E-ISSN 2250-0944
ISSN 2250-1150
doi: 10.31838/ijprt/16.02.194

**Research Article**

# Impact of Artificial Intelligence-Assisted Feedback on Clinical Reasoning Skills among Postgraduate Medical Trainees

**Palwasha Zahid[1], Shomaila[2*], Madiha Akhwand[3], Syeda Hanaa Fatima[4], Sadia Awais[5], Saadia Rehman[6]**

[1]Lecturer, Department of Physiology, Peshawar Medical College, Peshawar, Pakistan.
[2*]Assistant Professor, Department of Medical Education & Research, Women Medical College, Abbottabad, Pakistan.
[3]Assistant Professor, Department of Medical Education, Rawal Institute of Health Sciences, Islamabad, Pakistan.
[4]Assistant Professor, Department of Health Professions Education, NUMS (National University of Medical Sciences), Islamabad, Pakistan.
[5]Assistant Professor, Department of Oral Biology, De' Montmorency College of Dentistry, Lahore, Pakistan.
[6]Assistant Professor, Department of Dental Education, Abbottabad International Medical Institute, Abbottabad, Pakistan.

**Corresponding Author:** Shomaila
Assistant Professor, Department of Medical Education & Research, Women Medical College, Abbottabad, Pakistan.
**Email:** drshomaila1@gmail.com
Received: 29.04.2026, Revised: 08.07.2026, Accepted: 17.07.2026

**ABSTRACT**

**Background:** Clinical reasoning is an important competency in postgraduate medical education, but traditional feedback systems can be slow and intermittent due to faculty workload and the time available for supervision. In case-based learning, feedback can be immediate, individualized and structured with the support of systems supported by artificial intelligence.

**Objective:** To determine the impact of artificial intelligence-assisted feedback on clinical reasoning skills among postgraduate medical trainees.

**Methods:** The study was a prospective quasi-experimental comparative study carried out at Rawal Institute of Health Sciences Islamabad, from January 2025 to June 2025. In total, there were 76 postgraduate medical trainees, half of whom received feedback via artificial intelligence (38 trainees) and the other half received it from the faculty (38 trainees). Both groups answered clinical cases and had pre- and post-test clinical reasoning assessments. The main outcome was the difference from the overall clinical reasoning score. Secondary outcomes were diagnostic accuracy, generation of differential diagnosis, selection of investigations, planning for management, diagnostic errors, time to complete the case, confidence and satisfaction. Analysis of the data was done by suitable descriptive and inferential statistical tests, and the p-value of $< 0.05$ was taken as statistically significant.

**Results:** Baseline clinical reasoning scores were comparable between the AI-assisted and conventional-feedback groups ($61.8 \pm 7.4$ versus $62.3 \pm 7.1$; $p = 0.766$). After the intervention, the mean score was significantly higher in the AI-assisted group than in the conventional group ($78.6 \pm 6.8$ versus $70.4 \pm 7.3$; $p < 0.001$). The mean improvement was $16.8 \pm 6.1$ points in the AI group compared with $8.1 \pm 5.7$ points in the conventional group. The AI-assisted group also demonstrated greater diagnostic accuracy, fewer diagnostic errors, shorter case-completion time, and higher confidence and satisfaction scores.

**Conclusion:** Artificial intelligence-assisted feedback was associated with greater improvement in clinical reasoning performance than conventional feedback. It may be used as a supplementary educational strategy under appropriate faculty supervision.

**Keywords:** Artificial Intelligence, Clinical Reasoning, Feedback, Postgraduate Medical Education, Medical Trainees, Formative Assessment.

## INTRODUCTION

Clinical reasoning is the cognitive process through which healthcare professionals collect

and interpret patient information, generate possible diagnoses, select appropriate investigations, and formulate management decisions. It is a key component of clinical practice and one of the important learning outcomes of postgraduate medical education and is an important aspect of safety and effectiveness. Although postgraduate trainees have sufficient theoretical knowledge, they may find it difficult to integrate the clinical findings, prioritize differential diagnoses and apply evidence to complex patient situations. Hence, clinical reasoning needs to be repeated in the context of clinical problems, deliberate practice, reflection and timely feedback. Problems with reasoning can lead to inappropriate investigation or delay of diagnosis, clinical errors in diagnosis and management, and ineffective management of patients [1-3].

