Editorial

Artificial intelligence in radiology: Risks and the need for residency education

Samantha Fisher, BSc (Hons)[1]

1. Faculty of Medicine, Memorial University, St. John's, NL, Canada

Keywords: Artificial Intelligence, Radiology, Residency Training, Medical Education

The intersection of radiology and artificial intelligence (AI) is a rapidly evolving development in modern medicine. AI has the potential to enhance diagnostic accuracy, improve workflow efficiency, and support clinical decision-making[1,2], positioning it as a powerful adjunct to radiology. At the same time, the integration of AI into radiology introduces complex risks that require critical oversight. In particular, algorithmic bias, automation bias, and AI interpretability represent some of the risks associated with modern AI systems. Despite the increasing adoption of AI within radiology, AI education has not been widely incorporated into radiology residency training programs[3,4]. By examining these three challenges of AI in radiology, this discussion highlights the growing need for formal AI education to ensure future radiologists can appropriately assess and safely integrate these technologies into clinical practice.

Algorithmic bias in the context of radiology and AI is defined as systematic errors or prejudices in AI algorithms[5]. Because AI models are trained on large medical imaging datasets[6], the quality and composition of these datasets directly influence how AI systems perform. Simply put, AI systems become efficient at recognizing patterns within the data on which they are trained. However, medical imaging datasets frequently contain demographic imbalances, often underrepresenting certain groups based on factors such as race, ethnicity, sex, or socioeconomic status[7]. These imbalances may reflect existing healthcare disparities, unequal access to care, or limited demographic reporting within datasets themselves[7,8]. As a result, AI systems can learn and reproduce these biases, leading to performance disparities that disproportionately disadvantage historically underrepresented populations[9]. For example, a study evaluating racial bias in a cardiac MRI segmentation AI model found significant differences in model performance across racial groups when trained on the UK Biobank dataset, which consisted predominantly of White patients (81%). The authors concluded that the observed racial bias was likely related to imbalance within the training data itself, highlighting how non-representative datasets can contribute to inequitable AI performance[10]. This raises significant concerns that AI may perpetuate, or even exacerbate, existing healthcare inequities. Identifying AI algorithmic bias is further complicated by inconsistent demographic reporting within medical image databases[11,12]. The Canadian Association of Radiology (CAR) has recognized both the risk of algorithmic bias and the challenges associated with identifying these biases, emphasizing the important role that physicians and data scientists play in recognizing, evaluating, and mitigating them[9]. As AI becomes increasingly integrated into radiology, radiologists must be equipped with the knowledge to identify and evaluate algorithmic bias to ensure that AI systems do not perpetuate existing healthcare inequities.

Automation bias is another risk associated with the integration of AI into radiology and has been widely discussed in the literature. It refers to the tendency of individuals to place excessive trust in AI-generated outputs while disregarding or underutilizing their own clinical judgment and critical thinking skills[13]. Literature suggests that the risk of automation bias increases with task complexity and time constraints, making radiology particularly susceptible[14,15]. This overreliance on AI systems in radiology may contribute to missed or incorrect diagnoses[16]. In a study evaluating AI-assisted mammography interpretation, radiologists across all experience levels were more likely to assign an incorrect Breast Imaging Reporting and Data System (BI-RADS) classification when the AI system also provided an incorrect classification[17]. Because BI-RADS classifications are used to estimate malignancy risk and guide patient management, these findings are of major concern, as they demonstrate how radiologists may become susceptible to automation bias, which may negatively influence clinical decision-making and patient care. Importantly, the way AI is integrated into radiology workflows may further compound the risk of automation bias. In many radiology workflows that incorporate AI, AI outputs are directly integrated into Picture Archiving and Communication Systems (PACS) with limited opportunity for radiologist interaction or modification.[9] This passive integration of AI outputs may encourage overreliance without sufficient independent evaluation by radiologists, potentially contributing to diagnostic errors. Furthermore, incorrect AI interpretations accepted without sufficient scrutiny may reinforce future AI model inaccuracies and contribute to repeated diagnostic errors over time[15]. Given this, formal AI education within radiology residency programs is essential to prepare future radiologists to critically engage with AI systems and reduce the risk of automation bias influencing clinical decision-making.

