Cover Letter AI Reliance in Diagnostic Radiology Among Intern Doctors in Palestine: A Triple-Arm, Triple-Blind, Parallel-Design Randomized Controlled

Trial NCT Number: Not Yet Assigned

Date: 8/4/2026

This study is carried out as a triple arm trial investigating AI reliance among intern doctors in Palestine. The study will involve the use of sham AI prompts and an online exam with standardized questions and response options.

AI Reliance in Diagnostic Radiology Among Intern Doctors in Palestine: A Triple-Arm, Triple-Blind, Parallel-Design Randomized Controlled Trial

Authors

Peter Bael, Medical Research Club, Faculty of Medicine, Alquds University, Jerusalem, Palestine
peter.bael@students.alquds.edu

Iman Abu Nawa, Medical Research Club, Faculty of Medicine, Alquds University, Jerusalem, Palestine
iman.nawa@students.alquds.edu

Alaa Adawi, Medical Research Club, Faculty of Medicine, Alquds University, Jerusalem, Palestine
alaa3dawi2001@gmail.com

Tamara Lahham, Medical Research Club, Faculty of Medicine, Alquds University, Jerusalem, Palestine
lahhamtamara@gmail.com

Zakaria Abu Ghattas, Medical Research Club, Faculty of Medicine, Alquds University, Jerusalem, Palestine
Zakaria.ghattas@students.alquds.edu

Introduction

The process of patient diagnosis is essential in medical practice, guiding clinical decisions and determining treatment outcomes. Accurate and timely diagnoses reduce the risks of delays or errors and ensure effective therapy. Nonetheless, traditional diagnostic methods encounter numerous obstacles: the rising patient numbers and the limited availability of doctors typically extend the time needed for diagnosis. Moreover, reliance on subjective interpretation may lead to variations in diagnostic precision. (1)

These issues highlight the urgent need for advancements that enhance the speed and accuracy of medical diagnostics, such as Artificial Intelligence (AI), a branch of computer science focused on creating algorithms that can execute tasks usually dependent on human intelligence, including pattern recognition, decision-making, and problem-solving (2). AI represents a groundbreaking technology, particularly evident during the coronavirus disease 2019 (COVID-19) pandemic (1) , which highlights the ability of analytical processes to identify patterns in large datasets. This became increasingly apparent during the pandemic, helped reduce shortcomings of traditional diagnostic techniques, and hastened the implementation of AI technologies.

These innovations offer promise for enhancing diagnostic precision and efficiency. The domain of AI has revolutionized medical imaging, particularly in radiology, which has been among its earliest and most proactive users. Deep learning algorithms can now identify patterns in medical images with impressive speed and accuracy. Recent research across various subspecialties, such as emergency radiology, neuroradiology, and breast imaging, highlights how specialized AI systems can tackle the distinct diagnostic and workflow challenges in each area. These findings suggest that AI has moved beyond being a future goal; it is now in active use. (3)

Although it is crucial nowadays to integrate AI into planning the healthcare workforce to promote collaboration between technology and clinicians, ultimately enhancing patient care. The constant use of AI techniques carries risks such as distraction, excessive reliance on technology, and biased decision-making. Although explainable AI improves transparency, it might still disproportionately sway decisions. A prolonged dependence on AI, particularly among trainee clinicians or those handling low-volume cases, can jeopardize their skill development. (4)

Literature Review

<u>Artificial Intelligence (AI) in Radiology</u>

AI has become an essential tool in everyday life, and no one can deny its impact in the health sector. Lately, AI performance has become more “competitive” with specialists in diagnosing diabetic retinopathy on fundoscopic exams, detecting abnormalities in a mammogram, and reading cardiothoracic imaging, CT angiographies, and coronary CT. (5,6)

