Radiology

REVIEWS AND COMMENTARY • REVIEW

The Financial, Operational, and Clinical Advantages of Generalist Radiology AI

Siddhant Dogra, MD[1] • Xiaoman Zhang, PhD[2] • Ezequiel Silva III, MD[3,4] • Pranav Rajpurkar, PhD[2]

Author affiliations, funding, and conflicts of interest are listed at the end of this article.

Radiology 2025; 316(3):e242362 • https://doi.org/10.1148/radiol.242362 • Content codes: AI IN

Despite the rapid growth of Food and Drug Administration–cleared artificial intelligence (AI)– and machine learning–enabled medical devices for use in radiology, current tools remain limited in scope, often focusing on narrow tasks and lacking the ability to comprehensively assist radiologists. These narrow AI solutions face limitations in financial sustainability, operational efficiency, and clinical utility, hindering widespread adoption and constraining their long-term value in radiology practice. Recent advances in generative and multimodal AI have expanded the scope of image interpretation, prompting discussions on the development of generalist medical AI. In this context, this review proposes the concept of generalist radiology AI (GRAI) and introduces key features for its implementation. GRAI aims to *(a)* create reports based on positive diagnoses, *(b)* tailor reports to indications for normal studies, *(c)* compare findings with prior imaging, *(d)* incorporate patient characteristics, and *(e)* provide uncertainty-informed, interactive recommendations. By consolidating image interpretation and expanding the incorporation of patient context, GRAI has the potential to overcome the limitations of narrow AI solutions, improve financial sustainability, streamline operational efficiency, and enhance clinical utility. Appropriate development of GRAI, building on these proposed features, is crucial for realizing the full potential of AI in radiology and enhancing diagnostic performance while reducing the clinical burden on radiologists.

The number of artificial intelligence (AI)– and machine learning–enabled medical devices cleared by the U.S. Food and Drug Administration (FDA) for use in radiology has rapidly grown into the high hundreds (1). However, even the most advanced tools intended for computer-aided detection and diagnosis remain task-specific, such as identifying intracranial hemorrhage (2). Many of these tools already show benefits in clinical practice and will continue to improve and incorporate new features. However, these task-specific approaches, referred to herein as "narrow AI," face inherent limitations that constrain long-term value (3). Financially, the need for multiple solutions leads to unsustainable costs. Operationally, repeatedly integrating various tools results in inefficiencies. Clinically, narrow AI lacks the ability to provide comprehensive and flexible interpretation and decision support.

There is growing interest in developing generalist approaches to bridge gaps in image interpretation, including but not limited to lesion detection, quantification, and differential diagnosis (4). These generalist radiology AI (GRAI) models are designed to assist radiologists across a wide range of imaging tasks rather than focusing on a single disease or modality. GRAI should support radiologists throughout the entire imaging workflow, including detection, differential diagnosis, measurements, and comparisons, while integrating clinical data, answering questions, and providing recommendations. Moreover, GRAI could perform superhuman tasks such as opportunistic screening and predictive analytics, identifying patterns imperceptible to human readers (5). While GRAI can enhance efficiency and provide valuable insights, expert human oversight will remain essential for clinical decision-making and managing complex or ambiguous cases. Like generalist medical AI, GRAI will have three distinguishing capabilities: dynamic task specification, where it adapts to new tasks described in plain language without retraining; the ability to accept inputs and produce outputs using multimodality data (eg, any or all of images, laboratory values, and operative notes); and the capacity to reason through unfamiliar tasks and logically explain outputs (3,6). Consequently, GRAI has the potential to serve as a true augmentative intelligence, supporting radiologists across a wider array of clinical tasks while enhancing efficiency and diagnostic precision.

Recent advancements such as foundation models have paved the way for GRAI (7,8). Foundation models, trained on vast diverse unlabeled datasets, learn general patterns without requiring extensive expert annotation and can then be adapted to a wide range of downstream tasks with minimal additional training, making them highly versatile across different applications (9). In research settings, general models have met or exceeded narrow model performance in several tasks (10).

We propose that GRAI can potentially address fundamental limitations of narrow AI solutions by offering a more holistic approach to radiology AI. Table 1 shows key differences in capabilities of narrow AI solutions and GRAI. By more comprehensively detecting abnormalities and providing targeted reporting, GRAI moves beyond narrow solutions that must be trained for individual abnormalities. It incorporates context by considering available clinical information to provide more personalized insights. GRAI offers comprehensive decision support by delivering tailored recommendations to various stakeholders, such as radiologists, referring physicians, and patients. This approach enables GRAI to improve financial sustainability by reducing the need for multiple point solutions—narrowly scoped AI tools each designed to tackle a single imaging task or pathologic abnormality. GRAI also streamlines operational efficiency by better integrating into the radiology workflow and enhances clinical utility by providing actionable insights. This review article describes the financial, operational, and clinical limitations of the current approach for radiology AI and outlines a vision for GRAI. Our framework aims to guide GRAI development to realize its full potential.

Abbreviations

AI = artificial intelligence, FDA = Food and Drug Administration, GRAI = generalist radiology AI

Summary

Generalist radiology artificial intelligence (GRAI), incorporating both fundamental attributes of generalist medical artificial intelligence (AI) and radiology-specific features, can help address financial, operational, and clinical limitations of narrow radiology AI solutions.

