Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Optimizing AI Onboarding in Radiology: Impact of Training Case Number Using Multiple Sclerosis Lesions on 3D FLAIR, a pilot study

Ludovichetti Riccardo

riccardo.ludovichetti@usz.ch

University Hospital of Zurich

Karolina Anna Pawlus

University Hospital of Zurich

Arti Chhugani

University Hospital of Zurich

Yash Chhugani

University Hospital of Zurich

Nathalie Nierobisch

University Hospital of Zurich

Christian Federau

AI Medical AG

Zsolt Kulcsar

University Hospital of Zurich

Nicolin Hainc

University Hospital of Zurich

Research Article

Keywords: Artificial intelligence, multiple sclerosis, MRI, neuroradiology, human-AI interaction.

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10445073/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. Christian Federau is founder and CEO of AI Medical AG (Zollikon, Switzerland), the developer of the Jazz software used in this study. All other

authors declare no financial or non-financial interests to disclose.

Abstract

Purpose

To evaluate how the number of training cases during onboarding influences radiology clinicians' confidence, efficiency, and reliance on support when using an AI-based multiple sclerosis lesion detection tool.

Methods

Thirteen radiology clinicians were randomly assigned to onboarding with one, two, or three representative MS cases on 3D FLAIR MRI using Jazz (AI Medical AG, Zollikon). Each session included a 10-minute tutorial, guided case demonstrations, and independent review of three standardized test cases. Recorded metrics included onboarding duration, questions asked during training and independent use, and self-reported confidence (5-point Likert scale). Clinical experience and outcomes were correlated using Spearman rank correlation; group differences evaluated with Kruskal-Wallis test.

Results

Mean onboarding duration ranged from 12:04 (Group 1, one case) to 26:38 minutes (Group 3, three cases). Three-case training (Group 3) yielded the best outcomes: fewest training questions (mean 2.0), fewest independent-use questions (mean 3.5), and highest confidence (4.1/5). The one- and two-case groups (Groups 1 and 2) showed higher and less consistent support needs during independent use (mean 5.2 and 7.0 questions; confidence 3.6 and 3.8, respectively), without a clear dose–response between them. Between-group differences were not statistically significant. The 14-minute additional investment in three-case training yielded meaningful gains in autonomy and confidence. Prior clinical experience did not predict outcomes (Spearman $\rho = -0.37$, $p = 0.213$).

Conclusions

Exposure to multiple training cases—not duration alone—is critical for effective AI onboarding. Three-case training improves clinician confidence and autonomy without clinically meaningful time increases.

Introduction

Traditionally, radiologists have relied on their clinical expertise and visual interpretation skills to analyze imaging studies. However, the growing complexity and volume of medical imaging, together with increasing demands for faster and more accurate diagnoses, have highlighted the need for innovative solutions to support radiology staff in routine clinical practice [1–3]. Artificial intelligence (AI) is increasingly being integrated into diagnostic radiology workflows, supporting tasks such as lesion

detection, segmentation, quantification, and case prioritization. These technologies have demonstrated the potential to improve diagnostic accuracy and workflow efficiency, particularly in high-volume clinical environments where time and precision are critical [4, 5]. Among radiology subspecialties, neuroradiology represents a domain in which AI assistance may be especially valuable, given the complexity, heterogeneity, and often subtle presentation of neurological lesions. Despite the high diagnostic performance of modern AI algorithms, successful integration into routine radiology practice remains challenging, with insufficient consideration of human–computer interaction representing a major barrier to adoption. Clinicians may be reluctant to engage with AI tools when they lack a clear understanding of the system's capabilities, intended role, and added value within existing workflows [6–9]. In addition, algorithm aversion and uncertainty regarding appropriate reliance may further limit effective use.