Feedback is a significant aspect in enabling trainees to reflect on their clinical work and understand where they have fallen short of a desired standard and what their line of thinking was during the activity. Feedback should be prompt, specific, constructive and related to decisions that the learner has made that can be observed. Typical sources of conventional feedback are clinical supervision, teaching rounds, case presentations, workplace-based assessment or formal education sessions. While faculty feedback is crucial, it can be hindered by a limited number of teaching hours, inconsistent feedback from faculty supervisors, high clinical demands and a delay between the time of the performance and when feedback is provided. These restrictions are especially relevant in postgraduate training situations where trainees need to be guided in a large number of clinical situations. Feedback that uses technology can get people over some of these hurdles because it can provide opportunities for more individualized practice without the need for direct faculty supervision [4-6].

In recent years, innovative technologies have emerged in the field of medical education, notably generative AI and large language models, which have introduced novel avenues for case-based learning. AI can review written clinical reasoning, detect missing data, suggest other potential diagnoses and give instant feedback to individual learners that corresponds to their response. Researchers conducted a randomized controlled trial to determine whether feedback generated by ChatGPT is effective in helping to improve clinical reasoning, with the results showing that ChatGPT feedback provided some benefits, but expert feedback continued to have advantages in certain complex situations [7]. One randomized crossover study found that when the response was provided in a timely manner and the AI system was well-developed, it was feasible to use AI tools to deliver personalized feedback on medical questions and that the educational value was affected by the prompt quality and development of the AI system [8]. The use of AI for generating and feeding back to script concordance tests, testing reasoning under clinical uncertainty, has also been examined [9]. A multi-institutional research effort also showed that large language models could assist in the evaluation of documented clinical reasoning and provide opportunities for feedback [10].

Despite these potential benefits, AI-generated feedback may contain inaccurate, incomplete, or overly confident information. Easy access to suggested diagnoses may also encourage cognitive dependence when trainees accept outputs without critically evaluating them. Research involving physicians has shown that simply providing access to a large language model does not necessarily improve clinical reasoning, indicating that educational design, learner preparation, and human–AI interaction are important determinants of effectiveness [11]. Reviews of AI in medical education have consequently highlighted the need for further empirical studies examining educational outcomes, critical thinking, ethical safeguards, and appropriate integration with faculty teaching [12]. Evidence is particularly limited regarding the impact of structured AI-assisted feedback on the clinical reasoning performance of postgraduate trainees in Pakistan. Therefore, the present study was conducted to compare AI-assisted feedback with conventional faculty feedback and determine its effect on clinical reasoning scores, diagnostic accuracy, error frequency, case-completion time, confidence, and satisfaction among postgraduate medical trainees at Rawal Institute of Health Sciences, Islamabad.

## MATERIALS AND METHODS

This prospective quasi-experimental comparative study was conducted at Rawal Institute of Health Sciences, Islamabad, from January 2025 to June 2025. The study was designed to evaluate the impact of artificial intelligence-assisted feedback on the clinical reasoning skills of postgraduate medical trainees. A total of 76 postgraduate trainees

were included and divided into two equal groups of 38 participants each. One group received AI-assisted feedback, while the comparison group received conventional faculty-based feedback. The study was carried out within structured postgraduate teaching sessions using standardized clinical case scenarios. Ethical approval was obtained from the institutional review committee of Rawal Institute of Health Sciences before the commencement of data collection. Written informed consent was obtained from all participants, and confidentiality of their personal and academic information was maintained throughout the study.

All the postgraduate medical trainees, who were pursuing recognized residency or postgraduate training program at the institution were included. Non-probability consecutive sampling was used to approach trainees from different years of post-graduate training and clinical specialties. Those who were actively engaged in clinical training, were willing to attend all educational sessions and gave written informed consent were included. Trainees that did not complete the clinical cases, were not available for baseline or follow-up assessment or had previously completed a formal clinical reasoning program based on artificial intelligence were excluded. There were 76 trainees who completed all phases of the study and were included in the final sample. Participants were randomly assigned to the AI-assisted group and conventional feedback group, with reasonable similarity between groups in terms of postgraduate years and initial clinical reasoning performance.

Demographic and academic data was gathered through a proforma prior to the intervention. The variables that were recorded were age, gender, post-graduate training year, specialty, prior academic performance, previous experience in using AI tools, and how often they are using digital learning resources. All subjects underwent an initial clinical reasoning task using standardized written clinical scenarios. The assessment looked at the identification of relevant clinical information, generation of differential diagnoses, selection of investigations, interpretation of diagnostic findings, formulation of a management plan, and justification of clinical decisions. The domains were each assessed with a predetermined assessment rubric and the individual domain scores were added together to create an overall clinical reasoning score, which ranged from 0 to 100. The baseline assessment was conducted with the same instructions, time limits, and case complexity for both groups.