Substantial advances have been made; AI machines can now more accurately represent and interpret complex data, tasks that until a few years ago were done by us humans. Deep learning refers to machine learning techniques that leverage neural network structures inspired by the human brain. Earlier, AI methods were inferior to humans; they were not able to fully comprehend or match our brains, but recently, AI machines, especially deep learning, have dominated the human brain and surpassed humans in task-specific applications. (7)

In radiology, trained doctors visually assess medical images and describe findings to detect, characterize, and evaluate diseases. Their judgment is based on the level of education, experience, and subjectivity. In contrast, AI recognizes patterns, not humans; it provides a quantitative assessment in an automated fashion. The physician's role is to know how to use it and integrate it into their workflow. (7)

Interacting with patients prior to, during, and following their imaging procedure is a crucial part of a radiologist's job. Informing the patient about the procedure and verifying identification and indications for the requested examination are all part of the interaction. AI may help with double-checking clinical indications, assessing the appropriate imaging modality and techniques to be used, and verifying patient identification records through an electronic health record. However, direct human communication between a patient and a health professional is unlikely to be replaced by AI technologies. (8)

Taking the AI's side for a change. In thoracic imaging, lung cancer is one of the most common and deadly tumors. Pulmonary nodules can be found with lung cancer screening, and for many patients, early detection can save their lives. These nodules can be automatically recognized and classified as benign or malignant with the help of AI. In colonoscopy, colonic polyps that are undetected or misclassified carry a risk of colorectal cancer. Although most are benign, they can become malignant over time. Hence, early detection and consistent monitoring with trustworthy AI-based tools are critical. (7)

In breast cancer screening, mammography is often involved. It is easy to operate, yields high-resolution images, and is relatively pain-free. Mammograms are an accurate way to identify benign lesions, malignant breast tumors, and breast masses that are not detectable by physical examination. Both machine learning and deep learning techniques have been recently implemented to help with detecting breast cancer, which signifies the importance of developing the recognition techniques used. The development is still lagging in this domain due to the lack and imbalance of databases, due to privacy concerns, ethical issues, conservative mindsets regarding this new technology, and the sensitivity of the topic to patients themselves. Radiologists are well capable of analyzing breast images. However, fatigue brought on by their heavy workload and long hours may result in poor judgment, missed diagnoses, or misdiagnoses. Potential human errors can be reduced or even eliminated by using AI. (9)

But what about clinical judgment? Does it matter, or is it just a fake deception to convince ourselves that we are irreplaceable? What if the machine is wrong? Who's to blame? The radiologist? The machine? The software developer?

AI will probably always be in our lives, and after all, radiologists have earned their place, in most cases, so the role of AI must be to augment, not replace. We should not label AI as an intruder or an imposter; we should domesticate it and make use of it as a tool to save time as an assistant, not as a superior, because after all, we are irreplaceable, no matter how powerful it gets. Not letting it get the best of our clinical judgment just because it has "intelligence" in its name.

Further Insight into AI and Medicine

The application of artificial intelligence (AI) in medicine, especially radiology, has been developing quickly. AI has also been the driving force behind significant innovation, an important topic in radiology societies, and groundbreaking studies in recent years.The core value of an AI solution designed for radiology lies in its eventual role in patient care, even though early interest in the application of AI in radiology seemed to be primarily motivated by technological advancements in machine learning and deep learning algorithms (10). A good radiology AI study optimizes the technical potential of AI techniques within their technical limitations by providing an answer to a well constructed imaging-based request based on clinical demands. Define the AI solution's role in the diagnostic pathway and examine its performance in accordance with that role (in the intended role and compared with suitable standard of care) for studies that look at the accuracy of the diagnosis. The authors should, for instance, compare the performance of an AI solution with that of a radiologist if it is meant to serve as a replacement test; if it is meant to serve as a triage or add-on function, they should compare the performance of radiologists alone with that of an AI/radiologists combination. Furthermore, better clinical judgment or better patient outcomes are not always the result of superior technological efficacy or diagnostic accuracy alone. Investigating these therapeutically focused objectives beyond diagnostic precision might help illustrate the usefulness of radiology AI methods and promote their broad use. (11)

Although risks and problems with quality control, artificial intelligence (AI) presents enormous potential to transform the offer of radiological services. Besides being a helpful training tool for radiologist trainees, it is also possible that AI might turn into a dependable, diligent ally of the radiologist rather than a foe.

