



Are the Pilots Onboard? Equipping Radiologists for Clinical Implementation of AI

Umber Shafique¹ · Umar Shafique Chaudhry² · Alexander J. Towbin^{3,4}

Received: 3 May 2023 / Revised: 25 July 2023 / Accepted: 26 July 2023 / Published online: 9 August 2023
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2023

Abstract

The incorporation of artificial intelligence into radiological clinical workflow is on the verge of being realized. To ensure that these tools are effective, measures must be taken to educate radiologists on tool performance and failure modes. Additionally, radiology systems should be designed to avoid automation bias and the potential decline in radiologist performance. Designed solutions should cater to every level of expertise so that patient care can be enhanced and risks reduced. Ultimately, the radiology community must provide education so that radiologists can learn about algorithms, their inputs and outputs, and potential ways they may fail. This manuscript will present suggestions on how to train radiologists to use these new digital systems, how to detect AI errors, and how to maintain underlying diagnostic competency when the algorithm fails.

Keywords Pilots · Radiologists · AI · Safety education

Introduction

The number of yearly radiologic examinations continues to grow, straining the radiologist workforce. Artificial intelligence (AI) applications are seen as one potential solution to mitigate this challenge. The hope is that these newly developed applications can augment radiologists helping them to efficiently identify, annotate, define, and categorize disease. While AI applications are being developed, the rollout of solutions has been haphazard and slow. This difficulty in solution rollout is related to multiple factors, including but not limited to the developing marketplace, non-unified AI governance, and a lack of standards related to AI deployment. Because of this, most of the

implementation effort has been focused on the technical challenges. Unfortunately, this focus ignores the many educational needs of using AI in clinical practice. The purpose of this manuscript is to describe some of the current challenges related to radiologist training. Throughout this manuscript, we will make recommendations for how radiologists may be able to develop and maintain competency regarding this new type of software.

Background

The healthcare system has been likened to the aircraft industry [1, 2]. Both industries are complex and rely on cutting-edge technology to manage their workflow. In addition, both industries strive to function with high reliability due to the seriousness of any error. Generally, the airline industry is held up as the gold standard that healthcare should endeavor to achieve.

While air travel is extremely safe, mistakes have occurred. Healthcare leaders have learned from these mistakes and have adopted some of the airline industry's interventions to improve patient safety. Checklists are perhaps the most notable example of an intervention that has been translated from the airline industry to healthcare [3, 4].

Because the airline industry has a long history of using automated and autonomous systems, we can continue to learn from their experience. Recently, the crash of the Boeing 737 MAX airliners highlighted a new concern. In these

✉ Umber Shafique
radumber@yahoo.com

Umar Shafique Chaudhry
umchaudh@utmb.edu

Alexander J. Towbin
alexander.towbin@cchmc.org

¹ Indiana University School of Medicine, Indianapolis, IN, USA

² University of Texas Medical Branch, Galveston, TX, USA

³ Department of Radiology, Cincinnati Children's Hospital, Cincinnati, OH, USA

⁴ Department of Radiology, University of Cincinnati College of Medicine, Cincinnati, OH, USA

accidents, the pilots were unaware of how AI was implemented, its potential impact and risks to flight, and how to disable the feature when the tool malfunctioned [5]. In describing this failure, Mongan and Kohli describe five lessons for Radiology as the specialty develops and implements AI [5]. One of the lessons that our manuscript will elaborate upon is “The people in those workflows must be made aware of the AI and must receive training on the expected function and anticipated dysfunction of the system [5].” This manuscript will present suggestions on how to train radiologists to use these new digital systems, how to detect AI errors, and how to maintain underlying diagnostic competency when the algorithm fails (Fig. 1).

Educational Content Creation

We expect that hundreds of AI algorithms will be running in the near future, helping radiologists to identify, define, characterize, and report imaging findings [6]. As we prepare for this eventuality, we must design systems that allow radiologists to learn about the algorithm, its inputs and output, and potential ways it can fail. However, it is important to stress that we should not create an environment where radiologists are expected to complete individual training modules for each algorithm. We believe this type of solution would not only be an arduous task for the administrators but also lead to added fatigue for radiologists, already strained by the educational and work demands of their systems [7, 8].