In both the groups, during the intervention period, they completed the same number of standardized clinical cases which included the common diagnostic and management issues in postgraduate medical training. The participants completed an AI-supported educational system by providing their diagnostic reasoning, investigation plans, and proposed decisions for patient care. The participants used an AI-supported educational system to submit their diagnostic reasoning, investigation plans, and proposed decisions for patient care. It immediately gave the trainee structured, case-specific feedback for missing clinical information, inconsistencies in the differential diagnosis, areas that needed to be reconsidered, and compared the trainee's reasoning with the criteria of the evidence-based cases. The feedback was only for educational support and should not substitute for faculty supervision or clinical decision making. The conventional-feedback group performed the same clinical cases and were provided with feedback using the traditional way (lecturing style) of faculty feedback, which involved verbal or written comments on the diagnostic accuracy, selection of investigations and planning for management. The length of time, number of cases, learning goals, and overall instructional content were also maintained the same across both groups.

Both groups faced a post-intervention clinical reasoning test at the end of the intervention, with the same clinical cases of similar difficulty as the pre-intervention clinical reasoning test. The major outcome measure was the differences in the overall clinical reasoning score between pre- and post-test. Secondary outcomes were diagnostic accuracy, generation of appropriate differential diagnoses, selection of relevant investigations, quality of the management plan, number of diagnostic errors, case-completion time, confidence of clinical decision making and satisfaction with the feedback method. A five-pointed Likert scale was used as the instrument for measuring level of confidence and satisfaction. IBM SPSS Statistics was used to enter and analyse data. Continuous variables were expressed as means ± standard deviations whereas categorical variables were expressed as frequencies and percentages. Independent-samples t-test was performed to compare continuous outcomes across groups, and paired-samples t-test was

performed to examine the change in continuous outcomes over time within groups. The chi-square test or Fisher's exact test was used for categorical variables, as appropriate. Multiple linear regression was used to assess the independent effect of AI-assisted feedback on improvement in clinical reasoning scores, after controlling for relevant baseline characteristics. A p value of < 0.05 was considered statistically significant.

## RESULTS

There were 76 postgraduate medical trainees, of whom 38 received feedback from artificial intelligence while 38 received feedback from the faculty. Baseline and post assessment was administered to all participants. The overall mean age was 28.9 ± 2.6 years. There were 43 male trainees (56.6%) and 33 female trainees (43.4%). Neither group had any significant difference in terms of their age, gender, postgraduate training year, specialty distribution, previous academic performance or prior experience of using AI tools.

Table 1. Baseline Characteristics of Postgraduate Medical Trainees

| Variable | AI-assisted feedback group (n = 38) | Conventional feedback group (n = 38) | p-value |
|---|---|---|---|
| Age, years, mean ± SD | 28.7 ± 2.5 | 29.1 ± 2.7 | 0.506 |
| Male gender, n (%) | 22 (57.9) | 21 (55.3) | 0.818 |
| Female gender, n (%) | 16 (42.1) | 17 (44.7) | |
| PGY-1, n (%) | 11 (28.9) | 10 (26.3) | 0.972 |
| PGY-2, n (%) | 10 (26.3) | 11 (28.9) | |
| PGY-3, n (%) | 9 (23.7) | 9 (23.7) | |
| PGY-4, n (%) | 8 (21.1) | 8 (21.1) | |
| Previous use of AI tools, n (%) | 17 (44.7) | 15 (39.5) | 0.644 |
| Baseline clinical reasoning score, mean ± SD | 61.8 ± 7.4 | 62.3 ± 7.1 | 0.766 |

There was no difference between the groups in the base line clinical reasoning score. Prior to the intervention, there was no difference between the two groups in their clinical reasoning ability, with both groups showing a mean baseline score of 61.8 ± 7.4 (p = 0.766) in the conventional-feedback group, and 62.3 ± 7.1 (p = 0.766) in the AI-assisted feedback group. After the intervention, the clinical reasoning performance of the trainees who were provided with feedback using AI was higher. They showed a higher score at the end of the intervention (78.6 ± 6.8) than at baseline (61.8 ± 7.4). Conventional feedback group had a gain of 62.3 ± 7.1 to 70.4 ± 7.3. The AI-assisted group showed a mean improvement of 16.8 ± 6.1 points, while the conventional group showed a mean improvement of 8.1 ± 5.7 points. The difference in scores between the groups was significant (p<0.001).

