Achieving large-scale clinician adoption of AI-enabled decision support

Ian A. Scott [iD] ,[1,2] Anton van der Vegt,[3] Paul Lane,[4] Steven McPhail,[5] Farah Magrabi [iD] [6]

¹Internal Medicine and Clinical Epidemiology, Princess Alexandra Hospital, Brisbane, Queensland, Australia
²Centre for Health Services Research, The University of Queensland Faculty of Medicine and Biomedical Sciences, Brisbane, Queensland, Australia
³Digital Health Centre, The University of Queensland Faculty of Medicine and Biomedical Sciences, Herston, Queensland, Australia
⁴Safety, Quality and Innovation, The Prince Charles Hospital, Brisbane, Queensland, Australia
⁵Australian Centre for Health Services Innovation, Queensland University of Technology Faculty of Health, Brisbane, Queensland, Australia
⁶Macquarie University, Sydney, New South Wales, Australia

Correspondence to
Professor Ian A. Scott;
ian.scott@health.qld.gov.au

ABSTRACT

Computerised decision support (CDS) tools enabled by artificial intelligence (AI) seek to enhance accuracy and efficiency of clinician decision-making at the point of care. Statistical models developed using machine learning (ML) underpin most current tools. However, despite thousands of models and hundreds of regulator-approved tools internationally, large-scale uptake into routine clinical practice has proved elusive. While underdeveloped system readiness and investment in AI/ML within Australia and perhaps other countries are impediments, clinician ambivalence towards adopting these tools at scale could be a major inhibitor. We propose a set of principles and several strategic enablers for obtaining broad clinician acceptance of AI/ML-enabled CDS tools.

STANDFIRST

New artificial intelligence (AI)-enabled technologies for augmenting clinical decision-making are proliferating but clinicians will only use them if convinced of their worth. Dr Ian Scott and colleagues outline 10 principles and 5 enabling system strategies that could promote wider adoption by clinicians.

AI-enabled computerised decision support (CDS) tools seek to augment the accuracy and efficiency of clinician decision-making at the point of care. Currently, conventional, task-specific models developed using supervised machine learning (ML) underpin most current clinician-facing AI-enabled CDS tools. These are dominated by diagnostic imaging and risk prediction tools.[1] However, large language models (LLMs) and generative AI, such as ChatGPT, are poised to revolutionise care given their ability to converse with clinicians and perform multiple tasks, ranging from clinical documentation to multidomain decision support. However, despite hundreds of regulator-approved ML tools internationally,[2] large-scale uptake into routine clinical practice has proved elusive.[3] While many non-clinical factors may partly account for this adoption gap,[4] ambivalence of frontline clinicians towards using AI tools may also

contribute, principally due to a lack of understanding of and trust in, AI applications.[5,6] We propose a set of principles and strategic enablers for achieving broad clinician acceptance of AI tools embedded within electronic medical records (EMRs). As no LLM has yet received regulator approval in clinical care, our focus is on approved conventional ML tools, although we would contend all the principles discussed will pertain equally to LLMs. This work builds on previous experience with digitally-enabled rule-based CDS systems[7] and is informed by recent research into AI implementation barriers and enablers.[3,8,9] There was no patient or public involvement in writing this article as our focus was clinician-facing tools.

PRINCIPLES FOR PROMOTING ADOPTION
The tool must address a pressing clinical need

Tools must enhance decision-making for commonly encountered scenarios where current clinical judgement is suboptimal, such as early detection of sepsis[10] or timely diagnosis of stroke.[11] Use of AI tools by clinicians in such instances can improve patient care,[12,13] and these tools do not have to be perfectly accurate. A modestly accurate tool substantially better than current clinical judgement will be favoured over a highly accurate tool no better than current judgement.[14] AI tools must also perform better than current well-accepted, high-performing but simpler decision rules.[15] Tool developers, collaborating with clinicians, must first deeply understand the clinical task and the data sets being targeted and why, their amenability to AI, current clinical decisional performance, clinician end-user needs and the primary goal(s) to be achieved.[16] These goals should ideally be expressed as measurable targets in improved clinical processes and outcomes, patient and professional experience, economic and efficiency gains or greater equity and sustainability in care delivery.

