Radiology

REVIEWS AND COMMENTARY • EDITORIAL

Human-AI Symbiosis: A Path Forward to Improve Chest Radiography and the Role of Radiologists in Patient Care

Warren B. Gefter, MD • Mathias Prokop, MD • Joon Beom Seo, MD • Suhail Raoof, MD • Curtis P. Langlotz, MD, PhD • Hiroto Hatabu, MD, PhD, FACR

We are at the dawn of a new era of artificial intelligence (AI)–augmented chest radiography. With explosive developments in deep learning, large language models (LLMs), and multimodal vision-language foundation models, AI is poised to transform the field and the role of radiologists in profound ways. Already, AI algorithms can detect a variety of chest radiograph findings with performance comparable to or exceeding that of human radiologists. These results are so impressive that some are questioning if AI will replace humans altogether. The question we should be asking, however, is a different one: How can radiologists and AI work together to create a system greater than the sum of its parts (1,2)?

The Promise of AI

Studies have demonstrated AI's ability to detect diverse chest radiograph findings, including pneumonia, lung nodules, pneumothorax, tuberculosis, and many others (3–5). In a study of lung nodule detection on digital chest radiographs from the National Lung Screening Trial, for example, AI outperformed radiologists, with a sensitivity of 96.0% versus 88.0% and specificity of 93.2% versus 82.8% (6). Progress rapidly accelerated during the COVID-19 pandemic, as AI was shown to detect, quantify, and predict outcomes of COVID-19 pneumonia (7). Moving forward, deep learning algorithms might be used to identify and triage normal chest radiographs when there aren't enough radiologists to keep up with the workload (8,9). AI can assess image quality and improve lesion conspicuity by bone suppression. It can rapidly identify malpositioned tubes, lines, and cardiac devices. It can quantify chest radiograph abnormalities. It can offer prognoses (7,10,11). It can even extract new forms of information from chest radiographs not visible to the human reader (10,12,13).

AI even offers new chest radiograph screening opportunities, including the estimation of lung cancer risk to triage patients for CT lung cancer screening (13). Opportunistic screening with chest radiographs has been used to predict type 2 diabetes (14), osteoporosis, sarcopenia, cardiovascular disease, and chronic obstructive pulmonary disease, as well as a patient's future health care expenses (12), all-cause mortality, and mortality from lung disease (10,11). All of these uses of AI stand to benefit individual and population health.

Meanwhile, there have also been improvements in natural language processing. LLMs have the potential to extract data from electronic health records, review the entire medical literature in an instant, and draft information-rich and individualized reports. Vision transformer–based neural networks have been trained to integrate such nonimaging data with chest radiographs for more accurate diagnoses (15). Together, these can add substantial clinical value to chest radiograph reports, which would be prohibitively time-consuming for radiologists already inundated with data.

Chest radiography is the most frequently performed imaging examination worldwide, yet it is prone to errors primarily from missed findings, such as pulmonary nodules (16). The problem is exacerbated by waning skill in the art of interpreting chest radiographs, as the focus of the field has shifted toward more advanced techniques like CT or MRI. Some of these errors result in serious patient harm. AI stands to increase accuracy while improving efficiency—prioritizing urgent cases, speeding up interpretation and overall turnaround time, and saving radiologists from the growing existential threats of work dissatisfaction and burnout.

Autonomous Chest Radiograph Interpretation by AI Is Premature

Between the rapid progress of AI and the ever-growing workload of radiologists, there is increasing motivation to let AI take the reins. Indeed, there may be particular circumstances where the benefits of autonomy outweigh the risks, especially for underserved populations. The World Health Organization, for instance, supports the use of autonomous chest radiograph screening for tuberculosis in settings where the disease is rampant and radiologists are scarce (17). Triaging and interpretation of normal chest radiographs (8,9) and those stable from prior examinations (18) are other emerging applications of what we might call "narrow autonomous AI": limited, specific use cases without human supervision. In one study, AI confidently classified 28% of normal chest radiographs as such; while 10 of the 130 chest radiographs classified as

From the Department of Radiology, Penn Medicine, University of Pennsylvania, Philadelphia, Pa (W.B.G.); Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands (M.P.); Department of Radiology, Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea (J.B.S.); Department of Medicine and Radiology, Zucker School of Medicine, Hofstra/Northwell and Lung Institute, Lenox Hill Hospital, New York, NY (S.R.); Department of Radiology and Biomedical Informatics and Center for Artificial Intelligence in Medicine and Imaging, Stanford University, Palo Alto, Calif (C.P.L.); and Center for Pulmonary Functional Imaging, Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis St, Boston, MA 02215 (H.H.). Received October 13, 2023; revision requested November 13; revision received December 8; accepted December 18. **Address correspondence to** H.H. (email: *hhatabu@partners.org*).

