

Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How?



Dania Daye, MD, PhD • Walter F. Wiggins, MD, PhD • Matthew P. Lungren, MD, MPH • Tarik Alkasab, MD, PhD • Nina Kottler, MD, MS • Bibb Allen, MD • Christopher J. Roth, MD • Bernardo C. Bizzo, MD • Kimberly Durniak, PhD • James A. Brink, MD • David B. Larson, MD, MBA • Keith J. Dreyer, DO, PhD* • Curtis P. Langlotz, MD, PhD*

From the Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 55 Fruit St, GRB 297, Boston, MA 02155 (D.D., T.A., B.C.B., K.D., J.A.B., K.J.D.); Department of Radiology, Duke University, Durham, NC (W.F.W., C.J.R.); Department of Radiology, Stanford University, Stanford, Calif (M.P.L., D.B.L., C.P.L.); Radiology Partners, El Segundo, Calif (N.K.); and Department of Radiology, Grandview Medical Center, Birmingham, Ala (B.A.). Received September 29, 2021; revision requested November 8; revision received March 30, 2022; accepted April 12. Address correspondence to D.D. (email: ddaye@mgh.harvard.edu).

M.P.L. sponsored by National Institutes of Health, Philips, and GE. D.B.L. supported by Siemens Healthineers and the Gordon and Betty Moore Foundation. C.P.L. supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (75N92020C00008, 75N92020C00021).

* K.J.D. and C.P.L. are co-senior authors.

Conflicts of interest are listed at the end of this article.

Radiology 2022; 305:555–563 • <https://doi.org/10.1148/radiol.212151> • Content codes:  

As the role of artificial intelligence (AI) in clinical practice evolves, governance structures oversee the implementation, maintenance, and monitoring of clinical AI algorithms to enhance quality, manage resources, and ensure patient safety. In this article, a framework is established for the infrastructure required for clinical AI implementation and presents a road map for governance. The road map answers four key questions: Who decides which tools to implement? What factors should be considered when assessing an application for implementation? How should applications be implemented in clinical practice? Finally, how should tools be monitored and maintained after clinical implementation? Among the many challenges for the implementation of AI in clinical practice, devising flexible governance structures that can quickly adapt to a changing environment will be essential to ensure quality patient care and practice improvement objectives.

© RSNA, 2022

An earlier incorrect version appeared online. This article was corrected on August 2, 2022.

As artificial intelligence (AI) is adopted in clinical practice, new workflows, management structures, and governance processes need to be established to evaluate, use, and monitor AI algorithms. AI technology is evolving rapidly, requiring us to continuously adapt to an ever-changing state of the art. The greatest challenge facing AI governance may be the need for balancing agility and stability to ensure an adaptable yet robust system that ensures clinical quality and patient safety while realizing the tremendous promise of the emerging technology.

Essentials of Clinical AI

The components and best practices for establishing AI infrastructure for clinical imaging have been described at length (1–9). These descriptions often come from groups engaged in algorithm development and therefore focus on the data science infrastructure needed to produce AI algorithms, as follows: data access and security, cross-platform and cross-domain integration, clinical translation and delivery, and a culture of innovation and inclusive participation (10) (Fig 1). However, the rapid increase in the number of commercially available algorithms and the variety of ways in which each algorithm can affect clinical workflows adds complexity to the AI implementation process. Therefore, institutions and radiology practices are under pressure to define an AI governance structure to guide evaluation, selection, procurement, implementation, and ongoing support for internally developed and vendor-supplied solutions. The optimal use may vary depending on the intended purpose of the AI solution, the use case, the available domain expertise, and the arrangements with

development and implementation partners. In this setting, a clear process with adequate governance mechanisms is necessary to support necessary data access and system integrations without interfering with clinical processes.

AI Governance and Management Structures

An imaging AI governing body has the responsibilities of defining the purposes, priorities, strategies, and scope of the group; establishing a framework for operation; and linking those to the organizational mission, values, vision, and strategy. AI governance structures provide mechanisms to decide which tools should be deployed locally and how to best allocate institutional and/or departmental resources to support the clinical implementation of the most valuable and highest-impact applications to improve patient care. Governance committees can establish a robust process to score and evaluate AI-based solutions objectively.

Such a governing body may operate within an imaging informatics governance structure, such as a subcommittee of an imaging practice's informatics committee, or operate as a separate entity, such as in an academic research center where algorithms and tools are developed. Free-flowing, multidirectional communication should occur between the imaging AI governing body, the broader organization that empowers it, the system-wide informatics governing bodies, and the end users of each AI tool. For this reason, imaging AI governing bodies tend to have representation from clinical leadership (including from imaging departments such as cardiology, ophthalmology, pathology, or radiology, and from nonimaging electronic health record managers), health system administrative leadership, data

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

Abbreviations

AI = artificial intelligence, FDA = Food and Drug Administration, IMDRF = International Medical Device Regulators Forum

Summary

Successful clinical implementation of artificial intelligence is facilitated by establishing robust organizational structures to ensure appropriate oversight of algorithm implementation, maintenance, and monitoring.

Essentials

- Clinical imaging artificial intelligence (AI) programs require four components for successful implementation: data access and security, cross-platform and cross-domain integration, clinical translation and delivery, and leadership that supports innovation.
- Oversight of AI in medical imaging should consider stakeholders across multiple disciplines who use radiology services.
- AI governance should address the factors used when assessing an algorithm for implementation, different implementation models, and model monitoring and maintenance after implementation.

scientists, compliance representatives, legal representatives, ethics experts, AI experts, and information technology management and end users. Incorporating end users into the governance structure is of utmost importance to consider their needs and concerns about an algorithm and to include it into the decision-making process.

