REVIEW ARTICLE

**AI IN MEDICINE**

Jeffrey M. Drazen, M.D., *Editor*, Isaac S. Kohane, M.D., Ph.D., *Guest Editor*, and Tze-Yun Leong, Ph.D., *Guest Editor*

# The Current and Future State of AI Interpretation of Medical Images

Pranav Rajpurkar, Ph.D., and Matthew P. Lungren, M.D., M.P.H.

THE INTERPRETATION OF MEDICAL IMAGES — A TASK THAT LIES AT THE heart of the radiologist's work — has involved the growing adoption of artificial intelligence (AI) applications in recent years. This article reviews progress, challenges, and opportunities in the development of radiologic AI models and their adoption in clinical practice. We discuss the functions that AI-based algorithms serve in assisting radiologists, including detection, workflow triage, and quantification, as well as the emerging trend of the use of medical-imaging AI by clinicians who are not radiologists. We identify the central challenge of generalization in the use of AI algorithms in radiology and the need for validation safeguards that encompass clinician–AI collaboration, transparency, and post-deployment monitoring. Finally, we discuss the rapid progress in developing multimodal large language models in AI; this progress represents a major opportunity for the development of generalist medical AI models that can tackle the full spectrum of image-interpretation tasks and more. To aid readers who are unfamiliar with terms or ideas used for AI in general or AI in image interpretation, a Glossary is included with this article.

From the Department of Biomedical Informatics, Harvard Medical School, Boston (P.R.); the Center for Artificial Intelligence in Medicine and Imaging, Stanford University, Stanford, and the Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco — both in California (M.P.L.); and Microsoft, Redmond, Washington (M.P.L.). Dr. Rajpurkar can be contacted at pranav_rajpurkar@hms.harvard.edu.

Drs. Rajpurkar and Lungren contributed equally to this article.

N Engl J Med 2023;388:1981-90.
DOI: 10.1056/NEJMra2301725


In recent years, AI models have been shown to be remarkably successful in interpretation of medical images.[1] Their use has been extended to various medical-imaging applications, including, but not limited to, the diagnosis of dermatologic conditions[2] and the interpretation of electrocardiograms,[3] pathological slides,[4] and ophthalmic images.[5] Among these applications, the use of AI in radiology has shown great promise in detecting and classifying abnormalities on plain radiographs,[6] computed tomographic (CT) scans,[7] and magnetic resonance imaging (MRI) scans,[8] leading to more accurate diagnoses and improved treatment decisions.

Even though the Food and Drug Administration (FDA) has approved more than 200 commercial radiology AI products, substantial obstacles must be overcome before we are likely to see widespread successful clinical use of these products. The incorporation of AI in radiology poses both potential benefits and challenges for the medical and AI communities. We expect that the eventual resolution of these issues and more comprehensive solutions, including the development of new foundation models, will lead to broader adoption of AI within this health care sector.

## AI USE IN RADIOLOGY

Radiology as a specialty is well positioned for the application and adoption of AI because of several key factors. First, AI excels in analyzing images,[9] and unlike other specialties that use imaging, radiology has an established digital workflow and universal standards for image storage, so that it is easier to integrate AI.[10]

**Glossary**

**Continual learning:** A process in which an AI model learns from new data over time while retaining previously acquired knowledge.

**Data set shift:** The shift from data used to train a machine-learning model to data encountered in the real world. This shift can cause the model to perform poorly when used in the real world, even if it performed well during training.

**Federated learning:** A distributed machine-learning approach that enables multiple devices or nodes to collaboratively train a shared model while keeping their individual data local, thereby preserving privacy and reducing data communication overhead.

**Foundation models:** AI models that serve as a starting point for developing more specific AI models. Foundation models are trained on large amounts of data and can be fine-tuned for specific applications, such as detecting lesions or segmenting anatomical structures.

**Generalist medical AI models:** A class of advanced medical foundation models that can be used across various medical applications, replacing task-specific models. Generalist medical AI models have three key capabilities that distinguish them from conventional medical AI models. They can adapt to new tasks described in plain language, without requiring retraining; they can accept inputs and produce outputs using various combinations of data types; and they are capable of logically analyzing unfamiliar medical content.

**Large language models:** AI models consisting of a neural network with billions of weights or more, trained on large amounts of unlabeled data. These models have the ability to understand and produce human language and may also apply to images and audio.

**Multimodal models:** AI models that can understand and combine different types of medical data, such as medical images and electronic health records. Multimodal models are particularly useful in medicine for tasks that require a comprehensive understanding of the patient, such as diagnosis and individualized treatment planning.

**Self-supervised models:** AI models that can learn from medical data without the need for explicit annotations. These models can be used to learn representations of medical data that are useful for a wide range of tasks, such as diagnosis and patient monitoring. Self-supervised models are particularly useful in medicine when labeled data are scarce or expensive to obtain.

**Zero-shot learning:** The capability of an AI model to perform a task or solve a problem for which it has not been explicitly trained, without the need for any additional training data. In medicine, this can be particularly useful when there is a shortage of labeled data available for a specific medical task.

