Advances in Medical Education and Practice

Dovepress
Taylor & Francis Group

Open Access Full Text Article

PERSPECTIVES

When Clinical AI and Learner Reasoning Conflict: An Emerging Educational Blind Spot and a Framework for Pedagogical Response

Samita M Heslin

Department of Emergency Medicine, Stony Brook University, Stony Brook, New York, USA

Correspondence: Samita M Heslin, Department of Emergency Medicine, Stony Brook University, Stony Brook, New York, USA, Email samita.heslin@stonybrookmedicine.edu

Abstract: Increasingly, artificial intelligence (AI)-enabled clinical decision support has been incorporated into healthcare settings where trainees of medicine learn. A major pedagogical problem is emerging: at times, when an AI-generated recommendation is presented to learners as part of their training, there is disagreement between the clinical decision support recommendation and learner clinical judgment, reflecting the need to interpret probabilistic AI outputs within clinical reasoning. This article suggests that the frequency of this phenomenon represents a growing educational blind spot; if not addressed by educators, it could negatively affect the development of clinical reasoning among learners, especially in situations where AI outputs are perceived as objective or inherently authoritative rather than probabilistic or fallible, and therefore could potentially hinder the development of the learner's professional identity. This article presents illustrative educational scenarios that show how learner-AI conflicts may occur in educational settings. After presenting the examples, it describes the educational risks that exist when there is no structure to supervise responses to these types of learner-AI conflicts. Finally, it suggests an approach using a 4-stage framework called SEED (Surface-Explore-Evaluate-Decide) to transform learner-AI conflict into an intentional opportunity to learn and provides specific ideas on how to use the SEED framework to create structures for teaching, assessment, and faculty development. In light of the growing presence of technology in all areas of education, it is essential for educators to be prepared to respond to these types of tensions to preserve the values of accountability, thoughtfulness, and humanness in education.

Keywords: automation bias, trust calibration, epistemic authority, algorithmic transparency, clinical decision support systems, health professions education

Introduction

Clinical decision support tools, including those utilizing artificial intelligence (AI), are becoming part of many health care environments, such as clinical settings in which medical students, residents, and fellows are trained.[1,2] Clinical decision support tools provide diagnostic, risk stratification, and treatment recommendations that are used for decision-making in the clinical setting.[3–5] At the same time, educators have been charged with preparing future clinicians who can utilize AI responsibly, critically, and effectively.[6,7] As such, much of the current discussion regarding education has been centered around literacy, technical knowledge, and ethical considerations for AI.[8–10]

A less frequently studied, but increasingly prevalent occurrence in the clinical environment has been identified: clinical decision support tools generating recommendations that oppose a learner's clinical reasoning or proposed management plan. The presence of this type of clinical decision-making discordance raises important pedagogical questions. When an educator encounters a situation in which there is a difference between the output of a clinical decision support tool and a learner's clinical reasoning or proposed management plan, how does the educator determine which recommendation to consider authoritative? Hypothetical clinical scenarios are used throughout this article to demonstrate these teaching strategies. These scenarios were constructed solely for illustrative educational purposes and are not derived from real students, educators, nor patient encounters.

https://doi.org/10.2147/AMEP.S592114
Received: 27 December 2025
Accepted: 25 February 2026
Published: 7 May 2026

Growing evidence suggests that discordance between clinical decision support recommendations, including both traditional rule-based and AI-generated outputs, and clinicians' professional judgment occurs in clinical practice. Studies show that clinicians dismiss or override a large proportion of clinical decision support alerts, such as drug–drug interactions and dose alerts. These are often dismissed due to alert fatigue, workflow disruption, and the perception that the recommendation is not relevant to the specific patient.[11,12] Research on the development and external validation of sepsis prediction models shows high false-positive rates in real-world clinical deployment.[13,14] Additionally, studies comparing agreement between AI-assisted diagnostic imaging and expert interpretation have found clinically meaningful rates of disagreement.[15,16] Research in AI-assisted mammography has shown that automation bias lowers reader specificity when clinicians are exposed to incorrect AI outputs.[17] Thus, recommendations, even when incorrect, may influence clinical decision making. Collectively, these findings indicate that discordance between clinical decision support recommendations and clinician judgment is not rare, but an expected part of technology-assisted clinical decision making. As AI becomes more integrated into the delivery of healthcare, clinicians will need to use, critically evaluate, and override AI systems when appropriate. The above realities support the need for educationally based frameworks that emphasize AI-clinician discordance as a core skill in current medical training.

Emergence of AI-Learner Discordance in Clinical Education

Historically, clinical education has had many forms of authority, such as textbooks, clinical guidelines, clinician judgments, and institutional protocols. The emergence of AI systems introduces a novel and distinct source of authority to the mix. In contrast to traditional decision-aids, AI tools typically provide probabilistic recommendations that are presented as personalized, data-driven, and objective, which may be perceived as having superior knowledge-based authority.[1,18,19] However, this authority is perceived rather than inherently present because AI-based outputs are model-based, limited by context, and susceptible to error.

Typology of Common Scenarios of AI-Human Learner Discordance

Discordance between AI-based systems and human learners can occur in a variety of clinical situations across different clinical domains. Understanding the nature of the discordance will help educators to determine when to intervene in the educational process.

The following hypothetical cases provide examples of how AI-learner discordance might occur in clinical settings. These scenarios are conceptual teaching exemplars designed to describe common forms of AI-learner discordance.