The road map for clinical AI implementation should answer four key questions (Fig 2): Who decides which tools to implement? What should be considered when assessing a tool for implementation? How should each application be implemented in clinical practice? And, finally, how should tools be monitored and maintained for implementation?

Radiology-led versus Organization- or Enterprise-led Governance and Management

Organizational leaders should design their AI governance structures based on the local leadership and the institutional structure in which they will be implemented. Some systems may adopt a governance structure at the health care system level, others may adopt a structure centered within the radiology department or practice, and others will adopt a hybrid structure. Because AI tools can be used at many points in the clinical workflow, these integration points must be considered when determining the composition of management and governance groups. In many health care organizations, radiology practices have led imaging AI implementation, in part because radiology algorithms represent the largest share of U.S. Food and Drug Administration (FDA)-cleared models to date. Accordingly, most organization-wide AI governance structures have radiology representation. Implementing an algorithm that processes multimodal data in the electronic health record (eg, imaging, genomics, telemetry, and patient-reported outcomes) adds complexity and requires leaders with broader informatics expertise.

Both radiology-led and practice-led AI governing bodies have the opportunity to lead by authority and can better guide the decisions about which algorithms are implemented, how they are implemented, and the timeline for implementation. Local control over AI implementation enables more seamless

implementation and integration at the local level. System-level governance tends to lead by influence as it strives to implement programs across departments and ensure institutional resources are transparently and fairly appropriated. A health system-wide AI governance structure supports the entire life cycle of clinical AI: from educational opportunities, research and development, clinical validation, use, and continuous monitoring to long-term maintenance and support. In academic practices, the governance structure may also include clinical investigators, data science researchers, and experts in AI development. Hospital-level and system-level governance structures allow for more complex and broader AI algorithm implementation beyond radiology and enable imaging AI models recommended by other stakeholders to be formally reconciled with radiology practice priorities. System-level AI governance structures allow for broader implementation of AI across the enterprise but are more complex, require more resources, and can impede algorithm implementation compared with departmentally led governance structures. The number of stakeholders complicates this model in that representatives from many clinical groups outside radiology leadership across the health care system will be included.

AI governance bodies must interface both with AI industry partners and enterprise information technology support teams to ensure successful implementation of AI models. The AI intake process will still require formal assessment, preferably quantified, with consideration of clinical safety and benefits, implementation complexity, and business aspects to assess the expected impact of implementation. After approval by the governance committee, a group of end users should have the opportunity to provide formal feedback on the approved algorithm, especially relating to the user interface and user experience, the use of the algorithms, and integration into the clinical workflow.

Hybrid Governance and Management

A hybrid approach to AI governance and management is important to consider when a practice-led or system-led group may not be feasible because of the heterogeneity of radiology practice arrangements. This model incorporates components from both system-wide governance and departmental-led governance, a model including a system-wide governance body with practice or department-centered subcommittees to guide individual decision-making and policy. Because of the challenges to medical imaging AI, the subcommittee to oversee these domains would retain responsibilities similar to the existing imaging informatics and information technology infrastructure committees in most imaging practices. The hybrid approach provides different pathways for AI algorithm implementation depending on complexity levels and the appropriate entry points into the clinical workflow.

Most private and community radiology practices have a good working relationship with their hospitals but are financially independent. This dichotomy makes a hybrid model between the health system and the radiologists most likely to be effective. Well-defined governance structures for AI development, purchase, and implementation in private and community practice are less prevalent than in academic practices. However, as adoption of AI in the community becomes more widespread, structured AI oversight within these radiology practices will be