Furthermore, AI fits naturally in the workflow of image interpretation and can replicate well-defined interpretive tasks effectively.[11]

## AI USE FOR RADIOLOGISTS

AI can be used in the field of radiology to analyze images from a wide range of techniques, including radiography, CT, ultrasonography, and MRI. Radiologic AI algorithms serve a number of narrow image-analysis functions to assist radiologists, such as quantification, workflow triage, and image enhancement (Fig. 1).[1,12-17] Quantification algorithms perform segmentation and measurements of anatomical structures or abnormalities. Common examples include measuring breast density, identifying anatomical structures in the brain, quantitating cardiac flow,[18] and assessing local lung-tissue density. Workflow triage involves flagging and communicating suspected positive findings, including, but not limited to, intracranial hemorrhage, intracranial large-vessel occlusion,[19] pneumothorax,[20] and pulmonary embolism. AI is also used for the detection, localization, and classification of conditions such as pulmonary nodules and breast abnormalities. In addition, AI algorithms enhance preinterpretive processes, including image reconstruction, image acquisition, and mitigation of image noise.[17]

There is promise in exploring radiologic AI models that can expand interpretive capabilities beyond those of human experts. For instance, AI algorithms can accurately predict clinical outcomes on the basis of CT data in cases of traumatic brain injury[21] and cancer.[22] In addition, AI-derived imaging biomarkers can help to quickly and objectively assess structures and pathological processes related to body composition, such as bone mineral density, visceral fat, and liver fat, which can be used to screen for various health conditions.[23] When applied to routine CT imaging, these AI-derived biomarkers are proving useful in predicting future adverse events.[24] Moreover, recent research has shown that coronary-artery calcium scores, which are typically obtained on the basis of CT scanning, can be determined by means of cardiac ultrasonography.[25] These findings point to the value of radiologic AI models for patients (e.g., no radiation exposure).

