Clinical *Standard*

CLINICAL STANDARD · 2026
DOI PENDING · CC BY 4.0

CLINICAL REVIEW · SURGERY · PUBLIC HEALTH · METHODS & STATISTICS

Artificial intelligence in clinical research: a narrative review of reporting guidelines, regulation and publication ethics

Ahmed Eldesouky[1*]

[1] Department of Trauma and Orthopaedics, Leeds Teaching Hospitals NHS Trust, Leeds, UK

[*] Correspondence: Ahmed Eldesouky, ahmed.eldesoki@nhs.net

ABSTRACT

Artificial intelligence (AI) is now used throughout clinical research, both as the subject of studies and as a tool for carrying them out. This narrative review summarises the reporting guidelines, appraisal tools, regulatory frameworks and publication rules that apply to AI research in medicine as of September 2026. Sources were identified through the EQUATOR Network library, PubMed and the websites of regulatory and editorial bodies. AI-specific reporting guidelines now cover trial protocols and reports, early clinical evaluations, diagnostic accuracy studies, prediction models, large language models, chatbot evaluations, medical imaging and radiomics, and an updated tool published in 2025 appraises risk of bias in prediction models built with regression or AI. In the European Union, the high-risk obligations of the AI Act for AI in medical devices were postponed in July 2026 and now apply from August 2028, while US Food and Drug Administration (FDA) guidance on AI-enabled devices remains in draft. Editorial bodies agree that AI tools cannot be authors and that their use must be disclosed. The evidence behind deployed tools is limited: of 1,357 FDA-cleared AI devices, 34 (2.5%) were linked to a registered prospective trial and 3 (0.2%) had been evaluated for patient-centred outcomes. Researchers should apply the relevant guideline from the protocol stage and give priority to external validation and outcome studies.

Keywords: Artificial Intelligence · Machine Learning · Large Language Models · Guidelines as Topic · Research Design · Device Approval · Ethics · Research

Key points

- Choose the AI reporting guideline by study design, and apply it when writing the protocol rather than at submission.
- Validate models externally, in a different site, period or population, before claiming clinical usefulness.
- Report calibration and subgroup performance alongside discrimination, and state where the model performs poorly.
- In studies of large language models, report the model version, settings, exact prompts and date of access.
- Disclose any AI used in the research or the writing, and check every reference against its original source.

Introduction

Most clinicians first met artificial intelligence (AI) through papers reporting an area under the receiver operating characteristic curve (AUC) somewhere in the 0.90s. Far fewer of those models were ever tested in a second hospital, and fewer still in a trial with patients. A recent census of every AI-enabled device cleared by the US Food and Drug Administration (FDA) up to December 2025 makes the point starkly. Of 1,357 devices, 34 (2.5%) were linked to a registered prospective

trial, and only three (0.2%) had been evaluated for patient-centred outcomes such as mortality, morbidity or readmission [1]. Radiology accounted for 78% of the cleared devices [1]. An earlier analysis of FDA device summaries found similar gaps in how devices were tested and described [2].

The research literature behind these tools has the same weakness. In a meta-analysis of AI in digital pathology, almost all of the included diagnostic accuracy studies had at least one domain at high or unclear risk of bias [3]. Real-world failures show why this matters. When a widely deployed proprietary sepsis model was validated externally, it missed about two thirds of patients with sepsis and generated alerts for 18% of all hospitalised patients [4].

Good evidence is possible, and it is persuasive when it arrives. In the MASAI randomised trial in Sweden, AI-supported screen reading of mammograms was compared with standard double reading; the AI arm detected at least as many cancers while cutting radiologists' screen-reading workload by 44% [5]. The field needs more studies of this kind: prospective, randomised where possible, and measured against what happens to patients and services.

Over the past two years, the rules for designing, reporting and publishing this research have changed quickly. This review brings together the current reporting guidelines, appraisal tools, regulatory frameworks and publication rules, and describes the errors that most often undermine AI studies.

Sources and methods

This is a narrative review. The EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network library [6] and PubMed were searched for AI-specific reporting guidelines and appraisal tools, and the websites of the FDA, the European Commission, the European Medicines Agency (EMA), the International Council for Harmonisation (ICH), the World Health Organization (WHO) and international editorial bodies were searched for regulatory and publication guidance, up to September 2026. PubMed and the websites of medical journals were also searched for recent empirical studies of the evidence behind AI devices. Documents were selected for their relevance to clinicians who design, report or appraise AI research; no formal quality assessment was performed.