One potential solution is to develop a tier-based system to guide training. This type of system could be based on the expected output of the algorithm, its impact on the decision-making of the radiologist, and the urgency of the findings

to direct patient care. In this type of tiered approach to risk stratification, algorithms that guide emergent triage and patient management would require more rigorous training than algorithms that improve image quality or guide routine lesion detection (Fig. 2).

For example, a stroke detection algorithm quantifies brain perfusion. The algorithm may trigger an alert before the radiologist can fully inspect the AI output, leading to the activation of the stroke team and subsequent therapy. Because delivery of harmful therapy may be initiated quickly, we believe radiologists should have detailed training and demonstrate competency in detecting failures before using the algorithm. The radiologists may be incentivized by earning continuing education (CE) and continuing medical education (CME) credits by participating in the education program. An example of an algorithm that may require less rigorous training is the segmentation of an organ. If this algorithm provides an overlay showing how the organ was segmented, the radiologist can evaluate the boundaries and decide if an error occurred. In this instance, because the output is easily explainable, is not time-sensitive, and does not drive urgent therapy, minimal training is needed.

Even with this tiered approach, radiology departments may struggle to determine the training needed. Additionally, if institutions differ from vendors in their training needs assessment, they may be forced to develop suitable training materials. This would lead to multiple instances of duplicated effort and non-uniformity of educational content. Thus, we believe an external body is needed to help determine the need for training related to a commercial algorithm. In other situations, the government has provided this regulatory oversight [9]. However, recent regulation changes have placed this responsibility on the vendors. Other authors have