ESSENTIALS FOR CLINICAL AI

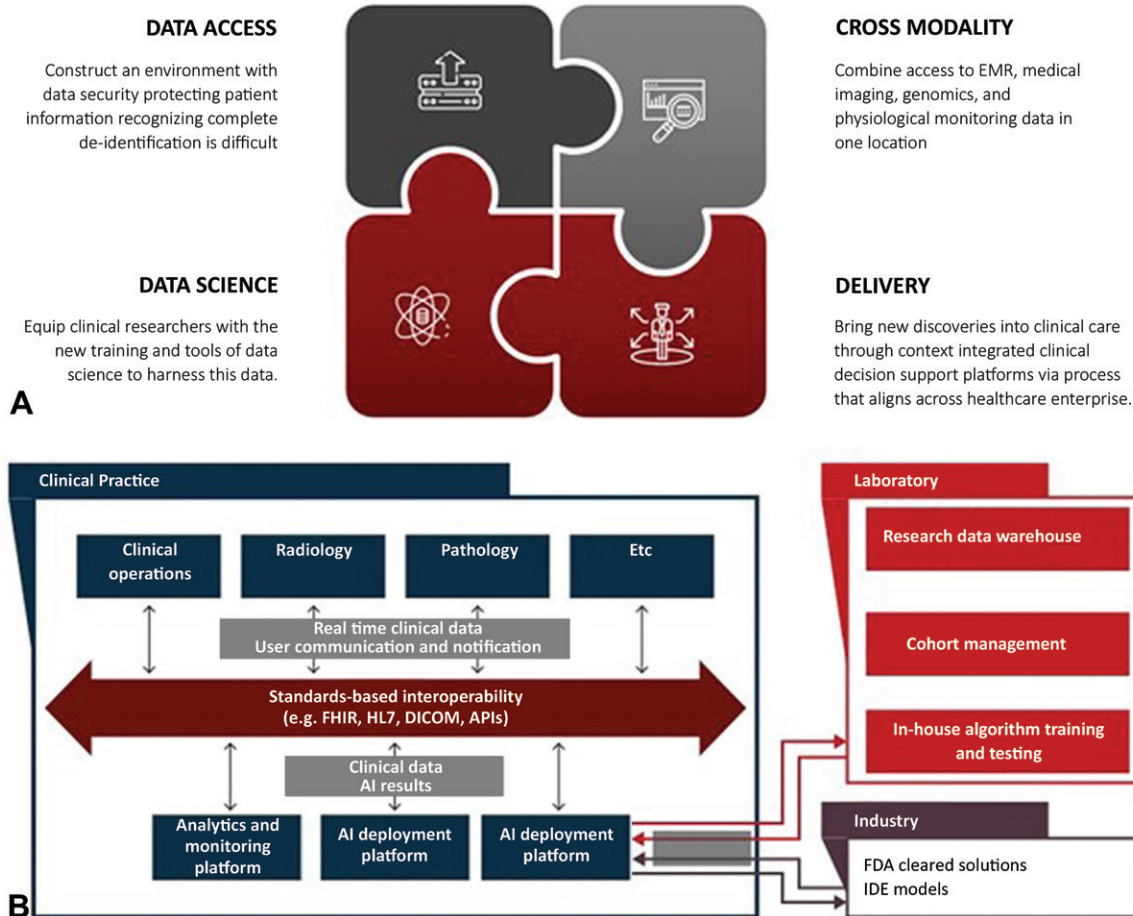


Figure 1: Essential components of clinical artificial intelligence (AI). **(A)** Successful implementation of clinical AI has four components for successful execution: data access and security, cross-platform and cross-domain integration, clinical translation and delivery, and supportive leadership who fosters innovation. EMR = electronic medical records **(B)** With the increased complexity of AI applications, a well-established infrastructure is needed for algorithm implementation. The infrastructure needs to integrate clinical data and should interface with industry and laboratory-built AI solutions, as shown. API = application programming interface, DICOM = Digital Imaging and Communications in Medicine, FDA = Food and Drug Administration, FHIR = Fast Healthcare Interoperability Resources, HL7 = Health Level 7 International, IDE = investigational device exemption.

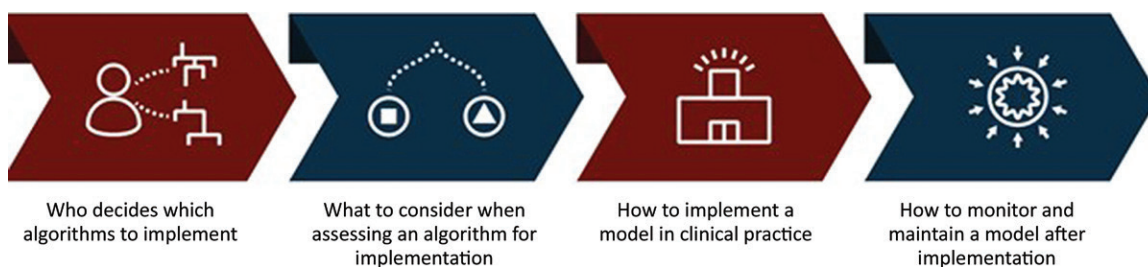


Figure 2: Artificial intelligence (AI) governance road map. Successful oversight of clinical AI implementation can be achieved by following a four-step road map, as shown.

equally important. Results of the American College of Radiology 2019 radiologist workforce survey demonstrated less than 17% of radiology group practices are part of academic university practices, with the majority of the remaining practices falling into the categories of private practice (47%), multispecialty clinic

(12%), and hospital-based practice and corporate practice (4%) (settings considered independent private practice group) (11). Additionally, diversity in practice models exists across the spectrum of nonacademic practices. A recent survey of members of the American College of Radiology showed that larger practices,

whether academic university based or not and especially those with over 50 members, had a higher likelihood of having used AI in their practice (12). However, 18% of groups with one to five members reported the use of at least one AI model. Because of the diversity of community practice size, it is unlikely a single model of AI governance will prevail across all practices.

Issues that might be less important for academic practices, such as which entity pays to install and maintain the AI models, will be important in community hospital settings, and radiologists will need to develop win-win scenarios with AI models that are valuable both to the radiologists and to the health system. Although AI vendors may engage in direct marketing to clinical staff or hospital administration without consulting the radiology department, the models will be implemented and monitored within the radiology department. Therefore, radiologists must have a role in the decision-making process before these models are used. If a health system bears the financial burden for AI, radiologists must develop the value proposition for each model. If models are seen as only improving radiologist efficiency or accuracy, then the radiology group may be asked to bear some or all of the financial cost. Articles have been published previously (13–15) that address the business case of AI in radiology and its role in improving quality and reducing costs.

Governance throughout the Algorithm Life Cycle

Important initial considerations for any governance structure are the scope of applications governed, the extent of review required for different use cases, and the point at which an AI algorithm is integrated into clinical workflow. For imaging algorithms, integration points include image acquisition (eg, CT scanner, MRI scanner, radiography unit, and US unit), image viewing (eg, patient triage, worklist prioritization, and image overlays), and interpretation and reporting. Beyond imaging, tools with the electronic health record can include inpatient-specific clinician decision support, population health management programs, or risk stratification in health plans. Certain imaging AI tools may require hybrid governance because they are designed to be used outside the radiologist workflow. For example, a tool may deliver imaging results directly to the referring clinician. In such cases, implementation requires governance not only from the radiology practice but also hospital leadership, information technology, information security, and compliance.