Table 2. Comparison of Pre-Intervention and Post-Intervention Clinical Reasoning Scores

| Clinical reasoning outcome | AI-assisted feedback group (n = 38) | Conventional feedback group (n = 38) | p-value |
|---|---|---|---|
| Pre-intervention score, mean ± SD | 61.8 ± 7.4 | 62.3 ± 7.1 | 0.766 |
| Post-intervention score, mean ± SD | 78.6 ± 6.8 | 70.4 ± 7.3 | <0.001 |
| Mean change in score | 16.8 ± 6.1 | 8.1 ± 5.7 | <0.001 |
| Percentage improvement | 27.2 ± 10.4 | 13.0 ± 9.1 | <0.001 |
| Within-group p-value | <0.001 | <0.001 | — |

There was no difference between the groups in the base line clinical reasoning score. Prior to the intervention, there was no difference between the two groups in their clinical reasoning ability, with both groups showing a mean baseline score of 61.8 ± 7.4 (p = 0.766) in the conventional-feedback group, and 62.3 ± 7.1 (p = 0.766) in the AI-assisted feedback group. After the intervention, the clinical reasoning performance of the trainees who were provided with feedback using AI was higher. They showed a higher score at the end of the intervention (78.6 ± 6.8) than at baseline (61.8 ± 7.4). Conventional feedback group had

a gain of 62.3 ± 7.1 to 70.4 ± 7.3. The AI-assisted group showed a mean improvement of 16.8 ± 6.1 points, while the conventional group showed a mean improvement of 8.1 ± 5.7 points. The difference in scores between the groups was significant (p<0.001).

Table 3. Post-Intervention Clinical Reasoning Component Scores

| Clinical reasoning component | AI-assisted feedback group | Conventional feedback group | p-value |
|---|---|---|---|
| Identification of relevant clinical information | 80.3 ± 7.2 | 72.5 ± 7.8 | <0.001 |
| Generation of differential diagnoses | 82.1 ± 7.5 | 71.8 ± 8.2 | <0.001 |
| Selection of appropriate investigations | 78.9 ± 7.0 | 70.6 ± 7.7 | <0.001 |
| Interpretation of diagnostic findings | 77.8 ± 7.6 | 70.9 ± 8.0 | <0.001 |
| Development of an appropriate management plan | 77.6 ± 7.1 | 69.8 ± 7.9 | <0.001 |
| Justification of clinical decisions | 75.8 ± 7.8 | 68.9 ± 8.1 | <0.001 |

Diagnostic accuracy was significantly higher in the AI-assisted feedback group. A correct final diagnosis was reached by 32 trainees (84.2%) in the AI group compared with 24 trainees (63.2%) in the conventional group (p = 0.037). The mean number of diagnostic errors was also lower among trainees receiving AI-assisted feedback. Furthermore, the AI group completed the clinical cases in a shorter mean duration than the conventional-feedback group.

Table 4. Diagnostic Accuracy, Errors, and Case-Completion Time

| Outcome | AI-assisted feedback group | Conventional feedback group | p-value |
|---|---|---|---|
| Correct final diagnosis, n (%) | 32 (84.2) | 24 (63.2) | 0.037 |
| Accurate differential diagnosis, n (%) | 30 (78.9) | 21 (55.3) | 0.029 |
| Appropriate investigation plan, n (%) | 31 (81.6) | 23 (60.5) | 0.043 |
| Appropriate management plan, n (%) | 30 (78.9) | 22 (57.9) | 0.049 |
| Diagnostic errors per trainee, mean ± SD | 1.2 ± 0.8 | 2.1 ± 1.0 | <0.001 |
| Case-completion time, minutes, mean ± SD | 18.4 ± 4.3 | 22.7 ± 5.1 | <0.001 |

The AI-assisted feedback group showed a greater improvement in the confidence of the trainees in making clinical decisions. They rated their mean confidence on a five-point Likert scale as 3.1 ± 0.6 before and 4.2 ± 0.5 after the intervention. In the conventional-feedback group, the mean confidence score increased from 3.0 ± 0.7 to 3.6 ± 0.6. There was a statistically significant difference between the groups in the post-intervention (p<0.001). Trainees who received feedback with AI also mentioned increased satisfaction with the clarity, timeliness, specificity and usefulness of feedback. Of these trainees, 34 (89.5%) agreed or strongly agreed that the feedback helped them find things wrong with their reasoning, and 32 (84.2%) indicated they would like to use AI-supported feedback in the future in clinical learning activities.