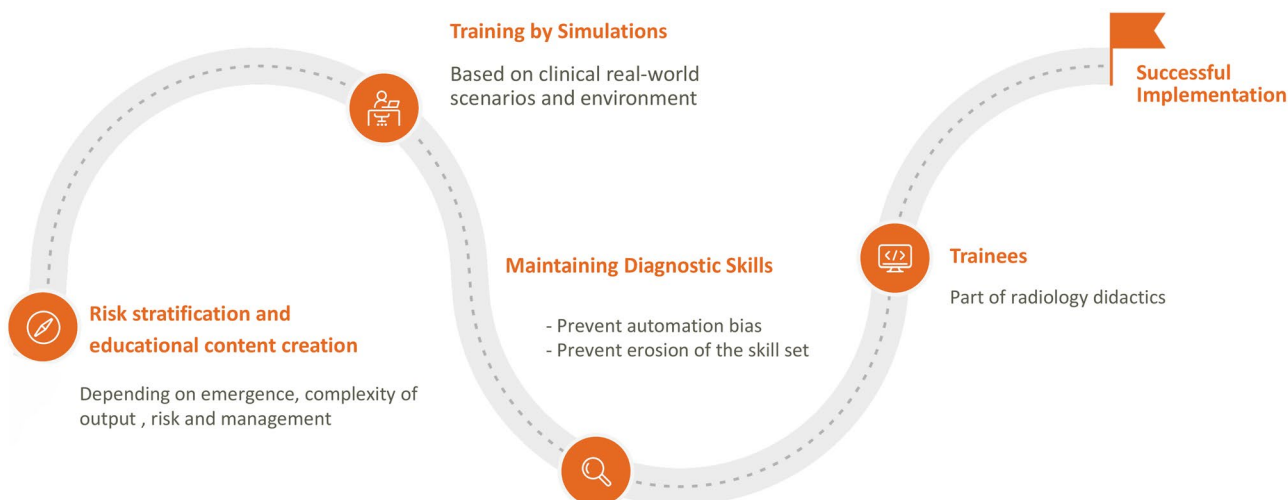
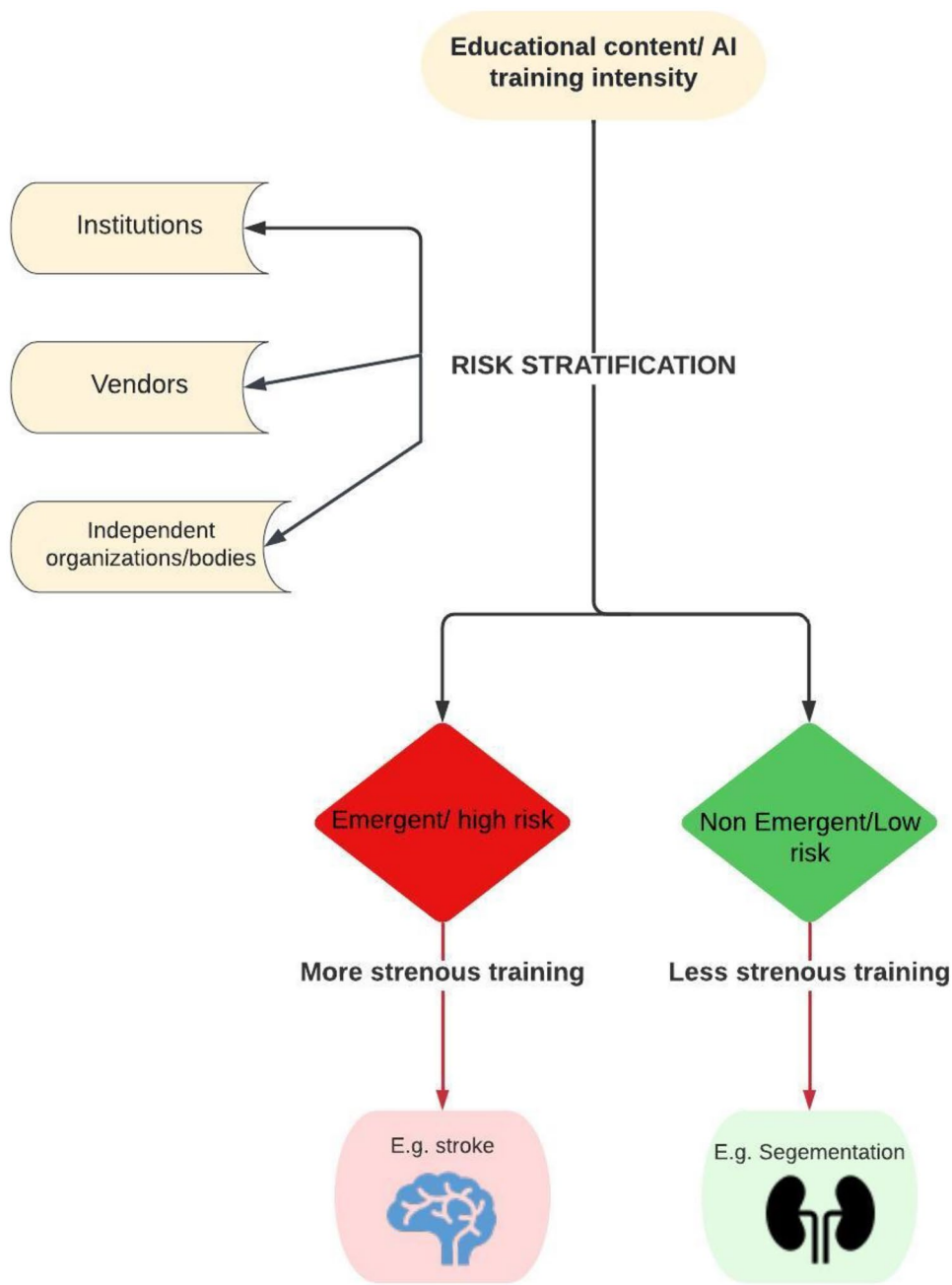


Fig. 1 Waypoints for AI. Pilots in the real-world use waypoints that guide them for flight. Similar markers can be obtained in radiology for successful implementation

Fig. 2 Educational content creation will depend on the risk stratification of AI applications to determine the intensity of training radiologists should undergo



highlighted vendors’ challenges in prioritizing safety and education over corporate profit [5].