Essentials

- Narrow radiology artificial intelligence (AI) solutions suffer from inherent financial (unsustainable price scaling, market fragmentation), operational (repetitive due diligence and integration needs), and clinical (lack of clinical context and flexibility) limitations.
- Generalist radiology AI (GRAI) can help address these limitations by consolidating image interpretation assistance into one package while incorporating context and providing tailored recommendations.
- By producing more clinically useful and comprehensive reports, GRAI can better help reduce radiologist effort and unlock new value propositions by truly helping address workload and cognitive burden concerns.
- GRAI systems should incorporate the following five key features to move beyond narrow task tools: *(a)* reporting with multifinding detection and characterization, *(b)* indication-tailored reporting for normal studies, *(c)* longitudinal image comparison, *(d)* integration of patient characteristics and clinical context, and *(e)* uncertainty-aware, interactive recommendations.

Limitations and Barriers to Adoption of Current Narrow Radiology AI

Short-term Limitations

Many FDA-cleared solutions triage certain emergent findings. These tools have merits, particularly in emergency departments or intensive care units where reduced turnaround time can have a major impact. However, evidence on whether these tools actually meaningfully shorten turnaround time remains inconsistent; institutions may not be willing to pay and make operational changes to achieve relatively small benefits from the handful of existing point solutions (11–14).

As detection labels for different pathologic abnormalities have become more comprehensive, multipathology narrow AI tools could conceivably be used to screen out normal studies; however, abnormal studies will remain challenging to comprehensively evaluate with existing narrow tools. For example, while many solutions for chest radiographs are specific for thoracic abnormalities, such as pneumothorax, they are less specific for others; a generic humeral bone lesion detection solution would need to provide additional characterization to be clinically useful (15). Even as vendors improve narrow AI solutions, these tools have inherent financial, operational, and clinical limitations (Fig 1).

Financial

AI and machine learning devices are often offered on a subscription basis, such as per site, per workstation, per radiologist, per study, or sometimes even per narrow AI solution (16). As vendors incorporate an increasing number of these solutions into their offerings, costs will also scale up. Although pricing structures vary, some vendors who use a per-solution structure charge up to $100 000 per solution. Moreover, the number of narrow solutions required for comprehensive coverage equivalent to a radiologist would be much higher for cross-sectional imaging studies. At some point, health systems could be priced out of these tools and may find it cheaper to hire an additional radiologist (17).

There are additional costs beyond initial licensure. Infrastructure and hardware may need to be adjusted or upgraded as tools evolve. As increasing numbers of solutions from different vendors are used, chances of malfunctions are higher. Upgrades, repairs, and other overhead costs will also increase alongside point solution availability (18). Point solutions complicate budget planning, as purchasers cannot predict new offerings (19).

Many imaging studies have anatomic coverage overlap; for example, chest CT scans capture the upper abdomen, while abdominal CT scans show the lung bases. Trying to achieve comprehensive coverage using narrow tools trained for particular types of studies could result in redundant coverage across different tools. For example, a comprehensive chest radiograph tool must identify abnormalities in the upper abdomen, but a dedicated abdominal radiograph tool captures the same abnormalities.

Table 1: Key Differences in Capabilities of Narrow and Generalist Radiology AI

Area	Narrow Radiology AI	GRAI
Adaptability	Will need to be retrained with large amounts of new examples to adapt to data distribution shifts, such as from new scanners or changing patient populations	Uses in-context and few-shot learning to adapt to shifting data distributions without needing a large amount of new data
Flexibility	Typically needs a rigid set of predefined inputs used in training	Combines data from any combination of available modalities for a given patient
Medical knowledge	Lacks true domain knowledge and relies on learned statistical associations	Has a representation of medical knowledge that can be used to apply concepts to new problems and explain evidence-based recommendations
Interactivity	Offers a predetermined type of output, such as detection or classification of pathologic abnormalities	Tailors output depending on user preference, using a combination of large language and vision language models
Training	Must be trained on many examples of a specific pathologic abnormality or task before it can learn to perform it well	Performs new tasks by being provided a natural language explanation (dynamic task specification) without needing to be retrained

Note.—In-context and few-shot learning refers to the ability of a model to generalize to a task after seeing a handful of examples (3). AI = artificial intelligence, GRAI = generalist radiology AI.

