups and cost savings, particularly for large-scale AI workloads running on TensorFlow-based frameworks. The operations of these data servers include strong cybersecurity (data encryption), data privacy, controlled access and updated regulatory policies on the proper use of the data, and supervised incremental learning of the AI/ML tools. In addition to security and proper governance, the ease and speed of access to different users are paramount. The physical and virtual infrastructure, and computer systems' architecture must be scalable to meet the increasing needs and demands of the tools. Alternatively, cloud-based infrastructures offer more feasible services in AI tool development from algorithm building to deployment and scale AI applications by providing access to a rich ecosystem of resources and tools. The three main cloud service providers are Amazon Web Services (AWS), the Google Cloud Platform (GCP) and Microsoft Azure (Table 7). In addition to providing scalable infrastructure, it also provides robust data storage, management solutions and a wide range of AI development tools and frameworks, such as TensorFlow and Azure Machine Learning, and application programming interfaces (APIs) to streamline the development workflow. Cloud services also facilitate collaboration among team members and simplify the deployment of AI models in production environments. Additionally, they offer

Table 7 The three main cloud service providers

No	Cloud Infrastructure Providers	Cloud Services
1.	Amazon Web Services (AWS) https://aws.amazon.com	AWS offers a comprehensive set of cloud-based AI services and tools, including Amazon SageMaker and AWS AI services. AWS also provides standalone AI frameworks like TensorFlow and PyTorch for users who prefer to develop and deploy AI models on their own infrastructure
2.	Google Cloud Platform (GCP) https://cloud.google.com	GCP provides a range of cloud-based AI services, including Google Cloud AI Platform and AutoML, as well as TensorFlow Enterprise. GCP also supports standalone AI frameworks like TensorFlow and PyTorch for users who prefer to develop and deploy on their own infrastructure
3.	Microsoft Azure https://azure.microsoft.com	Azure offers a range of cloud-based AI services including Azure Machine Learning and Cognitive Services. Azure also supports standalone AI frameworks like TensorFlow and PyTorch for users who prefer to develop and deploy on their own infrastructure. Additionally, OpenAI's GPT- 3.5-Turbo, GPT- 4 and DALL-E, Whisper, Babbage and Davinci are also available

The content of the table was adapted from that given by ChatGPT3.5

monitoring and optimisation tools to ensure the optimal performance of AI applications.

AI neural networks on open access platforms

Many AI algorithms are readily available on open-access platforms (Table 8), with similar algorithms often already developed and tested. Choosing appropriate AI/ML models and methods is essential for resolving clinical challenges. The model selection framework should balance performance requirements with cost, risk, deployment needs and stakeholder expectations [60]. The choice of algorithm depends on the input type, whether speech, language, vision, decision-making, or a combination of these. For example, convolutional neural networks are ideal for image data, whereas recurrent neural networks are best suited for text and numerical data [61]. The development of new AI neural networks requires data scientists with advanced skill sets and is time-consuming.

AI/ML model validation

The selected or newly developed algorithm must undergo training, validation and testing on a curated dataset (Table 5). Its performance should be evaluated and compared with that of the baseline model or standard of care before external validation, especially if the model is applied in different settings from where it was developed and then deployed in practice [37]. Table 9 shows the classification tasks and ML strategies on data [60, 61]. Both nonclinical and clinical validation are essential to establish its performance, ensuring its integration into routine clinical workflows, usability and positive effects on clinical outcomes. Properly designed clinical research, including clinical trials, may be necessary to assess its real-world clinical impact. Table 3 outlines recommendations for evaluating AI tools in clinical settings, whether as diagnostic or prognostic tools. Once finalised, the results are published for broad dissemination and peer scrutiny. It is also critical to explore and address any ethical and legal implications associated with using these tools in healthcare, as liability risks may arise from sources of error, error identification, potential harm and legal redress [62].

Registration

The registration of tools with relevant authorities is typically carried out by the manufacturer, developer or the organisation responsible for the AI tool. In many cases, this involves collaboration between technical and regulatory teams within the organisation to ensure compliance with the regulatory requirements of the target market. Tools registration with relevant authorities such as medical device authorities could increase the likelihood of successful implementation and deployment in the real world [63]. The evaluation criteria differ across countries,

Table 8 Examples of open sources for AI algorithms

No	Open-source algorithm	Characteristics and examples
1.	TensorFlow https://www.tensorflow.org	Developed by Google Brain Team. It is well-suited for complex deep learning tasks and large-scale projects. Consider using TensorFlow for tasks such as image classification, natural language processing and reinforcement learning. Widely used for deep learning tasks such as medical imaging analysis, genomics and clinical natural language processing
2.	PyTorch https://pytorch.org	Developed by Facebook's AI Research lab (FAIR). PyTorch offers dynamic computational graphs, making it suitable for research and experimentation. Consider PyTorch for prototyping and implementing cutting-edge deep learning models. Leverage PyTorch's flexibility and simplicity for rapid model iteration and debugging. PyTorch is popular in academia and research due to its ease of use and Pythonic syntax
3.	scikit-learn https://scikit-learn.org/stable/	Built on top of NumPy, SciPy and matplotlib. It is ideal for traditional machine learning tasks such as classification, regression and clustering. Consider scikit-learn for projects with structured data and well-defined features. Provides simple and efficient tools for data mining and data analysis tasks such as disease prediction, diagnosis and drug discovery
4.	Keras https://keras.io	High-level neural networks API, running on top of TensorFlow or Theano. It offers a user-friendly interface for building and training neural networks, and simplifies the process of building, training and deploying neural networks. Take advantage of Keras's simplicity and modularity to iterate quickly on model architectures and hyperparameters. Keras seamlessly integrates with TensorFlow for production deployment
5.	Apache MXNet https://mxnet.apache.org/versions/1.9.1/	Developed by Apache Software Foundation. It supports multiple programming languages including Python, R, Scala, and Julia. Known for its scalability, efficiency and flexibility, making it suitable for projects requiring scalability and performance optimisation across multiple programming languages such as in distributed deep learning and model serving in production environments

Other Popular Open-Source AI Frameworks: 1) JAX – Google's high-performance framework for scientific computing and neural networks; popular in academic circles for its composability and speed via XLA compilation. 2) Hugging Face Transformers – A hub and library for state-of-the-art pretrained models, especially for NLP and multimodal AI; works with PyTorch, TensorFlow, and JAX. 3) FastAI – A high-level wrapper around PyTorch designed for rapid prototyping and educational purposes. 4) ONNX – A standard exchange format for model interoperability across different frameworks; critical for deployment and hardware optimization. The content of the table was adapted from that given by ChatGPT3.5, and 4o