AI inside the research process

AI is no longer only the thing we study. Researchers now use it to search and screen the literature, draft protocols, extract variables from free-text records, segment images, write analysis code and edit manuscripts. Each of these uses has a typical way of going wrong. Literature tools can miss studies or invent citations. Protocol drafts can contain plausible but incorrect methods. Extraction and segmentation can carry labelling errors into a dataset unnoticed, and generated code can run without error while computing the wrong quantity.

These failures are usually quiet. The output looks finished, so the checking falls to the researcher, and the more fluent the tool, the easier the checking is to skip. In practice this means a human second screener for systematic reviews, a statistician who reads any generated code, a description of who labelled the data and how, and a check of every reference against PubMed or its digital object identifier.

Choosing a reporting guideline

Choosing a guideline is simpler than the list of acronyms suggests. Start from the study design, then add the AI extension. Most extensions build on a parent guideline, so authors complete both. Table 1 matches common designs to the relevant guideline, with full references, and Figure 1 presents the same choices as a decision guide.

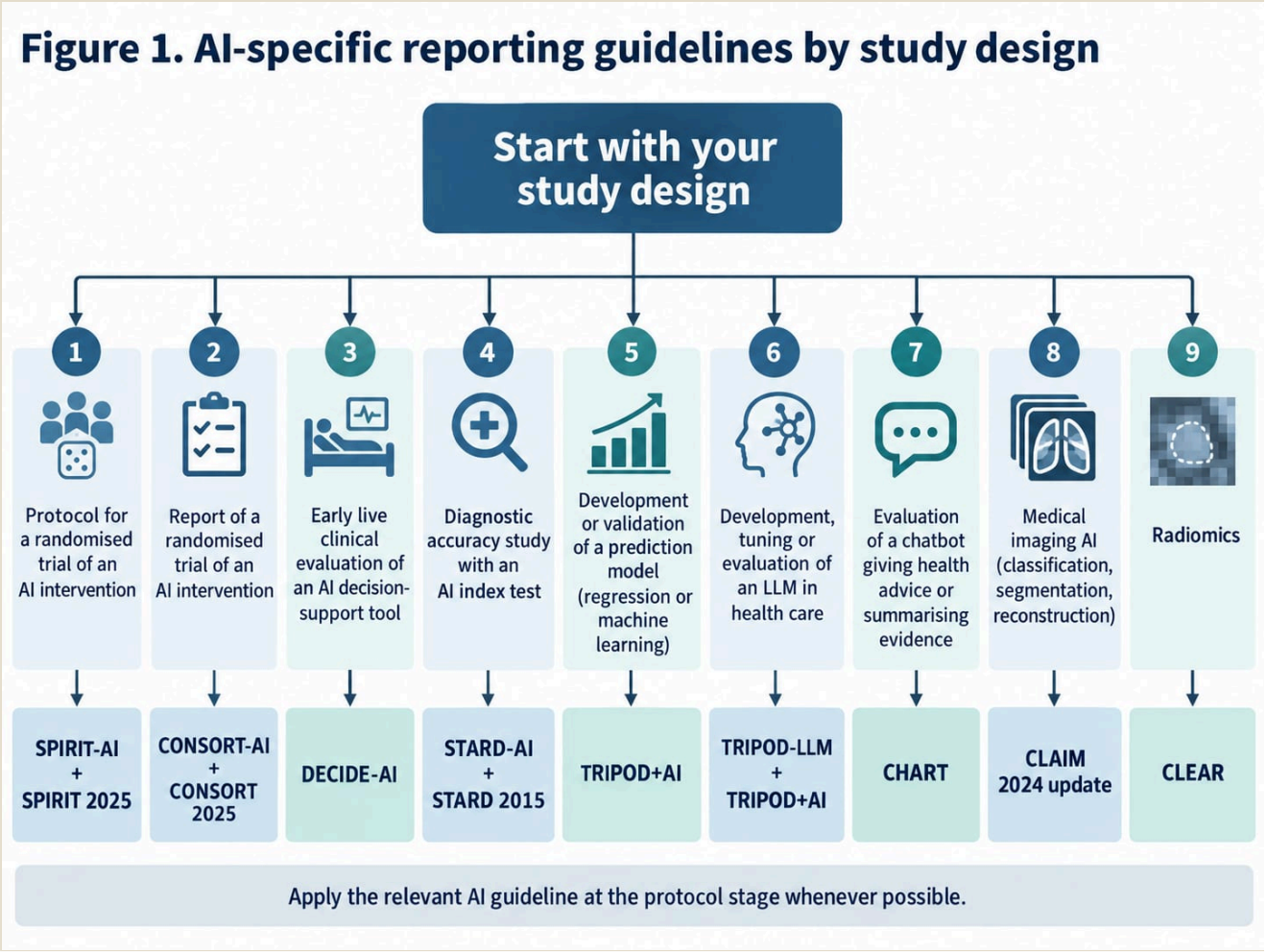


Figure 1. Artificial intelligence-specific reporting guidelines by study design Decision guide matching nine common study designs to the relevant AI reporting guideline and, where one exists, its parent guideline. Full references for each guideline are given in Table 1. AI, artificial intelligence; CHART, Chatbot Assessment Reporting Tool; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; CLEAR, CheckList for EvaluAtion of Radiomics research; CONSORT, Consolidated Standards of Reporting Trials; DECIDE-AI, Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence; LLM, large language model; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; STARD, Standards for Reporting of Diagnostic Accuracy Studies; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Table 1. Artificial intelligence-specific reporting guidelines by study design

Study design	AI guideline	Parent guideline
Protocol for a randomised trial of an AI intervention	SPIRIT-AI [7]	SPIRIT 2025 [8]
Report of a randomised trial of an AI intervention	CONSORT-AI [9]	CONSORT 2025 [10]
Early live clinical evaluation of an AI decision-support tool	DECIDE-AI [11]	None
Diagnostic accuracy study with an AI index test	STARD-AI [12]	STARD 2015 [13]
Development or validation of a prediction model (regression or machine learning)	TRIPOD+AI [14]	Replaces TRIPOD 2015