> **Box 1 Shortcomings in comparative studies of artificial intelligence versus clinician[21 22]**
>
> A systematic review of 82 studies compared the diagnostic accuracy of deep learning tools versus clinicians in classifying diseases using medical images.[21] Most studies had several limitations that biased against clinicians:[21 22]
> ⇒ Model accuracy was assessed in isolation in ways that do not reflect clinical practice.
> ⇒ Very few studies reported comparisons with clinicians using the same test data set.
> ⇒ Clinicians were rarely provided with additional clinical information, as they would have been in usual clinical practice.
> ⇒ Diagnostic criteria for disease were often poorly defined.
> ⇒ Performance metrics varied greatly across studies, and many were under-reported.
> ⇒ External validation was not done for both the tool and the clinician.
> ⇒ Very few prospective studies performed using live data in real-world clinical environments.
> ⇒ No randomised trials.

The tool must demonstrate clinically meaningful improvements in care

Clinicians need to know if deployed AI tools will improve patient care and outcomes to an extent they and their patients would regard as clinically relevant, irrespective of the statistical significance of reported results. Whether an effect is clinically important depends on the nature of the condition, the effect, and the context such as patient population and clinical setting. Minimally important absolute effects may range from a 5% decrease in deaths[17] to as high as a 40% decrease in pain.[18] Prospective impact studies of clinically deployed tools are few and incomplete. In one review, only one-third of 51 studies examined patient outcomes, with mixed results (8 positive effects, 6 no change).[1] In a more recent review of 32 studies, only 8 (25%), 10 (31%) and 12 (38%) assessed effects on decision-making, care delivery and patient outcomes, respectively, in all cases reporting mixed results.[19] Randomised trials are even fewer, mostly involving imaging tools and limited by high variability in adherence to current reporting standards, risk of bias, under-representation of minority groups, small samples and single site designs.[20] Other studies contain methodological flaws that bias against clinician judgement (box 1).[21 22] Training data must be representative of populations to which the tool will be applied and models must undergo rigorous external validation. Impact effects in absolute terms are also often small, with a review of 122 trials of CDS tools showing the proportions of patients receiving recommended care increasing by an average of only 5.8 percentage points.[23]

The tool is, and remains, accurate and safe for the chosen task

Tools may generate inaccurate and unsafe advice if their models have been trained on inadequate or unrepresentative (biased) data,[24] used in an inappropriate clinical

> **Box 2 Calibrating artificial intelligence tools in optimising clinical utility**
>
> A failure to recognise clinical deterioration in the hospital due to sepsis or other potentially life-threatening conditions is a leading cause of in-hospital death and unplanned transfers to intensive care units. Early warning systems (EWS) can predict a patient's risk of clinical deterioration, and potentially allow clinicians to intervene earlier. Current EWS comprise simple prediction rules to estimate risk based on a combination of a small number of input variables, usually fewer than 10, such as vital signs. The rules only offer a narrow time window, usually less than 12 hours, to trigger an alert prior to overt deterioration that activates a medical emergency team response. The rules are also prone to false-positive alerts which induce alert fatigue. An EWS that uses machine learning could make more accurate and timely predictions given its ability to input hundreds of variables.
> The ideal prediction tool should miss very few cases of clinical deterioration (high sensitivity) and not overall cases with no deterioration (high specificity). Clinicians may decide the tool should aim for no more than two false alerts for every true positive alert in order to balance the time required to assess alert patients with other competing demands. The data scientists would then set the threshold for categorising patients as high risk at a positive predictive value of around 30%. At this threshold, based on historical data, the sensitivity may be only 50%, but clinicians may decide this would be a useful proportion of cases to detect. Clinicians may find the tool more useful if it can predict events within the following 48 hours. A shorter window would not leave enough time to intervene, and a longer window would make it difficult for clinicians to know how to respond.
> In adjusting sensitivity thresholds and striking the right balance between clinician workload and patient safety, input from clinician users is required. Such adjustments will also vary according to the criticality of the event being predicted, for example, pressure sores versus septic shock.

setting or context, misinterpret minor data set shifts that clinicians know to ignore or account for (ie, changes in patient, clinical practice or equipment characteristics), or which under-sense (too few alerts resulting in harm) or over-sense (too many causing alert fatigue) (box 2). Data required to operate the tool must be accurate, representative and readily accessible when needed and models must be resilient to class imbalance (ie, outcomes being predicted are infrequent) and label leakage (ie, using image background or other artefacts to make predictions rather than clinically relevant features).