The intensity of oversight for a given application should be based on the stage of development and whether the algorithm has been cleared by the FDA, the clinical use case, the complexity of the application, and the level of human supervision of the algorithm. The International Medical Device Regulators Forum (IMDRF) Software as a Medical Device working group proposed a framework for risk categorization and for considering the safety and regulation of medical devices, citing the increasing complexity of software, the increasing connectedness of systems, and the resultant emergent behaviors not seen in hardware medical devices (16,17). The IMDRF model serves as a useful template for assessing the risks of AI implementation. The IMDRF risk categorization uses the impact of the medical device software on medical decision-making and the severity and acuity of the intended use case to stratify risk into

Table 1: International Medical Device Regulators Forum Software as a Medical Device Risk Categorization

State of Health Care Situation or Condition	Information Provided by SaMD to Health Care Decision		
	Treat or Diagnose	Drive Clinical Management	Inform Clinical Management
Critical	IV	III	II
Serious	III	II	I
Nonserious	II	I	I

Note.—SaMD categories are as follows: I, low impact; II, medium impact; III, high impact; and IV, very high impact. SaMD = Software as a Medical Device.

four categories (Table 1). Subsequent documents published by the IMDRF addressed how risk stratification might affect the clinical evaluation of software as a medical device (18–20). This framework may help to simplify governance around lower risk algorithms such as those that prioritize the radiology worklist or improve image reconstruction. These relatively low-risk tools could be locally handled within radiology departments.

Consider the case of an FDA-approved AI algorithm to detect and characterize large vessel thrombotic stroke on CT images and to inform nonradiologist clinicians whether emergent thrombectomy is indicated. This solution may be recommended by nonradiologist clinicians. A system-wide imaging AI governance structure would facilitate the necessary cross-departmental oversight regarding evaluation, installation, monitoring, and maintenance. Nonradiologists may also work with hospital leadership to acquire medical imaging AI solutions independent of radiology even if the solutions are not clinician-facing. For example, a pulmonary nodule detection algorithm might be requested by cardiologists or pulmonologists who would be less likely to engage with a governance structure contained within radiology without broader medical system oversight.

It is not always possible for radiology practices to engage in system-level decision-making regarding the implementation of AI algorithms, even when they leverage medical imaging data. Individuals with informatics expertise who work for the radiology practice can integrate the algorithm in the practice with vendor support; for simple applications, the process can be analogous to other software installations. However, more complex AI tools might require not only domain expertise but also computing resources, monitoring processes, and methods for data access. The governance team will have to expand to include diverse experts who can review evidence, perform utility analysis, estimate risk, assess technical and clinical readiness, and predict economic effects. Ethics and fairness review should be incorporated into the algorithm assessment process. Whereas this complete portfolio of expertise will be beyond the resources of many smaller practices, a governing body (in cooperation with the hospital system) could include clinical and scientific domain experts, technical leads, quality experts, an information technology operations coordinator, and medical ethics experts.

MOCK SCORING RUBRIC



Figure 3: Artificial intelligence (AI) algorithm evaluation for clinical implementation mock scoring rubric. When evaluating an AI solution for clinical implementation, a number of areas should be assessed and scored. These include data security risk, clinical readiness, technical readiness, available evidence for validation and performance metrics, and clinical value the tool will bring. UX = user experience.

AI Governance Process

When a user or group identifies a clinical need, they may identify an AI application that addresses the need and request implementation of that tool to the governance group. The request should be evaluated by using a rubric-based analysis that gathers information about several domains, including local performance metrics and ease of use, evidence basis for efficacy, technical readiness, value to patient safety, quality of care or cost efficiency, and a detailed clinical impact. A structured scoring rubric can inform decisions to adopt or reject the request (Fig 3). Representative applications of the scoring rubric are included in Table 2, summarizing the scores for the bone age assessment algorithms and for the COVID-19 mortality prediction algorithm. The former received a high score on the rubric and was eventually implemented, whereas the latter was not implemented. As the number of AI applications grows, this process will provide institutional memory to avoid duplication when other solutions are proposed to address similar clinical use cases.

Preimplementation Considerations

Many consumers of AI believe they can rely on FDA clearance (21) and vendor marketing (22) to ensure AI applications will work as expected in clinical practice. But even those tools cleared by the FDA may not perform as expected outside of the environments where they are trained. Local evaluation of AI tools by radiology practices will be equally important in community practices as in academic practices. Multiple components are typically included in the scoring rubrics established by the AI governance committee when evaluating an application for implementation.

Data security risks that may result from application implementation should be considered first. In a recent report by the National Academy of Medicine regarding health data sharing

(23), developing and maintaining a trustworthy data sharing environment was identified as a key component of any data sharing strategy. Specifically, appropriate data oversight, disclosure, and consent should be adopted to ensure suitable data security and sharing practices at every step of the implementation process. These factors should be carefully vetted for all AI tools, especially when tools are used in the cloud and in cases where practices or research entities develop their own AI tools for clinical deployment. Many governing committees have representatives from compliance, ethics, and the institutional review board to ensure that these conditions are met. The current practice is for governing committees to design data flow maps that identify data assets and data flows pertaining to each project. These maps describe the physical location (on-site or off-site) and encryption status of the data at all steps. They provide an opportunity to identify steps where added security measures such as de-anonymization, encryption, or additional data processing are needed, especially if the data are passed and processed by an outside entity. This detailed analysis and accountability for data security and robust governance of data use are vital steps toward ensuring patient protection and public trust.