Development, tuning or evaluation of an LLM in health care	TRIPOD-LLM [15]	TRIPOD+AI
Evaluation of a chatbot giving health advice or summarising evidence	CHART [16,17]	None
Medical imaging AI (classification, segmentation, reconstruction)	CLAIM 2024 update [18]	Modelled on STARD
Radiomics	CLEAR [19]	None

AI, artificial intelligence; CHART, Chatbot Assessment Reporting Tool; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; CLEAR, CheckList for EvaluAtion of Radiomics research; CONSORT, Consolidated Standards of Reporting Trials; DECIDE-AI, Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence; LLM, large language model; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; STARD, Standards for Reporting of Diagnostic Accuracy Studies; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Several of these guidelines are recent. STARD-AI, published in September 2025, adds 18 new or modified items to STARD 2015 and drew on more than 240 international stakeholders [12]. The Chatbot Assessment Reporting Tool (CHART), published jointly in six journals in 2025, has 12 items and 39 sub-items agreed through a Delphi process with 531 participants [16]. The SPIRIT and CONSORT statements for trial protocols and reports were themselves updated in 2025 [8,10], so AI trials should now be read against the 2025 parent versions.

Two habits prevent most problems. The first is to open the guideline when writing the protocol, because items such as the reference standard or the handling of poor-quality inputs cannot be fixed after the data are collected. The second is to complete both checklists when a study spans two designs, for example an imaging model that predicts a clinical outcome, and to say so in the methods.

Appraisal of AI studies

Reporting guidelines tell authors what to write down. Appraisal tools tell reviewers and systematic reviewers how far to trust it, and designing a study with those tools in mind is the cheapest way to avoid a high risk-of-bias rating later. For prediction models, PROBAST+AI replaced the 2019 Prediction model Risk Of Bias ASsessment Tool in 2025. It assesses model development with 16 signalling questions and model evaluation with 18, and applies equally to regression and machine learning [20]. Diagnostic accuracy studies are still appraised with QUADAS-2 [21], randomised trials with the revised Cochrane risk-of-bias tool, RoB 2 [22], and radiomics studies with the METhodological RadiomICs Score (METRICS) [23]. For teams building tools intended for deployment, the FUTURE-AI guideline, agreed by 117 experts from 50 countries, sets out 30 recommendations organised around fairness, universality, traceability, usability, robustness and explainability [24]. Figure 2 summarises which tool applies to each study type and what it assesses.

