International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Quality-Control and Multimodal Clinical Intelligence (QC-MCI) Framework for Improving Reliability and Clinical Integration in AI-Enabled Radiology Workflows

Oyindamola Adebayo

Abstract

Artificial intelligence (AI) is moving from image classification experiments into radiology worklists, reporting systems and clinical decision support. The central implementation problem is no longer whether an algorithm can produce a strong score on a retrospective test set. It is whether the combined system—data, software, radiologist, workflow and governance—continues to produce safe and useful decisions when it encounters routine clinical variation. This article proposes a Quality-Control and Multimodal Clinical Intelligence (QC-MCI) framework for that problem. The framework treats reliability as a property of the complete clinical pathway rather than of the model alone. It combines five linked functions: input quality control; multimodal clinical context integration; model and subgroup validation; human–AI interaction control; and continuous post-deployment monitoring with escalation. The framework is grounded in pre-April 2023 evidence and guidance, including radiology workflow studies, external-validation research, studies of hidden stratification and underdiagnosis bias, and the CLAIM, CONSORT-AI, SPIRIT-AI and DECIDE-AI reporting standards. Practical radiology studies have shown that AI assistance can reduce reading time or improve reader performance in defined settings, but they also show why local validation, clear operating procedures and monitoring are necessary. QC-MCI therefore recommends a staged pathway: verify the input, interpret image and clinical context together, test performance across relevant subgroups and sites, expose AI output at a controlled point in the reader's workflow, record disagreement and overrides, and pause or recalibrate the system when safety signals emerge. The framework is a proposal for implementation and evaluation, not evidence that a particular AI product improves patient outcomes. Its value lies in making the quality-control and clinical-integration tasks explicit, measurable and assignable.

Keywords: radiology; artificial intelligence; multimodal clinical intelligence; quality control; clinical workflow; external validation; human–AI interaction; patient safety; dataset shift

1. Introduction

Radiology is a particularly visible setting for clinical AI because imaging data are digital, large in volume and often accompanied by structured reports and clinical metadata. This creates favourable conditions for machine learning, but it does not make deployment straightforward. A model is trained on a defined sample, with a particular acquisition protocol, prevalence pattern, label process and patient population. Clinical practice is less tidy. Scanners change, protocols vary, image quality is uneven, referrals are selective, and the meaning of an image depends on the indication and the patient's wider record.

That distinction changes the implementation question. A model may perform well in a retrospective evaluation and still be unreliable when used in a new hospital, with a different scanner, a different patient mix or a different reporting pathway. External-validation work in medical imaging has found that published external validations are relatively infrequent and that performance commonly falls outside the development environment.[1] Hidden stratification creates a further problem: a model can achieve an acceptable aggregate result while failing on clinically important subgroups that are not represented clearly in the headline metric.[2] Underdiagnosis bias in chest-radiograph algorithms illustrates the patient-safety consequence when performance differences are not examined across populations.[3]

At the same time, AI should not be judged only as a stand-alone classifier. It is usually introduced as a component of a sociotechnical system. The radiologist must understand what the output means, decide when to trust it, recognise when the input is outside the tool's intended use, and communicate the final interpretation. The system must deliver the result at a useful point in the worklist or viewer. The institution must define responsibility, document incidents and monitor whether the tool continues to behave as expected. DECIDE-AI describes this translation problem directly: early live evaluation must address clinical utility, safety, human factors, user variability, dataset shift and generalisability rather than relying on preclinical accuracy alone.[4]

Several practical studies show the potential of workflow-integrated assistance. In breast tomosynthesis, AI support improved cancer detection and reduced reading time in a multi-reader, multi-case study.[5] A prospective single-centre study evaluated commercial AI assistance during chest CT interpretation.[6] Another study examined the effect of concurrent AI tools on radiologists' reading time for non-contrast chest CT.[7] These findings are encouraging, but they are context-specific. They do not establish that any AI tool is safe in every department, nor do they remove the need to understand input quality, alert timing, user behaviour and local prevalence.