Trust, transparency, and user competence are key determinants of successful AI adoption in radiology [10, 11]. Inadequate training may lead to underutilization of AI tools, whereas excessive or uncritical reliance may introduce risks such as automation bias—the tendency to favor algorithmic outputs over independent clinical judgment, potentially leading to diagnostic errors [12]. Consequently, structured onboarding and training are essential components of safe and effective AI integration. Radiology training programs are increasingly expected to incorporate AI literacy, enabling clinicians to critically evaluate and appropriately integrate AI outputs into diagnostic workflows [10, 13]. However, current onboarding approaches remain heterogeneous, and standardized, evidence-based training strategies are lacking. In real-world settings, adoption of AI tools remains inconsistent, and gaps in training and user preparation remain [14]. These limitations are exacerbated in radiology training hospitals, where residents rotate amongst subspecialties every few months.

One important but underexplored aspect of AI onboarding is the number and diversity of training cases presented to users. From an educational perspective, exposure to multiple representative examples may facilitate schema formation and improve generalization to new clinical scenarios. At the same time, excessively prolonged or complex training may increase cognitive load, reduce efficiency, and not be feasible in daily routine, posing practical barriers to implementation. Identifying an optimal balance between training depth and usability is therefore critical for successful integration into routine clinical workflows. To address this gap, we conducted a simulated prospective study to evaluate how the number of onboarding cases influences radiology clinicians' confidence, efficiency, and reliance on support when using an AI-based lesion detection tool. Our study cohort included clinicians with a wide range of experience levels, reflecting the diversity of real-world users. We hypothesized that exposure to a small but diverse set of representative cases would improve user confidence and autonomy without substantially increasing onboarding time. To our knowledge, this study is among the first to systematically investigate the impact of onboarding case quantity on human–AI interaction outcomes in a clinical radiology setting.

Materials and Methods

Participants

This simulated prospective, single-center pilot study enrolled thirteen radiology clinicians representing a broad range of clinical experience. The term “simulated prospective” refers to the prospective, standardized recording of clinician onboarding interactions using retrospectively acquired imaging data. General radiology experience ranged from less than one month to 108 months, and neuroradiology-specific experience ranged from less than one month to 38 months. The cohort included radiology residents, fellows, and one senior attending physician.

This retrospective study involving human participants was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This retrospective study used anonymized brain MRI examinations acquired during routine clinical care. All included patients had provided either general consent for the use of their coded health-related data for research purposes, or specific consent for the use of their imaging data, and the retrospective use of these anonymized images was approved by the Ethics Committee of the Canton of Zurich (BASEC 2022-00041). Radiology clinicians took part voluntarily in an internal educational and quality-improvement onboarding activity; no patient-identifiable or sensitive personal data were collected from participating clinicians.

AI Tool and Training Protocol

AI Tool and Training Protocol
The onboarding process was built around an AI software (Jazz, AI Medical AG, Zollikon, Switzerland) designed for semi-automated multiple sclerosis lesion detection and follow-up on brain MRI. Jazz allows for human correction of AI-based MS lesion detection and volume measurements, thereby improving output accuracy on a per-patient basis and preventing exclusion of failed AI cases. All imaging data consisted exclusively of three-dimensional fluid-attenuated inversion recovery (3D FLAIR) sequences obtained from patients with a confirmed clinical diagnosis of multiple sclerosis.

To ensure exposure to a representative spectrum of lesion burden, three cases were pre-selected by consensus of the study investigators, reflecting the range of presentations typically encountered in clinical neuroradiology practice. Participants were randomly assigned using simple randomization to one of three groups, differing in the number of training cases completed during onboarding:

- • Group 1: one training case
- • Group 2: two training cases
- • Group 3: three training cases

Group sizes were n = 5 (one-case), n = 4 (two-case), and n = 4 (three-case), totalling 13 participants. Each onboarding session began with a standardized tutorial of approximately 10 minutes, covering software navigation, tool functions, and interpretation of AI-generated outputs. During both the tutorial and

subsequent case demonstrations, participants were encouraged to ask questions and request clarification freely.