Table 9 ML learning, techniques and evaluation metrics

No	ML learning	Techniques	Key evaluation metrics
1.	**Supervised Learning**: Models trained on labelled data to predict outcomes	• **Regression**: Support Vector Machines, Neural Networks, Ridge Regression, Lasso, Random Forest • **Classification**: Decision Trees, Random Forest, Support Vector Machines, Discriminant Analysis, Naïve Bayes, Nearest Neighbour, Neural Networks	**Accuracy**: Proportion of correct predictions or overall correctness **Sensitivity (Recall)**: Measures the model's ability to correctly identify positive cases **Specificity**: Indicates the model's ability to correctly identify negative cases **Positive Predictive Value (Precision)**: Helps understand the likelihood that a positive prediction is correct **F1 Score**: Harmonic mean of precision and recall **ROC-AUC (Receiver Operating Characteristic —Area Under Curve)**: Assesses the trade-off between sensitivity and specificity. Think of it as an overall ability to distinguish between positive and negative cases
2.	**Unsupervised Learning**: Models that identify patterns in unlabelled data	• **Clustering**: K-Means, K-Medoids, Fuzzy C-Means, Hierachical, DBSCAN, Gaussian Mixture, Hidden Markov Model, Neural Network), • **Dimensionality Reduction**: PCA, LDA, Isomap, Autoencoder	
3.	**Semi-Supervised Learning**: Combines a small amount of labelled data with a large amount of unlabelled data during training	Both of the above	
4.	**Reinforcement Learning**: Models learn by interacting with an environment to achieve a goal	Markov Decision Process	

DBSCAN Density-Based Spatial Clustering of Applications with Noise, *Isomap* Isometric Mapping is a nonlinear dimensionality reduction technique, *LDA* Linear Discriminant Analysis, *PCA* Principal Component Analysis. The content of the table was adapted from that given by ChatGPT-4o

which may include an effectiveness trial [32]. AI tools in the UK are classified as medical devices and therefore require Medicines and Healthcare products Regulatory Agency (MHRA) approval bearing the "United Kingdom Conformity Assessed" (UKCA) logo. However, AI tools in Europe are regulated by the EU Medical Device Regulation (EU MDR) and bearing the "Conformité Européenne" logo to be marketed in Europe. In the USA, AI tools are regulated by the Food & Drug Administration (FDA) (https://www.fda.gov/science-research/science-and-research-special-topics/artificial-intelligence-and-medical-products). FDA classifies AI tools based on their risk level and intended use following pathways such as 510(k) premarket notification, De Novo classification or premarket approval. Many AI tools are categorised as Software as a Medical Device and must meet rigorous criteria for safety, effectiveness and transparency including Good Machine Learning Practices. Post-market surveillance is often required to monitor real-world performance while labelling must clearly define intended use, performance metrics and limitations. European AI Act [64] prohibits AI systems that collect sensitive personal information that causes discrimination, manipulates human behaviour or exploits vulnerabilities of certain groups of people at all social places. The core principle is that AI "… should be a human-centric technology. It should serve as a tool for people, with the ultimate aim of increasing human well-being." Clinicians play a vital role in evaluating AI tools' suitability for their practice and ensuring their safe and effective use.

While regulatory frameworks vary across regions, a unifying principle among global authorities is the emphasis on ensuring that AI tools align with ethical standards, prioritising human well-being, fairness and accountability. These principles not only guide the evaluation and approval processes but also ensure that the implementation of AI tools remains consistent with societal values and promotes trust among users and stakeholders. The ethical principles of nonmaleficence, beneficence, autonomy and justice with added governance and associated principles of privacy, diversity, inclusiveness, transparency, reliability, fairness, social good, well-being, sustainability, auditability, explicability, interpretability and quality data are referred to in high-level policy documents [65–68] (see Fig. 3, and Additional File 2: AI Ethics and Policy Frameworks from the United Nations Educational, Scientific and Cultural Organization 2022 [7, 65, 69], UN Resolution on AI 2024 [66], International Scientific Report on the Safety of Advanced AI: Interim Report 2024 [70], Diversity, INclusivity and Generalisability: STANDING Together project team 2023 [58], US Executive Order on AI 2023 [67], Artificial Intelligence Act European Parliament/2024 [68], Harmonised Standards for the European AI Act: European Parliament 2024 [71], Ethics and governance of artificial intelligence for health: World Health Organization guidance 2021 [72], Organization for Economic Cooperation and Development AI Principles 2019 [73], Universal Guidelines for AI: Center for AI and Digital Policy 2018 [74], Asilomar AI Principles: Future of Life Institute 2017[75]). These