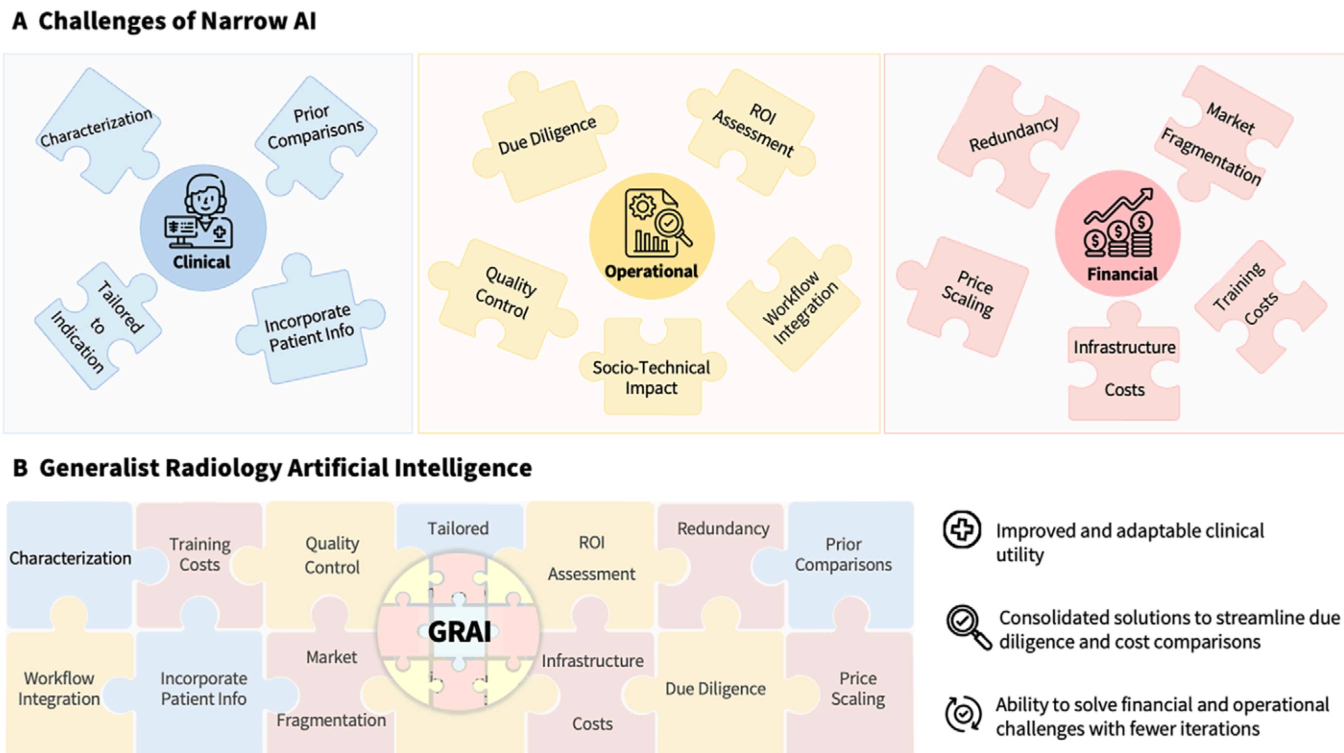

Figure 1: A visual representation comparing **(A)** narrow artificial intelligence (AI) and **(B)** generalist radiology AI (GRAI). Narrow AI algorithms in radiology suffer from many challenges. These challenges include limited scope, lack of contextual understanding, and potential for missed findings, which lead to clinical, operational, and financial challenges (puzzle pieces in **A**). GRAI offers more comprehensive and integrated solutions (the completed puzzle in **B**). GRAI provides improved and adaptable clinical utility, consolidated solutions to streamline due diligence and cost comparisons, and the ability to solve financial and operational challenges with fewer iterations. Sociotechnical impacts include organizational, behavioral, and cultural consequences of new technology interacting with users and workflows. ROI = return on investment.

Market fragmentation resulting from a wide spectrum of vendors and tools also increases the complexity of financial decisions. Vendors have different pricing structures and service agreements, making it difficult to evaluate disparate offers. Fragmentation also reduces the ability of providers to negotiate bulk discounts, as they are more likely to engage with multiple vendors.

Operational

Deploying even a single new tool requires buy-in and monitoring from several stakeholders, including information technology staff, radiologists, clinicians, and administrators. They must agree on the clinical indication and expected return on investment for any device under consideration (20). Carrying out due diligence among narrow tools is especially onerous due to fragmentation, as many vendors differ in the imaging modalities and examinations for which they offer solutions.

Incorporating tools into the workflow is also challenging. Information technology needs to be involved early to ascertain needs in terms of graphics and central processing units and to review network security and performance for cloud-based solutions. If a new tool is integrated into a picture archiving and communication system (PACS), then PACS specialists also need to be involved (19). Tool efficacy in the target population typically needs internal and external testing before deployment due to potential decreases in performance as a result of technical parameters (scanners, algorithms, protocols) and patient characteristics (21).

Workflow optimization is necessary to ensure radiologists are not slowed once these tools are implemented. Sociotechnical impacts, namely the organizational, behavioral, and cultural consequences of new technology interacting with users and workflows, have historically been more decisive barriers to adoption than purely technical shortcomings, such as with early sepsis alerts that clashed with work patterns and were ultimately often ignored. These impacts need to be considered from perspectives of both radiologists and other clinicians (22). After a tool has been integrated into the workflow, stakeholders monitor for data drift, where the real-world cases the model sees start to differ from the training data, along with performance and overall return on investment.

Iterating through this process for an increasing number of solutions over time magnifies the operational effort required manyfold, particularly when multiple vendors are involved but even when a single vendor expands its array of solutions. Platforms integrating multiple AI tools will help reduce financial and logistical implementation costs, simplify the user workflow, allow for easier solution additions or swaps, and make it easier to continuously monitor performance metrics. However, certain challenges posed by the diversity of different narrow solutions cannot be fully solved by platforms, such as comparison of price structures and scaling costs. Moreover, as described in the following section, platforms do not address inherent clinical limitations of narrow radiology AI.