In assessing clinical readiness and local performance metrics, several facets must be considered. The reproducibility and generalizability of the model are of primary importance (24). The AI governance committee should assess whether the data set used for model development is congruent with the setting in which the model will be used. Any inclusion or exclusion criteria applied during model development should be carefully evaluated for the presence of selection bias. Because imaging AI algorithms are especially prone to overfitting or overparameterization, only external validation can ensure that the predictive accuracy of the model is sufficiently robust beyond the cohort that was used for

Table 2: Representative Scoring of Two Artificial Intelligence Algorithms

Parameter	Algorithm 1: Automated Bone Age Assessment (Total Points, 115)	Algorithm 2: COVID-19 Mortality Prediction on Chest Radiograph (Total Points, 45)
Local performance ease of use	20—Good local performance, good UX	10—Modest local performance, useable UX
Scientific evidence	30—Level I evidence	0—None/pilot data
Fairness, bias, and/or harm	30—Lowest risk	0—Unknown
Technical readiness workflow impact	20—Minor technical modification	20—Minor technical modification
Value and cost	5—Net negative	5—Net negative
Clinical impact	10—Moderate volume	10—Moderate volume

Note.—The representative scoring uses the proposed scoring rubric. The first algorithm was approved for implementation, whereas the second was not. UX = user experience.

model development (25). Furthermore, the accuracy and reproducibility of the tool should be quantified and compared with sensitivity and specific thresholds chosen to enable clinical utility. Institutions may consider preparing an unbiased local data set to test the algorithm and confirm its performance before use.

The available scientific evidence supporting the technical readiness and robustness of the proposed application should also be considered (26). Standard metrics include the area under the receiver operating characteristic curve; sensitivity and specificity; the precision-recall curve; and, when applicable, regression metrics (root mean square error, mean absolute error, R^2). Several useful guidelines have been published for evaluating and reporting the results of AI models, including Transparent Reporting of a Multivariable Prediction Model of Individual Prognosis or Diagnosis (known as TRIPOD-AI) (27), Standards for Reporting of Diagnostic Accuracy Studies (known as STARD-AI) (28), and Consolidation Standards of Reporting Trials (known as CONSORT-AI) (29); any or all of these might be appropriate for evaluating tools. Finally, the committee should evaluate whether the tool has been tested in a comparable operational environment to its intended local use (25). Performance of any tool should be closely scrutinized against both retrospective and prospective data before clinical use. Technical readiness for deployment across a variety of imaging facilities with potentially diverse acquisition parameters, postprocessing methods, and Digital Imaging and Communications in Medicine metadata must also be considered and assessed before implementation (30).

For tools that meet these baseline technical and operational performance requirements, the effect on patient care and outcomes should be assessed. A tool that minimally impacts patient care or creates data irrelevant to clinical decision-making will deliver minimal clinical value. For example, an AI governing body could assess whether the output labels generated by the tool reflect the true disease states by comparing AI results to human expert consensus or a reference standard from the electronic health record (25).

Next, the AI application should be assessed regarding whether its output is easy to use from a user interface and user experience perspective (ie, user interface and user experience) (26). Tools with difficult-to-interpret results should be modified to provide results that are more easily understood and that facilitate rapid clinical interpretation. Next, the application should be assessed based on whether it will complement the current clinical

workflow. Finally, to establish the value of the tool, the impact on patient care and outcomes should be assessed (26). The main predictor of a successful clinical deployment is when clinicians believe that it improves patient care. Before implementation, the governing committee should establish safeguards to prevent patient harm, especially in high-risk scenarios, such as screening applications in otherwise healthy populations or for tools that make drug or treatment recommendations. Legal considerations should also be addressed. Although the regulation of AI is in its early stages, a legal representative should be included in every AI governance committee. For every algorithm, the risk for liability and malpractice should be discussed. The FDA defines four areas where diagnostic devices may be harmful, as follows: increasing false-positive rates, increasing false-negative rates, providing incorrect outputs, and being misused by humans. It is currently unclear who is liable for an error made by an AI algorithm. The topic of legal considerations for AI algorithms is covered at length elsewhere (14,31,32). Beneficence, doing no harm, and patient autonomy should remain at the center of all decisions made by an AI governance committee.

For every AI application, the committee must also consider fairness and other ethical issues (32,33). The algorithm should be assessed for fairness, and safeguards put in place to eliminate systemic biases that may influence the model's outcomes including those for disadvantaged communities.

Recommendations for Implementing an AI Algorithm in Clinical Practice

Once an algorithm has been approved by the governance group, responsible resources must work with vendors or internal developers for robustness and integration testing, ideally with staged “shadow” and pilot implementation, respectively (Fig 4). In shadow deployment, clinical data are fed to the algorithm in real time and results are gathered to assess performance and safety, but the generated results are not provided to clinical users. In a pilot deployment, chosen clinical users test the model in a limited part of the practice and provide production feedback before full clinical deployment. After these staged deployments, prechosen metrics are reviewed to determine if the application warrants a fuller implementation, further assessment, or rejection.