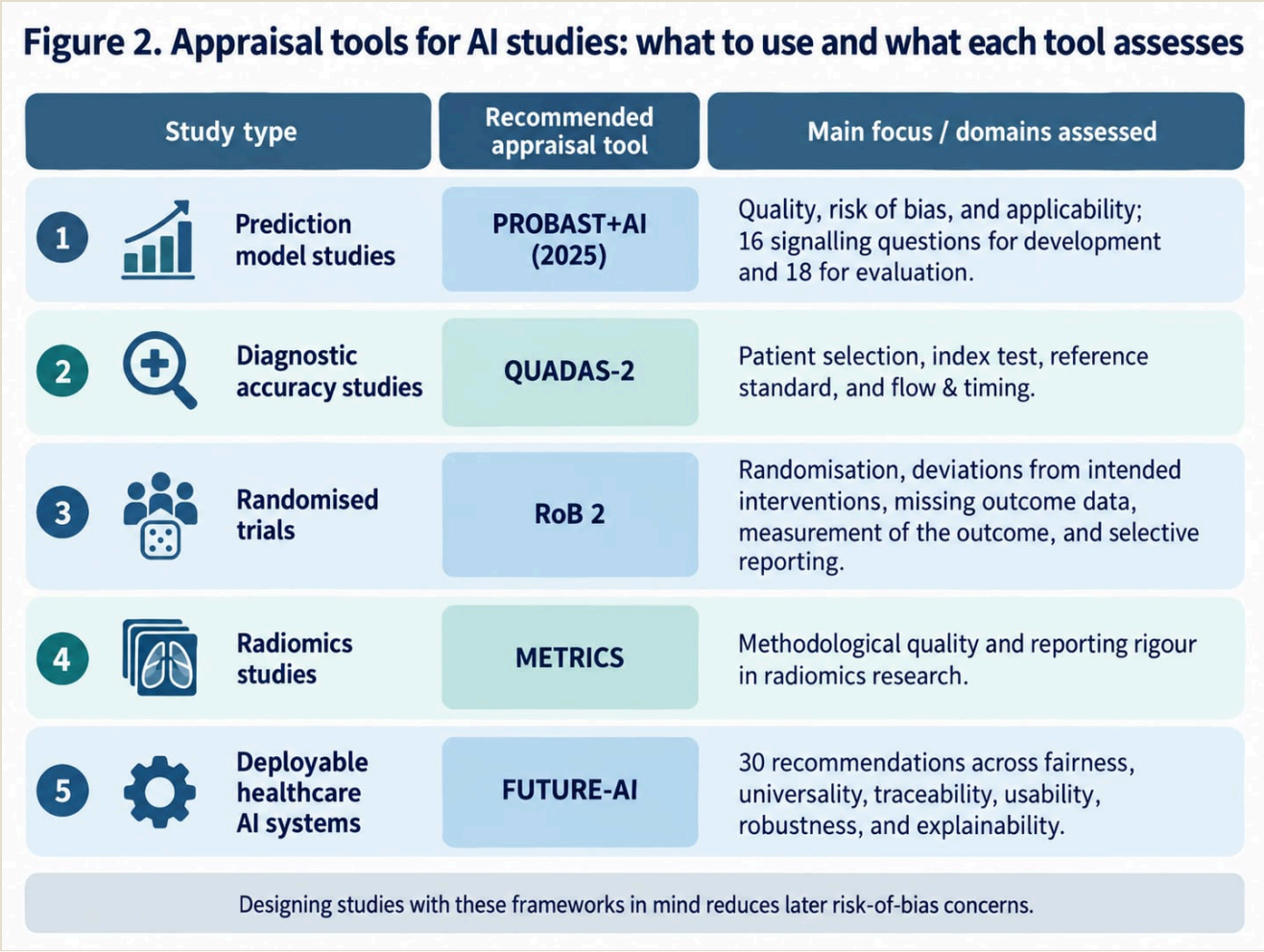


Figure 2. Appraisal tools for artificial intelligence studies Recommended appraisal tool for each study type and the main domains each assesses: PROBAST+AI [20], QUADAS-2 [21], RoB 2 [22], METRICS [23] and FUTURE-AI [24]. AI, artificial intelligence; METRICS, METhodological RadiomICs Score; PROBAST, Prediction model Risk Of Bias ASsessment Tool; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2; RoB 2, revised Cochrane risk-of-bias tool for randomised trials.

The regulatory position in 2026

Regulation matters to researchers because studies done today may later be used to support a product, and because ethics committees increasingly ask about it. The main frameworks are summarised in Table 2. The most important change this year came in the European Union (EU). The Digital Omnibus on AI entered into force on 27 July 2026 and postponed the high-risk obligations of the AI Act; for AI embedded in regulated products such as medical devices, they now apply from 2 August 2028 [25,26]. The timelines of the EU Medical Device Regulation and In Vitro Diagnostic Medical Devices Regulation did not change.