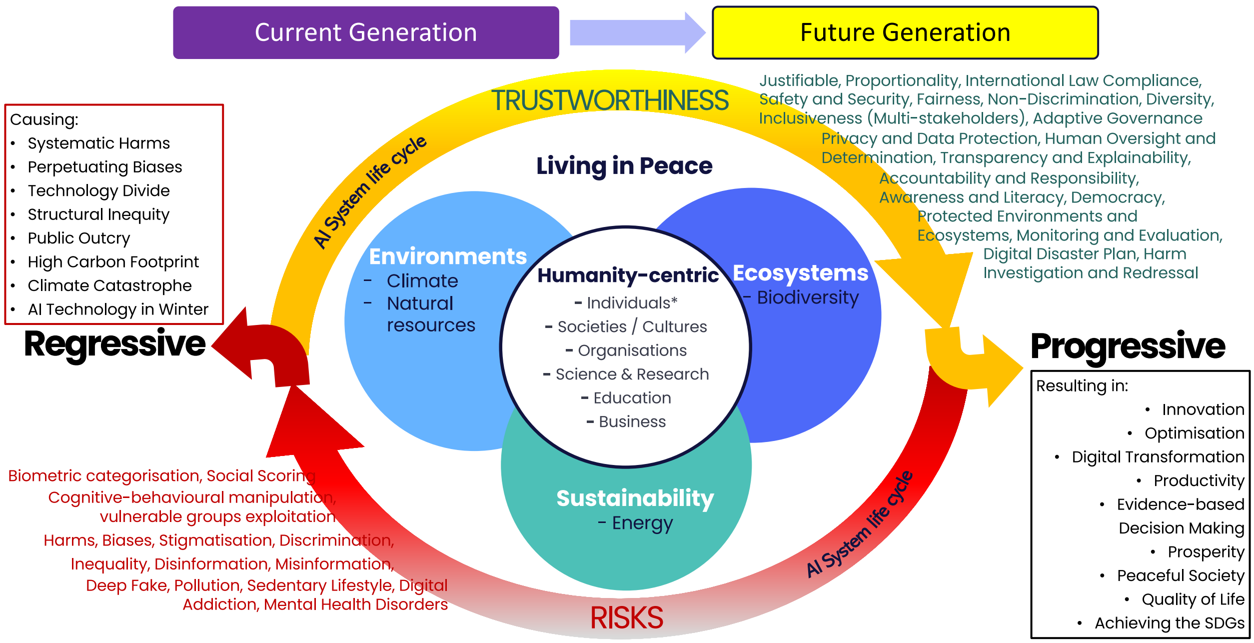

Fig. 3 AI ethics that may determine progressive or regressive outcomes. *Different age groups, cultural systems and language groups, persons with disabilities, girls and women, and disadvantaged, marginalised and vulnerable people or people in vulnerable situations. SDG = Sustainable development goals.

principles, in the form of typology according to the different stages of the AI life cycle and sources, are available here (https://ricardo-ob.github.io/tools4responsibleai/#title-cite) [76] and foster the development of responsible AI tools and systems by technical and nontechnical persons, balancing the risk and benefits to the public [77]. AI tools and systems are prohibited by the European Union's AI Act if they manipulate cognitive behaviours, classify the traits and status of people through facial or emotion recognition and collect sociobiological characteristics such as sexual orientation or religious beliefs into various forms of social scoring or biometric categorisation [68].

Clinical deployment and monitoring of performance

Deployment is the method by which the tested AI tools are integrated into an existing clinical workflow to make practical healthcare decisions (outputs) on the basis of data (input). The best deployment strategy would consider the software systems or applications environment where the AI tools are to be deployed. If this system is a web service or electronic health records system, then it will require an API to enable data pipeline integration where the input and output could be executed. The easier the deployment process, which includes having the same API endpoint references, the faster the model improvements are. The design of the user interfaces must allow alerts or notifications to be displayed noninterruptively but effectively to achieve practice efficiency and provider acceptance and adoption. This could be tested via "silent" or "shadow" deployment, which is deployed in the actual environment but not fully for routine use. Another important step before deployment to production is the quality testing of scalability and performance optimisation in scenarios when high data flow occurs. The final deployment approach is likely a decision between the budget, availability of the infrastructures and the required performance of the AI tools.

Another important task is to train clinicians and team members in healthcare facilities where tools are used or integrated into the electronic medical records system. Promoting human-AI teaming would augment performance and safeguard autonomy but require calibrated design, support and monitoring [34]. Be prepared to explain the decision process of the AI/ML tools and be present to support their use. This responsibility often lies with a collaborative effort between the developers and the clinical staff, with oversight from regulatory bodies. Neural networks in AI are often called "black boxes" because their internal workings are complex and not easily understood. This lack of transparency can be problematic in healthcare where understanding how decisions are made is crucial. To address this, techniques like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) help identify which input features most influence the AI's decisions. Visualisation tools such as Grad-CAM (Gradient-weighted

Class Activation Mapping) can highlight areas in medical images that the AI focuses on, providing insights into its decision-making process [78].

The tool's effect in clinical care should be measured, and the incremental learning of the tool should be supervised. Additionally, monitoring should continue for any unwanted effects arising from its use including data integrity, cybersecurity and impact mitigation in cases of breach and functional recovery [79]. A medical algorithmic audit could and should be conducted if it was not performed before real-world deployment [80]. Changes in the clinical practices such as changing guidelines, treatments or diagnostic protocols can render previously trained models less effective over time. Be vigilant and prepared to retrain the AI/ML algorithm if the performance has drifted to below expectations. Clinicians, developers and IT teams must remain vigilant in monitoring AI performance for signs of drift, with clinicians reporting inconsistencies and developers tracking key metrics. This includes prediction accuracy, response time and resource utilisation. Developers are primarily responsible for retraining models, using updated data and collaborating with clinical staff to ensure relevance, while compliance teams ensure adherence to regulatory standards. Retraining involves collecting new data, refining the model, validating updates and carefully redeploying the system with ongoing performance monitoring. Any update done to the tools may require a notification to the regulatory body, and it is essential to consult specific guidelines and maintain open communication with the relevant regulatory body.

AI ecosystem improvement

Actively engage stakeholders and the public with the tool throughout the development-deployment-monitoring process. This could improve confidence and progress the AI ecosystem in the country [17]. This approach aims to improve broader and better perceptions of AI/ML technologies, invite more keen interest and training from experts, and establish central governance, trusted custodian, ethical value and proper regulation. This may increase the investment and uptake of AI technologies in healthcare and research that are supported by sufficient funding and infrastructure to allow freedom to innovate and implement more AI/ML tools. Ethical considerations around data privacy and patient safety are well-known challenges that must be addressed [81, 82]. Similarly, traditional medical principles remain crucial as they uphold patient dignity and foster mutual trust among doctors, patients and society. Establishing a successful partnership between technology companies, which provide technological expertise, and healthcare facilities which offer data and expert input, is essential. This partnership must be both regulatory compliant and economically beneficial to ensure the effective implementation and deployment of algorithms [83].

Experiences in National University Health System, Singapore

This research shares the experience and valuable lessons of the National University Health System, Singapore (NUHS), in obtaining AI tools for production in healthcare services. NUHS's experience in implementing AI-driven healthcare systems offers valuable insights for institutions pursuing similar transformations. Success in AI implementation extends beyond the technology itself, requiring four critical elements: (1) establishing robust data infrastructure, (2) building organisational trust, (3) ensuring continuous human oversight through committees and (4) committing to long-term engagement with AI technology [84].

They developed the ENDEAVOUR AI platform, a comprehensive AI system that integrates various tools to streamline operations [85]. Additionally, they established DISCOVERY AI, a private AI training cloud featuring NVIDIA DGX A100 s to support the development of AI models, which functions as both a production system and a research sandbox for modular machine learning tools [16, 85]. It adheres to local and international regulatory guidelines, with data anonymized by removing identifiers such as names, addresses and identification numbers. A robust master governance structure ensures equitable data access, centralised anonymisation and differential data linkage. Data access and sharing are overseen by custodians of specific databases and a dedicated committee. This governance framework also manages research administration, including institutional review board processes and collaborative agreements. Integrated with the electronic health record system, the platform leverages multiple clinical data and research databases through embedded algorithms to enable many AI predictions.