Clinical

Clinical limitations of narrow tools will ultimately limit their impact on patient outcomes. The most advanced tools now detect a remarkable number of primary pathologic abnormalities. However, even if an abnormality is identified, fully characterized findings often need to be communicated to guide further management. For example, detection of a humeral bone lesion is useful, but clinical value is lost by only reporting the presence of an abnormal bone lesion. This is because certain lesions can be classified as benign findings that do not need additional work-up while others can immediately be classified as aggressive lesions that need biopsy. Training a narrow AI model to detect and classify every possible abnormality is daunting already for radiographs; scaling such an approach to CT and MRI is much more challenging.

Current narrow tools are limited by their inability to draw on multimodal data, most prominently through comparisons with prior imaging, which is key to ensuring high-quality reporting (23). Even for narrow triage assistance tools, comparisons with recent studies are important to avoid repeatedly flagging studies as positive for acute findings when they are stable follow-up examinations, such as of intracranial hemorrhage, to avoid alarm fatigue (24).

Longitudinal image comparison tools will likely emerge for narrow solutions in the coming years. However, clinical data limitations may be more challenging. Radiologists often rely on electronic health record information such as patient presentation or medical and surgical history. Certain data such as operative reports or laboratory values are not always available. Multimodal tools will need to flexibly use whatever data are present, as opposed to rigid models that always require the same inputs. Models also cannot output reasoned reports and recommendations tailored to different indications, reducing their ability to answer targeted questions posed by different physicians. For example, liver volume should not be reported in all abdominopelvic examinations but is critical to report for transplant planning.

Ultimately, narrow AI models' inability to incorporate medical context and provide utility beyond initial diagnosis caps their value to the end user. The fundamental approach underlying narrow AI development, particularly the need for task-specific training and inflexibility with regard to inputs, makes it unlikely that these features will emerge, particularly for CT and MRI.

Objectives for GRAI

We propose features that diagnostic GRAI should incorporate (Table 2, Fig 2). These features are meant to supplement the three core aspects of a generalist medical AI model previously referenced in the introduction (3). Some of these new features are likely to be introduced to narrow AI tools as well but should be universal features for GRAI.

Table 2: Key Features That Should Be Incorporated into GRAI

Proposed Key GRAI Features
Reporting with multifinding detection and characterization
Reports tailored to indications for normal studies
Longitudinal image comparison
Incorporation of patient characteristics
Uncertainty-informed and interactive recommendations

Note.—GRAI = generalist radiology artificial intelligence.

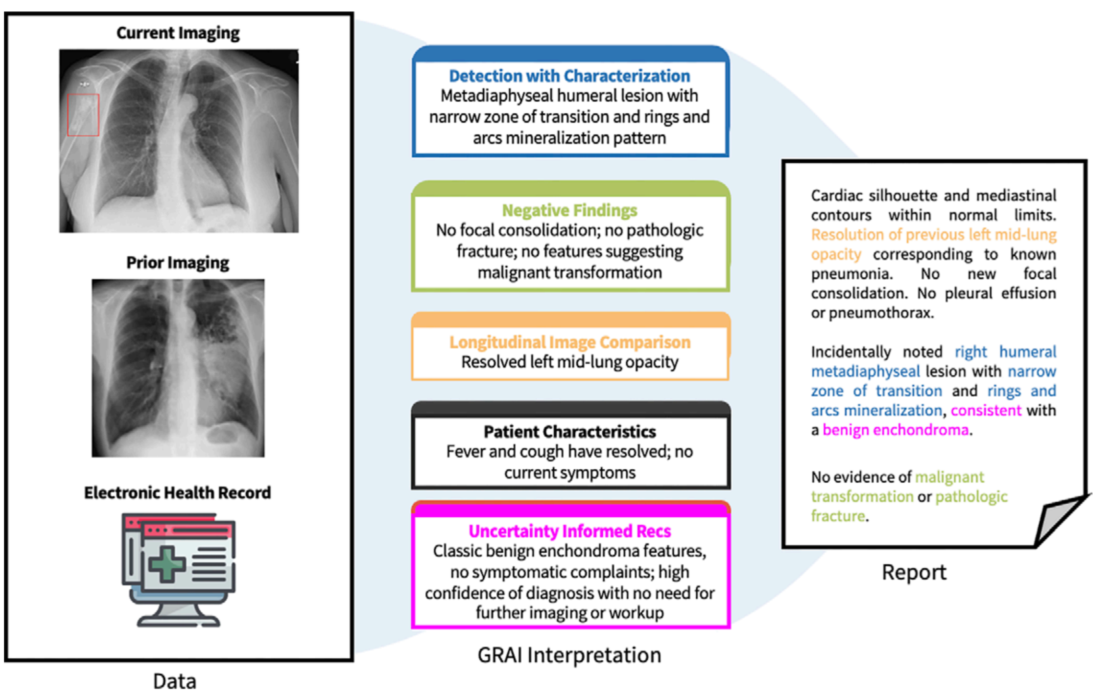

Figure 2: Example of generalist radiology artificial intelligence (GRAI). In this hypothetical example, a patient with recent pneumonia seen on a prior chest radiograph underwent a repeat chest radiography examination for a given indication of "pneumonia follow up." The current radiograph shows no pulmonary findings but an incidental benign enchondroma (red rectangle). GRAI assists the radiologist by consolidating current and prior imaging interpretation while incorporating context from the electronic health record to provide uncertainty-informed and interactive recommendations (recs) in a report. This is an advantage over current narrow artificial intelligence tools, which are limited by their inability to draw on comparisons with prior imaging, fully characterize positive findings and pertinent negative findings, incorporate patient data, and provide recommendations, which are key to ensuring high-quality reporting.