This article proposes the Quality-Control and Multimodal Clinical Intelligence (QC-MCI) framework. The framework responds to a practical gap: radiology departments need a way to organise the checks that occur before, during and after an AI tool enters clinical use. QC-MCI defines those checks as a connected pathway. It also makes a distinction that is often blurred in discussion: multimodal does not mean that an AI system should automatically consume every available patient datum. It means that imaging findings are interpreted in relation to the clinical question and the relevant, authorised context, with explicit controls for missing, conflicting or misleading information.

2. Basis and scope of the framework

QC-MCI is an evidence-informed implementation framework. It is not a report of a new clinical trial, and it does not claim that the framework itself has been prospectively validated. Its design draws on four sources of pre-April 2023 evidence.

First, radiology workflow research shows that AI can affect reader performance and reading time when introduced into a defined task.[5–7] Second, studies of external validation, hidden stratification and subgroup bias show that aggregate performance can conceal local or clinically important failures.[1–3] Third, reporting standards provide operational requirements for describing the input data, AI intervention, human–AI interaction and clinical evaluation. CLAIM was introduced for transparent reporting of AI in medical imaging.[8] CONSORT-AI asks clinical-trial reports to describe input-data acquisition and selection, handling of poor-quality or unavailable inputs, and the way clinicians interact with the AI intervention.[9] SPIRIT-AI extends similar expectations to trial protocols.[10] DECIDE-AI adds a stage-

specific focus on early, small-scale, live clinical evaluation.[4] Fourth, the World Health Organization's guidance places safety, transparency, accountability, equity and human autonomy within the governance of health AI.[11]

The framework is intended for AI-enabled radiology workflows in which the output may influence prioritisation, interpretation, reporting, referral or follow-up. It can be used for a stand-alone image-analysis tool, a worklist-prioritisation tool, a reporting assistant or a system that combines images with selected clinical information. It is not a substitute for regulatory approval, local information-governance review, clinical risk management or professional judgement.

3. The QC-MCI framework

QC-MCI has five functions. They are shown as a cycle rather than a one-time checklist because the clinical environment changes and because each function generates information for the next review.

3.1. Function 1: Input quality control

The first question is whether the case is suitable for the tool. Input quality control should be performed before the model's output is treated as clinically informative. Depending on the examination, the checks may include modality, body region, laterality, acquisition protocol, slice thickness, contrast phase, field of view, artefact burden, incomplete series and the presence of the required image views.

The output of this function should not be a vague warning that the image is "poor quality." It should produce a defined state, such as acceptable, acceptable with limitation, not suitable, or unable to determine. The system should record the reason for rejection or limitation. A case with motion artefact may be suitable for one task but not another. A missing contrast phase may not invalidate a non-contrast task, but it may make a contrast-dependent assessment inappropriate.

Input control also applies to clinical data. If a multimodal tool uses age, sex, symptoms, laboratory values, prior imaging or the clinical indication, the system should verify whether those fields are present, current and internally consistent. Missing data should not be silently converted into a plausible value. If the model was not trained for a particular missing-data pattern, the safe response may be to withhold the output or display the limitation prominently.

This function is closely aligned with CONSORT-AI's requirement to describe how input data are acquired and selected and how poor-quality or unavailable inputs are handled.[9] In practice, it creates a quality gate between image acquisition and clinical interpretation. The gate should be auditable and should not be designed to turn an unsafe output into a reassuring score.

3.2. Function 2: Multimodal clinical context integration

Radiology interpretation is already multimodal in the clinical sense. The radiologist considers the image together with the indication, relevant history, previous examinations, laboratory data and the intended clinical decision. QC-MCI therefore treats multimodal integration as a controlled clinical process rather than as an uncontrolled data-collection exercise.

The framework recommends four rules. First, include only variables with a defined clinical purpose. Second, record the source and time of each variable. Third, identify conflicts between the image and the clinical context instead of allowing one modality to silently dominate. Fourth, preserve the clinician's ability to review the source information rather than presenting an unexplained composite score.