It is challenging to forecast if AI will eventually outperform humans or where AI application in radiology will go in the future. Nonetheless, according to a recent study of AI specialists by Grace et al., AI will surpass humans in a number of activities over the course of the next ten years, including language translation (by 2024), truck driving (by 2027), and surgery (by 2053). According to the experts who took part in the poll, there is a 50% possibility that artificial intelligence would surpass human abilities in 45 years (12). Regarding whether AI will benefit radiologists or hinder them, it is also hard to say if it will ever grow to the point where it can outperform humans. According to some scary predictions, that day will come soon, and AI will replace entire fields like radiology and dermatology. However, the actual results of using AI in medicine are far less certain. (13)

SO Clinical doctors' and radiologists' perceptions of AI range from complete acceptance and enormous excitement to significant doubt, anxiety, fear of job losses, and apprehension.

Aims & Objectives

Aim

We aim to measure intern doctor reliance on AI in diagnostic radiology among intern doctors in Palestine. For the sake of our study, this will be done by measuring specific test scores for equivalent intern doctors, and using the mean of decrement with “incorrect AI vs control” and increment with “correct AI vs control”.

Objectives

1. Characterize the risk factors of reliance on AI in radiology
2. Measure the perception of intern doctors on AI use in radiology

Methodology

Study Design

This study is a triple arm, triple blinded, parallel design randomized controlled trial. The study will measure how much intern doctors rely on AI assistance in radiologic interpretation and the behavioral impact of correct versus incorrect AI guidance. All interns will undergo a radiology exam with identical questions, and have their results compared across groups. This study will apply for ethical approval, and subsequently register on Clinicaltrials.Gov.

Study Protocol

The study will be split into 2 parts: recruitment and testing. Subjects will be recruited for an initial period of 2 weeks, based on our sampling protocol. Then, a researcher will encode the participants into a number list. The number list will be given to a non-researcher staff volunteer, who will input the list into a computer-based randomizer at a 1:1:1 ratio. Participants were recruited prior to data collection and, following random group allocation, were provided with a link to the online assessment along with standardized instructions, after which they completed the test according to their assigned condition.

The three test versions will be on identical online forms, though they will all be labeled identically as well. Upon administration of the test, the second volunteer will also be responsible for administering the test, by providing the participants with the link for their exam. The subjects will be blinded to the different exam types of the experiment until after the end of the exam, where they will be debriefed, and informed about the nature of the experiment and its findings. They will be given the option to opt out again.

Allocation concealment will be ensured through centralized automated randomization performed using a computer-generated sequence created by an independent volunteer who is not involved in participant recruitment, data collection, or outcome assessment. Participants will be randomly assigned in a 1:1:1 ratio to either Group A or Group B or Group C using block randomization with variable block sizes to maintain allocation balance while preventing prediction of upcoming assignments. The randomization sequence will be implemented through a secure digital allocation system that releases only the participant's group identifier after enrollment. Study investigators will not have access to the underlying allocation sequence. This study will employ triple blinding. Intern doctors will be blinded to the intervention condition and will only be aware that they are interacting with a diagnostic support system without knowledge of the study hypothesis or group allocation. Second, investigators responsible for administering the assessment and collecting responses will be blinded to group assignments and will only see anonymized participant identifiers linked to response submissions. Third, the data analysts will receive a de-identified dataset in which study groups are coded as neutral labels (e.g., Group X and Group Y) until all primary analyses are completed. All exams will be designed to appear visually identical to prevent participants from identifying their group allocation based on interface characteristics. All guidance will be delivered through preconfigured automated responses to ensure consistency across participants and eliminate investigator influence during the intervention. To address potential inadvertent unblinding, investigators interacting with participants will not have access to the AI output interface. If a participant reports suspected knowledge of their allocation, the event will be documented and the participant will remain in the analysis under the intention-to-treat principle. Any confirmed unblinding events will be reported transparently in the final study manuscript.