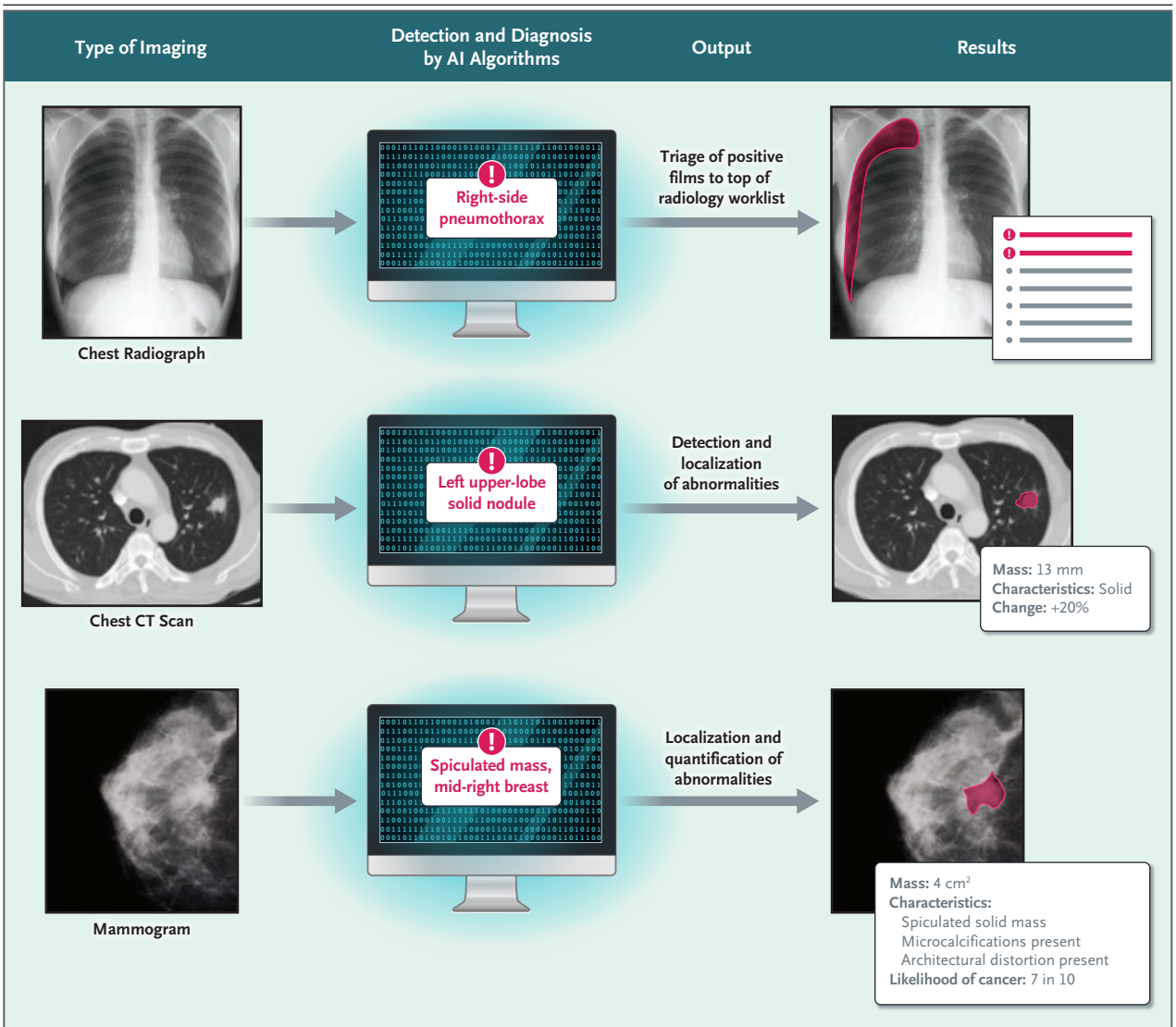


**Figure 1. Current Uses of Artificial Intelligence (AI) in Radiology.**

Shown are common clinical AI solutions for the functions of triage, detection, and diagnosis with CADt (computer-aided detection for triage), CADe (computer-aided detection for characterizing abnormalities), and CADx (computer-aided detection for diagnosis). Other AI applications for radiology include image reconstruction and noise reduction. Applications for nonimaging tasks are not shown. CT denotes computed tomography.

Radiologic AI has attracted global interest, and commercial AI algorithms have been developed by companies based in more than 20 countries. Studies have shown that some hospitals, as well as other point-of-care centers, already use AI products successfully, and larger practices are more likely than smaller practices to use AI currently. Radiologists who use AI in their practices are generally satisfied with their experience and find that AI provides value to them and their patients. However, radiologists have expressed concerns about lack of knowledge, lack of trust, and changes in professional identity and autonomy.[26] Local champions of AI, education, training, and support can help overcome these concerns. The majority of radiologists and residents expect substantial changes in the radiology profession within the next decade and believe that AI should have a role as a "co-pilot," acting as a second reader and improving workflow tasks.[27]

Although the penetration of AI in the U.S. market is currently estimated to be only 2%, the readiness of radiologists and the potential of the technology indicate that further translation into clinical practice is likely to occur.

### EMERGING USES FOR NONRADIOLOGISTS

Although many current radiologic AI applications are designed for radiologists, there is a small but emerging trend globally toward the use of medical-imaging AI for nonradiologist clinicians and other stakeholders (i.e., health care providers and patients). This trend presents an opportunity for improving access to medical imaging and reducing common diagnostic errors[28] in low-resource settings and emergency departments, where there is often a lack of around-the-clock radiology coverage.[29] For instance, one study showed that an AI system for chest radiograph interpretation, when combined with input from a nonradiology resident, had performance values that were similar to those for board-certified radiologists.[30] A popular AI application that is targeted for use by nonradiologist clinicians for detecting large-vessel occlusions in the central nervous system has resulted in a significant reduction in time to intervention and improved patient outcomes.[31] Moreover, AI has been shown to accelerate medical-imaging acquisition outside traditional referral workflows with new, clinician-focused mobile applications for notifications of AI results.[32] This trend, although not well established, has been cited as a potential long-term threat to radiology as a specialty because advanced AI models may reduce the complexity of technical interpretation so that a nonradiologist clinician could use imaging without relying on a radiologist.[33,34]

Portable and inexpensive imaging techniques are frequently supported by AI and have served to lower the barrier for more widespread clinical use of AI in medical imaging outside the traditional radiology workflow.[35,36] For example, the Swoop portable MRI system, a point-of-care device that addresses existing limitations in forms of imaging technology, provides accessibility and maneuverability for a range of clinical applications. The system plugs into a standard electrical outlet and is controlled by an Apple iPad. Portable ultrasound probes and smartphones in AI-enabled applications can be used to obtain diagnostic information even by users without formal training in echocardiography or the use of ultrasound in obstetrical care.[37] Overall, although the use of medical-imaging AI by nonradiologist clinicians is still in the early stages, it has the potential to revolutionize access to medical imaging and improve patient outcomes.

## SAFEGUARDS FOR EFFECTIVE GENERALIZATION

In considering the widespread adoption of AI algorithms in radiology, a critical question arises: Will they work for all patients? The models underlying specific AI applications are often not tested outside the setting in which they were trained, and even AI systems that receive FDA approval are rarely tested prospectively or in multiple clinical settings.[38] Very few randomized, controlled trials have shown the safety and effectiveness of existing AI algorithms in radiology, and the lack of real-world evaluation of AI systems can pose a substantial risk to patients and clinicians.[39]

Moreover, studies have shown that the performance of many radiologic AI models worsens when they are applied to patients who differ from those used for model development, a phenomenon known as "data set shift."[40-44] In interpretation of medical images, data set shift can occur as a result of various factors, such as differences in health care systems, patient populations, and clinical practices.[45] For instance, the performance of models for brain tumor segmentation and chest radiograph interpretation worsens when the models are validated on external data collected at hospitals other than those used for model training.[46,47] In another example, a retrospective study showed that the performance of a commercial AI model in detecting cervical spine fractures was worse in real-world practice than the performance initially reported to the FDA.[48] Patient age, fracture characteristics, and degenerative changes in the spine affected the sensitivity and false positive rates to an extent that limited the clinical usefulness of the AI model and aroused concerns about the generalization of radiologic AI algorithms across clinical environments.

There is a pressing need for the development of methods that improve the generalization of algorithms in new settings.[49-51] As the field matures, better generalization checks based on accepted standards must be established before the algorithms are widely applied. These checks encompass three related areas: clinician–AI collaboration, transparency, and monitoring (Fig. 2).