In addition to algorithmic and automation bias-related risks, the interpretability of AI systems remains a major concern in radiology. AI interpretability refers to the ability to understand how an AI system arrives at its conclusions.

However, many AI systems function as "black boxes", meaning they may generate conclusions without providing a clear understanding of how those conclusions were reached[18]. As a result, radiologists may have limited insight into how AI outputs are produced. This can be problematic, as AI systems may occasionally produce unexpected or unreliable outputs that do not align with conventional clinical reasoning[19], and the "black box" nature of these systems makes AI-related errors difficult to detect[20]. This creates significant challenges for radiologists who are expected to assess the reliability of AI-generated recommendations in clinical practice. Closely related to these interpretability challenges is the susceptibility of AI systems to adversarial inputs. Adversarial AI refers to techniques that manipulate AI models through subtle alterations in input data intended to deceive the system[21]. The potential impact of adversarial AI in radiology has been demonstrated in the literature. For example, one study demonstrated that adversarial AI attacks could add or remove evidence of lung cancer on CT imaging, thereby altering clinical interpretation[22]. This represents only one of many ways adversarial AI attacks may impact radiology practice. Because many AI systems function as "black boxes," adversarial AI can be challenging to recognize or understand. Consequently, these attacks may contribute to missed diagnoses, inappropriate or unnecessary treatments, and delayed urgent care[21,23]. Therefore, education regarding AI interpretability and the potential for adversarial AI is essential to ensure radiologists recognize the vulnerabilities and limitations of AI systems and maintain appropriate oversight[21].

Despite the increasing integration of AI into radiology practice and the risks associated with its use, formal AI education within radiology residency programs remains limited and inconsistent. This is particularly problematic given the scale of AI adoption in medicine. Over 1,000 FDA-cleared AI algorithms have been approved for clinical use as of 2025, with many of them designed for radiology tasks[24]. Given this, there is a responsibility to prepare radiology residents to critically evaluate and safely work with AI. However, in a 2023 study in the United States, nearly one-quarter of radiology residents reported having no AI or machine learning educational offerings within their residency program[3]. Similar concerns have been recognized in Canada. A 2023 study evaluating a pilot AI workshop within a Canadian radiology residency program noted that it represented one of the first formal AI educational initiatives integrated into radiology residency training in Canada[25]. Initiatives such as these demonstrate growing recognition of the importance of AI education. Despite this progress, there remains no standardized AI curriculum for radiology residents[25]. Furthermore, many of the existing AI educational opportunities remain external to residency curricula and rely on extracurricular learning[25]. As a result, AI education remains inconsistent across residency programs[25] and may be difficult to incorporate alongside residents' demanding clinical and academic responsibilities[26,27]. Importantly, evidence suggests that structured AI education can be effective. A systematic review published in 2025 found that roughly 93% of studies showed that AI training improves radiology residents' AI knowledge and diagnostic performance[28], suggesting that AI education is beneficial to radiology residency curricula. Overall, these findings make a compelling case that while AI education is limited, its integration into radiology residency curriculums can positively influence their training. Such a curriculum should not only address the foundations of AI, but also prepare residents to critically assess and mitigate challenges associated with its integration into radiology practice.

As AI becomes increasingly integrated into radiology, radiologists will be expected not only to use these technologies, but also to recognize their limitations and apply them responsibly to clinical practice. While AI shows significant potential in the field, it also introduces a wide range of risks, including algorithmic bias, automation bias, and limited AI interpretability. Without sufficient education regarding the risks of AI systems, patient safety may be compromised. Therefore, radiologists cannot function as passive users of AI. Instead, they must be equipped with the knowledge necessary to responsibly integrate AI into radiology practice, underscoring the need for formal AI education within radiology residency programs.