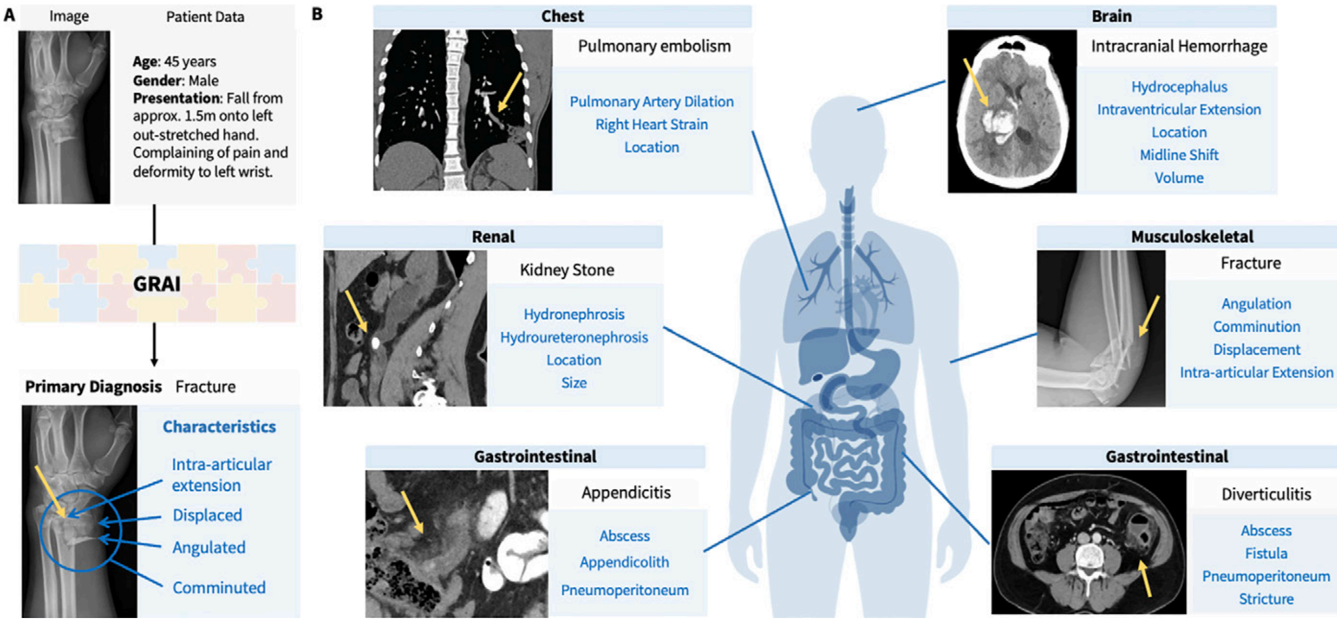

Figure 3: Generalist radiology artificial intelligence (GRAI) should be capable of providing fully characterized diagnoses with the reference standard based on the **(A)** input image and patient clinical history data. Diagram shows examples of six primary imaging diagnoses (indicated by yellow arrows) with pertinent characteristics that GRAI should describe if present in **(B)** the corresponding radiology report. GRAI offers an advantage over current narrow artificial intelligence tools, which may detect a primary diagnosis but not incorporate associated characteristics.

Feature 1: Reporting with Multidisease Detection and Characterization

Even when an initial diagnosis is made, an end user will often need to report associated characteristics; a useful model will need to provide all of this output in the form of a report. These findings may not be inherently pathologic, such as variant abdominal vascular anatomy that may not be noteworthy or mentioned in the impression on most abdominopelvic CT examinations, but are critical to report for patients who require the operating room, such as for transplant surgery or tumor resection. Other characteristics may only be relevant when there is a primary abnormality, such as a fracture for which the presence of intra-articular extension increases the necessity of surgical fixation. Finally, providing complete pathologic abnormality characterization can help instill confidence or provide reasoning behind a particular diagnosis, just as good radiology reports do (25). Figure 3 shows examples of pertinent GRAI characterization for primary diagnoses.

For each diagnosis, GRAI will need to provide imaging information important for directing clinical decision-making. An initiative that can be used to train models for this purpose is the development of common data elements (CDEs) as proposed by the RSNA and the American College of Radiology and seen on RadElement.org (26). The most relevant example is a collection of imaging features that should be reported for a disease entity. For example, elements that currently correspond to "acute appendicitis" are presence, appendiceal diameter, fat stranding, free fluid, obstructing focus and size, lumen contents, abscess, and opacification. While additional elements could be proposed, such as extraluminal air for acute appendicitis, the larger obstacle is that only 51 CDEs are currently available, and not all are for diagnoses.

Feature 2: Outputting Reports Tailored to Requested Indications for Normal Studies

Even for normal studies, GRAI will need to output reports to reflect provided indications. A major purpose of report generation documenting normal findings is to reassure referring physicians and specialists about the AI output. For example, a normal brain MRI report for an indication of seizure may mention the absence of cortical developmental malformations, heterotopic gray matter, or mesial temporal structural abnormalities that would not be mentioned in a report for an indication such as stroke. An epileptologist reading the brain MRI report would likely expect to see these specific normal findings explicitly mentioned; on the other hand, their absence might reduce the confidence in the report and result in a request for a more detailed review by a radiologist.