The initial deployment mode for each AI tool will vary based on risk and context. A triage algorithm that selects some studies as normal without clinician input might require extensive

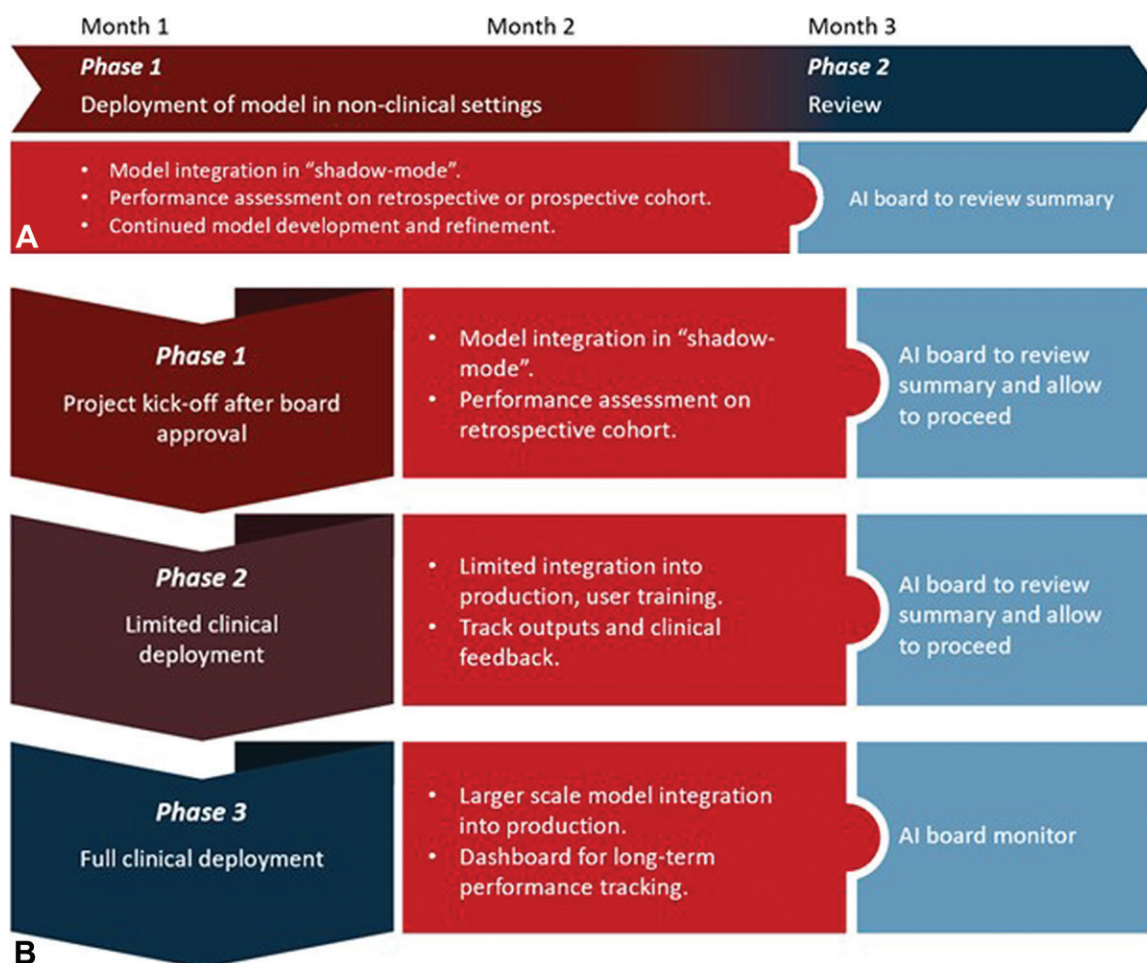


Figure 4: Modes of artificial intelligence (AI) implementation and integration in clinical practice. Two modes of clinical implementation are typically pursued by AI governance structures: shadow mode and canary clinical mode. **(A)** In shadow mode, tools are first implemented in the background with no impact on clinical decision-making. During this time, algorithms are being tested and refined prior to live clinical implementation. **(B)** In model integration into production, or canary mode implementation, tools are released in the live clinical environment using a phased approach while undergoing rigorous oversight by the AI governance structure.

shadow testing. However, a measurement algorithm requiring human interaction to approve its output might be used in a pilot mode based on good retrospective performance data. Algorithms in which the primary goal is to increase efficiency can often be tested in a simulation environment to measure effects on clinical workflow.

Monitoring and Maintaining an Application after Implementation

Whereas many radiology practices focus on implementation, the maintenance and monitoring of AI applications may be just as vital to long-term success and should be in place before the launch of any tool to ensure patient safety. Monitoring AI applications over time is difficult and resource intensive in any context. In the clinical imaging domain, the lack of established standards and best practices make such monitoring especially challenging.

Regardless of governance structure, once an application has been introduced clinically, metrics should be established before clinical implementation and monitored continuously for product improvement. For example, the quality team could monitor discrepancies between the AI algorithm and radiologists' reports.

This form of quality control and monitoring can mitigate the risk of prediction drift, feature drift, and input data errors, and can help identify problems early. Discrepancies could also be used to educate radiologists about discrepancies, both those driven by radiologists as well as those originating from shortcomings in the AI tool.