With both the ENDEAVOUR AI platform and DISCOVERY AI, a series of AI tools have been developed and successfully deployed for clinical care, while some have undergone internal validation within the institution and are pending full peer-reviewed publication [84, 85]. An AI-driven system, the Pathfinder Dashboard (Additional File 3 shares the experience of Pathfinder Dashboard AI tools development and challenges according to the nine steps mentioned in this review) predicts patient wait times and manages patient inflows at the emergency department, enhancing care quality and patient satisfaction. Should patients have to be admitted for inpatient care, the estimated length of stay model predicts patient hospital stays, ensuring timely and appropriate care and thus optimising the effective planning and allocation of resources [86]. When discharge is possible or decided upon, the 30-day readmission prediction model could personalise patient care to prevent readmission and reduce hospital costs. The Disease Progression Modelling tool enables earlier intervention by anticipating disease progression, particularly for chronic conditions, and the Pharmacogenomics Alerts System tailors medication recommendations on the basis of genetic profiles, enhancing precision medicine and reducing adverse drug reactions. NUHS has enhanced patient communication with various chatbot systems, including RUSSELL-GPT [87], which provides instant responses and personalised health information. These chatbots use advanced GPT models to cater to both patients and researchers while maintaining data security. For managing chronic diseases, all AI tools at NUHS are integrated with the Epic EMR System, providing a unified AI dashboard that offers comprehensive insights. This integration enhances decision-making and patient care by consolidating information and streamlining hospital operations.

There is the CURATE.AI to optimise chemotherapy treatments for prostate cancer [88] and solid tumours [89]. It has also been applied to personalise dose selection [90] and to tailor immunosuppressant drug dosages for liver transplant patients to prevent organ rejection [91]. NUHS introduced the Chronic Disease Management Programme (CHAMP) Chatbot System, which engages patients with reminders and follow-ups via WhatsApp. Compared with similar programmes, this tool aims to improve patient adherence to treatment plans, leading to higher enrollment rates and lower dropout rates.

Discussion

Developing and translating AI innovations from research to clinical practice faces significant challenges often referred to as the "valley of death" [92]. These include the complexity of identifying the right pain point, clinical validation, regulatory hurdles and the need for robust evidence of efficacy and safety, registration with the regulatory body and communication with trust with healthcare stakeholders for integration into an existing clinical workflow [93, 94]. Additionally, the lack of standardised reporting and evaluation frameworks complicates the explainability and interpretability for the integration of AI tools into healthcare settings [95]. Clinicians who are more than aware of the full cycle of AI tools development delineated in this paper could facilitate the development, reporting, assessment and smoother transitions from bench to bedside of these tools [96, 97].

The most important step and challenge to tackle is the biases in training data. This could perpetuate healthcare disparities such as the underrepresentation of specific demographic groups or the reinforcement of historical biases in data collection [98]. Compounding this issue would be a dataset shift post-deployment where the model's operational environment differs from its training environment causes a degrade performance and compromise generalisability [99]. Mitigating these challenges requires careful dataset curation to have diverse and representative samples, along with the deployment of bias detection and mitigation strategies [100]. Rigorous external validation across varied populations and settings is essential to ensure the reproducibility and generalisability of AI models, both of which are foundational to achieving fairness, equity and clinical adoption [101].

Beside the strategies alluded to when faced with a lack of data quantity or poor data quality, there are several strategies generative AI can offer. This includes creating synthetic data based on the electronic health records, omics data and bioimages to train diagnostic and predictive models [102, 103]. This transformative potential alleviates data scarcity, enhances patient privacy and enables the simulation of rare or complex clinical scenarios. However, challenges remain in ensuring that synthetic data maintains the variability and complexity of real-world datasets to achieve reproducibility [104]. For AI models trained on synthetic data, rigorous testing and validation are necessary to confirm that they generalise accurately to diverse populations and clinical realities. Addressing these challenges allows generative AI to significantly enhance the robustness and utility of healthcare AI systems.

Evaluating AI models against professional clinicians is crucial to understanding their clinical utility and assessing their algorithmic quality [105]. While some AI models achieve expert-level accuracy, a lack of rigorous study design often leads to overestimated performance claims. Comparative studies and standardised evaluation frameworks are critical for determining whether AI tools can complement or enhance human decision-making in healthcare [93]. Such evaluations are vital for building confidence among stakeholders and ensuring the safe and effective deployment of AI in clinical practice.

Reproducibility and generalisability are critical pillars for ensuring the effective translation, application, and evaluation of AI models that has been developed for healthcare. Reproducibility demands consistent results through transparent documentation of data collection, preprocessing and model training, fostering trust and reliability [106]. Generalisability ensures that AI models perform accurately and equitably across diverse populations, clinical settings and evolving medical practices, addressing key challenges such as dataset biases and shifts [107]. These principles are essential for validating all AI tools and ensuring robust, scalable solutions. By integrating these considerations across the entire AI lifecycle, from development to deployment, clinicians and developers can create innovative, equitable and reliable tools that meet the demands of real-world healthcare.

All AI systems, regardless of type, follow a similar life cycle as described in this review. The primary differences lie in complexity, scalability, interpretability, resource demands and training methodology. Traditional AI/ML models that utilise established statistical or rule-based techniques with manually chosen features are typically simpler, more interpretable and less computationally demanding. Neural networks and deep learning models that employ layered neuron-like architectures that learn patterns from raw data could scale well with large data but require more computational resources and are harder to interpret. Generative AI models implement advanced frameworks that generate original outputs by modelling data distributions do push these challenges further, often requiring massive data and compute resources, more complex training regimes (pretraining plus fine-tuning), and specialised evaluation and monitoring strategies.

Conclusions

This review presents a straightforward explanation of the entire development process of AI tools, outlined in nine cyclical and iterative steps, which could enhance understanding among clinicians. More importantly, the presentation with many infographics and examples, combined with adequate technical details, has the potential to reach a broader audience, particularly in countries that face greater inequities in the health AI/ML literature [108, 109] and are at risk of health disparities from this technology [110, 111]. Notably, other great challenges include win–win partnerships between technology companies such as technological know-how, health-care facilities, as data sources, and expert inputs to algorithms [112, 113]. This is to be regulatory, acceptable and economically rewarding to the two stakeholders [83, 114]. Robust AI tools are those that resolve real-world clinical problems, are developed by a team of relevant stakeholders, are trained on broad-based high-quality data; are validated externally, prospectively and in controlled trials or equivalently. They perform in real time, are unbiased, safe and trustworthy with acceptable human–AI tool interactions [20], are quick in algorithm updates to cover emerging diseases, are controllable by human agents, are acceptable to target users who are either explainable or unexplainable [5, 115], and are ethically justifiable and legally compliant [116]. Challenges in attaining high-performing AI systems include having high-quality infrastructures in terms of computing power, memory and storage capacity, high-speed internet connectivity, low-latency networking, more energy efficient computing technology (quantum computing and optical computing) [117], and scalability and elasticity that are supported by ethics and regulatory compliance data governance [118].