After onboarding, participants completed an independent case review phase, during which they analyzed three additional MS cases using the AI tool without direct instruction. The independent review cases demonstrated different lesion burdens, with 25, 16, and 30 lesions and lesion volumes ranging from 2.46 to 11.68 mL, reflecting a clinically relevant spectrum of disease severity. The AI tool allows for human correction of erroneous segmentation of white matter lesions. This phase allowed assessment of autonomous tool use and any residual need for support following training.

The following metrics were recorded for each participant:

- • Total onboarding duration, encompassing the tutorial and case demonstration phases
- • Number and type of questions asked during onboarding (clarification questions) and during the independent review phase (follow-up questions), categorized as: technical inquiries, lesion interpretation questions, and software usage questions
- • Self-reported confidence in using the AI tool, assessed immediately after onboarding using a 5-point Likert scale (1 = not confident, 5 = very confident), with half-point ratings permitted
- • Volumetric lesion measurements generated by the AI tool and corrected by the human reader for each training case, used to assess consistency and accuracy of AI-assisted quantification across participants

Data Collection and Analysis

Participant-level data — including overall and neuroradiology-specific experience, onboarding duration, question frequency and type, self-reported confidence scores, and AI-generated lesion volume measurements — were compiled and analyzed. These data were used to evaluate the influence of training case quantity on key human–AI interaction outcomes: user confidence, onboarding efficiency, and reliance on support. To evaluate whether prior clinical experience independently influenced onboarding outcomes, Spearman rank correlation coefficients were calculated between total clinical experience (combined radiology and neuroradiology months) and the number of clarification questions asked during independent use, as well as self-reported confidence ratings. Spearman correlation was performed given the small sample size and non-normal distribution of experience values. Differences in questions asked and confidence ratings across training groups were assessed using the Kruskal-Wallis test. A p-value < 0.05 was considered statistically significant. All statistical analyses were performed using Python (SciPy library). As this study describes a retrospective, observational pilot evaluation of an internal training intervention rather than a randomized clinical trial or the development or validation of a diagnostic prediction model, it does not fall within the scope of an existing trial- or model-specific reporting checklist; it is reported following general principles of transparency for observational

comparative studies, including clear description of participants, the training intervention, outcome measures, and statistical methods.

Results

Training Efficiency

Mean onboarding duration increased with the number of assigned cases: Group 1 (one case) 12:04 minutes, Group 2 (two cases) 18:20 minutes, and Group 3 (three cases) 26:38 minutes (Table 1). The absolute difference between the least and most intensive onboarding conditions was approximately 14 minutes — a modest investment relative to the breadth of case complexity covered. Importantly, although Group 3 required the longest onboarding time, this group also demonstrated the lowest mean question burden during the subsequent independent review of three standardized cases, suggesting that additional upfront training was associated with more efficient autonomous tool use thereafter. Conversely, Group 1 participants, despite completing onboarding in the shortest time, required the greatest number of follow-up questions during independent use, indicating that brevity of training did not translate into greater overall efficiency when the full onboarding-to-independent-use timeline is considered.

Interaction Patterns

Clear differences in question frequency were observed across groups. During training, question frequency did not follow a simple dose–response: Group 3 (three cases) asked the fewest questions (mean 2.0, range 0–5), reflecting a low and relatively homogeneous support requirement, whereas the one- and two-case groups asked more (Group 1, mean 4.6, range 0–12; Group 2, mean 5.2, range 2–10). During the subsequent independent review of three standardized cases, Group 3 continued to demonstrate the lowest support requirement (mean 3.5 questions, range 3–4). Groups 1 and 2 showed higher question burdens (means of 5.2 and 7.0, respectively), with considerable individual variability in Group 2 (range 1–17), indicating that two-case onboarding produced inconsistent levels of preparedness across users. Notably, prior radiology or neuroradiology experience did not consistently predict lower question frequency or greater autonomy. No significant correlation was found between total clinical experience and the number of clarification questions during independent use (Spearman $\rho = -0.37$, $p = 0.213$) (Fig. 1), suggesting that accumulated clinical experience does not substitute for structured case exposure during AI tool onboarding. Similarly, no significant differences were observed across groups in questions asked (Kruskal–Wallis $H = 2.56$, $p = 0.278$) or confidence ratings ($H = 0.69$, $p = 0.707$), consistent with the pilot nature of this study.