For all these reasons, rigorous external validation of acceptable model performance when used in different populations by different clinicians[25] is paramount, together with an ability to retrain models on local data if performance is found to be suboptimal. Importantly, clinicians want to know when and for whom a tool should, and should not, be used (ie, clear, transparent task specification). Ideally, information should be forthcoming about how the model was trained, who was included in the data set, what its performance is like, who funded its development and what assumptions or conditions should be satisfied for its use.[26] Tool developers should share model code and input features to allow other researchers

---

Box 3 Limitations of attempts to render artificial intelligence (AI) models and tools fully explainable[28–31]

⇒ There is a lack of agreement on the different levels of explainability, no clear guidance on how to choose among different explainability methods and an absence of standardised methods for evaluating explainability.[28]

⇒ The value to clinicians of any explanation will vary according to the specific model and its task (or use case) and the expertise (ie, level of AI or domain knowledge), preferences for accuracy relative to explainability and other contextual values of the clinician user.[29]

⇒ The more complex the model, especially deep learning models, the less explainable it becomes and hence expecting clinicians (and patients) to master the technical and statistical intricacies of most models is unrealistic.

⇒ Explainability methods commonly used to identify model input features strongly influencing its predictions,* while useful in making input–output relationships clearer, are imperfect post hoc approximations of model functions rather than precise explanations of the inner workings of the model.

⇒ Explainability methods may present plausible but misleading explanations, do not ensure the model has considered all relevant features,[30] and may hamper human ability to detect model mistakes, resulting in decreased vigilance and auditing of AI tools and over-reliance on their outputs.[29 30]

⇒ Clinician experts will question the clinical plausibility of implied causal relationships involving predictive input features identified by explainability methods, will assess how well tool outputs align with observable clinical features and prioritise established knowledge and experience over finding novel but potentially spurious associations.[30]

⇒ Citizen jurors, when faced with two healthcare scenarios in one UK study, favoured accuracy over explainability of AI tools because of the potential for harm from inaccurate predictions and the potential of accurate tools to increase the efficiency of, and access to, services.[31]

*These methods comprise Locally Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP) and heat or saliency maps.

Box 4 Human factor principles applicable to artificial intelligence/tools[33 34]

Any tool must:

⇒ Sit and operate seamlessly within existing digital platforms such as an electronic medical record already familiar to users and be readily accessible.

⇒ Be automated and not incur unacceptable delays in providing necessary advice for time-sensitive decision-making and operate at the right time in the clinical trajectory.

⇒ Have a standardised visualisation and delivery of outputs that is minimally interruptive.

⇒ Require no or very little manual data entry by clinicians.

⇒ Minimise clerical tasks and added work generated by its use (eg, extra clicks, menu navigation, more documentation).

⇒ Be able to operate on mobile devices where required.

⇒ Reflect a 'human-centred' design approach that adapts to user needs rather than a 'technology-centred' approach that expects users to adapt to the technology.

---

to reproduce and reconfirm model performance using different data sets from different settings.

The tool outputs must be comprehensible and actionable, but not necessarily fully explainable in how they were derived

Tools should produce user-friendly visualisations of outputs that are readily understood and clinically actionable, especially for more inexperienced clinicians. Evidence suggests clinicians desire graphical or numerical displays of probabilities or alert thresholds for a diagnosis or event, confidence scores for these outputs and links to relevant, consistent recommendations for tests or treatments.[27] However, decisional discretion must remain based on clinician/patient preferences and clinician judgement about possible model bias or clinical and situational factors unknown to the model. In comparing simpler and more explainable models with complex but more accurate ones, clinicians will likely trade-off model explainability for greater accuracy, as full explainability is, in many instances, neither possible[28] nor necessary

for both clinician[29 30] and patient acceptance[31] (box 3). Greater explainability may be warranted for high-stakes, nuanced decision-making such as choosing the right antibiotic in a septic, immunosuppressed patient or determining organ donor and recipient matches.