Table 5. Trainee Confidence and Perception of the Feedback Method

| Outcome | AI-assisted feedback group | Conventional feedback group | p-value |
|---|---|---|---|
| Post-intervention confidence score | 4.2 ± 0.5 | 3.6 ± 0.6 | <0.001 |
| Feedback clarity score | 4.4 ± 0.5 | 3.8 ± 0.6 | <0.001 |
| Feedback specificity score | 4.3 ± 0.6 | 3.5 ± 0.7 | <0.001 |
| Timeliness of feedback score | 4.6 ± 0.5 | 3.7 ± 0.7 | <0.001 |
| Feedback usefulness score | 4.5 ± 0.5 | 3.9 ± 0.6 | <0.001 |
| Overall satisfaction score | 4.4 ± 0.6 | 3.8 ± 0.7 | <0.001 |

Multiple linear regression analysis was conducted to determine the factors that were associated with improvement in the clinical reasoning scores. These results did not change

after adjustment for age, gender, postgraduate year, baseline CRS, or prior experience with AI tools, with AI-assisted feedback group still being an independent predictor of higher score increase. An adjusted increase of 8.2 points was observed with AI-assisted feedback versus conventional feedback for the clinical reasoning score.

Table 6. Predictors of Improvement in Clinical Reasoning Scores

| Predictor | Adjusted β coefficient | 95% CI | p-value |
|---|---|---|---|
| AI-assisted feedback | 8.20 | 5.48 to 10.92 | <0.001 |
| Baseline clinical reasoning score | −0.21 | −0.38 to −0.04 | 0.016 |
| Previous experience with AI tools | 1.34 | −0.72 to 3.40 | 0.198 |
| Postgraduate training year | 0.86 | −0.18 to 1.90 | 0.104 |
| Age | 0.12 | −0.29 to 0.53 | 0.561 |
| Male gender | 0.74 | −1.31 to 2.79 | 0.474 |

Overall, AI-assisted feedback was associated with significantly greater improvement in clinical reasoning scores, diagnostic accuracy, recognition of clinical errors, confidence in decision-making, and satisfaction with the Feedback process. These findings suggest that AI-supported feedback may provide an effective supplementary approach for strengthening clinical reasoning skills among postgraduate medical trainees.

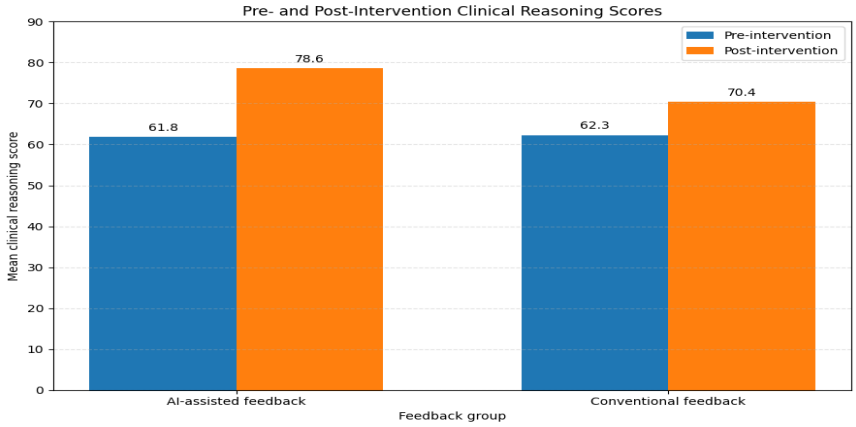


**Figure 1.** Comparison of Mean Pre-Intervention and Post-Intervention Clinical Reasoning Scores Between the AI-Assisted Feedback and Conventional Feedback Groups.

The AI-assisted feedback group showed a greater increase in mean clinical reasoning score, from 61.8 to 78.6, compared with an increase from 62.3 to 70.4 in the conventional feedback group.