Self-Reported Confidence

Post-onboarding confidence increased with the number of training cases: Group 1 (one case) reported the lowest mean confidence (3.6 out of 5), followed by Group 2 (3.8) and Group 3 (three cases; 4.1)

(Table 1, Fig. 2). Within-group variability was most pronounced in Group 1, where scores ranged more widely, reflecting heterogeneous preparedness following single-case onboarding. Group 3 participants showed a more consistent confidence profile, with ratings clustering toward the upper end of the scale across users of varying experience levels. As with question frequency, confidence did not show a straightforward relationship with prior clinical experience, further supporting the role of case diversity in shaping user readiness independently of prior expertise.

Lesion Volume Coverage

The lesion volume ranges encountered during independent use with human correction were broadly comparable across groups: 2.46–10.25 mL in Group 1, 2.40–11.68 mL in Group 2, and 2.46–10.70 mL in Group 3. The similarity in volume ranges across groups — despite differences in case number — reflects the pre-selection strategy aimed at ensuring representative lesion complexity even for participants assigned to fewer training cases, and limits the extent to which outcome differences can be attributed to lesion volume variability alone.

Table 1
Summary of Training Metrics by Group Training metrics by group, stratified by number of cases reviewed during onboarding. Training time in mm:ss; questions and confidence ratings (1–5 scale) shown as mean values.

Group	Number of Cases	Number of participants	Mean Training Time (mm:ss)	Mean Questions During Training	Mean Questions During Independent Use	Mean Confidence Rating (1–5)
Group 1	1	5	12:04	4.6	5.2	3.6
Group 2	2	4	18:20	5.2	7.0	3.8
Group 3	3	4	26:38	2.0	3.5	4.1

Discussion

In this pilot study, we demonstrate that onboarding with three representative MS training cases was associated with greater clinician confidence, reduced reliance on support, and improved efficiency during independent use of an AI-based lesion detection tool, compared to onboarding with two or one case. Importantly, these benefits were achieved without a clinically relevant increase in overall training time, suggesting that targeted adjustments to onboarding design may yield meaningful improvements in human–AI interaction outcomes.

One possible explanation for these findings, consistent with Cognitive Load Theory, is that learning is optimized when instructional design balances intrinsic task complexity with the learner's cognitive

capacity [15]. Exposure to multiple varied examples is thought to facilitate schema construction and support transfer of knowledge to new clinical situations [16]. We propose that participants who reviewed three cases spanning a range of lesion burdens and sizes may have developed more robust mental models of how the AI system behaves across different clinical scenarios. This could account for the reduced number of clarification questions and improved efficiency observed during independent use. Conversely, limited exposure to a single case may have resulted in incomplete schema formation, leading to greater uncertainty and increased reliance on support during independent review.

Beyond cognitive learning mechanisms, our findings are consistent with the broader literature on human–AI collaboration, particularly the concept of calibrated trust — the dynamic balance between appropriate reliance on and skepticism toward AI outputs [17]. Insufficient exposure to AI behavior may promote under-trust and underutilization, whereas uncritical exposure without adequate contextual understanding may increase the risk of automation bias [18]. By providing a small but diverse set of representative cases, three-case onboarding may have enabled participants to develop a more accurate appreciation of the AI system's behavior, thereby supporting more appropriately calibrated reliance. However, direct assessment of trust calibration was beyond the scope of this study, and this interpretation should be regarded as a hypothesis requiring prospective validation.