Explainable AI exploits the ability of some AI models to provide feedback on how input features drive generated output. This form of explanation can help identify the cause of observed problems, where exploration of the importance of each parameter—and in some cases, decision tree structure—can more readily provide a human-interpretable explanation of flawed model decision-making. However, whereas the ability to explain can be a valuable trait in an AI tool, it does not equate to trustworthiness (34). Unfortunately, unlike classic machine learning methods applied to electronic health record data, the ability to interpret and explain are typically more challenging for imaging AI tools, which often use deep learning methods with millions of parameters.

Most health care systems use periodic retesting to evaluate performance over time or rely on user-reported issues to

reveal problems. These strategies are used for other clinical enterprise software solutions but are insufficient for AI-based systems, wherein the performance is less predictable. In other industries, a monitoring paradigm, MLOps (machine learning operations), has emerged. This monitoring paradigm outlines patterns for setting and tracking leading and lagging metrics of model health and performance in production. Once a performance deviation has been identified, an investigation is launched to identify the root causes and propose solutions (eg, retrain the model on new data or repair data pipelines). Such robust maintenance and monitoring processes are essential for successful long-term implementation of AI tools in clinical practice.

Whereas large private practice groups may have well-established pipelines for AI implementation, establishing monitoring processes in smaller practices and those without sophisticated informatics resources can be more challenging, both because of a lack of a dedicated informatics team specifically for radiology and because of scarce resources and end users' intolerance of workflow disruptions. Registries that are automatically populated by reporting or viewing software can reduce the radiologist burden of performance monitoring. Developers should also be encouraged to incorporate seamless data capture for performance monitoring into their products.

Key Considerations and Future Directions

Whether based at the practice, at the institutional level, or as a hybrid, effective AI governance requires interdisciplinary organization and collaboration among key stakeholders to ensure successful clinical implementation of AI tools. Effective governance structures establish a standardized implementation timeline with processes for evaluation, initial implementation, and monitoring of AI applications, with gates or checkpoints for each phase. A robust monitoring plan—and continuous learning procedure, if applicable—detects performance degradation and allows for early intervention. Regardless of governance structure, equitable allocation of available resources should be considered when evaluating tools for implementation, and capacity and efficiency must also be considered. AI governance bodies assume responsibility for balancing the desire to drive innovation forward with processes to ensure quality and safety in clinical implementation.

As more AI tools are considered for implementation, standard strategies should be adapted to the life cycle of each tool. The governing body may become aware of tools at different development stages, leveraging different technology platforms, and requiring different points of clinical integration. A set of useful tools that support a variety of scenarios will allow for more effective and seamless clinical implementation of AI applications. Governing bodies should establish processes to stratify risk and determine the appropriate path for initial implementation and monitoring. The stage of tool development, FDA clearance status, and method of clinical implementation may inform risk and guide the implementation strategy. Implementation of early-stage tools may favor governance decentralization and shadow implementation, whereas validated applications in

low-risk settings may immediately undergo clinical implementation after expedited approval from central governance structures. The governance process we propose herein may require financial and resource investment at the institutional level. However, this investment is essential to ensure the quality and safety of AI implementation in clinical practice.

The complexity of governance and management of AI implementation likely has contributed to the slow introduction of AI into clinical practice. As the number of FDA AI algorithms increases and the cost of evaluating and monitoring decreases, the use of AI likely will begin to grow. Radiologists and their institutional partners in both community and academic practice will need to address these AI governance issues. We acknowledge that building the governance structure we propose in our review may not be feasible in nonacademic centers, specifically community hospitals and small private radiology practices. As governance of AI evolves, national organizations such as the American College of Radiology may take a larger role in vetting AI algorithms for clinical use, which may make implementation more practical for smaller practices.

From its inception, radiology has led many technological revolutions. Despite the challenges that are brought by major changes, transformation is not foreign to our specialty. We have always adapted. The advent of artificial intelligence (AI) marks yet another such watershed moment in the history of our specialty. Appropriate governance and management structures will empower us to adapt to this change and fully embrace it, even though it may be years before mature AI tools are routinely integrated into daily practice. Although the infrastructure and investment required for clinical implementation of AI is daunting and imminent changes will bring many challenges—uncertainty and fear not least among them—we believe this transformation will immensely benefit our specialty and presents us with an opportunity to lead as AI augments the practice of medicine.

Disclosures of conflicts of interest: **D.D.** Grant from RSNA, Society of Interventional Radiology, and CRICO; consulting fees from Sigilon Therapeutics, Medtronic; data safety board participation from Massachusetts General Hospital. **W.F.W.** Honoraria from Stanford AIMI Symposium, University of Wisconsin GE CT Protocol Project and Medical Advisory Board. **M.P.L.** Consulting fees from Microsoft, Nines Radiology, Bayer, Segmed; stock/stock options from Nines Radiology, BunkerHill, SegMed, Centaur. **T.A.** Consulting fees from Nuance Communications; data safety board participation from Nuance Communications. **N.K.** Meeting/travel support from Radiology Partners; patents planned, issued, or pending from Radiology Partners; steering committee member with RadXX, SIIM, American College of Radiology; stock/stock options, Radiology Partners. **B.A.** Meeting support from the American College of Radiology; leadership role, ACR, Data Science Institute, International Society of Radiology. **C.J.R.** No relevant relationships. **B.C.B.** No relevant relationships. **K.D.** No relevant relationships. **J.A.B.** Board of directors, Accumen. **D.B.L.** Stock/stock options from Bunkerhill Health. **K.J.D.** Chief science officer for ACR Data Science Institute. **C.P.L.** Board of directors, RSNA; board of directors and shareholder, Bunkerhill Health; advisor and option holder, whiterabbit.ai, Nines, GalileoCDS, Sirona Medical, Adra.