A multimodal system should be tested in the conditions in which it will actually be used. A model that receives complete structured data during development may encounter incomplete free-text referrals in practice. A model that benefits from a prior report may behave differently when the prior examination is unavailable. These are not minor technical details; they alter the information presented to the algorithm and to the radiologist.

The framework does not require that the AI system produce a final diagnosis. In many workflows, the safer use is narrower: identify a limited finding, prioritise a case, provide a measurement or prompt the reader to examine a defined region. The clinical context can then be used by the radiologist to decide whether that finding matters for the patient in front of them.

3.3. Function 3: Model, site and subgroup validation

Validation should answer three separate questions. Can the model perform the task? Does it perform acceptably in this site and workflow? Does it remain acceptably safe across the patient and examination subgroups that matter here? A single area under the curve cannot answer all three.

Local validation should use a dataset that reflects the intended indication, referral pathway, equipment, protocols and prevalence. It should be separated from the development data and should include cases that expose likely failure modes. The evaluation plan should prespecify the reference standard, unit of analysis, missing-data rules, thresholds, confidence intervals and subgroup analyses.

QC-MCI recommends reporting at least sensitivity, specificity, positive and negative predictive values where appropriate, calibration, indeterminate or rejected-case rates, and the effect of the AI output on the human reader. For prioritisation tools, time-to-review and time-to-report may be more meaningful than classification accuracy alone. For measurement tools, agreement and clinically important error should be reported. For reporting assistants, omissions, unsupported statements and downstream amendments require attention.

Subgroup validation should be based on clinical relevance rather than convenience. Relevant strata may include age, sex, pregnancy status where applicable, disease severity, body habitus, language or referral pathway, scanner vendor, protocol, site, inpatient or outpatient status and the presence of prior disease. The purpose is not to create an endless list of stratifications. It is to identify situations in which a failure would change care or disproportionately affect a group of patients.

Hidden stratification is a reason to inspect errors, not simply to enlarge the aggregate test set. A model may distinguish normal from abnormal cases while failing to separate clinically important subtypes within the abnormal group.[2] Review of false negatives and clinically discordant cases should therefore be part of validation. The department should also decide what constitutes a stop condition, such as unacceptable performance in a high-risk subgroup, a rise in rejected inputs, or a pattern of clinically serious false reassurance.

3.4. Function 4: Human–AI interaction control

The clinical effect of an AI tool depends partly on when and how its output is shown. A correct output can be unhelpful if it arrives after the report is signed. An uncertain output can be harmful if it appears authoritative and the interface does not make the uncertainty visible. Human–AI interaction must therefore be specified as part of the intervention.

The local procedure should state whether the radiologist sees the AI output before independent review, after an initial read, or only when a discrepancy or high-risk finding is present. It should specify which

actions remain the radiologist's responsibility, how an output can be overridden, and how disagreement is recorded. Users should receive training that covers the tool's intended use, known failure modes, input limitations and escalation process. Training should not be reduced to a product demonstration.

A useful control is to separate three concepts in the interface: model output, clinical interpretation, and final report. The model may identify a candidate finding. The radiologist decides whether the finding is present and clinically relevant. The final report communicates the professional interpretation. This separation reduces the risk that a preliminary model signal will be copied into the record without verification.

The department should monitor automation bias and over-reliance, but it should also monitor the opposite problem: alert fatigue and routine dismissal of useful outputs. The appropriate balance cannot be assumed from a laboratory reader study. It must be examined in the intended workflow, with attention to workload, interruptions, case mix and user experience. DECIDE-AI specifically identifies human factors and user variability as core elements of early live evaluation.[4]

3.5. Function 5: Continuous monitoring, escalation and learning

Deployment is the beginning of monitoring, not the end of validation. QC-MCI recommends a dashboard that combines technical, clinical and workflow signals. Technical signals include input rejection, missing-data patterns, latency, software errors and changes in output distribution. Clinical signals include false-negative and false-positive reports identified through review, discordance with pathology or follow-up where available, subgroup performance and clinically significant overrides. Workflow signals include turnaround time, worklist position, user adoption, alert acknowledgement, report amendments and incidents.