Feature 3: Longitudinal Image Comparison

Longitudinal image comparison is already needed but generally not available for current narrow tools. Longitudinal image comparison avoids misleading worklist prioritization or false alarms for "acute" stable findings from prior studies. Given that image comparisons are a core feature of imaging guidelines (and considering the historical medicolegal complications from lack of proper comparison), long-term image comparisons will need to be a prerequisite feature of GRAI (27,28).

Value is lost by looking only at static images, particularly for patients who already have a diagnosis. For example, oncologists need comparisons with baseline and prior scans to assess response to treatment. Orthopedic surgeons obtain follow-up radiographs after fractures to assess proper healing in appropriate alignment. Depending on the abnormality, different features will need to be compared across the prior studies, ranging from three-dimensional size (tumors) to angulation and displacement (fractures).

Feature 4: Incorporation of Patient Characteristics

Like radiologists, GRAI will need to use clinical characteristics such as demographics (eg, race, sex, and age), clinical and family history, and symptoms to optimize reports. Multimodal narrow AI could also incorporate some of this information; however, more advanced types of data such as key laboratory markers, pathologic findings, or operative reports are not always available. GRAI, by virtue of its second key capability, can use any combination of available data to modify its output.

Incorporating clinical context can improve performance. For example, imaging findings carry different differential diagnoses depending on age; for instance, white matter hyperintensities on brain MRI scans are often ascribed to chronic microvascular ischemic changes in older adults, while the same appearance might be interpreted as sequela of chronic migraines in young patients. Certain findings may be expected depending on the details in the operative report produced by the surgeon. GRAI must take nonimaging details into account to avoid clinically dubious outputs.

Feature 5: Uncertainty-informed and Interactive Recommendations

While GRAI may be able to improve on human performance, it will not always have 100% confidence in its outputs. Therefore, GRAI will need to convey confidence level while offering corresponding recommendations (25) and be able to tailor recommendations based on the end user. Clinicians from different specialties will often ask a radiologist questions regarding a single individual imaging examination. For example, an oncologist and neurosurgeon will ask different questions for a patient with metastatic prostate cancer with new pathologic spinal fractures at CT. Users should be able to interact with GRAI, potentially by prompting a large language model that collaborates with a vision language model. These multimodal models can interpret both images and text to generate a contextually relevant response.

Value Proposition of GRAI

Through consolidation of image interpretation assistance into one package with an expanded ability to incorporate context and provide recommendations, GRAI provides tremendous financial, operational, and clinical value propositions.

By having the choice to pay for one comprehensive tool rather than debate merits of mix-and-match approaches using multiple point solutions, users could better compare costs of competing investments. GRAI's multitask capabilities will also increase its appeal to diverse stakeholders. These stakeholders may range from private groups seeking efficiency gains to academic centers seeking more value from tools that can identify incidental and opportunistic findings while driving downstream care (29). Adopting GRAI would add the benefits of integrated AI platforms and reduce the effort required for due diligence, infrastructure planning, workflow integration, and sociotechnical adjustments, as these adaptations would happen once rather than repeatedly as new solutions are added. Model procurement and due diligence would also be more straightforward, as disparate vendor offerings are consolidated into GRAI. Performance review and data drift monitoring would still be required, but generalist models could be expected to better adapt to data drift than narrow models given their inherent adaptability using in-context and few-shot learning whereby a model generalizes to a task after being shown a handful of examples (3,30).

GRAI will improve clinical utility and unlock new value propositions. Clinical limitations of narrow tools necessitate that radiologists ultimately review and edit imaging examinations for each patient (31). GRAI will be more likely to produce viable reports, decreasing radiologist work. Consequently, radiologist efficiency would be expected to truly improve with downstream effects of decreased cognitive burden, helping address burnout, worsening imaging backlogs, and delayed reporting (32,33).

The Road Ahead

As GRAI expands to new tasks, we must refine task-specific performance metrics. For example, report generation has moved beyond metrics like BLEU (BiLingual Evaluation Understudy), which only scores surface-level word overlap, to more relevant measures like RadGraph F1 and RadCliQ, which assess whether outputs preserve clinical entities and relations (34). The Conversational Reasoning Assessment Framework for Testing in Medicine, or CRAFT-MD, is used to assess large language model output for patient interactions (35). Additional metrics are needed for tasks such as explaining outputs and making recommendations. Error classification must also evolve alongside technical advancements. Large language models, for instance, struggle with hallucinations, where they produce confident but factually incorrect statements. Mitigation approaches such as retrieval-augmented generation, which inserts verifiable external evidence at inference time, and chain-of-thought prompting, which elicits step-by-step reasoning, reduce but do not eliminate these hallucinations (36). Some hallucinations, such as fabricating prior studies versus nonexistent lesions, pose distinct risks and should be categorized separately to identify model deficiencies.