Trust in AI is particularly relevant in radiology, where studies have shown that clinicians may either disregard AI outputs or over-rely on them depending on their level of familiarity and confidence with the system [19]. Prior work has emphasized that structured training and transparency are essential to mitigate automation bias and improve decision-making accuracy in AI-assisted diagnostics [20]. Our findings contribute to this literature by suggesting that even brief, structured onboarding interventions may meaningfully influence interaction patterns and user confidence.

From an implementation perspective, our results have practical implications for the integration of AI tools into clinical radiology workflows. While increased emphasis has been placed on validating algorithmic performance, comparatively less attention has been paid to human factors, including training design, usability, and workflow integration [21–23]. Our findings suggest that onboarding protocols should prioritize case diversity over training duration, as exposure to multiple representative scenarios was associated with improved downstream efficiency and reduced need for ongoing support. If confirmed in larger studies, this principle could translate into reduced institutional training burden and faster, more consistent adoption of AI tools in routine practice.

Notably, prior clinical experience did not consistently predict confidence, question frequency, or interaction patterns in our cohort. Several participants with greater overall radiology experience assigned to single-case onboarding required more support during independent use than less experienced peers who had completed three-case onboarding. This finding aligns with prior literature suggesting that familiarity with a specific AI system, rather than general clinical expertise, is a key determinant of effective human–AI interaction [19, 24], and further underscores the importance of structured onboarding even for experienced clinicians.

Limitations and Future Directions

Several limitations of this study should be acknowledged. First, the sample size was small (participants = 13), limiting statistical power and the generalizability of findings. Second, this was a single-center study conducted with a single AI tool in a specific clinical context (multiple sclerosis lesion detection on 3D FLAIR MRI), and results may not generalize to other AI applications or imaging modalities. Third, although exploratory inferential tests (Spearman rank correlation and Kruskal–Wallis) were performed, the small sample size left the analyses substantially underpowered; none of the between-group comparisons reached statistical significance, and all findings should therefore be interpreted as hypothesis-generating rather than confirmatory.

Fourth, confidence was assessed via self-report on a Likert scale, which may not fully capture objective competence or actual reliance behavior. Fifth, the absence of a control group with no structured onboarding limits causal inference regarding the absolute benefit of the onboarding protocol itself. Future studies with larger, multicenter cohorts, standardized outcome measures, and formal statistical frameworks are needed to confirm and extend these findings.

Conclusion

Onboarding radiology clinicians with three representative multiple sclerosis training cases on 3D FLAIR MRI was associated with greater user confidence, reduced reliance on support, and improved post-training autonomy compared with limited case exposure, without a proportional increase in onboarding time. These findings suggest that case diversity, rather than training duration alone, is a critical design parameter for effective AI onboarding protocols. As AI-based lesion detection tools become increasingly integrated into routine neuroradiology practice, structured and representative onboarding may represent a practical, low-cost strategy to improve human–AI interaction outcomes and support safer, more consistent clinical adoption.

Declarations

Ethics approval and consent to participate

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This study was approved by the Cantonal Ethics Committee Zurich (Kantonale Ethikkommission Zürich; BASEC-Nr. 2022-00041, approved 05 April 2022), covering retrospective analysis of neuroradiology MRI data. All patients had provided either general consent for the use of their coded health-related data for research purposes, or specific consent for the use of their imaging data, and this study falls within the scope of that consent. The clinician onboarding component was an internal educational and quality-improvement activity; radiology clinicians took part voluntarily, and no patient-identifiable or sensitive personal data were collected from participating clinicians.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form.

Competing Interests

Christian Federau is founder and CEO of AI Medical AG (Zollikon, Switzerland), the developer of the Jazz software used in this study. All other authors declare no financial or non-financial interests to disclose.

Funding

The authors did not receive support from any organization for the submitted work.