Ultimately, AI/ML tools offer significant benefits by reducing systemic, sporadic medical errors and enhancing patient care quality. These tools streamline healthcare processes, integrate seamlessly into health systems and are continuously monitored to ensure safety and effectiveness. Having legal framework that ensure compliance with data security, protection and privacy policies, positive economic impacts or at least an oversight by an established data governance body including representation from the public and patients could further strengthens accountability and trust in their use [119]. Accordingly, success clinical integration and implementation of AI tools must include building trust and confidence among clinicians in the development process, having quality data, and risk levels are understood by all stakeholders and mitigated as a team with clinicians [120], satisfying fairness, equity, robustness, privacy, safety, transparency, explainability and accountability with assured benefits for patients, healthcare providers and the organisation involved [121].

As the field of AI is anticipated to evolve quickly with new technologies and algorithms, it is essential for all stakeholders including clinicians, to stay informed about new guidelines, reporting standards for AI tools and systems, and the application of AI in medicine (see Table 2 on some of the important organisations on AI-related matters in Additional File 1) [116].

Abbreviations

AI	Artificial Intelligence
ALTAI	Assessment List for Trustworthy AI
API	Application Programming Interface
ASIC	Application-Specific Integrated Circuit
AUC	Receiver Operating Characteristic – Area Under Curve
AWS	Amazon Web Services
BERT	Bidirectional Encoder Representations from Transformers
CAM	Gradient-weighted Class Activation Mapping
CHAMP	Chronic Disease Management Programme
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CLAIM	Checklist for Artificial Intelligence in Medical Imaging
CNN	Convolutional Neural Networks
CONSORT-AI	Clinical Trial Reports For Interventions Involving Artificial Intelligence
CODE-EHR	Best practice checklist to report on the use of structured electronic healthcare records in clinical research
CPS	Clinical Practice Statement
CPT	Current Procedural Terminology
DECIDE-AI	Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence
DL	Deep Learning
ECG	Electrocardiogram
EHR	Electronic Health Record
EITCA	European Information Technologies Certification Academy
FAFM	Foundation for Advancing Family Medicine
FAIR	Artificial Intelligence Research (lab)
FDA	Food & Drug Administration
FPGA	Field-Programmable Gate Array
FUTURE-AI	International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare

GC	Google Cloud Platform
GPT	Generative Pre-trained Transformer
HPC	High-Performance Computing
ICD	International Classification of Diseases
IDEAL	Innovation, Development, Exploration, Assessment, and Long-term Framework
IPU	Intelligence processing units
LIME	Local Interpretable Model-Agnostic Explanations
MHRA	Medicines and Healthcare products Regulatory Agency
MI-CLAIM	Minimum information about clinical artificial intelligence modelling
MINIMAR	MINimum Information for Medical AI Reporting
ML	Machine Learning
NLG	Natural Language Generation
NLP	Natural Language Processing
NUHS	National University Health System
OMA	Obesity Medicine Association
OPTICA	Organisational PerspecTIve Checklist for AI solutions adoption
ROC-AUC	Receiver Operating Characteristic—Area Under Curve
SHAP	Shapley Additive Explanations
SPIRIT-AI	Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence
STARD-AI	Standards for Reporting of Diagnostic Accuracy Studies AI Extension
TPU	Tensor Processing Unit
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04076-0.

Additional file 1. Table 1–2, and Fig. 1. Table 1- Glossary. Figure 1- AI Landscape. Table 2- Important organisations on AI-related matters.

Additional file 2. Table 1- AI ethics, safety and policy frameworks.

Additional file 3. Table 1: The Experience of Pathfinder Dashboard AI tools development and challenges in National University Health System (NUHS), Singapore.

Acknowledgements

We would like to thank Ms. Lin Jing, a data scientist in the Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore who assisted the preparation of Additional File 3 on the experience of Pathfinder Dashboard AI tools development and challenges in National University Health System (NUHS), Singapore. I am grateful to Dr. Judice, Dr. Mohammad Shaheryar Furqan and Professor Dr. Ngiam Kee Yuan for their hospitality when hosting me at NUS, where this work was completed. I was supported and sponsored by UPM to attend some of the course and for my sabbatical leave to NUS for a period of nine months when a few other works were also produced beside this one. I would like to express my appreciation to the Foundation for Advancing Family Medicine (FAFM) for the grant of the Besrour Centre Family Medicine Early Career Researcher Award 2022 that supported my attendance to the online courses of the EITCA/Artificial Intelligence Academy. I greatly appreciate my colleagues in the Department of Family Medicine, UPM, who were kind to spare me and to stand in for me to make my sabbatical and research leaves possible. Lastly, the authors acknowledge the use of ChatGPT 3.5, 4o and o1 (OpenAI, San Francisco, CA, USA) to assist in drafting and language editing of portions of this manuscript. The authors have reviewed and edited the content produced by ChatGPT for accuracy and integrity, and accept full responsibility for the final version of the manuscript.

Authors' contributions

BHC conceived the work, completed the acquisition and analysis of data, wrote and drafted the manuscript. BHC and KYN involved in the interpretation of data, read and approved the final version of the manuscript.