The test will run for 30 minutes, with 20 questions, all of which will be spot diagnoses (None of the questions are straightforward, with moderate difficulty, maintaining that the AI will have a convincing prompt, as to maintain the blinded nature of the study, and without rendering any of the answers obvious). The images will all be taken from the NIH database, and the test will be reviewed by 2 senior radiologists for validation. Additionally, the participants will undergo a separate data collection form for further data collection upon recruitment.

The groups will all have identical tests, with one arm having no AI assistance, one arm having AI assistance that is correct for all questions, and one arm having AI assistance that is incorrect for all questions. The test, in its 3 formats, is available as an attachment to this proposal. Test questions will use images uploaded from Radiopedia and reviewed by two senior radiologists to ensure accuracy. Images were only included if the diagnosis was at least near certain. The validation process included 2 visits to the radiologists 1 week apart, where they completed the questions blinded to the answers, and then reviewed the questions, the prompts and the data collection tool for quality, accuracy and applicability among our population. We will further run a pilot study with 9 students to assess the study for reliability and validity following ethical approval. Not that answers will either be scored as correct or incorrect.

The AI assistance has been described in the test bank, and are all in fact sham statements written by the authors, and validated by 2 senior radiologists. They include correct AI: which are statements that are correctly describing the image, and incorrect AI: which are statements that are incorrectly describing the image.

The exact tests are available within the attached test bank file. The first version will include a question, followed by an imaging modality picture, and multiple choice question, of which one is correct, and three are incorrect. Answers, and time-to-answer will be collected. Following this, we will collect self-reported confidence with and without AI support. The second version will have an identical exam, with the AI sham prompts provided below the image, but the prompt will be the correct prompt. The third version will replace the correct prompts with the incorrect prompts. These include explanations where applicable to be convincing whether correct or incorrect.

Population & Sampling

Our population includes all intern doctors in Palestine. Specifically, we'll hold the study in 3 centers; The Bethlehem Arab Society for Rehabilitation, the Caritas Baby Hospital, and the Red Crescent Hospital in Hebron. We will have an initial recruitment period where we will allocate the participants into groups as well. The sample size calculated for 3 groups, at an alpha of 0.05, a power of 0.80, and an effect size of 0.25, yields a total sample size of 159, with 53 participants per group. We will be using this effect size to detect small, large and medium effect sizes, considering that there is no specific data predating this study. There is no need to account for dropouts, as there is no follow-up for the study, and drop-outs can be immediately replaced with new participants. We will nonetheless report dropouts within the final study report.

Our inclusion criteria is any intern doctor at any of the 3 centers who is at least 3 months into their 1-year internship. They have to provide informed consent and have previous exposure to radiologic interpretation during training. Any physician who has already completed their internship, has knowledge of the study, or hasn't completed the initial 3 months of their internship will be excluded. Recruitment will be done via the internship coordinator of each center. Once the study goal sample is achieved, we will allocate participants into different groups.

Data Collection

Beyond our outcome data, we will collect demographic data including age, gender, region, hospital, working hours, propensity to read imaging independently, and data on interest in radiology in general and as a specialty path, and perception of AI use in radiology based on the tool used by Chen et al. “Radiology Residents’ Perceptions of Artificial Intelligence: Nationwide Cross-Sectional Survey Study.", and the “Assessing diagnostic radiology knowledge among Syrian medical undergraduates” scale for interest in radiology. Both these scales are already published and validated studies that we are integrating into our AI reliance study [9,10, 14]. Our study does not aim to characterize knowledge with a questionnaire as that information is better studied in a separate study, alternatively we look into perception and interest as these points are relevant to the literature gap of the role of AI in radiology. These data points will be collected following the administration of the test, and the disclosure of our findings. We will reassess the validity of the tool using Cronbach alpha, but we will administer the test and tool in English, considering all participants are physicians and are used to being tested in English.