Author Contribution

R.L.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft K.P.: Investigation, Data curation, Writing – review & editingA.C.: Investigation, Data curation, Writing – review & editing Y.C: Investigation, Data curation, Writing – review & editing N.N.: Supervision, Validation, Writing – review & editing C.F.: Conceptualization, Resources, Software, Supervision, Writing – review & editing Z.K.: Conceptualization, Supervision, Writing – review & editing N.H.: Conceptualization, Supervision, Validation, Writing – review & editingAll authors have reviewed and approved the final version of this manuscript and agree with its submission to BMC Medical Education.

Acknowledgement

We acknowledge the radiology clinicians who participated in this study.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to institutional privacy restrictions and patient confidentiality requirements, but de-identified data are available from the corresponding author upon reasonable request.

References

1. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 2018;2:35. https://doi.org/10.1186/s41747-018-0061-6.

2. Pinto-Coelho L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioeng (Basel). 2023;10:1435. https://doi.org/10.3390/bioengineering10121435.
3. Pierre K, Haneberg AG, Kwak S, Peters KR, Hochhegger B, Sananmuang T, et al. Applications of Artificial Intelligence in the Radiology Roundtrip: Process Streamlining, Workflow Optimization, and Beyond. Semin Roentgenol. 2023;58:158–69. https://doi.org/10.1053/j.ro.2023.02.003.
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56. https://doi.org/10.1038/s41591-018-0300-7.
5. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18:500–10. https://doi.org/10.1038/s41568-018-0016-5.
6. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis. JMIR Med Inf. 2018;6:e24. https://doi.org/10.2196/medinform.8912.
7. Kohli A, Jha S. Why CAD Failed in Mammography. J Am Coll Radiol. 2018;15:535–7. https://doi.org/10.1016/j.jacr.2017.12.029.
8. Musen MA, Middleton B, Greenes RA. Clinical Decision-Support Systems. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Cham: Springer; 2021. pp. 795–840. https://doi.org/10.1007/978-3-030-58721-5_24.
9. Yang Q, Steinfeld A, Zimmerman J, Unremarkable AI. Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems [Internet]. 2019; https://api.semanticscholar.org/CorpusID:127989976
10. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol. 2020;30:5525–32. https://doi.org/10.1007/s00330-020-06946-y.
11. Verwey J, Zwart B, IJzerman M, Visser JJ, Sülz S. Factors influencing AI acceptance in radiology: a systematic review across the radiology workflow. Eur Radiol. 2026;36:3687–96. https://doi.org/10.1007/s00330-025-12129-4.
12. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med. 2021;4:31. https://doi.org/10.1038/s41746-021-00385-9.
13. van Kooten MJ, Tan CO, Hofmeijer EIS, van Ooijen PMA, Noordzij W, Lamers MJ, et al. A framework to integrate artificial intelligence training into radiology residency programs: preparing the future radiologist. Insights Imaging. 2024;15:15. https://doi.org/10.1186/s13244-023-01595-3.
14. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol. 2021;31:3797–804. https://doi.org/10.1007/s00330-021-07892-z.
15. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. Perspect Med Educ. 2015;4:119–27. https://doi.org/10.1007/s40037-015-0192-x.

16. Paas F, Renkl A, Sweller J. Cognitive Load Theory and Instructional Design: Recent Developments. Educ Psychol. 2003;38:1–4. https://doi.org/10.1207/S15326985EP3801_1.
17. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors. 2004;46:50–80. https://doi.org/10.1518/hfes.46.1.50_30392.
18. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inf Assoc. 2017;24:423–31. https://doi.org/10.1093/jamia/ocw105.
19. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med. 2024;30:837–49. https://doi.org/10.1038/s41591-024-02850-w.
20. Buçinca Z, Malaya MB, Gajos KZ. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. Proc ACM Hum-Comput Interact. 2021;5:188. https://doi.org/10.1145/3449287.
21. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. Insights Imaging. 2019;10:44. https://doi.org/10.1186/s13244-019-0738-2.
22. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25:30–6. https://doi.org/10.1038/s41591-018-0307-0.
23. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195. https://doi.org/10.1186/s12916-019-1426-2.
24. Kocak B, Cuocolo R. Human-AI interaction and collaboration in radiology: from conceptual frameworks to responsible implementation. Diagn Interv Radiol. 2026. https://doi.org/10.4274/dir.2026.263780.