References

1. Lui YW, Geras K, Block KT, Parente M, Hood J, Recht MP. How to Implement AI in the Clinical Enterprise: Opportunities and Lessons Learned. *J Am Coll Radiol* 2020;17(11):1394–1397.
2. Driver CN, Bowles BS, Bartholmai BJ, Greenberg-Worisek AJ. Artificial Intelligence in Radiology: A Call for Thoughtful Application. *Clin Transl Sci* 2020;13(2):216–218.

3. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA* 2016;316(22):2353–2354.
4. Sohn JH, Chillakuru YR, Lee S, et al. An Open-Source, Vendor Agnostic Hardware and Software Pipeline for Integration of Artificial Intelligence in Radiology Workflow. *J Digit Imaging* 2020;33(4):1041–1046.
5. Dikici E, Bigelow M, Prevedello LM, White RD, Erdal BS. Integrating AI into radiology workflow: levels of research, production, and feedback maturity. *J Med Imaging (Bellingham)* 2020;7(1):016502.
6. Bizzo BC, Almeida RR, Michalski MH, Alkasab TK. Artificial Intelligence and Clinical Decision Support for Radiologists and Referring Providers. *J Am Coll Radiol* 2019;16(9 Pt B):1351–1356.
7. Kapoor N, Lacson R, Khorasani R. Workflow Applications of Artificial Intelligence in Radiology and an Overview of Available Tools. *J Am Coll Radiol* 2020;17(11):1363–1370.
8. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *Eur Radiol* 2020;30(6):3576–3584.
9. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology* 2018;286(3):800–809.
10. Willemink MJ, Koszek WA, Hardell C, et al. Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020;295(1):4–15.
11. Bender CE, Bansal S, Wolfman D, Parikh JR. 2019 ACR Commission on Human Resources Workforce Survey. *J Am Coll Radiol* 2020;17(5):673–675.
12. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR Data Science Institute Artificial Intelligence Survey. *J Am Coll Radiol* 2021;18(8):1153–1159.
13. Golding LP, Nicola GN. A Business Case for Artificial Intelligence Tools: The Currency of Improved Quality and Reduced Cost. *J Am Coll Radiol* 2019;16(9 Pt B):1357–1361.
14. Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Can Assoc Radiol J* 2018;69(2):120–135.
15. Martin-Carreras T, Chen P-H. From Data to Value: How Artificial Intelligence Augments the Radiology Business to Create Value. *Semin Musculoskelet Radiol* 2020;24(1):65–73.
16. International Medical Device Regulators Forum. Software as a Medical Device (SaMD). <http://www.imdrf.org/workitems/wi-samd.asp>. Accessed May 25, 2021.
17. IMDRF Software as a Medical Device (SaMD) Working Group. Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations. International Medical Device Regulators Forum; 2014. <http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.docx>. Accessed May 25, 2021.
18. IMDRF SaMD Working Group. Software as a Medical Device (SaMD): Application of Quality Management System. International Medical Device Regulators Forum; 2015. <http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-151002-samd-qms.docx>. Accessed May 25, 2021.
19. IMDRF Software as a Medical Device Working Group. Software as a Medical Device (SaMD): Clinical Evaluation. International Medical Device Regulators Forum; 2016. <http://www.imdrf.org/docs/imdrf/final/consultations/imdrf-cons-samd-ce.pdf>. Accessed May 25, 2021.
20. Moshi MR, Parsons J, Tooher R, Merlin T. Evaluation of Mobile Health Applications: Is Regulatory Policy Up to the Challenge? *Int J Technol Assess Health Care* 2019;35(4):351–360.
21. Harvey HB, Gowda V. How the FDA Regulates AI. *Acad Radiol* 2020;27(1):58–61.
22. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol* 2021;31(6):3797–3804.
23. National Academy of Medicine. Health Data Sharing to Support Better Outcomes: Building a Foundation of Stakeholder Trust. 2020. https://nam.edu/wp-content/uploads/2020/11/Health-Data-Sharing-to-Support-Better-Outcomes_prepub-final.pdf. Accessed May 25, 2021.
24. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health* 2020;2(9):e489–e492.
25. Faes L, Liu X, Wagner SK, et al. A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. *Transl Vis Sci Technol* 2020;9(2):7.
26. Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Health Care Inform* 2021;28(1):e100251.
27. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019;393(10181):1577–1579.
28. Sounderajah V, Ashrafian H, Aggarwal R, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat Med* 2020;26(6):807–808.
29. Liu X, Rivera SC, 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. *BMJ* 2020;370:m3164.
30. Baker B, Chilamkurthy S, Rao P. Engineering Radiology AI for National Scale in the US. 2021. <https://blog.que.ai/notes/engineering-radiology-AI-US-national-scale>. Accessed May 25, 2021.
31. Walters R, Novak M. Artificial Intelligence and Law. In: *Cyber Security, Artificial Intelligence, Data Protection & the Law*. Singapore: Springer, 2021; 39–69.
32. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging* 2018;9(5):745–753.
33. Geis JR, Brady AP, Wu CC, et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *Radiology* 2019;293(2):436–440.
34. Kitamura FC, Marques O. Trustworthiness of Artificial Intelligence Models in Radiology and the Role of Explainability. *J Am Coll Radiol* 2021;18(8):1160–1162.