Monitoring should have an owner, a review schedule and thresholds that trigger action. A threshold without a response plan is not a safety control. Possible actions include targeted case review, temporary restriction of the tool, recalibration, updated user guidance, rollback to an earlier version, or suspension while an incident is investigated.

The system should maintain a versioned audit trail. Each AI output should be linked to the software version, input-data status, time of inference, user interaction and final clinical outcome when that outcome is available. This record supports incident analysis and prevents the common problem of trying to reconstruct what the tool displayed months after a concern is raised.

Monitoring should also look for changes in the population served. A new scanner, a protocol change, a new referral service or a change in disease prevalence can alter performance without any change to the algorithm. External-validation evidence and the clinical-evaluation guidance both support treating dataset shift as an operational safety issue.[1,4]

4. Findings from practical radiology studies

The practical value of AI-assisted radiology has been examined in reader studies and early clinical-workflow evaluations. These studies do not validate QC-MCI as a complete intervention, but they provide measurable findings that support the need to evaluate accuracy, reading time, input limitations and human–AI interaction together.

4.1 Breast tomosynthesis reader study

van Winkel et al. conducted a multi-reader, multi-case study using 240 bilateral digital breast tomosynthesis examinations, including 71 breasts with cancer lesions, 70 breasts with benign findings and

339 normal breasts. Eighteen radiologists interpreted the examinations with and without AI support. Mean area under the receiver operating characteristic curve increased from 0.833 without AI to 0.863 with AI support ($p = 0.0025$). Mean reading time also decreased from 41 seconds to 36 seconds per examination ($p < 0.001$). The stand-alone AI system had an area under the curve that was non-inferior to the average radiologist, with a difference of 0.007 ($p = 0.8115$). These findings suggest that, in this defined breast-imaging task, AI assistance can improve reader discrimination while reducing reading time. They do not show that the same effect will occur with a different population, scanner, prevalence pattern or workflow.[5]

4.2 Prospective chest CT workflow study

Yacoub et al. evaluated a commercial AI platform prospectively in a single-centre clinical workflow. The study included 390 outpatients who underwent chest CT. The scans were allocated equally to AI-assisted and non-AI-assisted arms and were interpreted by three cardiothoracic radiologists. The combined mean interpretation time was 328 seconds with AI assistance compared with 421 seconds without AI assistance, a mean reduction of 93 seconds and 22.1% in relative terms ($p < 0.001$). The reduction was observed in contrast-enhanced and non-contrast examinations and in both negative and positive examinations. This is a practical workflow finding rather than proof of improved patient outcomes: the study measured interpretation time, so the result should be interpreted alongside the manuscript's proposed controls for accuracy, disagreement, escalation and monitoring.[6]

4.3 Prospective feasibility study of concurrent AI use

Müller et al. conducted a prospective feasibility study in which an AI tool was integrated into the PACS workflow for non-contrast chest CT examinations. One consultant radiologist and one resident were observed. Each reader interpreted 25 cases without AI and 20 cases with AI. For the resident, mean reading time was 370 seconds without AI and 437 seconds with AI, a difference of 67 seconds that was not statistically significant ($p = 0.16$). For the consultant, mean reading time was 366 seconds without AI and 380 seconds with AI, a difference of 13 seconds that was also not statistically significant ($p = 0.70$). The AI tool identified additional actionable findings in 5 of 40 examinations (12.5%), and the resident reported a better overview in 18 of 20 AI-assisted examinations (90%). This study is important because it shows that workflow effects are not uniformly positive: AI may add useful findings and improve perceived overview without reducing reading time.[7]

4.4 Synthesis of the practical findings

Taken together, these studies show three practical points. First, AI assistance may improve reader performance in a specific task, as shown in breast tomosynthesis. Second, integration into a real clinical workflow may reduce interpretation time, as shown in the prospective chest CT study. Third, the effect can be neutral or mixed when the tool is used concurrently, as shown by the feasibility study in which reading time did not significantly decrease but additional actionable findings and improved overview were reported. The results therefore support QC-MCI's central position: AI should be evaluated as part of a complete sociotechnical pathway. A department should measure not only model discrimination or time savings, but also case suitability, rejected inputs, clinically meaningful additional findings, user interaction, disagreement, subgroup performance and escalation events.