The tool must align with clinical workflows

Tools must be easy to use with intuitive human-computer interfaces that standardise output visualisation, blend seamlessly into clinical workflows, avoid creating workarounds and alert fatigue, customise the alert sensitivity to local populations and prevent cognitive overload and over-reliance on automated decisions. The requirements for integration may vary according to whether tools are assistive (ie, offering predictions for clinicians to consider) or more autonomous (explicit determinations directly influencing clinical actions). Involving clinician end-users is critical in providing current operational context, pre-empting training and support needs and raising awareness of how incorrect tool use by clinicians, such as inputting data errors, misinterpreting information displays or clicking wrong options, can incur patient harm.[32] All these human factors relevant to AI tool use have to date been underemphasised[33 34] (box 4).

The tool must operate within a governance and regulatory framework

Clinicians will want an organisational governance framework that guarantees all the previously stated principles are met at inception, and continue to be met over the life cycle of the tool.[35] Such a framework will determine when adoption should proceed or be revoked if the model proves valueless, is not implementable, does not operate across sites, fails in prospective evaluations or leads to potentially unsafe over-reliance. Clinicians will also demand a regulatory framework that determines, under software as a medical device legislation, when liability for errors and resultant patient harm from tool

use lies primarily with them and their personal indemnity insurer (eg, negligent, reckless or 'off-label' use), or their employing organisation, or tool developers and vendors.[36] Liability may extend to 'failure to use' if using a specific AI tool becomes a practice standard for certain clinical scenarios. Such frameworks remain works in progress in most jurisdictions, trying to balance regulation with innovation and aligning it with evolving clinical governance procedures. More autonomous tools or those directly impacting critical clinical decisions will require greater regulatory oversight and higher levels of safety evidence for approval.[37] Ongoing monitoring of tool performance, tool auditing processes[38] and in-built self-improvement feedback loops will be needed in ensuring tool resilience to data set shifts, noise and cyberattacks.[39]

The tool must not compromise the clinician-patient relationship

Using AI tools, especially LLMs, to produce evidence syntheses, clinical letters and discharge summaries may free up cognitive time and space for clinicians to engage more in empathetic, person-centred shared decision-making (SDM). However, more information is needed about the true impacts of AI tools on clinician-patient interactions in different contexts,[40] tool designs that best support each step of SDM,[41] how to obtain patient consent to AI being used to assist SDM and the circumstances in which care is not compromised if patients may not want to know, or are able to comprehend, model predictions.

The tool must not promote overdiagnosis and overtreatment

Tools used in screening programmes may promote overdiagnosis of benign or indolent disease by the inclusion of a loose disease definition in the model, overdetection of minor abnormalities or misinterpretation of normal physiological variation as pathological due to continuous monitoring of multiple variables over prolonged time periods. For example, increased AI detection of non-progressive ductal carcinoma in situ on screening mammograms[42] may incite overtreatment which carries ethical and economic implications. Clinical studies are needed that assess outcome impacts according to different definitions of disease and patient risk, and which should prompt greater collaborative efforts at rendering disease definitions more explicit. Over time, models need to become more capable of differentiating between benign variations and true disease.

The tool must promote health equity

AI tools must alleviate, not exacerbate, health disparities. Model bias is often disproportionately distributed to underserved populations with poorer health, reinforcing the need for representative training data. The tools and required digital infrastructure must be accessible to such populations, as well as treatments and interventions for treating identified diseases or risks. Such equity requirements go beyond the tool itself to the capacity and responsiveness of the healthcare system more broadly.

The tool must not incur excessive opportunity cost or environmental impacts

Developing, testing and deploying tools cost money: data scientists for gathering and pre-processing input data; clinicians for labelling data sets; information and communication technology (ICT) staff for converting models into software-embedded tools and training staff. Added to this are ongoing life-cycle costs of maintaining the tool and hardware and redressing the effects of tool-induced errors. Carbon emissions from training and deploying AI models must also be weighed against the potential for models to reduce emissions through improved process efficiency and changing models of care.[43] The few economic evaluations of AI tools are of limited quality,[44] mostly cost minimisation analyses of specific cost elements within single-use cases over short time horizons and with no emissions quantification. For clinicians, a key consideration is estimating, for the outcome being predicted, the number of patients the tool flags as being positive, thereby incurring costs of preventive or therapeutic interventions, versus the number of true-positives.[45] This equation and the estimated costs will vary according to what clinicians perceive as the most clinically appropriate sensitivity and specificity thresholds or cut-off points for the tool which, using simulation methods, determine the net monetary benefit.[46]

Strategic enablers for greater adoption

Several cross-cutting strategic enablers may facilitate the enactment of these 10 principles.