While a governing body could recommend the level of training needed for any one application, we believe that vendors should support the cost and creation of educational materials. However, we must also be mindful of the lessons from the airline industry. In the Boeing 737 MAX disasters, the training material did not describe the system sufficiently and did not provide pilots with the guidance needed to disable the application [5]. To address this concern, we recommend the development of standards guiding educational modules and the content each module should include.

To address this challenge, we believe medical societies, staffed by subspecialty experts, are best suited to guide training needs and collaborate with vendors. Thus, there is also a need to recruit and develop expert leadership in AI and its clinical implementation. These AI experts will need adequate time, training, and funding to provide this service [10, 11].

Radiologists should not be held accountable for quality assurance of individual algorithms as this will become an unwarranted burden added to the job responsibilities. While radiologists can aid in identifying the problem, they should have a robust support system to resolve such issues. We recommend the creation of a new position, the AI technologist. These technologists would be

responsible for testing algorithms, analyzing their outputs, and evaluating system performance over time, thus helping to ensure the quality and integrity of the generated data.

Training by Simulation

The number of hours required to train an airline pilot with and without simulation varies depending on multiple factors, including the aviation authority's regulations, the airline's training program, and the pilot's proficiency. According to the United States Federal Aviation Administration (FAA), the minimum flight experience required for an Airline Transport Pilot (ATP) certificate is 1500 h [12]. In the past, it was not uncommon for pilots to accumulate several thousand flight hours before obtaining an ATP certificate. However, many pilots complete their training within structured programs that integrate simulator time [13]. These programs can take anywhere from 12 to 24 months to complete. The integration of flight simulation technology in pilot training programs has led to a decrease in hazards and improved instructional quality, ultimately promoting overall flight security while minimizing expenses related to both training and aircraft operations [14–16].

Simulation is already commonly used in healthcare to help caregivers practice complex or stressful techniques such as cardiopulmonary resuscitation [17]. Simulation has also shown promise in radiology to provide training and demonstrate ongoing competency [18–20]. Radiology is well-suited to simulation-based training due to its entirely digital workflow, its reliance on standards, and its potential ability to re-use anonymized patient data to mimic a newly acquired study. These features mean that with the appropriate permissions, the same data could be used to train radiologists at multiple sites. Because of these factors, we believe simulation-based training is an essential tool to help guide radiologist AI education.

During initial training, the simulation would allow radiologists to learn while actively “interpreting” curated cases. Ideally, this type of training would include instances where the algorithm is functioning normally and other instances where the algorithm fails. If this type of training were to be implemented, radiologists could experience the modes of failure and learn how this failure could be detected.

The simulation could be used for more than initial training. It could also be used as a tool for radiologists to demonstrate ongoing competency and as a mechanism to deliver urgent training to a large group of radiologists. For example, if a simulator was integrated into the picture archiving and communication system (PACS), simulated cases could appear randomly as part of daily work. If this were to occur, radiologists could continuously learn how new algorithms work and maintain their skills, even with infrequently used algorithms.

While the ideal simulator would be integrated into the clinical PACS, allowing radiologists to interpret and learn from simulated cases as part of their daily work, this concept is not possible with today's viewers. Thus, there is an opportunity for third-party simulation tools to be developed using human-in-the-loop machine learning.