Data and hardware needs remain crucial. Data sharing is evolving as we see greater participation by technology companies in radiology AI. We expect to see increasing academic-industry relationships wherein data-rich institutions share resources with commercial partners that have greater computational capacity. At the same time, there is an increasing number of large datasets available for foundation model pretraining and tuning (9). Technical approaches will also improve; indeed, improved self-supervised learning algorithms have been shown to better approximate human performance on rare diseases even without human annotation labels (37). Emerging technology such as quantum computing may also ease development of GRAI models by reducing overall computational cost (38).

The GRAI timeline will depend on required features. While GRAI should ultimately be modality-agnostic and available for all modalities, early iterations will likely focus on specific modalities due to variations in data availability, interpretive complexity, and technical challenges unique to each modality. In the next 1–2 years, we expect rapid progress in multidisease computer-aided diagnosis for radiographs and CT, with MRI lagging due to its more complex sequences and findings. Comprehensive detection (bounding boxes, segmentation) will take longer (2–4 years), requiring more human annotation, although automated segmentation tools like TotalSegmentator may assist (39,40). Basic report generation focused on single-study findings without incorporating prior imaging findings or recommendations is already feasible with growing evaluation metrics and will likely be achieved within 2–3 years as multidiagnostic models improve. More advanced capabilities, such as longitudinal image comparison, clinical data integration, and uncertainty estimation, are already research priorities in narrow AI, accelerating their integration into GRAI, although refining these for report generation will likely take 3–5 years. Finally, explainability remains the most challenging feature, as even state-of-the-art large language models still demonstrate faulty reasoning. As other features integrate, explainability will become more complex, likely requiring 5–7 years to develop and externally test for robust, transparent GRAI.

The current regulatory framework for radiology AI is primarily designed for narrow tools and is incompatible with GRAI's flexible and evolving capabilities that can handle multiple distinct tasks. With the increasing focus on foundation models by both researchers and vendors, the FDA will inevitably need to adapt its regulatory pathways to accommodate these broader systems. The FDA has already adapted its approach to evolving AI technologies, as demonstrated by the Predetermined Change Control Plan (PCCP), which allows manufacturers to prespecify intended updates to AI models while maintaining regulatory oversight (41). This concept, initially designed to accommodate continual updating in AI-based Software as a Medical Device, could serve as a foundation for developing a regulatory framework for GRAI that permits controlled adaptation without requiring repeated full-scale reapproval. To ensure safe and effective deployment, regulators will need to expand on PCCP principles by incorporating scenario-based testing, real-world performance monitoring, and risk-based stratification mechanisms that differentiate low-risk decision support functions from high-stakes autonomous analyses. Regulatory approval will likely follow a task-based approach, where individual functionalities such as multipathologic detection or report generation are cleared separately rather than approving an entire GRAI system at once.

Reimbursement pathways also require a major overhaul to support the integration of GRAI into clinical practice (42). Current reimbursement structures are designed for specific diagnostic and procedural tasks, making them ill-suited for a system that augments radiologists across multiple domains of imaging interpretation. To address this, policymakers and payers must develop new reimbursement models that account for the broad, adaptable nature of GRAI, potentially introducing performance-based payment structures that reward AI's contribution to efficiency, diagnostic accuracy, and patient outcomes.

Conclusion

The transition from narrow radiology artificial intelligence (AI) to generalist radiology AI (GRAI) frameworks will be a pivotal evolution in medical imaging. As ongoing technical advancements are integrated into developing GRAI, we anticipate a shift toward versatile image interpretation that more closely aligns with the multifaceted nature of diagnostic imaging. Proper development of GRAI, building on the features we have proposed, is key to realizing the full potential of AI in radiology.

Deputy Editor: Michael Lev
Scientific Editor: Sarah Atzen

Author affiliations:

[1] Department of Radiology, New York University Langone Health, New York, NY

[2] Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, MA 02115

[3] South Texas Radiology, San Antonio, Tex

[4] University of Texas Health, Long School of Medicine, San Antonio, Tex

Received August 8, 2024; revision requested September 19; final revision received June 3, 2025; accepted June 18.

Address correspondence to: P.R. (email: Pranav_Rajpurkar@hms.harvard.edu).

Funding: Authors declared no funding for this work.

Author contributions: Guarantors of integrity of entire study, **S.D., P.R.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **S.D., E.S.**; statistical analysis, **E.S.**; and manuscript editing, all authors

Disclosures of conflicts of interest: **S.D.** Consulting fees from a2z Radiology AI; deputy editor for *Radiology* In Training. **X.Z.** No relevant relationships. **E.S.** No relevant relationships. **P.R.** Co-founder of a2z Radiology AI.