The studies reviewed here are published external evidence and are not an original dataset collected for the present framework. Accordingly, they should be described as practical evidence informing QC-MCI, not as findings generated by the author. An original empirical evaluation would require a separate local sample, defined methods, ethics or exemption information, and results collected by the study team.

4.5 Summary tables and practical comparison

The following tables summarise the published practical studies discussed above. They are included to make the evidence and its limitations visible to the reader. The data are not an original dataset collected for the present framework.

Study	Modality / task	Design and sample	Main measure
van Winkel et al. [5]	Breast tomosynthesis	Multi-reader, multi-case; 240 exams; 18 radiologists	AUC and reading time
Yacoub et al. [6]	Chest CT	Prospective single-centre; 390 outpatients; 3 radiologists	Interpretation time
Müller et al. [7]	Non-contrast chest CT	Prospective feasibility; 1 consultant and 1 resident	Reading time and added findings

Table 1. Design and sample of the practical radiology studies.

5. A staged implementation pathway

QC-MCI can be implemented in five stages.

Stage 1: Define the intended use**.** The department specifies the clinical question, patient population, examination type, output, user, decision point, unacceptable uses and escalation route. A tool should not be approved because it is generally described as "an AI solution." Its intended action must be precise.

Stage 2: Test the local environment. The department checks data pathways, image routing, viewer integration, access permissions, latency, input quality and failure messages. A technically accurate model can still be unusable if its output arrives late or is not visible in the correct worklist.

Stage 3: Conduct silent evaluation**.** The AI runs without changing clinical care. Its outputs are compared with the local reference standard and reviewed for subgroup and failure-mode behaviour. This stage should include cases that the tool is likely to reject or misinterpret.

Stage 4: Conduct controlled live evaluation**.** The tool is introduced to a limited group of trained users with predefined monitoring, incident reporting and stop criteria. The evaluation should examine clinical utility, safety, human factors and workflow effects, consistent with DECIDE-AI.[4]

Stage 5: Sustain and review. If the tool proceeds to broader use, the department maintains monitoring, periodic revalidation and version control. A change in software, scanner, protocol, indication or population should trigger a documented impact assessment.