**Figure 2. Generalization Checks for AI Systems in Radiology.**
The three essential components of generalization checks for radiologic AI systems are clinician–AI collaboration, transparency, and postdeployment monitoring. Clinician–AI collaboration reflects the need to move from evaluations of the stand-alone performance of AI models to evaluations of their value as assistive tools in real-world clinical workflows. Transparency with regard to lack of information about an AI model requires greater rigor through the use of checklists and public release of medical-imaging data sets. Postdeployment monitoring involves mechanisms to incorporate feedback from clinicians and continual learning strategies for regular updating of the models.

### CLINICIAN–AI COLLABORATION

The successful use of AI in radiology depends on effective clinician–AI collaboration. In theory, the use of AI algorithms to assist radiologists allows for a human–AI collaboration workflow, with humans and AI leveraging complementary strengths.[52] Studies have shown that AI assistance in interpretation of medical images is more useful to some clinicians than to others and generally provides more benefit to less experienced clinicians.[53,54]

Despite some evidence that clinicians receiving AI assistance can achieve better performance than unassisted clinicians,[53,55,56] the body of research on human–AI collaboration for image interpretation offers mixed evidence regarding the value of such a collaboration. Results vary according to particular metrics, tasks, and the study cohorts in question, with studies showing that although AI can improve the performance of radiologists, sometimes AI alone performs better than a radiologist using AI.[57,58]

Many AI methods are "black boxes," meaning that their decision-making processes are not easily interpretable by humans; this can pose challenges for clinicians trying to understand and trust the recommendations of AI.[59] Studies of the potential for explainable AI methods to build trust in clinicians have shown mixed results.[59,60] Therefore, there is a need to move from evaluations centered on the stand-alone performance of models to evaluations centered on the outcomes when these algorithms are used as assistive tools in real-world clinical workflows. This approach will enable us to better understand the effectiveness and limitations of AI in clinical practice and establish safeguards for effective clinician–AI collaboration.

### TRANSPARENCY

Transparency is a major challenge in evaluating the generalization behavior of AI algorithms in medical imaging. Scientific, peer-reviewed evidence of efficacy is lacking for most commercially available AI products.[38] Many published reports on FDA-cleared devices omit information on sample size, demographic characteristics of patients, and specifications of the equipment used to acquire the images to be interpreted. In addition, only a fraction of device studies offer data on the specific demographic subgroups used during algorithm training, as well as the diagnostic performance of these algorithms when applied to patients from underrepresented demographic subgroups. This lack of information makes it difficult to determine the generalizability of AI and machine-learning algorithms across different patient populations.

The limited independent validation of these models has generated a call for greater transparency and rigor with the use of checklists to verify the proper implementation of AI models in medical imaging and to ensure adequate reproducibility and clinical effectiveness.[61-63] One solution for transparency is the curation and public release of medical-imaging data sets to serve as a common benchmark and show algorithm performance.[64-67] The availability of publicly released chest radiograph data sets has already provided support for marked advances in improving AI validation.[68,69] However, there are challenges in curating public medical-imaging data sets, including privacy concerns about sharing data,[70] costs of data infrastructure,[71] and overrepresentation of data from academic medical centers with substantial resources.[72] Federated learning, another approach to data sharing, involves training an AI model on decentralized data sources without transferring the data to a central repository.[73,74] Streamlined processes for curating and sharing diverse medical data sets are necessary for transparency in establishing clinical usefulness.

### POSTDEPLOYMENT MONITORING

Even after a model is deployed, its performance in the real world may degrade over time. In interpretation of medical images, these shifts can occur as a result of various factors such as changes in disease prevalence, advances in medical technology, and alterations in clinical practices.[38,75-77] Failure to update the model to reflect these changes can lead to poor model performance and misuse. However, regulatory requirements may restrict updating of models after they have been approved.

Continuous monitoring of model performance and proactive measures to address data set shifts over time can improve the accuracy and reliability of AI models in medical-imaging interpretation. Regular updates of the training data and retraining of the model on new data through continual learning can help maintain model performance over time.[78] In addition, incorporating feedback from clinicians can help improve the performance of the model by providing real-world insights and identifying areas for improvement. Ultimately, postdeployment monitoring is essential to ensure that AI models remain effective and reliable in clinical settings.[79,80]

## GENERALIST MEDICAL AI MODELS FOR RADIOLOGY

The current generation of AI models in radiology can handle only a limited set of interpretation tasks, and they rely heavily on curated data that have been specifically labeled and categorized.[81] Although focusing on the image as an isolated model input has some value, it does not reflect the true cognitive work of radiology, which involves interpreting medical-imaging examinations comprehensively, comparing current and previous examinations,[82] and synthesizing this information with clinical contextual data to make diagnostic and management recommendations.[83,84] The narrow focus of existing AI solutions on interpretation of individual images in isolation has contributed to the limited penetration of radiologic AI applications in practice.

However, there is a trend toward a more comprehensive approach to the development of radiologic AI, with the aim of providing more value than simply automating individual interpretation tasks. Recently developed models can identify dozens or even hundreds of findings on chest radiographs and brain CT scans obtained without contrast material,[85] and they can provide radiologists with specific details about each finding. More and more companies are offering AI solutions that address the entire diagnostic and clinical workflow for conditions such as stroke and cancer, from screening to direct clinical referrals and follow-up. Although these comprehensive AI solutions may make it easier for medical professionals to implement and use the technology, the issues of validation and transparency remain a concern.