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

Radiology 2024; 310(1):e232778 • https://doi.org/10.1148/radiol.232778 • Content codes: AI CH •

normal by AI had findings, only one was actionable (8). More generally, though, we are far from deploying "totally autonomous AI"—interpretation of all chest radiographs without a human in the loop (19,20).

To start, we need more rigorous testing of algorithms with prospective, pragmatic, real-world clinical trials in diverse settings to assure robust generalizability, lack of biases, and a high level of accuracy and reliability. Such testing should include chest radiographs with subtle findings, atypical manifestations of common diseases, rare diseases, and complex cases. It should also include normal chest radiographs with a range of ages and normal variants mimicking disease. Ideally, the reference standard used should not only be the radiologist's interpretation but a superior standard that includes cross-sectional imaging or follow-up. All of this is necessary to ensure that these algorithms do not miss any actionable findings. We also need to carefully consider testing design. In trying to determine whether an AI qualifies for autonomy, current studies often test to see if it can outperform a human radiologist. But this is the wrong metric. Rather, we should compare the AI with and without the assistance of the radiologist. If the AI alone falls short of the human-AI pair, then it hasn't met a strict standard for autonomy. Whether it should be used as such anyway is a matter of benefit versus risk.

Currently, protocols to evaluate the safety and effectiveness of autonomous AI for chest radiograph interpretation are still evolving as regulatory agencies collaborate with expert radiologists to develop methodologies of approval. We need continuous postmarket surveillance and monitoring of U.S. Food and Drug Administration–cleared models to detect performance drift (19). We need systems in place for compliance with regulatory, legal, ethical, and privacy standards. We need careful studies of the application of LLMs to electronic health records (21). Last, we need value-based AI reimbursement strategies and randomized clinical trials to assess the impact of AI on patient outcomes.

The fact is, algorithm performance in standalone research settings may not translate to diverse, real-world clinical care settings, where patient demographics, imaging acquisition parameters, and disease prevalence can vary widely (22). Even within the same data set, algorithm performance is not always reproducible (23). Many of the algorithms do not utilize lateral chest radiographs or prior examinations, which are essential components of a complete chest radiograph interpretation. Moreover, the ground truth used in training AI chest radiograph algorithms frequently has semantically inconsistent and overlapping labels with a mixture of radiographic findings and diagnoses, such as "consolidation," "pneumonia," and "opacity."

AI models still make errors (24). They can be biased, unpredictable, and at times nonsensical. A chest radiograph foundation model built from more than 800 000 images showed poorer performance in female and Black patients (25). LLMs are known to fabricate and "hallucinate." Lacking common sense, genuine understanding, reasoning, and intuition, AI algorithms are particularly susceptible to getting it wrong when faced with novel or rare cases. Even the most "comprehensive" multiclass algorithms (4) are trained on far fewer than the nearly 300 potential chest radiograph findings and nearly 3000 chest disorders (26). These need improvement; but no matter how well-trained the algorithms are on past data, the world is always changing. New diseases and new therapies inevitably arise, bringing unforeseen pulmonary manifestations with them. AI needs to prove itself capable of using drift detection methods, now under development, to know when to retrain on more recent data. It needs to show that it can adapt to unexpected changes and variations in imaging technique before we can consider letting it interpret chest radiographs without our help.

In contrast to narrow autonomy, which may be inevitable, achieving totally autonomous AI is highly unlikely. Even if we could, it doesn't mean we should. The urge for autonomy comes from a fundamental misunderstanding of what radiologists do. Radiologists don't just detect findings. They don't just integrate those findings with clinical history, prior examinations, laboratory data, and more to arrive at differential diagnoses and assess interval change, all tasks that multimodal AI models (15,27) may eventually be able to do. Rather, human radiologists assess the clinical significance of the findings. They use their medical judgment, intuition, common sense, and contextual knowledge to synthesize it all into a coherent picture. Beyond a mere list of findings, they provide meaning. They can elaborate on the basis for their conclusions, answer questions, and engage in conversation; they can even change their minds. In contrast, while an algorithm may be able to output more than a hundred abnormalities, it is unable to contextualize them, which can be anxiety-producing for patients and time-consuming for clinicians trying to sort through the AI results.