References

1. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. US Food and Drug Administration. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Updated March 25, 2025. Accessed March 2025.
2. McLouth J, Elstrott S, Chaibi Y, et al. Validation of a deep learning tool in the detection of intracranial hemorrhage and large vessel occlusion. Front Neurol 2021;12:656112.
3. Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature 2023;616(7956):259–265.
4. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. Nat Med 2022;28(1):31–38.
5. Gauriau R, Bizzo BC, Comeau DS, et al. Head CT deep learning model is highly accurate for early infarct estimation. Sci Rep 2023;13(1):189.
6. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. N Engl J Med 2023;388(21):1981–1990.
7. Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications. Radiology 2024;310(1):e232756.
8. Langlotz CP. The future of AI and informatics in radiology: 10 predictions. Radiology 2023;309(1):e231114.
9. Paschali M, Chen Z, Blankemeier L, et al. Foundation models in radiology: what, how, why, and why not. Radiology 2025;314(2):e240597.
10. Tu T, Azizi S, Driess D, et al. Towards generalist biomedical AI. NEJM AI 2024;1(3).
11. Rothenberg SA, Savage CH, Abou Elkassem A, et al. Prospective evaluation of AI triage of pulmonary emboli on CT pulmonary angiograms. Radiology 2023;309(1):e230702.
12. Topff L, Ranschaert ER, Bartels-Rutten A, et al. Artificial intelligence tool for detection and worklist prioritization reduces time to diagnosis of incidental pulmonary embolism at CT. Radiol Cardiothorac Imaging 2023;5(2):e220163.
13. O'Neill TJ, Xi Y, Stehel E, et al. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. Radiol Artif Intell 2021;3(2):e200024.
14. Lobig F, Subramanian D, Blankenburg M, et al. To pay or not to pay for artificial intelligence applications in radiology. NPJ Digit Med 2023;6(1):117.
15. Jones CM, Danaher L, Milne MR, et al. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. BMJ Open 2021;11(12):e052902.
16. Tadavarthi Y, Vey B, Krupinski E, et al. The state of radiology AI: considerations for purchase decisions and current market offerings. Radiol Artif Intell 2020;2(6):e200004.
17. van Leeuwen KG, de Rooij M, Schalekamp S, et al. Clinical use of artificial intelligence products for radiology in the Netherlands between 2020 and 2022. Eur Radiol 2024;34(1):348–354.
18. Filice RW, Mongan J, Kohli MD. Evaluating artificial intelligence systems to guide purchasing decisions. J Am Coll Radiol 2020;17(11):1405–1409.
19. Omoumi P, Ducarouge A, Tournier A, et al. To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines). Eur Radiol 2021;31(6):3786–3796.
20. Saw SN, Ng KH. Current challenges of implementing artificial intelligence in medical imaging. Phys Med 2022;100:12–17.
21. Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. A multi-society

statement from the ACR, CAR, ESR, RANZCR and RSNA. Radiol Artif Intell 2024;6(1):e230513.
22. Lluch M. Healthcare professionals' organisational barriers to health information technologies—a literature review. Int J Med Inform 2011;80(12):849–862.
23. Acosta JN, Falcone GJ, Rajpurkar P. The need for medical artificial intelligence that incorporates prior images. Radiology 2022;304(2):283–288.
24. Lewandowska K, Weisbrot M, Cieloszyk A, et al. Impact of alarm fatigue on the work of nurses in an intensive care environment—a systematic review. Int J Environ Res Public Health 2020;17(22):8409.
25. Hartung MP, Bickle IC, Gaillard F, et al. How to create a great radiology report. RadioGraphics 2020;40(6):1658–1670.
26. Rubin DL, Kahn CE Jr. Common data elements in radiology. Radiology 2017;283(3):837–844.
27. European Society of Radiology (ESR). Good practice for radiological reporting. Guidelines from the European Society of Radiology (ESR). Insights Imaging 2011;2:93–96.
28. Berlin L. Reporting the "missed" radiologic diagnosis: medicolegal and ethical considerations. Radiology 1994;192(1):183–187.
29. Trivedi H. The business of artificial intelligence in radiology has little to do with radiologists. J Am Coll Radiol 2022;19(4):564–566.
30. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. Adv Neural Inf Process Syst 2020:14165. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.
31. Huisman M, van Ginneken B, Harvey H. The emperor has few clothes: a realistic appraisal of current AI in radiology. Eur Radiol 2024;34(9):5873–5875.
32. Canon CL, Chick JFB, DeQuesada I, et al. Physician burnout in radiology: perspectives from the field. AJR Am J Roentgenol 2022;218(2):370–374.
33. Omofoye TS, Vlahos I, Marom EM, et al. Backlogs in formal interpretation of radiology examinations: a pilot global survey. Clin Imaging 2024;106:110049.
34. Yu F, Endo M, Krishnan R, et al. Evaluating progress in automatic chest x-ray radiology report generation. Patterns (N Y) 2023;4(9):100802.
35. Johri S, Jeong J, Tran BA, et al. An evaluation framework for clinical use of large language models in patient interaction tasks. Nat Med 2025;31(1):77–86.
36. Farquhar S, Kossen J, Kuhn L, et al. Detecting hallucinations in large language models using semantic entropy. Nature 2024;630(8017):625–630.
37. Agarwal N, Huang R, Moehring A, et al. Comparative advantage of humans versus AI in the long tail. AEA Pap Proc 2024;114:618–622.
38. Guenot M. Can quantum computing crack the biggest challenges in health? Nat Med 2025;31(1):4–7.
39. Wasserthal J, Breit HC, Meyer MT, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol Artif Intell 2023;5(5):e230024.
40. Akinci D'Antonoli T, Berger LK, Indrakanti AK, et al. TotalSegmentator MRI: robust sequence-independent segmentation of multiple anatomic structures in MRI. Radiology 2025;314(2):e241613.
41. Zhang K, Khosravi B, Vahdati S, et al. FDA review of radiologic AI algorithms: process and challenges. Radiology 2024;310(1):e230242.
42. Dogra S, Silva EZ, Rajpurkar P. Reimbursement in the age of generalist radiology artificial intelligence. NPJ Digit Med 2024;7(1):350.