A new generation of generalist medical AI models with the potential to tackle the entire task of radiologic image interpretation and more is on the horizon.[86] These models will be capable of accurately generating the full radiologic report by interpreting a wide range of findings with degrees of uncertainty and specificity based on the image, by fusing the clinical context with the imaging data, and by leveraging previous imaging in the decision of the model.[84,87-91] This comprehensive approach is more closely aligned with the overall cognitive work in radiology. Early studies of such models have shown that they can detect several diseases on images at an

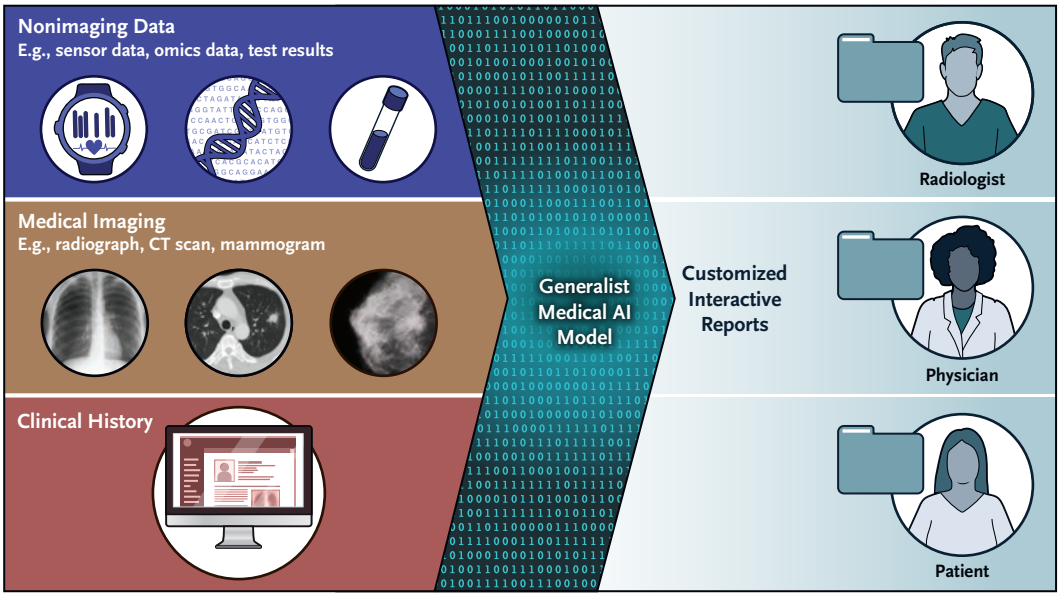


**Figure 3. Potential of Generalist Medical AI Models in Radiology.**

Generalist medical AI models have great potential for transforming radiology. These models can produce complete radiologic reports containing interpretive and descriptive findings derived from various sources such as imaging, clinical context, and previous imaging. They could also potentially link specific image regions with language descriptions, adjust to end users, and generate findings within an interpretive context. In addition, these models may possess the ability to adapt to new environments and technological advancements. "Omics" data are derived from studies such as genomics, epigenomics, proteomics, and metabolomics.

expert level without requiring further annotation, a capability known as zero-shot learning.[92]

Rapid developments in AI models, including self-supervised models,[92,93] multimodal models,[82] foundation models, and particularly large language models for text data and for combined image and text data,[94,95] have the potential to accelerate progress in this area. Large language models are AI models consisting of a neural network with billions of weights or more, trained on large amounts of unlabeled data. Early studies of large language models for text-based tasks in medicine have included chatbots such as GPT-4 (Generative Pre-trained Transformer 4) and have shown that these models are capable of clinical expert–level medical note-taking, question answering, and consultation.[96,97] We anticipate that future AI models will be able to process imaging data, speech, and medical text and generate outputs such as free-text explanations, spoken recommendations, and image annotations that reflect advanced medical reasoning. These models will be able to generate tailored text outputs based on medical-image inputs, catering to the specific needs of various end users, and will enable personalized recommendations and natural-language interactions on the imaging study. For instance, given a medical image and relevant clinical information, the model will produce a complete radiologic report for the radiologist,[98] a patient-friendly report with easy-to-understand descriptions in the preferred language for the patient, recommendations regarding a surgical approach that are based on best practices for the surgeon, and evidence-based follow-up suggestions and tests for the primary care provider — all derived from the imaging and clinical data by a single generalist model (Fig. 3). In addition, these models may be able to generalize easily to new geographic locations, patient populations, disease distributions, and changes in imaging technology without requiring substantial engineering effort or more than a handful of new data.[99]

Given the capabilities of large language models, training new multimodal large language models with large quantities of real-world medical imaging and clinical text data, although challenging, holds promise in ushering in trans-

formative capabilities of radiologic AI. However, the extent to which such models can exacerbate the extant problems with widespread validation remains unknown and is an important area for study and concern. Overall, the potential for generalist medical AI models to provide comprehensive solutions to the task of interpretation of radiologic images and beyond is likely to transform not only the field of radiology but also health care more broadly.