Completely autonomous AI would be not only imprudent, but unsustainable. Trainees would fail to develop skills in chest radiograph interpretation. Without replenishing a steady pool of experienced radiologists, we would lose the very expertise on which AI trains and thrives.

By setting total autonomy as a benchmark for AI, we not only raise unnecessary fears of radiologists being replaced, but also obscure the real path forward. We've been drawn into a binary mindset, pitting human against machine. In doing so, we've missed the deeper point: There are cases where an AI can outperform a human, and vice versa, but, if it's done right, neither the AI nor the radiologist can outperform a human-AI partnership (1,3,9,28).

The Need for Human-AI Symbiosis

More than 60 years ago, J.C.R. Licklider, a pioneer in computer science, proposed the notion of human-computer symbiosis (29). He predicted that through cooperation, humans and AI together would give rise to a new form of intelligence. More recently, "human-centered AI" has emerged as a practical paradigm for human-AI collaboration (30). Human-centered AI seeks to maximize machine automation but with human control and oversight. Applying the goals of human-centered AI and symbiosis to the design and implementation of AI in chest radiography will help ensure that AI remains safe, reliable, and trustworthy (30).

There are fundamental differences in human cognition and computer-based intelligence. Radiologists come to the workstation with real-world wisdom, understanding, and common sense, not to mention creativity, empathy, and flexibility. AI

Key Points for Human-AI Symbiosis

- Developments in deep learning, large language models, and multimodal vision-language foundation models have positioned AI to transform chest radiography and the role of radiologists in profound ways
- AI can already extract new forms of information and detect a diverse range of chest radiograph findings, with performance comparable to or exceeding that of human radiologists, raising fears that radiologists will be replaced
- While narrow autonomous AI—the unsupervised use of AI in limited, specific scenarios where benefits outweigh the risks—is emerging, totally autonomous AI, in which AI would interpret all chest radiographs without a human in the loop, is highly unlikely
- By setting total autonomy as a benchmark for AI, thus pitting human against machine, we miss the deeper point: There may be cases where an AI can outperform a human, and vice versa, but when done right, neither the AI nor the radiologist can outperform a human-AI partnership
- Human-AI symbiosis—not totally autonomous AI—should be the goal for AI in chest radiography
- Together with the principles of human-centered AI, human-AI symbiosis can harness the vast potential of AI for chest radiography while helping to ensure that AI is safe, reliable, and effective
- We can maximize the power of human-AI symbiosis by designing user interfaces and collaborative workflows that play to the complementary strengths and weaknesses of human cognition and AI
- If radiologists proactively adapt and evolve together with AI, we will not be replaced but rescued; symbiosis with AI will lessen workloads and provide more time for direct engagement with patients and referring clinicians—allowing radiologists to be less like machines and more human

Note.—AI = artificial intelligence.

systems come with pattern recognition, a vast memory, blazing processing speed, and the ability to manipulate huge amounts of data without ever stopping for lunch. Humans can be distracted or tired. AI systems can be confused by a lack of context or a biased training set.

These contrasting strengths and weaknesses yield the power of human-AI symbiosis. Radiologists can correct AI's mistakes, and AI can alert radiologists to pending missteps, improving one another through continuous and mutual feedback. The key is to maximize the differences. When there is too much overlap between human and AI errors, the benefits of symbiosis are lost. In the worst-case scenario, co-interpretation, even with human experts, can bring down the performance of AI, which has been shown for pulmonary nodules on chest radiographs (5). Alternatively, overreliance on AI, known as automation bias, can hamper the performance of the radiologist. More often, human and AI errors diverge, such as when an AI mistakes an obvious skin fold for a pneumothorax or a fatigued radiologist misses a lesion. In such cases, each partner will strengthen the other. Together, they can not only boost detection accuracy (3,9,20,28) but create a new form of collective intelligence.