REFERENCES

1. Khalifa M, Albadawy M. AI in diagnostic imaging: Revolutionising accuracy and efficiency. Computer Methods and Programs in Biomedicine Update. 2024 Jan 1;5:100146. doi:10.1016/j.cmpbup.2024.100146
2. Briody H, Hanneman K, Patlas MN. Applications of Artificial Intelligence in Acute Thoracic Imaging. Can Assoc Radiol J. 2025 Aug;76(3):454–65. doi:10.1177/08465371251322705 PubMed PMID: 39973060.
3. Salastekar NV, Maxfield C, Hanna TN, Krupinski EA, Heitkamp D, Grimm LJ. Artificial Intelligence/Machine Learning Education in Radiology: Multi-institutional Survey of Radiology Residents in the United States. Academic Radiology. 2023 Jul 1;30(7):1481–7. doi:10.1016/j.acra.2023.01.005 PubMed PMID: 36710101.
4. Volin J, Assen M van, Bala W, Safdar N, Balthazar P. Artificial Intelligence and Its Effect on Radiology Residency Education: Current Challenges, Opportunities, and Future Directions. Journal of the American College of Radiology. 2025 Nov 1;22(11):1264–70. doi:10.1016/j.jacr.2025.07.004 PubMed PMID: 40659237.
5. Koçak B, Ponsiglione A, Stanzione A, Bluethgen C, Santinha J, Ugga L, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance,