## CONCLUSIONS

AI is a prime instance of a technological breakthrough that has widespread current and future possibilities in the field of medical imaging. Radiology has witnessed the adoption of these tools in everyday clinical practice, albeit with a modest impact thus far. The discrepancy between the anticipated and actual impact can be attributed to various factors, such as the absence of data from prospective real-world studies, limited generalizability, and the scarcity of comprehensive AI solutions for image interpretation. As health care professionals increasingly use radiologic AI and as large language models continue to evolve, the future of AI in medical imaging appears bright. However, it remains uncertain whether the traditional practice of radiology, in its current form, will share this promising outlook.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

## REFERENCES


1. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28:31-8.
2. Jones OT, Matin RN, van der Schaar M, et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review. Lancet Digit Health 2022;4(6):e466-e476.
3. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021;18:465-78.
4. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature 2021;594:106-10.
5. Abràmoff MD, Cunningham B, Patel B, et al. Foundational considerations for artificial intelligence using ophthalmic images. Ophthalmology 2022;129(2):e14-e32.
6. Nam JG, Hwang EJ, Kim J, et al. AI improves nodule detection on chest radiographs in a health screening population: a randomized controlled trial. Radiology 2023;307(2):e221894.
7. Eng D, Chute C, Khandwala N, et al. Automated coronary calcium scoring using deep learning with multicenter external validation. NPJ Digit Med 2021;4:88.
8. Astuto B, Flament I, K Namiri N, et al. Automatic deep learning-assisted detection and grading of abnormalities in knee MRI studies. Radiol Artif Intell 2021;3(3):e200165.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
10. Brady AP. The vanishing radiologist — an unseen danger, and a danger of being unseen. Eur Radiol 2021;31:5998-6000.
11. Dreyer KJ, Geis JR. When machines think: radiology's next frontier. Radiology 2017;285:713-8.
12. European Society of Radiology (ESR). Current practical experience with artificial intelligence in clinical radiology: a survey of the European Society of Radiology. Insights Imaging 2022;13:107.
13. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR Data Science Institute artificial intelligence survey. J Am Coll Radiol 2021;18:1153-9.
14. Yuba M, Iwasaki K. Systematic analysis of the test design and performance of AI/ML-based medical devices approved for triage/detection/diagnosis in the USA and Japan. Sci Rep 2022;12:16874.
15. Tariq A, Purkayastha S, Padmanaban GP, et al. Current clinical applications of artificial intelligence in radiology and their best supporting evidence. J Am Coll Radiol 2020;17:1371-81.
16. Richardson ML, Garwood ER, Lee Y, et al. Noninterpretive uses of artificial intelligence in radiology. Acad Radiol 2021;28:1225-35.
17. Wang G, Ye JC, De Man B. Deep learning for tomographic image reconstruction. Nat Mach Intell 2020;2:737-48.
18. Tao Q, Yan W, Wang Y, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. Radiology 2019;290:81-8.
19. Elijovich L, Dornbos Iii D, Nickele C, et al. Automated emergent large vessel occlusion detection by artificial intelligence improves stroke workflow in a hub and spoke stroke system of care. J Neurointerv Surg 2022;14:704-8.
20. Hillis JM, Bizzo BC, Mercaldo S, et al. Evaluation of an artificial intelligence model for detection of pneumothorax and tension pneumothorax in chest radiographs. JAMA Netw Open 2022;5(12):e2247172.
21. Pease M, Arefan D, Barber J, et al. Outcome prediction in patients with severe traumatic brain injury using deep learning from head CT scans. Radiology 2022;304:385-94.
22. Jiang Y, Zhang Z, Yuan Q, et al. Predicting peritoneal recurrence and disease-free survival from CT images in gastric cancer with multitask deep learning: a retrospective study. Lancet Digit Health 2022;4(5):e340-e350.
23. Lee MH, Zea R, Garrett JW, Graffy PM, Summers RM, Pickhardt PJ. Abdominal CT body composition thresholds using automated AI tools for predicting 10-year adverse outcomes. Radiology 2023;306(2):e220574.
24. Pickhardt PJ, Graffy PM, Perez AA, Lubner MG, Elton DC, Summers RM. Opportunistic screening at abdominal CT: use of automated body composition biomarkers for added cardiometabolic value. Radiographics 2021;41:524-42.
25. Yuan N, Kwan AC, Duffy G, et al. Prediction of coronary artery calcium using deep learning of echocardiograms. J Am Soc Echocardiogr 2022 December 23 (Epub ahead of print).
26. Huisman M, Ranschaert E, Parker W, et al. An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education. Eur Radiol 2021;31:8797-806.
27. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol 2020;30:5525-32.
28. Yang D, Fineberg HV, Cosby K. Diagnostic excellence. JAMA 2021;326:1905-6.
29. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020;395:1579-86.