## DISCUSSION

The present study evaluated the impact of artificial intelligence-assisted feedback on clinical reasoning skills among postgraduate medical trainees. Trainees who received AI-assisted feedback demonstrated a significantly greater increase in their mean clinical reasoning score than those who received conventional faculty feedback. The mean score in the AI-assisted group increased from 61.8 ± 7.4 at baseline to 78.6 ± 6.8 after the intervention, whereas the conventional-feedback group improved from 62.3 ± 7.1 to 70.4 ± 7.3. Although both groups showed significant improvement, the mean increase was approximately twice as high in the AI-assisted group. These findings suggest that immediate, individualized, and structured AI-generated feedback may help postgraduate trainees recognize weaknesses in their reasoning and correct them more effectively during case-based learning [13-15].

The present study is consistent with emerging evidence of the use of large language models in education. Brügge et al. carried out a randomized controlled study that involved learners taking part in medical history-taking practice sessions with simulated patients using AI technology. Structured AI generated feedback resulted in higher clinical decision making scores for participants than those who went through the simulations without the

feedback. One area of improvement was in providing a clinical context and gathering important information [16]. Likewise, AI-driven learning platforms have demonstrated their ability to offer learners repeated opportunities to engage in deliberate practice, which enables them to change their practice based on feedback, without having to wait for a clinical supervisor to be available. Another randomized controlled trial that tested the performance of clinical reasoning in ChatGPT-generated feedback versus expert-written feedback found no overall effect of the type of feedback, but an improvement in clinical reasoning performance for some complex clinical cases with expert feedback [7]. This suggests that AI feedback could potentially be a useful supplement to education but it could be dependent on the complexity of the case, the quality of feedback, and how it is incorporated into the curriculum.

The AI group scored better when it came to identifying the relevant information, making differential diagnoses, selecting investigations, interpreting findings, and formulating management plans in the present study. The greatest improvement was observed in the generation of differential diagnoses. AI systems can help with this improvement by quickly finding possibilities that were not mentioned and suggesting alternative answers, as well as prompting trainees to consider opposing diagnoses. AI-generated script concordance tests have also been found to be useful in providing educationally valuable feedback, and in mimicking some aspects of expert clinical reasoning, when properly trained [9]. A systematic review of virtual patients has also found that technology-based case-based learning can help develop clinical reasoning, especially when the students are provided with feedback regarding learning content and the explanation of the case, which can be more descriptive than just a right/wrong response [17]. The value of AI, then, might not only be in giving answers, but in making the learner's thought process transparent and highlighting any missing or conflicting components [18].

The accuracy of the diagnosis was also greater in the group of trainees who had the AI coaching help them. The AI-assisted group outperformed the conventional group with a correct final diagnosis rate of 84.2% vs 63.2%. The AI group had fewer misdiagnoses and cases were handled in a shorter time. Overall, the results indicate that AI-related feedback can enhance the precision and efficiency of clinical problem-solving. Immediate feedback can be used to correct an incorrect reasoning pathway, which is still fresh in the learner's memory. It can also give the same feedback for a number of cases and learners, thereby minimizing the variation that occurs if faculty lack sufficient time or adopt a variety of feedback methods. The postgraduate trainees have generally indicated interest in receiving more formal training in the application of AI, but there has been concern that there might be a decrease in opportunities for them to learn clinical judgment for themselves if they rely too heavily on AI [19]. Feedback from AI should therefore prompt trainees to explain and defend the diagnosis and/or management plan that the AI has proposed, not just blindly accept it [20].

The AI-assisted trainees had more confidence, clarity of feedback, timeliness, usefulness, and satisfaction overall. Feedback was immediate and this may have helped learners to spot the particular mistake and make changes in real time. However, higher self-confidence does not necessarily mean higher competency. AI systems can create incomplete or inaccurate, or outdated or misleading information; learners may trust the information produced by the AI system to be correct, even when it is not so. AI assisted feedback should, therefore, be subject to the appropriate faculty supervision and supervision should involve trainees critically assessing outputs against clinical guidance, patient context and expert judgement. There were several limitations in the present study. It was carried out in one institution, with limited sample size (76 trainees) and short intervention duration. The differences in the clinical reasoning performance may also have been influenced by differences between specialties and years of postgraduate training. It was not possible to blind the participants and satisfaction outcomes were self-reported. Furthermore, long-term retention and transfer to real-world practice and the accuracy of all AI-generated recommendations were not measured. Larger samples, delayed evaluations, workplace-based outcomes, validated reasoning instruments, and detailed evaluation of the AI errors should be included in future multicentre studies. Comparing the effects of feedback with AI only, faculty only, and combined AI–faculty feedback would be especially helpful.