1. Enhance AI literacy of clinicians: Clinicians need to have an understanding of the basic concepts of AI/ML tool design and evaluation in gauging its appropriate use.[47] This requires the provision of educational resources,[48] sets of AI competencies[49] and interdisciplinary training programmes involving AI specialists and clinicians. When a tool is deployed, there must be adequate training, technical support and onboarding of new clinician users.[50] Healthcare institutions will need to provide the time, money and personnel required for such activities.

2. Establish interdisciplinary AI teams: At the local organisational level, clinicians must partner with data and computer scientists, ICT personnel, vendor representatives and consumers in forming multistakeholder co-design groups tasked to select, develop, test, deploy and monitor AI tools most relevant to addressing locally prioritised needs.[51] Such collaboration must also be extended to regulators in formulating workable regulatory frameworks, all of which promote clinician receptivity to AI.

3. Streamline and harmonise data access and sharing procedures: Collaborative, multistakeholder efforts are needed to build and curate large repositories of diverse, accurate, multimodal data from EMRs and other sources necessary for training high-performing models acceptable to clinicians and applicable to different populations and clinical settings. Siloing of data and

cumbersome data access approval processes involving multiple data custodians must be replaced by efficient, standardised processes for accessing and sharing data from EMR and other sources which is rendered interoperable using data exchange standards (eg, Health Level Seven Fast Healthcare Interoperability Resource and Observational Medical Outcomes Partnership).[52] Concurrently, data privacy and security must be safeguarded under umbrella instruments such as the General Data Protection Regulation.

4. Establish platforms for integrating and testing tools within EMRs: A testing infrastructure is needed whereby prototype AI tools can be integrated into current EMRs, using application programming interfaces, and their performance compared with standard care in 'silent trials' or 'shadow mode' conducted within live-data clinical environments. These activities and subsequent clinical trials should be conducted with clinician oversight, prior to full roll-out.[53] This approach avoids delays in undertaking full-platform EMR reconfigurations to facilitate such testing, while allowing clinician-informed customisation in prototype design and functionality. It also facilitates trialability of the tool in that, even without a deep understanding of AI, clinicians can build trust through experience in using it, seek expert endorsement and validation and help design a tool that accommodates their autonomy and expertise, while providing a 'second pair of eyes' and supporting them across their entire workflow, not just for a one-off task.[54]

5. Invest in and use implementation science targeting AI tools: Research into successful translation of AI tools into clinical practice is nascent with few examples of applied implementation science.[55] There is a critical need for metrics and methods to measure success and identify areas for improvement. Only recently have step-by-step implementation frameworks been developed and validated which clearly delineate the different phases, both clinical and technical, of tool development and deployment and the decision points, enablers and barriers at each phase.[8 9] Such frameworks sit under overarching system issues related to organisational readiness for AI and the broader ethical, legal and policy environment in which AI tools will operate.

CONCLUSION

The current adoption gap for the ever-increasing number of AI-enabled CDS tools will persist if clinicians remain unconvinced of their utility in clinical decision-making. While not intended to be an exhaustive list, the principles and enablers enunciated here may help guide actions all stakeholders will need to take in closing the gap and which align with modern concepts of ethically responsible use of AI.

ORCID iDs
Ian A. Scott http://orcid.org/0000-0002-7596-0837
Farah Magrabi http://orcid.org/0000-0002-8426-5588