Authors' twitter handles

BHC: https://twitter.com/chewboonhow.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 September 2024 Accepted: 11 April 2025
Published online: 24 April 2025

References

1. Elenko E, Underwood L, Zohar D. Defining digital medicine. Nat Biotechnol. 2015;33:456–61.
2. Tian S, Yang W, Grange JML, Wang P, Huang W, Ye Z. Smart healthcare: making medical care more intelligent. Glob Health J. 2019;3:62–5.
3. Adler-Milstein J, Aggarwal N, Ahmed M, Castner J, Evans BJ, Gonzalez AA, et al. Meeting the moment: addressing barriers and facilitating clinical adoption of artificial intelligence in medical diagnosis. NAM Perspect. 2022:10.31478/202209c.
4. Rosen R. How is technology changing clinician-patient relationships? BMJ. 2024;384:q574.
5. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak. 2023;23:73.
6. Trinkley KE, An R, Maw AM, Glasgow RE, Brownson RC. Leveraging artificial intelligence to advance implementation science: potential opportunities and cautions. Implement Sci. 2024;19:17.
7. UNESCO. Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO; 2023. Available from: https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence. Accessed 21 Apr 2025.
8. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med. 2023;388:1201–8.
9. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. Silver Spring (MD): FDA; 2023. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Accessed 21 Apr 2025.
10. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. Lancet Digit Health. 2023;5:e618–26.
11. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. J Biomed Inform. 2019;100: 103311.
12. Bays HE, Fitch A, Cuda S, Gonsahn-Bollie S, Rickey E, Hablutzel J, et al. Artificial intelligence and obesity management: An Obesity Medicine Association (OMA) Clinical Practice Statement (CPS) 2023. Obes Pillars. 2023;6: 100065.
13. Mohaideen K, Negi A, Verma DK, Kumar N, Sennimalai K, Negi A. Applications of artificial intelligence and machine learning in orthognathic surgery: A scoping review. J Stomatol Oral Maxillofac Surg. 2022;123:e962–72.
14. Yin J, Ngiam KY, Teo HH. Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. J Med Internet Res. 2021;23: e25759.

15. Ghaddaripouri K, Ghaddaripouri M, Mousavi AS, Mousavi Baigi SF, Rezaei Sarsari M, Dahmardeh Kemmak F, et al. The effect of machine learning algorithms in the prediction, and diagnosis of meningitis: A systematic review. Health Sci Rep. 2024;7: e1893.
16. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20:e262–73.
17. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research. Asian Bioeth Rev. 2019;11:299–314.
18. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare. Elsevier; 2020. p. 25–60.
19. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. Npj Digit Med. 2019;2:69.
20. Bach TA, Kristiansen JK, Babic A, Jacovi A. Unpacking Human-AI Interaction in Safety-Critical Industries: A Systematic Literature Review. 2023. https://doi.org/10.48550/ARXIV.2310.03392.
21. Fehr J, Citro B, Malpani R, Lippert C, Madai VI. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. Front Digit Health. 2024;6:1267290.
22. Yu K-H, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf. 2019;28:238–41.
23. Adler-Milstein J, Redelmeier DA, Wachter RM. The Limits of Clinician Vigilance as an AI Safety Bulwark. JAMA. 2024. https://doi.org/10.1001/jama.2024.3620.
24. Robertson C, Woods A, Bergstrand K, Findley J, Balser C, Slepian MJ. Diverse patients' attitudes towards Artificial Intelligence (AI) in diagnosis. PLOS Digit Health. 2023;2: e0000237.
25. Shaffer VA, Probst CA, Merkle EC, Arkes HR, Medow MA. Why Do Patients Derogate Physicians Who Use a Computer-Based Diagnostic Support System? Med Decis Making. 2013;33:108–18.
26. Miller A, Moon B, Anders S, Walden R, Brown S, Montella D. Integrating computerized clinical decision support systems into clinical work: A meta-synthesis of qualitative research. Int J Med Inf. 2015;84:1009–18.
27. Olakotan OO, Mohd YM. The appropriateness of clinical decision support systems alerts in supporting clinical workflows: A systematic review. Health Informatics J. 2021;27:146045822110075.
28. Kennedy G, Gallego B. Clinical prediction rules: A systematic review of healthcare provider opinions and preferences. Int J Med Inf. 2019;123:1–10.
29. Knop M, Weber S, Mueller M, Niehaves B. Human Factors and Technological Characteristics Influencing the Interaction of Medical Professionals With Artificial Intelligence-Enabled Clinical Decision Support Systems: Literature Review. JMIR Hum Factors. 2022;9: e28639.
30. Artificial Intelligence (AI) in Healthcare Market (By Component: Software, Hardware, Services; By Application: Virtual Assistants, Diagnosis, Robot Assisted Surgery, Clinical Trials, Wearable, Others; By Technology: Machine Learning, Natural Language Processing, Context-aware Computing, Computer Vision; By End User) - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2022 – 2030. Precedence Research; 2023.
31. Hassan N, Slight R, Morgan G, Bates DW, Gallier S, Sapey E, et al. Road map for clinicians to develop and evaluate AI predictive models to inform clinical decision-making. BMJ Health Care Inform Online. 2023;30: e100784.
32. Kwong JCC, Khondker A, Lajkosz K, McDermott MBA, Frigola XB, McCradden MD, et al. APPRAISE-AI Tool for Quantitative Evaluation of AI Studies for Clinical Decision Support. JAMA Netw Open. 2023;6: e2335377.
33. Bakker L, Aarts J, Uyl-de Groot C, Redekop K. How can we discover the most valuable types of big data and artificial intelligence-based solutions? A methodology for the efficient development of the underlying analytics that improve care. BMC Med Inform Decis Mak. 2021;21:336.
34. Committee on Human-System Integration Research Topics for the 711th Human Performance Wing of the Air Force Research Laboratory, Board on Human-Systems Integration, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine. Human-AI Teaming: State-of-the-Art and Research Needs. Washington, D.C.: National Academies Press; 2022.
35. Tan M, Lee H, Wang D, Subramonyam H. Is a Seat at the Table Enough? Engaging Teachers and Students in Dataset Specification for ML in Education. 2023. https://doi.org/10.48550/ARXIV.2311.05792.
36. Ng FYC, Thirunavukarasu AJ, Cheng H, Tan TF, Gutierrez L, Lan Y, et al. Artificial intelligence education: An evidence-based medicine approach for consumers, translators, and developers. Cell Rep Med. 2023;4: 101230.
37. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med. 2020;26:1320–4.
38. Kotecha D, Asselbergs FW, Achenbach S, Anker SD, Atar D, Baigent C, et al. CODE-EHR best-practice framework for the use of structured electronic health-care records in clinical research. Lancet Digit Health. 2022;4:e757–64.
39. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28:924–33.
40. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, The SPIRIT-AI and CONSORT-AI Working Group, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26:1351–63.
41. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ. 2020;370:m3164.
42. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350 jan07 4:g7594–g7594.
43. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
44. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019;170:W1.
45. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, 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: e048008.
46. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505.
47. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open. 2021;11: e047709.
48. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc. 2020;27:2011–5.
49. Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol Artif Intell. 2020;2: e200029.
50. Hawksworth C, Elvidge J, Knies S, Zemplenyi A, Petykó Z, Siirtola P, et al. Protocol for the development of an artificial intelligence extension to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022. medRxiv. 2023. https://doi.org/10.1101/2023.05.31.23290788.
51. Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards for Interventions that use Artificial Intelligence (CHEERS-AI). Value Health. 2024;:S1098301524023660.
52. Bilbro NA, Hirst A, Paez A, Vasey B, Pufulete M, Sedrakyan A, et al. The IDEAL Reporting Guidelines: A Delphi Consensus Statement Stage Specific Recommendations for Reporting the Evaluation of Surgical Innovation. Ann Surg. 2021;273:82–5.

53. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. The Lancet. 2009;374:1105–12.
54. Marcus HJ, Bennett A, Chari A, Day T, Hirst A, Hughes-Hallett A, et al. IDEAL-D Framework for Device Innovation: A Consensus Statement on the Preclinical Stage. Ann Surg. 2022;275:73–9.
55. Lekadir K, Osuala R, Gallin C, Lazrak N, Kushibar K, Tsakou G, et al. FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging. 2021. https://doi.org/10.48550/ARXIV.2109.09658.
56. Dagan N, Devons-Sberro S, Paz Z, Zoller L, Sommer A, Shaham G, et al. Evaluation of AI Solutions in Health Care Organizations — The OPTICA Tool. NEJM AI. 2024;1(1):e2300269.
57. European Commission. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed 21 Apr 2025.
58. The STANDING Together collaboration. Recommendations for Diversity, Inclusivity, and Generalisability in Artificial Intelligence Health technologies and Health Datasets. 2023. Available from: https://zenodo.org/records/10048356.
59. Whicher D, Rapp T. The Value of Artificial Intelligence for Healthcare Decision Making—Lessons Learned. Value Health. 2022;25:328–30.
60. Simon GJ, Aliferis C, editors. Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls. Cham: Springer International Publishing; 2024.
61. Das S, Nayak SP, Sahoo B, Nayak SC. Machine Learning in Healthcare Analytics: A State-of-the-Art Review. Arch Comput Methods Eng. 2024. https://doi.org/10.1007/s11831-024-10098-3.
62. Mello MM, Guha N. Understanding Liability Risk from Using Health Care Artificial Intelligence Tools. N Engl J Med. 2024;390:271–8.
63. Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. Npj Precis Oncol. 2024;8:54.
64. Council of the European Union. 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—analysis of the final compromise text with a view to agreement. Brussels: Council of the European Union; 2024. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206. Accessed 21 Apr 2025.
65. Recommendation on the Ethics of Artificial Intelligence. programme and meeting document [184681]. France: UNESCO; 2022.
66. United Nations. Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. United Nations; 2024. Available from: https://digitallibrary.un.org/record/4040897. Accessed 21 Apr 2025.
67. Biden Jr JR. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. Accessed 22 Apr 2024.
68. European Parliament. Artificial Intelligence Act. Brussels: European Parliament; 2024. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.
69. UNESCO. Readiness assessment methodology: a tool of the Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO; 2023. Available from: https://www.unesco.org/en/articles/readiness-assessment-methodology-tool-recommendation-ethics-artificial-intelligence. Accessed 21 Apr 2025.
70. UK Government. International scientific report on the safety of advanced AI: interim report. London: UK Government; 2024.
71. Josep SG, Sarah DN, Elias B, Ignacio S, Tatjana E, André A-A, et al. Harmonised Standards for the European AI Act. Seville: European Commission; 2024. Available from: https://publications.jrc.ec.europa.eu/repository/handle/JRC139430. Accessed 21 Apr 2025.
72. Ethics and governance of artificial intelligence for health. WHO guidance. Geneva: World Health Organization; 2021.
73. OECD AI. OECD AI Principles overview. https://oecd.ai/en/ai-principles. Accessed 24 Apr 2024.
74. CAIDP. Center for AI and Digital Policy (CAIDP). Universal Guidelines for AI. 2018. https://www.caidp.org/universal-guidelines-for-ai/. Accessed 24 Apr 2024.
75. Asilomar AI Principles. Open Letters. 2017. https://futureoflife.org/open-letter/ai-principles/. Accessed 12 Oct 2024.
76. Ortega-Bolaños R, Bernal-Salcedo J, Germán Ortiz M, Galeano Sarmiento J, Ruz GA, Tabares-Soto R. Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. Artif Intell Rev. 2024;57:110.
77. Prainsack B, Forgó N. New AI regulation in the EU seeks to reduce risk without assessing public benefit. Nat Med. 2024. https://doi.org/10.1038/s41591-024-02874-2.
78. Chaddad A, Peng J, Xu J, Bouridane A. Survey of Explainable AI Techniques in Healthcare. Sensors. 2023;23:634.
79. Cohen IG, Evgeniou T, Gerke S, Minssen T. The European artificial intelligence strategy: implications and challenges for digital health. Lancet Digit Health. 2020;2:e376–9.
80. Liu X, Glocker B, McCradden MM, Ghassemi M, Denniston AK, Oakden-Rayner L. The medical algorithmic audit. Lancet Digit Health. 2022;4:e384–97.
81. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLOS Med. 2018;15: e1002689.
82. Klonoff DC. The Current Status of mHealth for Diabetes: Will it Be the Next Big Thing? J Diabetes Sci Technol. 2013;7:749–58.
83. Zeitoun J-D, Ravaud P. Artificial intelligence in health care: value for whom? Lancet Digit Health. 2020;2:e338–9.
84. Boulais W. Transforming Healthcare with AI: The NUHS Model and Its Global Implications. 2024. https://www.linkedin.com/pulse/transforming-healthcare-ai-nuhs-model-its-global-wayne-boulais-yukac/. Accessed 5 Sep 2024.
85. Meet the doctor whose healthcare innovations are 'out of this world.' NUHS+ Health Inside Out. 2024. https://nuhsplus.edu.sg/article/meet-the-doctor-whose-healthcare-innovations-are--out-of-this-world. Accessed 6 Sep 2024.
86. Shorter hospital waiting times with artificial intelligence. NUHS+ Health Inside Out. 2023. https://nuhsplus.edu.sg/article/shorter-hospital-waiting-times-with-artificial-intelligence. Accessed 6 Sep 2024.
87. Chua CE, Lee Ying Clara N, Furqan MS, Lee Wai Kit J, Makmur A, Tham YC, et al. Integration of customised LLM for discharge summary generation in real-world clinical settings: a pilot study on RUSSELL GPT. Lancet Reg Health - West Pac. 2024;51:101211.
88. Pantuck AJ, Lee D, Kee T, Wang P, Lakhotia S, Silverman MH, et al. Modulating BET Bromodomain Inhibitor ZEN-3694 and Enzalutamide Combination Dosing in a Metastatic Prostate Cancer Patient Using CURATE.AI, an Artificial Intelligence Platform. Adv Ther. 2018;1:1800104.
89. Blasiak A, Truong A, Tan WJ Lester, Kumar KS, Tan SB, Teo CB, et al. PRECISE CURATE.AI: A prospective feasibility trial to dynamically modulate personalized chemotherapy dose with artificial intelligence. J Clin Oncol. 2022;40 16_suppl:1574–1574.
90. Blasiak A, Tan LWJ, Chong LM, Tadeo X, Truong ATL, Senthil Kumar K, et al. Personalized dose selection for the first Waldenström macroglobulinemia patient on the PRECISE CURATE.AI trial. Npj Digit Med. 2024;7:223.
91. Zarrinpar A, Lee D-K, Silva A, Datta N, Kee T, Eriksen C, et al. Individualizing liver transplant immunosuppression using a phenotypic personalized medicine platform. Sci Transl Med. 2016;8.
92. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. Transl Med Commun. 2019;4(1):18. https://doi.org/10.1186/s41231-019-0050-7.
93. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368:m689. https://doi.org/10.1136/bmj.m689.
94. Rajagopal A, Ayanian S, Ryu AJ, Qian R, Legler SR, Peeler EA, et al. Machine Learning Operations in Health Care: A Scoping Review. Mayo Clin Proc Digit Health. 2024;2:421–37.
95. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. Cureus. 2023. https://doi.org/10.7759/cureus.46454.