**30.** Rudolph J, Huemmer C, Ghesu F-C, et al. Artificial intelligence in chest radiography reporting accuracy: added clinical value in the emergency unit setting without 24/7 radiology coverage. Invest Radiol 2022;57:90-8.
**31.** Karamchandani RR, Helms AM, Satyanarayana S, et al. Automated detection of intracranial large vessel occlusions using Viz.ai software: experience in a large, integrated stroke network. Brain Behav 2023; 13(1):e2808.
**32.** Mazurek MH, Cahn BA, Yuen MM, et al. Portable, bedside, low-field magnetic resonance imaging for evaluation of intracerebral hemorrhage. Nat Commun 2021;12:5119.
**33.** Brink JA, Hricak H. Radiology 2040. Radiology 2023;306:69-72.
**34.** Lee HW, Jin KN, Oh S, et al. Artificial intelligence solution for chest radiographs in respiratory outpatient clinics: multicenter prospective randomized study. Ann Am Thorac Soc 2022 December 12 (Epub ahead of print).
**35.** Narang A, Bae R, Hong H, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. JAMA Cardiol 2021;6:624-32.
**36.** Pokaprakarn T, Prieto JC, Price JT, et al. AI estimation of gestational age from blind ultrasound sweeps in low-resource settings. NEJM Evid 2022;1(5). DOI: 10.1056/EVIDoa2100058.
**37.** Baribeau Y, Sharkey A, Chaudhary O, et al. Handheld point-of-care ultrasound probes: the new generation of POCUS. J Cardiothorac Vasc Anesth 2020;34:3139-45.
**38.** Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021;27:582-4.
**39.** Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. JAMA Netw Open 2022;5(9):e2233946.
**40.** Pooch EHP, Ballester PL, Barros RC. Can we trust deep learning models diagnosis? The impact of domain shift in chest radiograph classification. In: Proceedings and abstracts of the Second International Workshop on Thoracic Image Analysis, October 8, 2020. Lima, Peru: Medical Image Computing and Computer Assisted Intervention Society, 2020.
**41.** Cohen JP, Hashir M, Brooks R, Bertrand H. On the limits of cross-domain generalization in automated X-ray prediction. In: Proceedings and abstracts of the Third Conference on Medical Imaging with Deep Learning, July 6–8, 2020. Montreal: Medical Imaging with Deep Learning Foundation, 2020.
**42.** Glocker B, Robinson R, Castro DC, Dou Q, Konukoglu E. Machine learning with multi-site imaging data: an empirical study on the impact of scanner effects. In: Proceedings and abstracts of the Medical Imaging Meets NeurIPS Workshop, December 14, 2019. Vancouver: Neural Information Processing Systems, 2019.
**43.** Koh PW, Sagawa S, Marklund H, et al. WILDS: a benchmark of in-the-wild distribution shifts. In: Proceedings of the 38th International Conference on Machine Learning, July 18–24, 2021. Virtual: International Conference on Machine Learning, 2021.
**44.** Hsu W, Hippe DS, Nakhaei N, et al. External validation of an ensemble model for automated mammography interpretation by artificial intelligence. JAMA Netw Open 2022;5(11):e2242343.
**45.** Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369: m1328.
**46.** AlBadawy EA, Saha A, Mazurowski MA. Deep learning for segmentation of brain tumors: impact of cross-institutional training and testing. Med Phys 2018;45:1150-8.
**47.** Rajpurkar P, Joshi A, Pareek A, Ng AY, Lungren MP. CheXternal: generalization of deep learning models for chest X-ray interpretation to photos of chest X-rays and external clinical settings. In: Proceedings of the Conference on Health, Inference, and Learning, April 8–10, 2021. New York: Association for Computing Machinery, 2021.
**48.** Voter AF, Larson ME, Garrett JW, Yu JJ. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of cervical spine fractures. AJNR Am J Neuroradiol 2021;42:1550-6.
**49.** Guan H, Liu M. Domain adaptation for medical image analysis: a survey. IEEE Trans Biomed Eng 2022;69:1173-85.
**50.** Castro DC, Walker I, Glocker B. Causality matters in medical imaging. Nat Commun 2020;11:3673.
**51.** Feng Y, Xu X, Wang Y, et al. Deep supervised domain adaptation for pneumonia diagnosis from chest X-ray images. IEEE J Biomed Health Inform 2022;26: 1080-90.
**52.** Langlotz CP. Will artificial intelligence replace radiologists? Radiol Artif Intell 2019;1(3):e190058.
**53.** Park A, Chute C, Rajpurkar P, et al. Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet model. JAMA Netw Open 2019;2(6):e195600.
**54.** Tschandl P, Rinner C, Apalla Z, et al. Human-computer collaboration for skin cancer recognition. Nat Med 2020;26: 1229-34.
**55.** Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. PLoS Med 2018;15(11):e1002699.
**56.** Ahn JS, Ebrahimian S, McDermott S, et al. Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. JAMA Netw Open 2022;5(8):e2229289.
**57.** Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health 2021;3(8):e496-e506.
**58.** Rajpurkar P, O'Connell C, Schechter A, et al. CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV. NPJ Digit Med 2020;3:115.
**59.** Saporta A, Gui X, Agrawal A, et al. Benchmarking saliency methods for chest X-ray interpretation. Nat Mach Intell 2022; 4:867-78.
**60.** Gaube S, Suresh H, Raue M, et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. Sci Rep 2023;13:1383.
**61.** Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020;2(2): e200029.
**62.** Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020;370: m3164.
**63.** Cruz Rivera S, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 2020;26:1351-63.
**64.** Irvin J, Rajpurkar P, Ko M, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings and abstracts of the 33rd AAAI Conference on Artificial Intelligence, January 27–February 1, 2019. Honolulu: Association for the Advancement of Artificial Intelligence, 2019.
**65.** Johnson AEW, Pollard TJ, Berkowitz SJ, et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Sci Data 2019;6:317.
**66.** Bustos A, Pertusa A, Salinas J-M, de la Iglesia-Vayá M. PadChest: a large chest x-ray image dataset with multi-label annotated reports. Med Image Anal 2020; 66:101797.
**67.** Nguyen HQ, Lam K, Le LT, et al. VinDr-CXR: an open dataset of chest X-rays with radiologist's annotations. Sci Data 2022; 9:429.
**68.** Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci U S A 2020;117: 12592-4.
**69.** Seyyed-Kalantari L, Zhang H, McDer-