REFERENCES

1 Yin J, Ngiam KY, Teo HH. Role of artificial intelligence tools in real-life clinical practice: systematic review. *J Med Internet Res* 2021;23:e25759.
2 Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med* 2020;3:118.
3 Magrabi F, Lyell D, Coiera E. Automation in contemporary clinical information systems: a survey of AI in clinical settings. *Yearb Med Inform* 2023;32:115–26.
4 Coiera EW, Verspoor K, Hansen DP. We need to chat about artificial intelligence. *Med J Aust* 2023;219:98–100.
5 Scott IA, Carter SM, Coiera E. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Health Care Inform* 2021;28:e100450.
6 Terry AL, Kueper JK, Beleno R, *et al*. Is primary health care ready for artificial intelligence? What do primary health care stakeholders say? *BMC Med Inform Decis Mak* 2022;22:237.
7 Magrabi F, Cresswell K, Fraser HS. International dimensions of clinical decision support systems. In: Greenes R, Fiol GD, eds. *Clinical decision support and beyond: progress and opportunities in knowledge-enhanced health and Healthcare*. London: Elsevier Academic Press, 2023: 145–88.
8 van der Vegt AH, Scott IA, Dermawan K, *et al*. Implementation frameworks for end-to-end clinical AI: derivation of the SALIENT framework. *J Am Med Inform Assoc* 2023;30:1503–15.
9 van der Vegt AH, Scott IA, Dermawan K, *et al*. Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework. *J Am Med Inform Assoc* 2023;30:1349–61.
10 Barker N, Scott IA, Seaton R, *et al*. Recognition and management of hospital-acquired sepsis among general medical inpatients in queensland public hospitals. *Int J Gen Med* 2023;16:1039–46.
11 Tarnutzer AA, Lee S-H, Robinson KA, *et al*. ED misdiagnosis of cerebrovascular events in the era of modern neuroimaging: a meta-analysis. *Neurology* 2017;88:1468–77.
12 Shimabukuro DW, Barton CW, Feldman MD, *et al*. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res* 2017;4:e000234.

13 Gunda B, Neuhaus A, Sipos I, *et al*. Improved stroke care in a primary stroke centre using AI decision support. *Cerebrovasc Dis Extra* 2022;12:28–32.

14 Sanders S, Doust J, Glasziou P. A systematic review of studies comparing diagnostic clinical prediction rules with clinical judgment. *PLoS One* 2015;10:e0128233.

15 Challen R, Denny J, Pitt M, *et al*. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28:231–7.

16 Huo W, Yuan X, Li X, *et al*. Increasing acceptance of medical AI: the role of medical staff participation in AI development. *Int J Med Inform* 2023;175:105073.

17 Abrams D, Montesi SB, Moore SKL, *et al*. Powering bias and clinically important treatment effects in randomized trials of critical illness. *Crit Care Med* 2020;48:1710–9.

18 Ferreira ML, Ferreira PH, Herbert RD, *et al*. People with low back pain typically need to feel 'much better' to consider intervention worthwhile: an observational study. *Aust J Physiother* 2009;55:123–7.

19 Susanto AP, Lyell D, Widyantoro B, *et al*. 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.

20 Plana D, Shung DL, Grimshaw AA, *et al*. Randomised clinical trials of machine learning interventions in health care. A systematic review. *JAMA Netw Open* 2022;5:e2233946.

21 Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;1:e271–97.

22 Nagendran M, Chen Y, Lovejoy CA, *et al*. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020;368:m689.

23 Kwan JL, Lo L, Ferguson J, *et al*. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *BMJ* 2020;370:m3216.

24 Gianfrancesco MA, Tamang S, Yazdany J, *et al*. Potential biases in machine learning models using electronic health record data. *JAMA Intern Med* 2018;178:1544–7.

25 Lyons PG, Hofford MR, Yu SC, *et al*. Factors associated with variability in the performance of a proprietary sepsis prediction model across 9 networked hospitals in the US. *JAMA Intern Med* 2023;183:611–2.

26 Sendak MP, Gao M, Brajer N, *et al*. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digit Med* 2020;3:41.

27 Tschandl P, Rinner C, Apalla Z, *et al*. Apalla Z, et al Human-computer collaboration for skin cancer recognition. *Nat Med* 2020;26:1229–34.

28 Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745–50.

29 Bienefeld N, Boss JM, Lüthy R, *et al*. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *NPJ Digit Med* 2023;6:94.

30 Tonekaboni S, Joshi S, McCradden MD, *et al*. What clinicians want: contextualising explainable machine learning for clinical end use. *Proc Mach Learn Res* 2019;106:359–80.

31 van der Veer SN, Riste L, Cheraghi-Sohi S, *et al*. Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries. *J Am Med Inform Assoc* 2021;28:2128–38.