How can we optimize this symbiosis for chest radiography? We can take a lesson from chess. According to chess grandmaster Garry Kasparov, the best predictor of a winning human-AI pair is not the expertise of the player nor the sophistication of the AI but the strength of their strategy for working together (31). For radiology, the key to human-AI integration will be well-designed, nondisruptive user interfaces and collaborative AI-radiologist workflows (28). These must include AI explainability techniques, which provide clues to the basis for the model's decisions, confidence levels for predictions (32,33), and editing functions to allow radiologists to immediately modify erroneous results. The AI, for instance, will display a warning signal when suspecting human error or an alert when its own confidence levels are low. (Uncertainty quantification methods, which allow AI models to provide such confidence levels for their predictions, are being actively investigated.) This can allow the symbiotic system to determine who should be the dominant reader: AI for clear-cut cases and radiologists for those about which AI is uncertain (28). Systems then need to be in place to provide ground truth feedback to the human-AI pair. With the right interface and workflow, we can harness the enormous benefits of AI for chest radiography while minimizing the dangers. In their symbiotic interaction, the radiologist and AI each can end up better than they might have been on their own (Table).

Coevolution of AI and Radiologists toward Improved Patient Care

As AI evolves, so too will the radiologist. In the coming years, radiologists will find themselves working in a reimagined diagnostic cockpit (2), increasingly interpreting chest radiographs and other imaging studies in symbiotic partnership with AI. While AI models are trained to maximize detection or classification accuracy, the AI-human symbiotic pair must be trained to maximize patient care. After all, the human truly at the center of human-centered AI is the patient. AI is key to lessening workloads and turnaround time, but it is up to radiologists to use this gained time for more direct engagement with patients and with referring clinicians (34).

If radiologists proactively adapt and evolve together with AI, we will not be replaced; we'll be rescued. With less time devoted to routine tasks, radiologists will be able to reclaim the role they were always meant to have: that of the caring and supportive physician (34,35). Symbiosis with AI will allow radiologists to be less like machines and more human.

Acknowledgment: The authors are grateful to Amanda Gefter for her generous assistance in editing the manuscript.

Disclosures of conflicts of interest: **W.B.G.** Grant to institution from Siemens Healthineers for salary support; consulting fees from Siemens Healthineers. **M.P.** Grants to institution from Canon Medical Systems and Siemens Healthineers; license fees from Canon Medical Systems; speaker bureau payment to institution from Canon Medical Systems and Siemens Healthineers; payment to institution for role as immediate past president of the Dutch Society of Radiology. **J.B.S.** Royalties from Vuno, Corelinesoft, and Promedius; patents planned, issued, or pending with Asan Medical Center; chief medical officer of Promedius; president of the Korean Society of Thoracic Radiology; stock in Corelinesoft and Promedius. **S.R.** Royalties from McGraw Hill. **C.P.L.** Grants to institution from the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (75N92020C00008 and 75N92020C00021); grants and gifts to institution from

Bunkerhill Health, Carestream, CARPL, Clairity, GE HealthCare, Google Cloud, IBM, Kheiron, Lambda, Lunit, Microsoft, Nightingale Open Science, Philips, Siemens Healthineers, Stability.ai, Subtle Medical, VinBrain, Visiana, Whiterabbit.ai, Lowenstein Foundation, and Gordon and Betty Moore Foundation; consulting fees from Sixth Street and Gilmartin Capital; honorarium for travel and presentations from Mayo Clinic; patent pending with GE HealthCare and Stanford University; president of the RSNA board of directors; board of directors member and shareholder in Bunkerhill Health, option holder in Whiterabbit.ai, and advisor for and option holder in GalileoCDS, Sirona Medical, Adra, and Kheiron. **H.H.** Grants from Canon Medical Systems and Konica Minolta; consulting fees from Canon Medical Systems and Boehringer Ingelheim.