Data Analysis

The study will be analyzed using R version 4.5.2. We will clean the data, correct for outliers, impute missing variables if they constitute less than 10% of the data, and otherwise delete it. Statistical significance will be defined as a two-sided p-value < 0.05. Univariate analysis will include means, medians, and modes, with adjunct standard deviations and interquartile ranges for continuous variables, as well as normality testing with the Shapiro Wilk test, and counts and percentages for categorical variables. We will also test several relationships using bivariate analysis as needed. Finally, we will test reliance on AI based on the differences of scores between the groups and the control arm. We will also use logistic regression to gauge perception of AI use in radiology and linear regression to gauge the extent of reliance. AI reliance, which is our main outcome, will be defined as the mean of decrement with incorrect AI vs control and increment with correct AI vs control. We will also report effect sizes and 95% confidence intervals, using Cohen’s D where applicable.

Baseline characteristics (age, gender, hospital, working hours, radiology exposure, and interest in radiology) will be compared between the three study groups to assess balance after randomization. The primary outcome is AI reliance, defined as the behavioral influence of AI guidance on diagnostic performance.

AI reliance will be operationalized as:

AI Reliance = Mean score improvement in the correct-AI group vs control

+ Mean score decrement in the incorrect-AI group vs control

This measure reflects both positive reliance (trust in correct AI) and negative reliance (misleading AI influence). Then, One-way ANOVA or Kruskal wallis will be used to compare mean test scores among the three arms (Control, Correct-AI, Incorrect-AI). Post-hoc pairwise analysis will then be used to confirm these findings. Multiple comparisons will be adjusted using the Benjamini–Hochberg false discovery rate procedure.

Diagnostic accuracy will be calculated as the proportion of correct responses out of the 20 exam questions. Multivariable regression analyses will be performed to explore predictors of AI reliance. Cronbach alpha testing will be used to confirm reliability in both the main and pilot studies.

Timeline Table

Milestone	**Timestamp**
Ethical Approval	1 Week
Trial Registration	1 Week
Recruitment	1 Week
Testing	1 Week
Analysis	1 Week
Writing	1 Week
Publication	1 Month

Clinical Significance :

This study is a parallel design, randomized controlled trial with three arms and three blinds. The study will assess the behavioral effects of accurate versus inaccurate AI guidance, as well as the extent to which intern physicians rely on AI support for radiologic interpretation. Every intern will take the same radiological test, and the findings will be compared between groups.

The findings are expected to contribute to a better understanding of the dependence of intern physician on AI on there decision making capacity

Ethical Considerations

Ethical approval will be sought from the Al-Quds University Research Ethics Committee. Formal permissions will also be obtained from the participating centers, including Bethlehem Arab Society for Rehabilitation, Caritas Baby Hospital, and The Red Crescent Hospital (Hebron). Participants' complete anonymity and confidentiality will be ensured, and no personal identifiers will be included. Participation will be entirely voluntary, after which informed consent will be obtained. The study poses no harm or risk to participants; on the contrary, it is expected to provide valuable benefits by guiding policy change.

This study is designed to maintain patient autonomy, beneficence, non-maleficence, justice and informed consent. As the study will offer appropriate policy guidance it is guaranteed to provide the population with direct benefit at no risk for harm. Additionally, we will adhere to a strategy of unbiased random enrollment into the study to cover different demographics. In terms of autonomy and informed consent, while our study will initially blind participants to the nature of the study, this will be brief, and will include complete disclosure including the right to withdraw at the end of the study.

Budget

Any incurred budgets will be covered by the Medical Research Club. No cost is expected for the study.