mitigation, challenges, ethics, and prospects. Diagn Interv Radiol. 31(2):75–88. doi:10.4274/dir.2024.242854 PubMed PMID: 38953330; PubMed Central PMCID: PMC11880872.
6. Kitamura FC, Prevedello LM, Colak E, Halabi SS, Lungren MP, Ball RL, et al. Lessons Learned in Building Expertly Annotated Multi-Institution Datasets and Hosting the RSNA AI Challenges. Radiology: Artificial Intelligence. 2024 May;6(3):e230227. doi:10.1148/ryai.230227
7. Yi PH, Bachina P, Bharti B, Garin SP, Kanhere A, Kulkarni P, et al. Pitfalls and Best Practices in Evaluation of AI Algorithmic Biases in Radiology. Radiology. 2025 May 20;315(2):e241674. doi:10.1148/radiol.241674 PubMed PMID: 40392092; PubMed Central PMCID: PMC12127964.
8. Jose O, Stoeckl EM, Miles RC, Mango VL, Reid NJ, Wagner ASB, et al. The Impact of Extreme Neighborhood Socioeconomic Deprivation on Access to American College of Radiology–accredited Advanced Imaging Facilities. Radiology. 2023 May;307(3):e222182. doi:10.1148/radiol.222182
9. Brady AP, Allen B, Chong J, Kotter E, Kottler N, Mongan J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: Practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. J Med Imaging Radiat Oncol. 2024 Feb;68(1):7–26. doi:10.1111/1754-9485.13612 PubMed PMID: 38259140.
10. Puyol-Antón E, Ruijsink B, Mariscal Harana J, Piechnik SK, Neubauer S, Petersen SE, et al. Fairness in Cardiac Magnetic Resonance Imaging: Assessing Sex and Racial Bias in Deep Learning-Based Segmentation. Front Cardiovasc Med. 2022 Apr 7;9. doi:10.3389/fcvm.2022.859310
11. Bachina P, Garin SP, Kulkarni P, Kanhere A, Sulam J, Parekh VS, et al. Coarse Race and Ethnicity Labels Mask Granular Underdiagnosis Disparities in Deep Learning Models for Chest Radiograph Diagnosis. Radiology. 2023 Nov;309(2):e231693. doi:10.1148/radiol.231693
12. Yi PH, Kim TK, Siegel E, Yahyavi-Firouz-Abadi N. Demographic Reporting in Publicly Available Chest Radiograph Data Sets: Opportunities for Mitigating Sex and Racial Disparities in Deep Learning Models. Journal of the American College of Radiology. 2022 Jan 1;19(1, Part B):192–200. doi:10.1016/j.jacr.2021.08.018
13. Khera R, Simon MA, Ross JS. Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support. JAMA. 2023 Dec 19;330(23):2255–7. doi:10.1001/jama.2023.22557
14. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012 Jan 1;19(1):121–7. doi:10.1136/amiajnl-2011-000089
15. Tejani AS, Ng YS, Xi Y, Rayan JC. Understanding and Mitigating Bias in Imaging Artificial Intelligence. RadioGraphics. 2024 May;44(5):e230067. doi:10.1148/rg.230067
16. Stogiannos N, Cuocolo R, D'Antonoli TA, Santos DP dos, Harvey H, Huisman M, et al. Recognising errors in AI implementation in radiology: A narrative review. European Journal of Radiology. 2025 Oct 1;191. doi:10.1016/j.ejrad.2025.112311 PubMed PMID: 40674942.
17. Dratsch T, Chen X, Rezazade Mehrizi M, Kloeckner R, Mähringer-Kunz A, Püsken M, et al. Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology. 2023 May;307(4):e222176. doi:10.1148/radiol.222176
18. Reyes M, Meier R, Pereira S, Silva CA, Dahlweid FM, Tengg-Kobligk H von, et al. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. Radiology: Artificial Intelligence. 2020 May;2(3):e190043. doi:10.1148/ryai.2020190043
19. Lang M, Bernier A, Knoppers BM. Artificial Intelligence in Cardiovascular Imaging: "Unexplainable" Legal and Ethical Challenges? Canadian Journal of Cardiology. 2022 Feb 1;Focus Issue: New Digital Technologies in Cardiology38(2):225–33. doi:10.1016/j.cjca.2021.10.009
20. Marey A, Arjmand P, Alerab ADS, Eslami MJ, Saad AM, Sanchez N, et al. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nucl Med. 2024 Sep 13;55(1):183. doi:10.1186/s43055-024-01356-2
21. Dietrich, N., & Patlas, M. N. (2025). Adversarial AI in radiology: A hidden threat. Canadian Association of Radiologists Journal. Advance online publication. https://doi.org/10.1177/08465371251331437
22. Mirsky Y, Mahler T. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning.
23. Dietrich N, Gong B, Patlas MN. Adversarial artificial intelligence in radiology: Attacks, defenses, and future considerations. Diagnostic and Interventional Imaging. 2025 Nov 1;106(11):375–84. doi:10.1016/j.diii.2025.05.006
24. Venugopal VK, Kumar A, Tan MO, Szarf G. Curriculum check, 2025—equipping radiology residents for AI challenges of tomorrow. Abdom Radiol. 2025 Dec 1;50(12):6251–61. doi:10.1007/s00261-025-05016-5
25. Hu R, Rizwan A, Hu Z, Li T, Chung AD, Kwan BYM. An Artificial Intelligence Training Workshop for Diagnostic Radiology Residents. Radiol Artif Intell. 2023 Feb 8;5(2):e220170. doi:10.1148/ryai.220170 PubMed PMID: 37035436; PubMed Central PMCID: PMC10077074.
26. Kijima S, Tomihara K, Tagawa M. Effect of stress coping ability and working hours on burnout among residents.

BMC Med Educ. 2020 Jul 13;20(1):219. doi:10.1186/s12909-020-02134-0

27. Volin J, Assen M van, Bala W, Safdar N, Balthazar P. Artificial Intelligence and Its Effect on Radiology Residency Education: Current Challenges, Opportunities, and Future Directions. Journal of the American College of Radiology. 2025 Nov 1;22(11):1264–70. doi:10.1016/j.jacr.2025.07.004 PubMed PMID: 40659237.
28. Keshavarz P, Mohammadigoldar Z, Bedayat A, Raman SS, Tai R. Artificial Intelligence Education in Radiology Training: A Systematic Review of Effectiveness, Barriers, and Future Directions. Academic Radiology. 2026 Mar 1;33(3):695–706. doi:10.1016/j.acra.2025.10.049