References

1. Langlotz CP. Will artificial intelligence replace radiologists? Radiol Artif Intell 2019;1(3):e190058.
2. Langlotz CP. The future of AI and informatics in radiology: 10 predictions. Radiology 2023;309(1):e231114.
3. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017;284(2):574–582.
4. 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.
5. Hwang EJ, Park S, Jin KN, et al. Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. JAMA Netw Open 2019;2(3):e191095.
6. Yoo H, Kim KH, Singh R, Digumarthy SR, Kalra MK. Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs. JAMA Netw Open 2020;3(9):e2017135.
7. Chamberlin JH, Aquino G, Nance S, et al. Automated diagnosis and prognosis of COVID-19 pneumonia from initial ER chest x-rays using deep learning. BMC Infect Dis 2022;22(1):637.
8. Plesner LL, Müller FC, Nybing JD, et al. Autonomous chest radiograph reporting using AI: estimation of clinical impact. Radiology 2023;307(3):e222268.
9. Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP. Assessment of convolutional neural networks for automated classification of chest radiographs. Radiology 2019;290(2):537–544.
10. Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts HJWL, Hoffmann U. Deep learning to assess long-term mortality from chest radiographs. JAMA Netw Open 2019;2(7):e197416.
11. Weiss J, Raghu VK, Bontempi D, et al. Deep learning to estimate lung disease mortality from chest radiographs. Nat Commun 2023;14(1):2797.
12. Sohn JH, Chen Y, Lituiev D, et al. Prediction of future healthcare expenses of patients from chest radiographs using deep learning: a pilot study. Sci Rep 2022;12(1):8344.
13. Lee JH, Lee D, Lu MT, et al. Deep learning to optimize candidate selection for lung cancer CT screening: advancing the 2021 USPSTF recommendations. Radiology 2022;305(1):209–218.
14. Pyrros A, Borstelmann SM, Mantravadi R, et al. Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. Nat Commun 2023;14(1):4039.
15. Khader F, Müller-Franzes G, Wang T, et al. Multimodal deep learning for integrating chest radiographs and clinical parameters: a case for transformers. Radiology 2023;309(1):e230806.
16. Gefter WB, Hatabu H. Reducing errors resulting from commonly missed chest radiography findings. Chest 2023;163(3):634–649.
17. WHO consolidated guidelines on tuberculosis: Module 2: screening – systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021.
18. Yun J, Ahn Y, Cho K, et al. Deep learning for automated triaging of stable chest radiographs in a follow-up setting. Radiology 2023;309(1):e230606.
19. Docket No. FDA-2019-N-5592: "Public Workshop – Evolving Role of Artificial Intelligence in Radiological Imaging": Comments of the American College of Radiology. American College of Radiology; Radiological Society of North America. https://www.acr.org/-/media/ACR/NOINDEX/Advocacy/acr_rsna_comments_fda-ai-evolvingrole-ws_6-30-2020.pdf. Published June 30, 2020. Accessed August 2023.
20. Hwang EJ, Goo JM, Yoon SH, et al. Use of artificial intelligence-based software as medical devices for chest radiography: a position paper from the Korean Society of Thoracic Radiology. Korean J Radiol 2021;22(11):1743–1748.
21. Wornow M, Xu Y, Thapa R, et al. The shaky foundations of large language models and foundation models for electronic health records. NPJ Digit Med 2023;6(1):135.
22. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiol Artif Intell 2022;4(3):e210064.
23. Cho Y, Kim YG, Lee SM, Seo JB, Kim N. Reproducibility of abnormality detection on chest radiographs using convolutional neural network in paired radiographs obtained within a short-term interval. Sci Rep 2020;10(1):17417.
24. Lind Plesner L, Müller FC, Brejnebøl MW, et al. Commercially available chest radiograph AI tools for detecting airspace disease, pneumothorax, and pleural effusion. Radiology 2023;308(3):e231236.
25. Glocker B, Jones C, Roschewitz M, Winzeck S. Risk of bias in chest radiography deep learning foundation models. Radiol Artif Intell 2023;5(6):e230060.
26. Filice RW, Kahn CE Jr. Biomedical ontologies to guide AI development in radiology. J Digit Imaging 2021;34(6):1331–1341. [Published correction appears in J Digit Imaging 2022;35(5):1419.]
27. Zhang X, Wu C, Zhang Y, Xie W, Wang Y. Knowledge-enhanced visual-language pre-training on chest radiology images. Nat Commun 2023;14(1):4542.
28. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. Lancet Digit Health 2022;4(7):e507–e519.
29. Licklider JCR. Man-computer symbiosis. IRE Trans Hum Factors Electron 1960;HFE-1(1):4–11.
30. Shneiderman B. Human-Centered AI. Oxford, UK: Oxford University Press, 2022.
31. Kasparov G. Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins. New York, NY: Public Affairs, Hachette Book Group, 2017.
32. Alves N, Bosma JS, Venkadesh KV, et al. Prediction variability to identify reduced AI performance in cancer diagnosis at MRI and CT. Radiology 2023;308(3):e230275.
33. Choi Y, Yu W, Nagarajan MB, et al. Translating AI to clinical practice: overcoming data shift with explainability. RadioGraphics 2023;43(5):e220105.
34. Wachter R. The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. New York, NY: McGraw-Hill Education, 2015.
35. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY: Basic Books, Hachette Book Group, 2019.