Appendices

Consent Form

Data Collection Sheet

Standardized Tests (3 Copies)

References

1. Jeong J, Kim S, Pan L, Hwang D, Kim D, Choi J, et al. Reducing the workload of medical diagnosis through artificial intelligence: A narrative review. Medicine (United States). 2025 Feb 7;104(6). doi:10.1097/MD.0000000000041470 PubMed PMID: 39928829.

2. Fontenele RC, Jacobs R. Unveiling the power of artificial intelligence for image-based diagnosis and treatment in endodontics: An ally or adversary? Int Endod J. 2025 Feb 1;58(2):155–70. doi:10.1111/iej.14163 PubMed PMID: 39526945.

3. NK RK, R.S. A, K S, K RKN, S AR, K S. Artificial Intelligence in Radiology: Augmentation, Not Replacement. Cureus. 2025 Jun 17;17(6). doi:10.7759/cureus.86247

4. Brunyé TT, Mitroff SR, Elmore JG. Artificial intelligence and computer-aided diagnosis in diagnostic decisions: 5 questions for medical informatics and human-computer interface research. Journal of the American Medical Informatics Association. 2026 Feb 1;33(2):543–50. doi:10.1093/jamia/ocaf123 PubMed PMID: 41101774.

5. Banerjee I, Bhattacharjee K, Burns JL, Trivedi H, Purkayastha S, Seyyed-Kalantari L, et al. “Shortcuts” Causing Bias in Radiology Artificial Intelligence: Causes, Evaluation, and Mitigation. J Am Coll Radiol. 2023 Sep 1;20(9):842. doi:10.1016/j.jacr.2023.06.025 PubMed PMID: 37506964.

6. Aquino GJ, Mastrodicasa D, Alabed S, Abohashem S, Wen L, Gill RR, et al. Radiology: Cardiothoracic Imaging Highlights 2023. Radiol Cardiothorac Imaging. 2024 Apr 1;6(2):e240020. doi:10.1148/ryct.240020 PubMed PMID: 38602468.

7. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018 Aug 1;18(8):500. doi:10.1038/s41568-018-0016-5 PubMed PMID: 29777175.

8. Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. Br J Radiol. 2020 Apr 1;93(1108):20190840. doi:10.1259/bjr.20190840 PubMed PMID: 31821024.

9. Al-Karawi D, Al-Zaidi S, Helael KA, Obeidat N, Mouhsen AM, Ajam T, et al. A Review of Artificial Intelligence in Breast Imaging. Tomography. 2024 May 1;10(5):705. doi:10.3390/tomography10050055 PubMed PMID: 38787015.

10. Nakaura T, Higaki T, Awai K, Ikeda O, Yamashita Y. A primer for understanding radiology articles about machine learning and deep learning. Diagn Interv Imaging. 2020 Dec 1;101(12):765–70. doi:10.1016/j.diii.2020.10.001 PubMed PMID: 33121910.

11. Chassagnon G, Dohan A. Artificial intelligence: from challenges to clinical implementation. Diagn Interv Imaging. 2020 Dec 1;101(12):763–4. doi:10.1016/j.diii.2020.10.007 PubMed PMID: 33187905.

12. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O. When Will AI Exceed Human Performance? Evidence from AI Experts.

13. Chen Y, Wu Z, Wang P, Xie L, Yan M, Jiang M, et al. Radiology Residents’ Perceptions of Artificial Intelligence: Nationwide Cross-Sectional Survey Study. J Med Internet Res. 2023 Jan 1;25(1):e48249. doi:10.2196/48249 PubMed PMID: 37856181.

14. Alchallah MO, Ismail H, Dia T, Shibani M, Alzabibi MA, Mohsen F, Turkmani K, Sawaf B. Assessing diagnostic radiology knowledge among Syrian medical undergraduates. Insights

Imaging. 2020 Nov 23;11(1):124. doi: 10.1186/s13244-020-00937-9. PMID: 33226458; PMCID: PMC7683655.