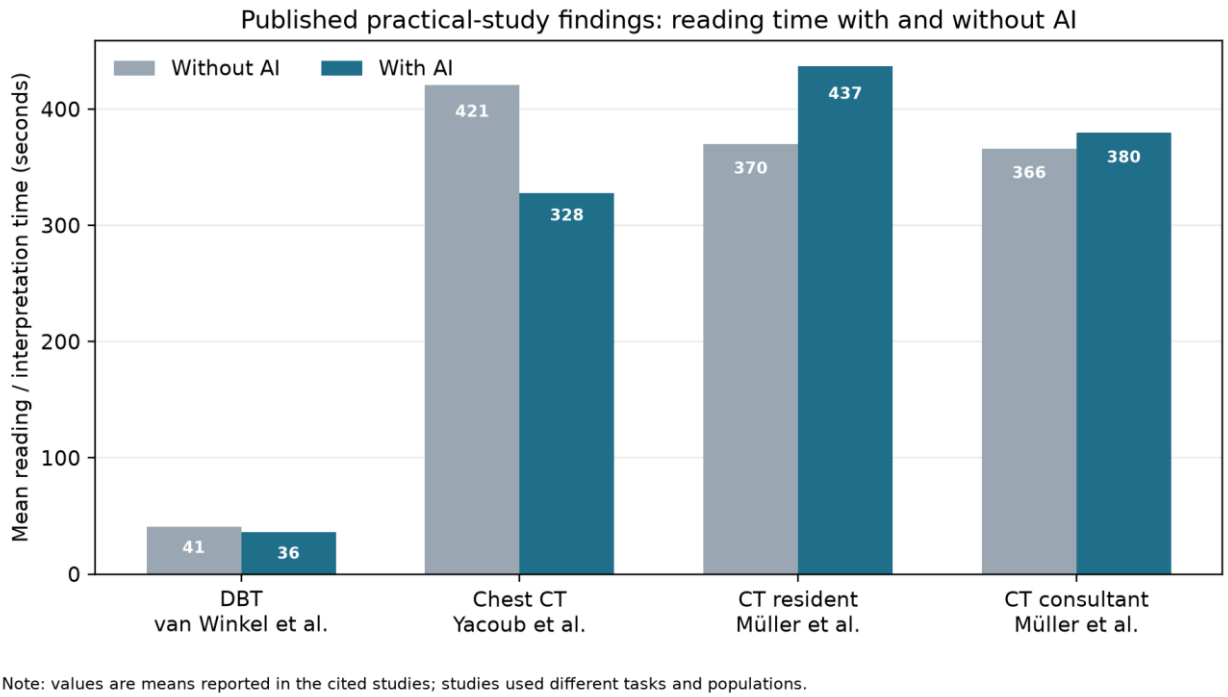


Figure 1. Comparison of reported mean reading or interpretation times with and without AI support. The studies used different modalities, tasks and reader groups; the chart is descriptive and should not be interpreted as a pooled comparison.

6. Practical measures for a radiology department

A department can convert the framework into a small set of operational records. The first is an intended-use statement that identifies the examination, clinical question, user and decision affected. The second is an input-quality log that records rejected or limited cases and the reason. The third is a validation report containing local performance, calibration, subgroup results, reference-standard details and clinically reviewed errors. The fourth is a human–AI interaction protocol covering display timing, user training, override and escalation. The fifth is a monitoring register with named owners, review frequency, thresholds and actions.

These records help distribute responsibility. The radiologist is not expected to solve a data-engineering problem alone, and the software team is not expected to define clinical significance without radiology input. A QC-MCI governance group should include radiology, medical physics or imaging informatics, clinical engineering or IT, data protection or information governance, quality and safety, and where appropriate patient or public representation. The group's role is to make decisions visible: what the tool is for, what it cannot do, how failures are found, and who acts when the pattern changes.

The framework also supports a more realistic definition of success. A tool may be successful if it reduces delay for a defined urgent pathway without increasing clinically important errors. Another tool may be useful because it standardises a measurement or reduces repetitive work. The outcome should be linked to the intended use. Faster reporting is not automatically better care, and a higher model score is not automatically clinical value.

6.1 Practical findings in a compact comparison

The table below presents the main numerical outcomes from the published studies in a compact form. These findings are external evidence informing QC-MCI, not original data collected for this framework.

Study	Without AI	With AI	Reported finding
van Winkel et al. [5]	AUC 0.833; 41 s	AUC 0.863; 36 s	Higher AUC and shorter reading time
Yacoub et al. [6]	421 s	328 s	93-second reduction; 22.1% lower time
Müller et al. [7]	Resident 370 s; consultant 366 s	Resident 437 s; consultant 380 s	No significant time reduction; 5/40 added actionable findings

Table 2. Key reported outcomes. Reading-time values are means; AUC means area under the receiver operating characteristic curve.

7. Discussion

QC-MCI places quality control at both ends of the AI pathway. At the front end, it asks whether the data are suitable and whether the clinical context is complete enough for the intended task. At the back end, it asks whether the output changed the human decision in a safe and useful way. Between those points, it requires local validation, interaction controls and monitoring.