If this tool were created, we believe that it should include multiple key features:

1. Full patient and study data sets to closely mimic the reading environment
2. Ability to integrate with AI platforms so that training material from newly purchased algorithms is immediately accessible
3. Ability to integrate with ancillary systems (such as the electronic health record and dictation software) so that radiologists can view all potential inputs and outputs of each algorithm
4. Provide radiologists with personalized feedback and teaching related to performance and algorithm use
5. Ability for radiologists to request additional training on particular use cases
6. Ability to link to reference material, including videos, checklists, articles, and websites
7. Maintain a portfolio of learning completed and current competencies
8. Ability to provide guidance highlighting expected response to system failures

Maintaining Diagnostic Skills

There may be unintended consequences of implementing AI. Two potential consequences are automation bias and erosion of a radiologist's skill. Automation bias occurs when the human observer favors the output of an automated system over contradictory information that they observe. If not checked, this bias could easily occur in radiology, particularly as AI performance improves and tools are developed that help radiologists complete tedious tasks.

Automation bias and improved AI performance could also lead to erosion of radiologist skill. For example, an algorithm designed to detect pulmonary nodules and present them to the radiologist for adjudication may be so successful that a radiologist stops performing their assessment to detect pulmonary nodules (automation bias). Because radiologists stop using their judgment, they may become less skilled in detecting additional nodules (erosion of skill).

A group of European researchers from Germany and the Netherlands presented a prospective multicenter study on automation bias in mammography [21]. In summary, when AI systems suggested an abnormality, inexperienced readers were more likely to make errors of commission by following

the false information provided. However, experienced readers showed no significant differences in their tendency to follow AI suggestions when they were false.

AI platforms and algorithms should be designed to prevent automation bias and skill erosion. Prior studies have shown that automation bias can be reduced by turning on the AI algorithm after independent interpretation, decreasing the prominence of algorithm output, providing algorithm confidence with a result, providing the rationale for the interpretation, and varying the reliance of the observer on the automation [22, 23]. Each of these interventions could be employed within radiology. Using the pulmonary module algorithm as an example, each of the following interventions could be employed to reduce automation bias:

1. Ability to turn on/off AI algorithm: Interpret the exam independently and then turn on the AI algorithm for result concordance.
2. Decreasing the prominence of algorithm output: The algorithm could include the output as the last series in PACS, encouraging the radiologist to view other series first.
3. Providing algorithm confidence with a result: As each nodule is displayed, a confidence score related to the nodule's likelihood of being cancer could be displayed.
4. Providing rationale for the interpretation: As each nodule is shown, features that make the nodule more or less likely to be malignant could be displayed.
5. Varying the reliance of the observer on automation: The radiologist could be instructed that the algorithm will not run on a random selection of cases. In these instances, the radiologist must be notified that the algorithm did not run and that a primary human review is needed.

We believe more studies are needed to determine the impact of automation bias and skill erosion within radiology. Additional recommendations may be made based on this future research.

Special Consideration—Trainees

AI and its impact on learning should be a focus for resident and fellow training. We believe that the training of junior learners should include the basics of AI in addition to the training described above for each algorithm. This added level of initial training is needed so that the next generation of radiologists can address the concepts of training and skill erosion as AI increasingly becomes integrated within clinical practice [24]. Given the rapidly changing landscape of radiology, we believe that this AI training is equivalent to the necessity of learning physics in that it impacts many elements of the daily workflow and is crucial to enable problem-solving.

It is important to mention that the risks of automation bias and skill erosion may be more pronounced for trainees [21]. Junior learners need time and practice to develop their skills and hone their experience. Using AI may significantly hinder this development. Therefore, the interventions used to reduce AI bias may take on even greater importance. For example, the frequency by which the algorithm's prompts are displayed may increase with increasing clinical experience such that a junior learner sees the automated results displayed only 5% of the time. In contrast, a radiologist with considerable experience sees the automated results revealed 95% of the time.

Conclusion

Integration of AI in the radiological clinical workflow is imminent. As these solutions are implemented, tools should be designed to train the radiologist, detect errors, and reduce bias and skill erosion. Potential solutions include external governing bodies, training platforms, competency-based learning portfolios, simulation, and changes to user interface and algorithm output. As these new solutions are developed, each type of user, from junior learners to experienced radiologists, should be considered so that our specialists can improve care and reduce potential harm.