Figures

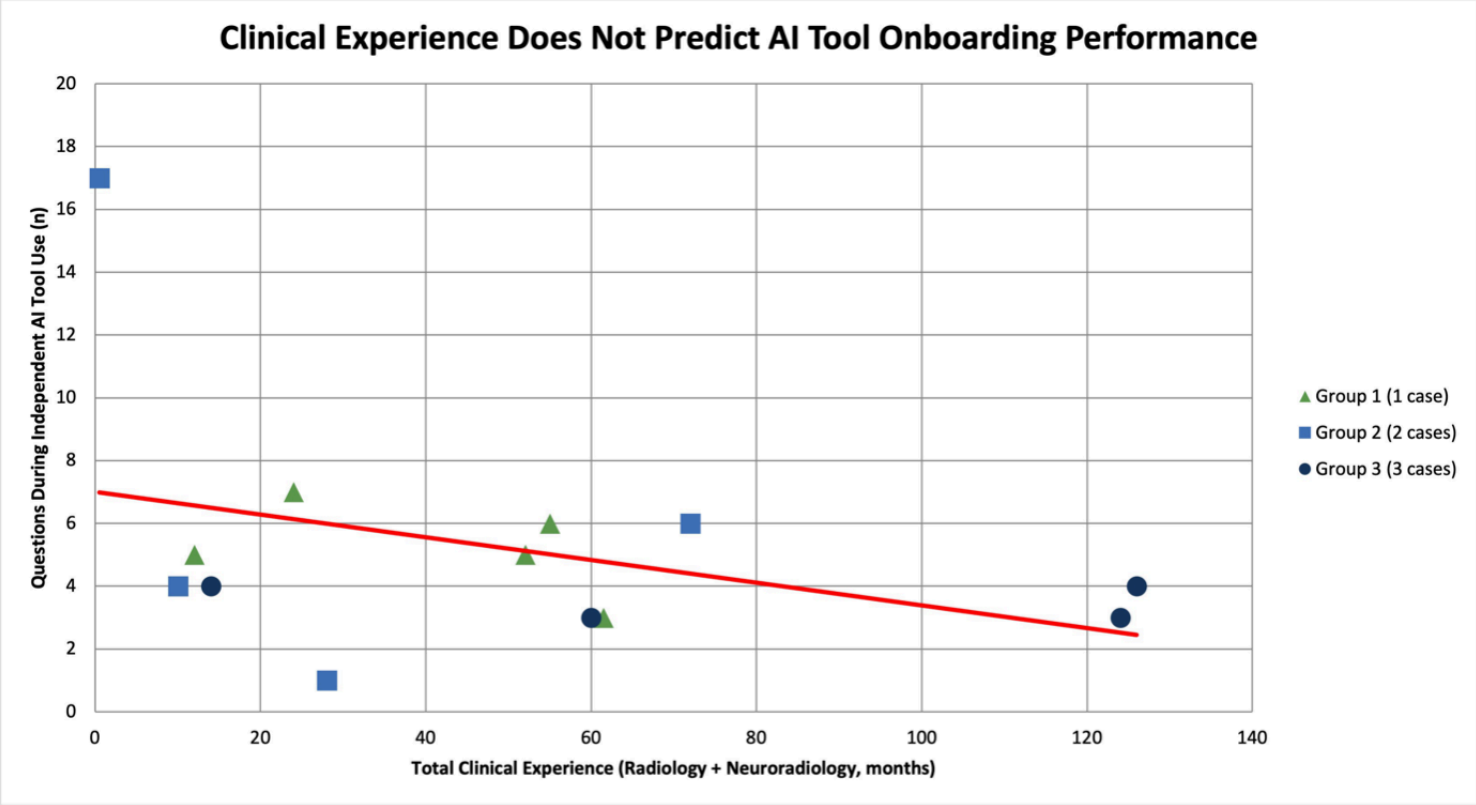


Figure 1

Scatter plot of total clinical experience in radiology and neuroradiology versus questions asked during independent AI tool use. Each point represents one participant. The red line shows the linear trend across all participants.