This approach addresses a recurring weakness in radiology AI implementation. Technical evaluations often report performance as if the model were independent of its setting. Workflow evaluations can report time savings without showing whether the tool changed diagnostic quality, error patterns or equity. Governance discussions may identify ethical principles without assigning measurable operational controls. QC-MCI connects these parts without claiming that they are interchangeable.

The framework is also deliberately conservative about multimodality. More data can improve a model, but more data can also introduce leakage, outdated information, proxy bias and confusion about responsibility. Multimodal integration should therefore be justified by the clinical task and monitored for missing or conflicting inputs. A model that cannot explain which context it used may be difficult to govern, even if its aggregate performance is strong.

The limitations are important. QC-MCI has not been tested as a complete intervention in a prospective multi-site study. Its components are supported by prior evidence and reporting standards, but the proposed combination remains a framework for evaluation. The measures and thresholds will need adaptation to the modality, disease, risk level and local governance environment. Some outcomes, such as long-term patient benefit, may take time to observe. Finally, monitoring can identify signals but cannot replace expert review, adequate staffing or a safe reporting culture.

8. Conclusion

Reliable clinical AI in radiology is not produced by model accuracy alone. It depends on the quality of the input, the relevance of the clinical context, the performance of the tool in the local population, the design of the human–AI interaction and the ability of the department to detect and respond to change.

The QC-MCI framework turns those requirements into a staged, auditable pathway. It asks departments to verify the case before inference, integrate only relevant clinical context, validate locally and across meaningful subgroups, control how the output enters the reader's workflow, and monitor technical, clinical and equity-related signals after deployment. The framework should be treated as a proposal to test, not as proof that any individual AI product is safe. Its practical contribution is to make reliability and clinical integration explicit responsibilities rather than assumptions.

References

1. Alice CY, et al. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. *Radiology: Artificial Intelligence*. 2022. https://pubs.rsna.org/doi/10.1148/ryai.210064
2. Oakden-Rayner L, et al. Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging. 2020. https://pmc.ncbi.nlm.nih.gov/articles/PMC7665161/
3. Seyyed-Kalantari L, et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in underserved patient populations. *Nature Medicine*. 2021;27:2176–2182. https://doi.org/10.1038/s41591-021-01595-0
4. Vasey B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. 2022;377:e070904. https://doi.org/10.1136/bmj-2022-070904
5. van Winkel SL, et al. Impact of artificial intelligence support on accuracy and reading time in breast tomosynthesis image interpretation: a multi-reader multi-case study. *European Radiology*. 2021. https://pubmed.ncbi.nlm.nih.gov/33948701/
6. Yacoub B, et al. Impact of Artificial Intelligence Assistance on Chest CT Interpretation: A Prospective, Single-Center Study. *Radiology: Artificial Intelligence*. 2022. https://pubmed.ncbi.nlm.nih.gov/35703413/
7. Müller FC, et al. Impact of Concurrent Use of Artificial Intelligence Tools on Radiologists' Performance and Reading Time in Chest CT. 2022. https://pubmed.ncbi.nlm.nih.gov/34801345/
8. Mongan C, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence*. 2020;2(2):e200029. https://pubs.rsna.org/page/ai/claim
9. Liu X, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Medicine*. 2020;26:1364–1374. https://doi.org/10.1038/s41591-020-1034-x
10. Cruz Rivera S, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nature Medicine*. 2020;26:1351–1363. https://doi.org/10.1038/s41591-020-1037-7
11. World Health Organization. *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization; 2021. https://www.who.int/publications/i/item/9789240029200

12. Dikici E, et al. Integrating AI into radiology workflow: levels of research, production, and feedback. *Journal of the American College of Radiology*. 2020. https://pubmed.ncbi.nlm.nih.gov/32064302/
13. Eche T, et al. Toward Generalizability in the Deployment of Artificial Intelligence in Radiology: A Clinical Practice Perspective. 2021. https://pmc.ncbi.nlm.nih.gov/articles/PMC8637230/
14. Hickman SE, et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology*. 2022;302(1):88–104. https://pubmed.ncbi.nlm.nih.gov/34665034/