32 Lyell D, Wang Y, Coiera E, *et al*. More than models: an analysis of safety events involving ML-enabled medical devices reported to the FDA. *J Am Med Inform Assoc* 2023;30:1227–36.

33 Asan O, Choudhury A. Research trends in artificial intelligence tools in human factors health care: mapping review. *JMIR Hum Factors* 2021;8:e28236.

34 Salwei ME, Carayon P. A Sociotechnical systems framework for the tool of artificial intelligence in health care delivery. *J Cogn Eng Decis Mak* 2022;16:194–206.

35 Reddy S, Allan S, Coghlan S, *et al*. A governance model for application of AI in healthcare. *J Am Med Inform Assoc* 2020;27:491–7.

36 Hwang TJ, Kesselheim AS, Vokinger KN. Life-cycle regulation of artificial Intelligence- and machine learning-based software devices in medicine. *JAMA* 2019;322:2285–6.

37 Bitterman DS, Aerts H, Mak RH. Approaching autonomy in medical artificial intelligence. *Lancet Digit Health* 2020;2:e447–9.

38 Feng J, Phillips RV, Malenica I, *et al*. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI models in healthcare. *NPJ Digit Med* 2022;5:66.

39 Liu X, Glocker B, McCradden MM, *et al*. The medical model audit. *Lancet Digit Health* 2022;4:e384–97.

40 Sauerbrei A, Kerasidou A, Lucivero F, *et al*. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. *BMC Med Inform Decis Mak* 2023;23:73.

41 Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, *et al*. Tool of artificial intelligence in shared decision making: scoping review. *JMIR Med Inform* 2022;10:e36199.

42 Lång K, Josefsson V, Larsson A-M, *et al*. Artificial intelligence-supported screen reading versus standard double reading in the mammography screening with artificial intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol* 2023;24:936–44.

43 Bloomfield PS, Clutton-Brock P, Pencheon E, *et al*. Artificial intelligence in the NHS: climate and emissions. *J Clim Change Health* 2021;4:100056.

44 Voets MM, Veltman J, Slump CH, *et al*. Systematic review of health economic evaluations focused on artificial intelligence in healthcare: the tortoise and the cheetah. *Value Health* 2022;25:340–9.

45 Liu VX, Bates DW, Wiens J, *et al*. The number needed to benefit: estimating the value of predictive analytics in healthcare. *J Am Med Inform Assoc* 2019;26:1655–9.

46 Parsons R, Blythe R, Cramb SM, *et al*. Integrating economic considerations into cutpoint selection may help align clinical decision support toward value-based healthcare. *J Am Med Inform Assoc* 2023;30:1103–13.

47 Scott I, Carter S, Coiera E. A clinician checklist for assessing the suitability of machine learning applications in healthcare. *BMJ Health Care Inform* 2021;28:e100251.

48 Scott IA. Demystifying machine learning: a primer for physicians. *Intern Med J* 2021;51:1388–400.

49 Russell RG, Lovett Novak L, Patel M, *et al*. Competencies for the use of artificial intelligence–based tools by health care professionals. *Acad Med* 2023;98:348–56.

50 Wisniewski H, GorrindoT, Rauseo-Ricupero N, *et al*. The role of digital navigators in promoting clinical care and technology integration into practice. *Digit Biomark* 2021;4:119–35.

51 Schwartz JM, Moy AJ, Rossetti SC, *et al*. Clinician involvement in research on machine learning-based predictive clinical decision support for the hospital setting: a scoping review. *J Am Med Inform Assoc* 2021;28:653–63.

52 Pereira T, Morgado J, Silva F, *et al*. Sharing biomedical data: strengthening AI development in healthcare. *Healthcare* 2021;9:827.

53 Corbin CK, Maclay R, Acharya A, *et al*. DEPLOYR: a technical framework for deploying custom real-time machine learning models into the electronic medical record. *J Am Med Inform Assoc* 2023;30:1532–42.

54 Henry KE, Kornfield R, Sridharan A, *et al*. Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *NPJ Digit Med* 2022;5:97.

55 Yousefi Nooraie R, Lyons PG, Baumann AA, *et al*. Equitable implementation of artificial intelligence in medical imaging: what can be learned from implementation science? *PET Clinics* 2021;16:643–53.