Table 2. Regulatory and governance frameworks relevant to clinical artificial intelligence research, status in September 2026

Framework	Status	Relevance to researchers
EU Artificial Intelligence Act [25], amended by the Digital Omnibus on AI [26]	In force; high-risk obligations for AI in medical devices apply from 2 August 2028	AI requiring a notified body under the MDR or IVDR is high risk; plan data governance, documentation and human oversight early
FDA draft guidance on AI-enabled device software functions [27]	Draft since January 2025	Expectations for validation, bias analysis and performance monitoring
FDA guidance on predetermined change control plans [28]	Final	How a cleared model can be updated without a new submission
Good machine learning practice guiding principles [29]	Published 2021	Ten principles on data, training, testing and monitoring
FDA draft guidance on AI for regulatory decisions on drugs and biologics [30]	Draft since January 2025	Risk-based credibility framework for AI used in trials and drug development
EMA reflection paper on AI in the medicinal product lifecycle [31]	Adopted 2024	European expectations from discovery to pharmacovigilance
ICH E6(R3) guideline for good clinical practice [32]	Adopted 2025	Risk-proportionate GCP covering computerised systems used in trials
WHO guidance on AI for health [33] and on large multimodal models [34]	Published 2021 and 2024	Ethical principles and recommendations for developers, providers and governments
EU GDPR [35]; UK GDPR and Data Protection Act 2018 [36]	In force	Lawful basis and safeguards for health data used to train or test models

AI, artificial intelligence; EMA, European Medicines Agency; EU, European Union; FDA, US Food and Drug Administration; GCP, good clinical practice; GDPR, General Data Protection Regulation; ICH, International Council for Harmonisation; IVDR, In Vitro Diagnostic Medical Devices Regulation; MDR, Medical Device Regulation; UK, United Kingdom; WHO, World Health Organization.

Generative AI remains the least regulated part of the field. According to the US Congressional Research Service, the FDA had not, as of June 2026, authorised any device based on generative AI, although it had granted breakthrough designation to one patient-facing application in March 2026 [37]. Patients and clinicians already consult chatbots every day, which makes careful evaluation studies, reported to TRIPOD-LLM [15] or CHART [16], especially valuable now.

Authorship, disclosure and peer review

On the use of AI in writing, the main editorial bodies have reached the same position. The International Committee of Medical Journal Editors (ICMJE) asks journals to require disclosure of AI-assisted technologies at submission and states that chatbots cannot be authors, because they cannot take responsibility for the accuracy, integrity and originality of the work [38]. The Committee on Publication Ethics (COPE) [39], the World Association of Medical Editors (WAME) [40] and the Council of Science Editors [41] take the same view, and WAME adds that editors and reviewers should disclose their own use. Confidentiality is the other concern. The US National Institutes of Health (NIH) prohibits its peer reviewers from using generative AI on grant applications, because uploading them to such tools breaches confidentiality [42].

A useful disclosure is short and specific: the tool and version, what it was used for, which parts of the manuscript it affected, and confirmation that the authors checked the output. Open use of AI is not the problem. The problem is hidden use, and in particular the fabricated reference, a citation that looks real but does not exist or does not say what the text claims. Journals now screen for these, and they are among the quickest routes to retraction.

Common errors in AI studies

Most AI manuscripts that fail peer review fail for familiar reasons, and few of them concern the

algorithm. The commonest is data leakage, in which images or visits from the same patient appear in both the training and the test set; splitting by patient, and by site where possible, prevents it. Close behind is the absence of external validation. Performance in the development data is nearly always optimistic, and a test in another hospital, time period or population is now the minimum that reviewers expect.

Models also learn shortcuts. A chest radiograph classifier can learn the scanner, the hospital's text label or the presence of a chest drain rather than the disease, and reporting performance by site and device is the simplest way to show that it has not. Accuracy alone is a weak measure in imbalanced data; discrimination should be reported with calibration, and with net benefit when the model is intended to guide a decision. Performance should be broken down by sex, age and other relevant groups, including the groups in which the model performs poorly. Labels deserve the same scrutiny as the model: who assigned them, how, and how disagreements were resolved. Case-control samples of clear-cut cases inflate accuracy, so consecutive or representative patients should be used wherever possible.

Studies of large language models bring problems of their own. Outputs depend on the model version, the settings and the exact wording of the prompt, and vendors update models without notice, so all of these should be reported with the date of access. Examination questions and published cases may already be in the training data, which is why new, unpublished test material is more convincing. Finally, a high AUC is not evidence of benefit. Authors should state which clinical decision the model is meant to change and what study would show that it does.