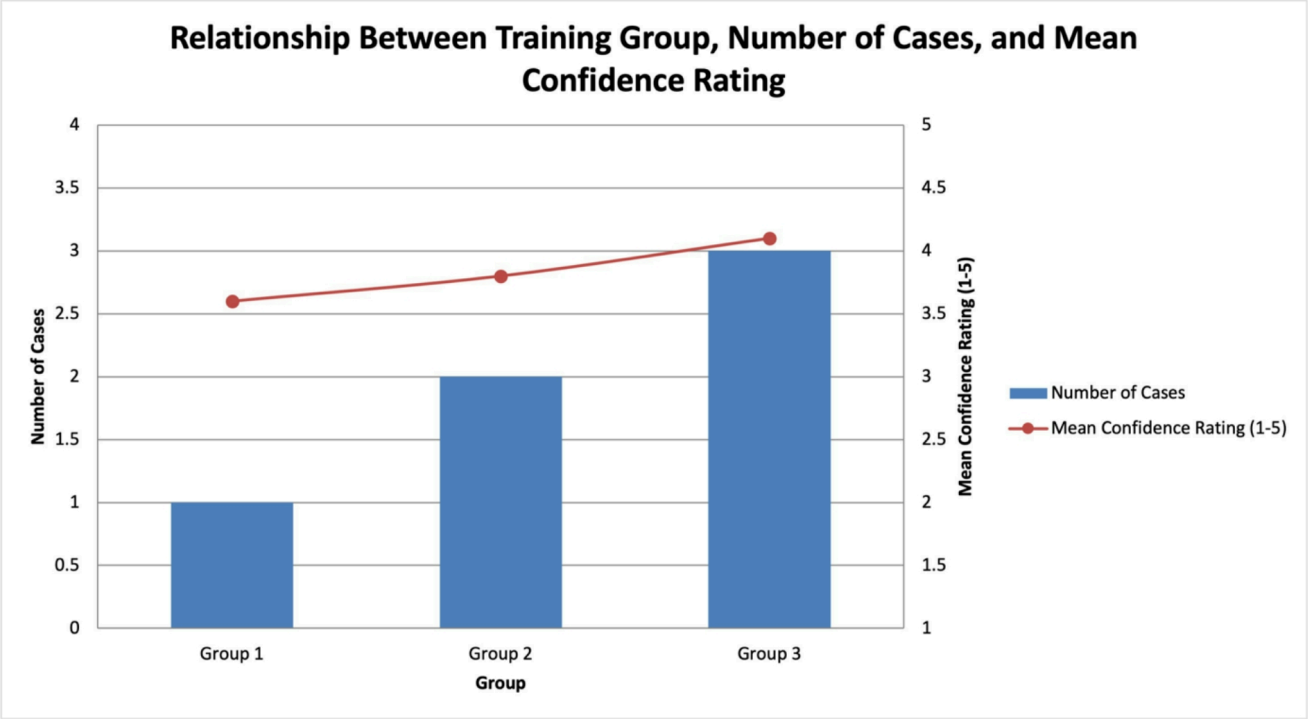


Figure 2

Mean self-reported confidence rating by training group. Confidence assessed on a 5-point Likert scale (1 = not confident, 5 = very confident), half-point ratings permitted. Error bars represent standard deviation. Group 1 (one case, n=5), Group 2 (two cases, n=4), Group 3 (three cases, n=4).