96. Teo ZL, Kwee A, Lim JC, Lam CS, Ho D, Maurer-Stroh S, et al. Artificial intelligence innovation in healthcare: Relevance of reporting guidelines for clinical translation from bench to bedside. Ann Acad Med Singapore. 2023;52:199–212.
97. Ayorinde A, Mensah DO, Walsh J, Ghosh I, Ibrahim SA, Hogg J, et al. Health Care Professionals' Experience of Using AI: Systematic Review With Narrative Synthesis. J Med Internet Res. 2024;26: e55766.
98. Celi LA, Cellini J, Charpignon M-L, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. PLOS Digit Health. 2022;1: e0000022.
99. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. Biostatistics. 2020;21(2):345–52. https://doi.org/10.1093/biostatistics/kxz041.
100. Ganapathi S, Palmer J, Alderman JE, Calvert M, Espinoza C, Gath J, et al. Tackling bias in AI health datasets through the STANDING Together initiative. Nat Med. 2022;28:2232–3.
101. Ratwani RM, Sutton K, Galarraga JE. Addressing AI Algorithmic Bias in Health Care. JAMA. 2024;332:1051.
102. Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. J Am Med Inform Assoc. 2024;31:1172–83.
103. Ghebrehiwet I, Zaki N, Damseh R, Mohamad MS. Revolutionizing personalized medicine with generative AI: a systematic review. Artif Intell Rev. 2024;57:128.
104. Ibrahim M, Khalil YA, Amirrajab S, Sun C, Breeuwer M, Pluim J, et al. Generative AI for synthetic data across multiple medical modalities: a systematic review of recent developments and challenges. arXiv. 2024. https://doi.org/10.48550/arXiv.2407.00116.
105. Takita H, Kabata D, Walston SL, Tatekawa H, Saito K, Tsujimoto Y, et al. Diagnostic performance comparison between generative AI and physicians: a systematic review and meta-analysis. medRxiv. 2024. https://doi.org/10.1101/2024.01.20.24301563.
106. Beam AL, Manrai AK, Ghassemi M. Challenges to the Reproducibility of Machine Learning Models in Health Care. JAMA. 2020;323:305.
107. Azad TD, Ehresman J, Ahmed AK, Staartjes VE, Lubelski D, Stienen MN, et al. Fostering reproducibility and generalizability in machine learning for clinical prediction modeling in spine surgery. Spine J. 2021;21:1610–6.
108. Alberto IRI, Alberto NRI, Altinel Y, Blacker S, Binotti WW, Celi LA, et al. A scientometric analysis of fairness in health AI literature. PLOS Glob Public Health. 2024;4: e0002513.
109. Yang R, Nair SV, Ke Y, D'Agostino D, Liu M, Ning Y, et al. Disparities in clinical studies of AI enabled applications from a global perspective. Npj Digit Med. 2024;7:209.
110. Serra-Burriel M, Locher L, Development VKN, Pipeline and Geographic Representation of trials for artificial intelligence, machine learning-enabled medical devices (2010 to 2023). NEJM AI. 2024;1(1):AIpc2300038. https://doi.org/10.1056/AIpc2300038.
111. Liu M, Ning Y, Teixayavong S, Mertens M, Xu J, Ting DSW, et al. A translational perspective towards clinical AI fairness. Npj Digit Med. 2023;6:172.
112. Susanto AP, Lyell D, Widyantoro B, Berkovsky S, Magrabi F. Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review. J Am Med Inform Assoc. 2023;30:2050–63.
113. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. Int J Med Inf. 2020;137: 104072.
114. Van De Sande D, Van Genderen ME, Smit JM, Huiskens J, Visser JJ, Veen RER, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health Care Inform. 2022;29: e100495.
115. Katharine M. Should AI Models Be Explainable? That depends. 2021. https://hai.stanford.edu/news/should-ai-models-be-explainable-depends. Accessed 20 Mar 2024.
116. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31–8.
117. Stiefel KM, Coggan JS. The energy challenges of artificial superintelligence. Front Artif Intell. 2023;6:1240653.
118. Jovanovic M, Mitrov G, Zdravevski E, Lameski P, Colantonio S, Kampel M, et al. Ambient Assisted Living: Scoping Review of Artificial Intelligence Models, Domains, Technology, and Concerns. J Med Internet Res. 2022;24: e36553.
119. Wolff J, Pauling J, Keck A, Baumbach J. Systematic Review of Economic Impact Studies of Artificial Intelligence in Health Care. J Med Internet Res. 2020;22: e16866.
120. Wolff J, Pauling J, Keck A, Baumbach J. Success Factors of Artificial Intelligence Implementation in Healthcare. Front Digit Health. 2021;3: 594971.
121. Saenz AD, Mass General Brigham AI Governance Committee, McCoy T, Mantha AB, Martin R, Damiano R, et al. Establishing responsible use of AI guidelines: a comprehensive case study for healthcare institutions. Npj Digit Med. 2024;7:348.

Publisher's Note