Checklist before submitting an AI study

- The correct reporting guideline has been completed and uploaded with page numbers.
- Trial protocols, and ideally diagnostic and prognostic study protocols, have been registered.
- Data have been split by patient and every data source is described.
- The model has been externally validated, or the reason it has not is stated.
- Calibration and subgroup performance are reported alongside discrimination.
- Code, and where lawful the data or a synthetic equivalent, are available.
- Every AI tool used in the research and the writing is disclosed.
- Every reference has been checked against its original source.

Implications for authors

The most useful AI research is research that helps a practising clinician make a better decision, and this is where AI studies most often fall short. The greatest needs are for external validation studies, including those showing that a model does not transfer; prospective and implementation studies, including early clinical evaluations reported to DECIDE-AI; randomised trials reported to CONSORT-AI; evaluations of language models and chatbots reported to TRIPOD-LLM or CHART; imaging studies reported to CLAIM; and studies of fairness, failure and harm. Work from low- and middle-income countries is especially needed, because tools developed elsewhere are often introduced there without local testing. Authors submitting an AI study should expect journals to ask for the completed reporting checklist, a clear statement of any AI used in the research or the writing, and access to code where possible, and should prepare these from the outset.

Conclusion

Figure 3 summarises the gap between the number of AI devices cleared for clinical use and the number evaluated in prospective trials or against patient outcomes. The value of clinical AI will be decided by the quality of the evidence behind it rather than by the performance figures in development papers. The guidelines, appraisal tools and regulatory expectations described here are now mature enough to support that evidence. What remains is for clinicians and researchers to test these tools properly, report what they find,

including what does not work, and publish it where others can read it.