Author Contribution All authors contributed to the study's conception and design. The first draft of the manuscript was written by Dr. Umer Shafique, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript."

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Doucette JN. View from the cockpit: what the airline industry can teach us about patient safety. *Nursing* 2021;36:50-53
2. Hudson P. Applying the lessons of high risk industries to health care. *BMJ Quality & Safety* 2003;12:i7-i12
3. Gawande A. The checklist. 2007
4. Clay-Williams R, Colligan L. Back to basics: checklists in aviation and healthcare. *BMJ quality & safety* 2015;24:428-431
5. Mongan J, Kohli M. Artificial intelligence and human life: five lessons for radiology from the 737 MAX disasters. *Radiology Artificial intelligence* 2020;2

6. Daye D, Wiggins WF, Lungren MP, et al. Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How? *Radiology* 2022;212:151
7. Harolds JA, Parikh JR, Bluth EI, et al. Burnout of radiologists: frequency, risk factors, and remedies: a report of the ACR Commission on Human Resources. *Journal of the American College of Radiology* 2016;13:411-416
8. Waldman JD, Kelly F, Arora S, et al. The shocking cost of turnover in health care. *Health care management review* 2010;35:206-211
9. Food, Administration D. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). 2019
10. Li D, Morkos J, Gage D, et al. Artificial Intelligence Educational & Research Initiatives and Leadership Positions in Academic Radiology Departments. *Current Problems in Diagnostic Radiology* 2022
11. Perchik J, Smith A, Elkassem A, et al. Artificial Intelligence Literacy: Developing a Multi-institutional Infrastructure for AI Education. *Academic Radiology* 2022
12. Congress U. Airline Safety and Federal Aviation Administration Extension Act of 2010. *Public Law* 2010:111-215
13. Longridge T, Burki-Cohen JS, Go TH, et al. Simulator fidelity considerations for training and evaluation of today's airline pilots. 2001
14. Socha V, Socha L, Szabo S, et al. Training of pilots using flight simulator and its impact on piloting precision. *Transport Means Juodkrante: Kansas University of Technology* 2016:374-379
15. Aragon CR, Hearst MA. Improving aviation safety with information visualization: a flight simulation study. *Proceedings of the SIGCHI conference on Human factors in computing systems*; 2005:441–450
16. Devices T. European Aviation Safety Agency. 2012
17. Sullivan NJ, Duval-Arnould J, Twilley M, et al. Simulation exercise to improve retention of cardiopulmonary resuscitation priorities for in-hospital cardiac arrests: a randomized controlled trial. *Resuscitation* 2015;86:6-13
18. Towbin AJ, Paterson BE, Chang PJ. Computer-based simulator for radiology: an educational tool. *Radiographics* 2008;28:309-316
19. Towbin AJ, Paterson B, Chang PJ. A computer-based radiology simulator as a learning tool to help prepare first-year residents for being on call. *Academic radiology* 2007;14:1271-1283
20. Shah C, Davtyan K, Nasrallah I, et al. Artificial Intelligence-Powered Clinical Decision Support and Simulation Platform for Radiology Trainee Education. *Journal of Digital Imaging* 2022:1–6
21. Baltzer PA. Automation Bias in Breast AI. *Radiological Society of North America*; 2023:e230770
22. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 2012;19:121-127
23. Do HM, Spear LG, Nikpanah M, et al. Augmented radiologist workflow improves report value and saves time: a potential model for implementation of artificial intelligence. *Academic radiology* 2020;27:96-105
24. Fischetti C, Bhattar P, Frisch E, et al. The evolving importance of artificial intelligence and radiology in medical trainee education. *Academic Radiology* 2022;29:S70-S75

Take-Home Points

1. Radiologists should be prepared for the implementation of AI into clinical workflow.
2. Training resources should be developed for radiologists at every level of experience.
3. AI software developers, governing bodies, and radiological societies should collaborate to develop solutions to standardize training and minimize the potential of automation bias and skill erosion.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.