mott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med 2021;27:2176-82.
**70.** Seastedt KP, Schwab P, O'Brien Z, et al. Global healthcare fairness: we should be sharing more, not less, data. PLOS Digit Health 2022;1(10):e0000102.
**71.** Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77.
**72.** Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. JAMA 2020;324:1212-3.
**73.** Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep 2020;10:12598.
**74.** Pati S, Baid U, Edwards B, et al. Federated learning enables big data for rare cancer boundary detection. Nat Commun 2022;13:7346.
**75.** Harvey HB, Gowda V. How the FDA regulates AI. Acad Radiol 2020;27:58-61.
**76.** Lacson R, Eskian M, Licaros A, Kapoor N, Khorasani R. Machine learning model drift: predicting diagnostic imaging follow-up as a case example. J Am Coll Radiol 2022;19:1162-9.
**77.** Feng J, Phillips RV, Malenica I, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. NPJ Digit Med 2022;5:66.
**78.** Yao H, Choi C, Cao B, Lee Y, Koh PW, Finn C. Wild-time: a benchmark of in-the-wild distribution shift over time. In: Proceedings and abstracts of the ICML 2022 Shift Happens Workshop, July 22, 2022. Baltimore: International Conference on Machine Learning, 2022.
**79.** Soin A, Merkow J, Long J, et al. CheXstray: real-time multi-modal data concordance for drift detection in medical imaging AI. March 17, 2022 (http://arxiv.org/abs/2202.02833). preprint.
**80.** Pianykh OS, Langs G, Dewey M, et al. Continuous learning AI in radiology: implementation principles and early applications. Radiology 2020;297:6-14.
**81.** Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. Radiology 2020;295:4-15.
**82.** Acosta JN, Falcone GJ, Rajpurkar P. The need for medical artificial intelligence that incorporates prior images. Radiology 2022;304:283-8.
**83.** Larson DB, Froehle CM, Johnson ND, Towbin AJ. Communication in diagnostic radiology: meeting the challenges of complexity. AJR Am J Roentgenol 2014;203:957-64.
**84.** Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med 2020;3:136.
**85.** Guo Y, He Y, Lyu J, et al. Deep learning with weak annotation from diagnosis reports for detection of multiple head disorders: a prospective, multicentre study. Lancet Digit Health 2022;4(8):e584-e593.
**86.** Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature 2023;616:259-65.
**87.** Ramesh V, Chi NA, Rajpurkar P. Improving radiology report generation systems by removing hallucinated references to non-existent priors. October 13, 2022 (http://arxiv.org/abs/2210.06340). preprint.
**88.** Endo M, Krishnan R, Krishna V, Ng AY, Rajpurkar P. Retrieval-based chest x-ray report generation using a pre-trained contrastive language-image model. PMLR 2021;158:209-19 (https://proceedings.mlr.press/v158/endo21a.html).
**89.** Bannur S, Hyland S, Liu Q, et al. Learning to exploit temporal structure for biomedical vision-language processing. March 16, 2023 (http://arxiv.org/abs/2301.04558). preprint.
**90.** Kather JN, Ghaffari Laleh N, Foersch S, Truhn D. Medical domain knowledge in domain-agnostic generative AI. NPJ Digit Med 2022;5:90.
**91.** Huang S-C, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. Sci Rep 2020;10:22147.
**92.** Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. Nat Biomed Eng 2022;6:1399-406.
**93.** Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. Nat Biomed Eng 2022;6:1346-52.
**94.** Fei N, Lu Z, Gao Y, et al. Towards artificial general intelligence via a multimodal foundation model. Nat Commun 2022;13:3094.
**95.** Zhang S, Xu Y, Usuyama N, et al. Large-scale domain-specific pretraining for biomedical vision-language processing. March 2, 2023 (http://arxiv.org/abs/2303.00915). preprint.
**96.** Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. December 26, 2022 (http://arxiv.org/abs/2212.13138). preprint.
**97.** Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388:1233-9.
**98.** Jeong J, Tian K, Li A, et al. Multimodal image-text matching improves retrieval-based chest X-ray report generation. March 29, 2023 (http://arxiv.org/abs/2303.17579). preprint.
**99.** Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. In: Proceedings and abstracts of the 2020 Conference on Neural Information Processing Systems, December 6–12, 2020. Virtual: Neural Information Processing Systems Foundation, 2020.