## CONCLUSION

Artificial intelligence-assisted feedback was associated with greater improvement in clinical

reasoning scores, diagnostic accuracy, differential diagnosis generation, investigation selection, management planning, and case-completion efficiency among postgraduate medical trainees. Participants receiving AI-supported feedback also reported higher confidence and satisfaction than those receiving conventional feedback. These findings indicate that AI can serve as a useful supplementary tool for formative assessment and individualized clinical reasoning practice. However, it should not replace faculty supervision, professional judgment, or direct clinical teaching. Its educational use should be supported by faculty review, learner training in critical appraisal, protection of patient information, and regular evaluation of the accuracy and appropriateness of AI-generated content.

## REFERENCES


1. Zhang, Z., et al., *Laparoscopic, Endoscopic and Robotic Surgery*. 2021.
2. Shaikh, F., et al., *Measuring the accuracy of cardiac output using POCUS: the introduction of artificial intelligence into routine care*. 2022. **14**(1): p. 47.
3. Natheir, S., *Utilizing Machine Learning and Electroencephalography to Assess Expertise in Virtual Neurosurgical Performance*. 2021, McGill University (Canada).
4. Lytras, M., et al., *Artificial intelligence and big data analytics for smart healthcare*. 2021: Academic Press.
5. Shaikh, F., J.-E. Kenny, and O.J.T.U.J. Awan, 14, *Measuring the accuracy of cardiac output using POCUS: the introduction of artificial*. 2022.
6. Hahn, M.G., et al., *A systematic review of the effects of automatic scoring and automatic feedback in educational settings*. 2021. **9**: p. 108190-108198.
7. Çiçek, F.E., et al., *ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial*. Postgraduate Medical Journal, 2025. **101**(1195): p. 458-463 DOI: 10.1093/postmj/qgae170.
8. Nissen, L., et al., *A randomised cross-over trial assessing the impact of AI-generated individual feedback on written online assignments for medical students*. 2025. **47**(9): p. 1544-1550.
9. Hudon, A., et al., *Teaching Clinical Reasoning in Health Care Professions Learners Using AI-Generated Script Concordance Tests: Mixed Methods Formative Evaluation*. JMIR Form Res, 2025. **9**: p. e76618 DOI: 10.2196/76618.
10. Schaye, V., et al., *Large language model–based assessment of clinical reasoning documentation in the electronic health record across two institutions: Development and validation study*. 2025. **27**: p. e67967.
11. Goh, E., et al., *Large language model influence on diagnostic reasoning: a randomized clinical trial*. 2024. **7**(10): p. e2440969.
12. Gordon, M., et al., *A scoping review of artificial intelligence in medical education: BEME Guide No. 84*. 2024. **46**(4): p. 446-470.
13. Chen, M., et al., *Physician and Medical Student Attitudes Toward Clinical Artificial Intelligence: A Systematic Review with Cross-Sectional Survey*. 2022.
14. Bhaskar, S., et al., *Designing futuristic telemedicine using artificial intelligence and robotics in the COVID-19 era*. 2020. **8**: p. 556789.
15. Shinners, L., et al., *Exploring healthcare professionals' perceptions of artificial intelligence: Validating a questionnaire using the e-Delphi method*. 2021. **7**: p. 20552076211003433.
16. Brügge, E., et al., *Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial*. BMC Medical Education, 2024. **24**(1): p. 1391 DOI: 10.1186/s12909-024-06399-7.
17. Ringwald, B.A. and J.L.J.A.M. Middleton, *Continuity of Care Explains Disparities Between Faculty and Trainee Practice in Residency Clinics*. 2024. **99**(12): p. 1321-1322.
18. Pandey, A., M.J.M.U.J.o.H. Misra, and S. Sciences, *Artificial Intelligence and Psychology: Developments and Future Potential*. 2022. **8**(1): p. 186-200.
19. Banerjee, M., et al., *The impact of artificial intelligence on clinical education: perceptions of postgraduate trainee doctors in London (UK) and recommendations for trainers*. 2021. **21**(1): p. 429.
20. Phillips, M., J. Vermylen, and H.J.A.M. Ballard, *From Bytes to Empathy: Can*

*ChatGPT Teach Anesthesiologists How to Deliver Bad News?* 2024. **99**(4): p. 347.