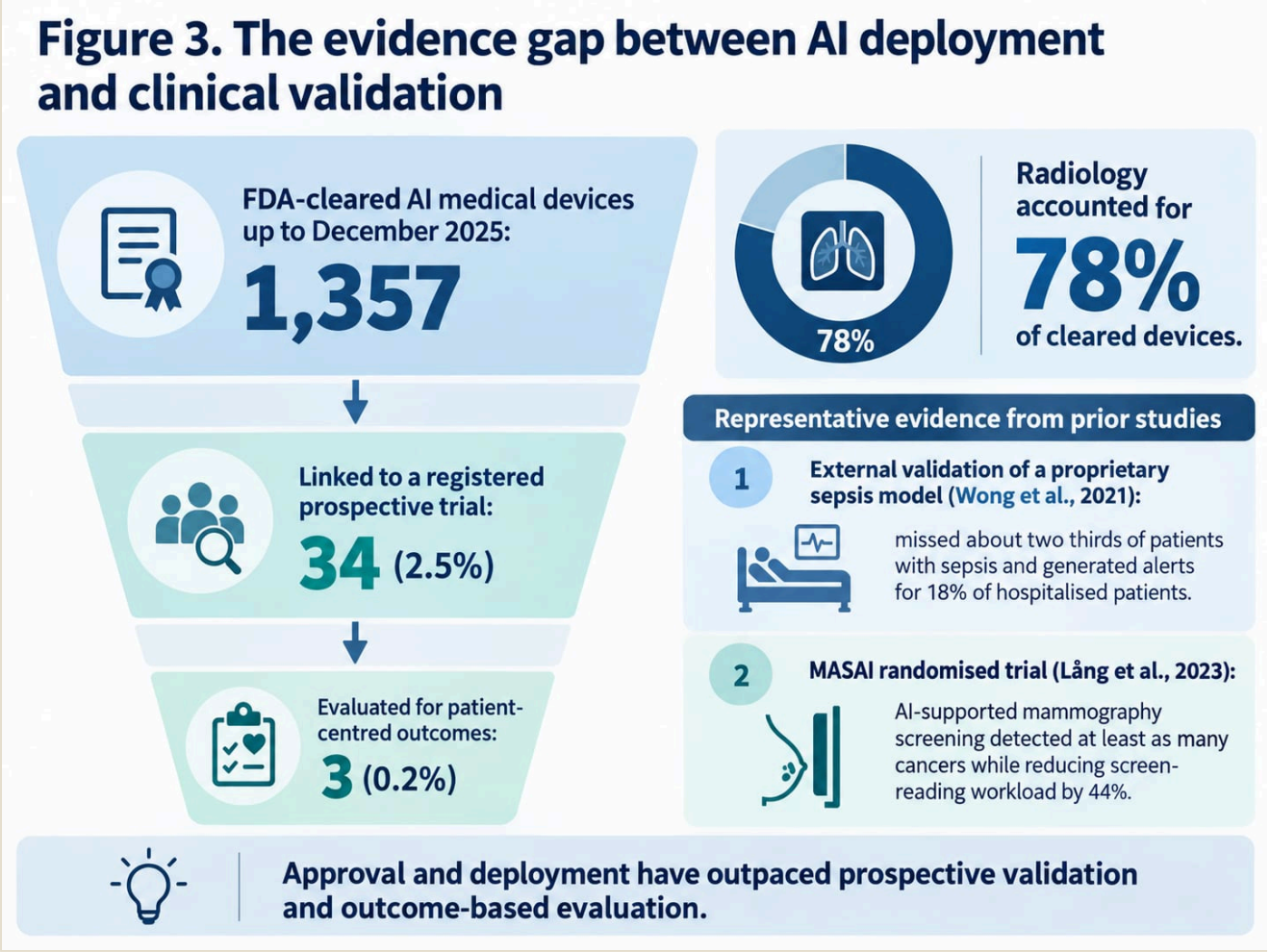


Figure 3. The evidence gap between deployment and clinical validation of artificial intelligence devices Left: attrition of evidence among AI-enabled devices cleared by the US Food and Drug Administration (FDA) up to 5 December 2025; percentages use all 1,357 cleared devices as the denominator [1]. Top right: share of cleared devices in radiology [1]. Bottom right: two examples from earlier studies, an external validation of a widely deployed proprietary sepsis prediction model [4] and the MASAI randomised trial of AI-supported mammography screening [5]. AI, artificial intelligence; MASAI, Mammography Screening with Artificial Intelligence trial.

Table 1. Artificial intelligence-specific reporting guidelines by study design

AI, artificial intelligence; CHART, Chatbot Assessment Reporting Tool; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; CLEAR, CheckList for EvaluAtion of Radiomics research; CONSORT, Consolidated Standards of Reporting Trials; DECIDE-AI, Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence; LLM, large language model; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; STARD, Standards for Reporting of Diagnostic Accuracy Studies; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

Table 2. Regulatory and governance frameworks relevant to clinical artificial intelligence research, status in September 2026

AI, artificial intelligence; EMA, European Medicines Agency; EU, European Union; FDA, US Food and Drug Administration; GCP, good clinical practice; GDPR, General Data Protection Regulation; ICH, International Council for Harmonisation; IVDR, In Vitro Diagnostic Medical Devices Regulation; MDR, Medical

Device Regulation; UK, United Kingdom; WHO, World Health Organization.

DECLARATIONS

Acknowledgements

None.

Funding

No specific funding.

Competing interests

The author declares no competing interests.

Ethical approval

Not applicable; this review did not involve human participants or patient data.

Data availability

Not applicable; no new data were generated.

Registration

Not applicable.

Use of AI tools

Claude (Opus 5.5; Anthropic) was used to help identify sources, draft text and format references. Figures 1 to 3 were produced with the image generation tool in ChatGPT (OpenAI) from content specified by the author, and were checked by the author against the text and the cited sources. The author reviewed and edited all content, checked every reference against its original source, and takes full responsibility for the article.

REFERENCES

1. Abulibdeh R, Cajas Ordóñez SA, Celi LA, Gorijavolu R, Izath N, Markussen Lunde T. 1,357 AI medical devices cleared, 3 actually tested on patient outcomes. PLOS Digit Health. 2026;5(8):e0001597. doi:10.1371/journal.pdig.0001597
2. Windecker D, Baj G, Shiri I, Kazaj PM, Kaesmacher J, Gräni C, et al. Generalizability of FDA-approved AI-enabled medical devices for clinical use. JAMA Netw Open. 2025;8(4):e258052. doi:10.1001/jamanetworkopen.2025.8052
3. McGenity C, Clarke EL, Jennings C, Matthews G, Cartlidge C, Freduah-Agyemang H, et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. NPJ Digit Med. 2024;7(1):114. doi:10.1038/s41746-024-01106-8
4. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181(8):1065-70. doi:10.1001/jamainternmed.2021.2626
5. Lång K, Josefsson V, Larsson AM, Larsson S, Högberg C, Sartor H, 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(8):936-44. doi:10.1016/S1470-2045(23)00298-X
6. EQUATOR Network. Library for health research reporting. Accessed 26 September 2026. https://www.equator-network.org
7. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26(9):1351-63. doi:10.1038/s41591-020-1037-7
8. Chan AW, Boutron I, Hopewell S, Moher D, Schulz KF, Collins GS, et al. SPIRIT 2025 statement: updated guideline for protocols of randomised trials. BMJ. 2025;389:e081477. doi:10.1136/bmj-2024-081477

9. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26(9):1364-74. doi:10.1038/s41591-020-1034-x
10. Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025;389:e081123. doi:10.1136/bmj-2024-081123
11. 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(5):924-33. doi:10.1038/s41591-022-01772-9
12. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31(10):3283-9. doi:10.1038/s41591-025-03953-8
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi:10.1136/bmj.h5527
14. 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. doi:10.1136/bmj-2023-078378
15. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med. 2025;31(1):60-9. doi:10.1038/s41591-024-03425-5
16. CHART Collaborative; Huo B, Collins GS, Chartash D, Thirunavukarasu A, Flanagin A, Iorio A, et al. Reporting guideline for chatbot health advice studies: the Chatbot Assessment Reporting Tool (CHART) statement. Ann Fam Med. 2025;23(5):389-98. doi:10.1370/afm.250386
17. CHART Collaborative. Reporting guidelines for chatbot health advice studies: explanation and elaboration for the Chatbot Assessment Reporting Tool (CHART). BMJ. 2025;390:e083305. doi:10.1136/bmj-2024-083305
18. Tejani AS, Klontzas ME, Gatti AA, Mongan JT, Moy L, Park SH, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. Radiol Artif Intell. 2024;6(4):e240300. doi:10.1148/ryai.240300
19. Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. 2023;14(1):75. doi:10.1186/s13244-023-01415-8
20. 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. doi:10.1136/bmj-2024-082505
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36. doi:10.7326/0003-4819-155-8-201110180-00009
22. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. doi:10.1136/bmj.l4898
23. Kocak B, Akinci D'Antonoli T, Mercaldo N, Alberich-Bayarri A, Baessler B, Ambrosini I, et al. METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. Insights Imaging. 2024;15(1):8.
24. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. doi:10.1136/bmj-2024-081554
25. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. Published 12 July 2024. Accessed 26 September 2026. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
26. Regulation (EU) 2026/1744 amending Regulation (EU) 2024/1689 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). Official Journal of the European Union. Published 24 July 2026.
27. US Food and Drug Administration. Artificial intelligence-enabled device software functions: lifecycle management and marketing submission recommendations. Draft guidance. Published January 2025. Accessed 26 September 2026. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device
28. US Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions. Final guidance. Accessed 26 September 2026. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence
29. US Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Good machine learning practice for medical device development: guiding principles. Published October 2021.

30. US Food and Drug Administration. Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products. Draft guidance. Published January 2025.
31. European Medicines Agency. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. EMA/CHMP/CVMP/83833/2023. Published September 2024.
32. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH E6(R3) guideline for good clinical practice. Adopted January 2025.
33. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization; 2021.
34. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. World Health Organization; 2024.
35. Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). Official Journal of the European Union. Published 4 May 2016.
36. Data Protection Act 2018 (c. 12). UK Public General Acts. Accessed 30 September 2026. https://www.legislation.gov.uk/ukpga/2018/12/contents
37. Congressional Research Service. FDA regulation of AI-enabled devices. In Focus IF13245. Published June 2026. Accessed 26 September 2026. https://www.everycrsreport.com/reports/IF13245.html
38. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated 2026. Accessed 26 September 2026. https://www.icmje.org/icmje-recommendations.pdf
39. Committee on Publication Ethics. Authorship and AI tools: COPE position statement. Published 2023. Accessed 26 September 2026. https://publicationethics.org/cope-position-statements/ai-author
40. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, Lapeña JF, et al. Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. World Association of Medical Editors. Published 2023. Accessed 26 September 2026. https://wame.org/page3.php?id=106
41. Jackson J, Landis G, Baskin PK, Hadsell KA, English M; CSE Editorial Policy Committee. CSE guidance on machine learning and artificial intelligence tools. Sci Ed. 2023;46(2):72. doi:10.36591/SE-D-4602-07
42. National Institutes of Health. The use of generative artificial intelligence technologies is prohibited for the NIH peer review process. Notice NOT-OD-23-149. Published June 2023.

Received 25 August 2026 **Accepted** **Published** 02 November 2026
Citation Ahmed Eldesouky. Artificial intelligence in clinical research: a narrative review of reporting guidelines, regulation and publication ethics. Clinical Standard. 2026.
Licence CC BY 4.0. This is an open-access article distributed under the terms of the Creative Commons Attribution licence.