Hypothetical Vignette Case I: Clinical Reasoning vs. Algorithmic Decision Making

A medical student evaluates a 52-year-old female who presents to the emergency department with chest pain. Following an extensive history taking and physical exam, the student notes that there are some atypical features (sharp, positionally related pain that is reproduced upon palpation), a normal ECG, and no concerning risk factors. Thus, the student proposes sending the patient home with an outpatient cardiology follow-up visit. However, when the AI-enabled risk stratification tool embedded within the electronic health record processes the chief complaint of "chest pain", it categorizes this patient as "intermediate" risk and recommends hospital admission to monitor serial troponins and further testing. This results in uncertainty regarding how to reconcile the discrepancy between the clinical decision-making process and the algorithm's recommendations. Additionally, it raises the question of how to approach these types of discrepancies clinically during the education and training of students.

In this case, a situation has developed which is an example of "AI-learner discordance" - when a clinical decision support tool (eg AI-enabled computer program) recommends something different than what the learner thinks the patient needs based on clinical experience. This type of conflict between an algorithm's output and clinical judgement represents a developing gap in the educational process. Without clear guidance on how to handle these situations, educators may choose to respond in inconsistent ways, or they may not be handled at all - both of which may have detrimental effects on the development of clinical decision-making skills, professional identity, and accountability of learners.[20,21] Transforming AI-learner disagreements into intentional opportunities for teaching will help bridge the gap in medical education and create alignment with what clinicians will need to do in their daily practice with AI-assisted technology.

Hypothetical Vignette Case 2: Diagnostic Imaging (AI-Human Discordance)

A resident evaluates a 68-year-old man who has community-acquired pneumonia. Using the CURB-65 criteria, the resident assesses that the patient has low-risk pneumonia, which would be appropriately managed with oral antibiotics as an outpatient. However, a diagnostic imaging AI system reviews the patient's chest radiograph and finds a small peripheral opacity. The AI system provides a recommendation for the resident to admit the patient for intravenous antibiotics, stating a 23% probability of treatment failure if the patient were treated as an outpatient with oral antibiotics. The resident's clinical gestalt suggests outpatient management. However, the use of a quantitative measure of risk generates uncertainty regarding the best course of action.

The two hypothetical cases illustrated above have common characteristics. First, they both illustrate how the algorithms used by the AI systems process information that may not account for the full clinical context. Second, each case illustrates how the output provided by the AI may appear as a certainty to the learner and clinician. Finally, each case illustrates how the AI-based recommendations may create uncertainty during critical periods of development of the learner's clinical judgment.[22,23]

In addition, the discordance created by the AI-human interaction does not indicate that one or the other has made an error. Rather, the discordance may result from a number of issues related to the environment in which the model was developed (eg lack of contextual information), issues related to the type of data available to the model (eg lack of relevant data), differences in the way that uncertainty is evaluated between humans and algorithms, or differences in the risk tolerance of humans vs. algorithms.[19,24–27] Figure 1 illustrates common categories in which AI-learner discordance may occur.

These scenarios may be prevalent in high-pressure settings (eg emergency medicine, inpatient services) that have a rapid pace. Time constraints in these areas do not allow for a lot of time for reflection on an issue or opportunity to discuss with other individuals involved in the care. In these environments, brief, workflow-integrated supervisory exchanges may represent a feasible approach to reflective teaching. Without structured pedagogy, the learner may use an ad-hoc method when they are faced with multiple potentially conflicting sources of guidance, including AI outputs and clinical judgments.

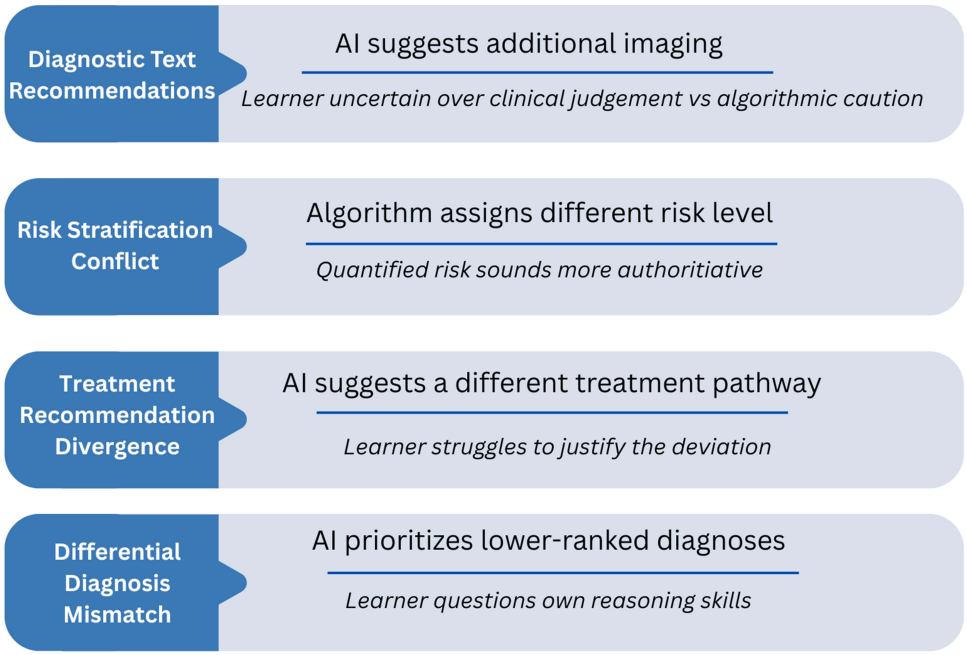

Figure 1 Types of AI-Learner Discordance in Clinical Teaching. Common categories in which artificial intelligence-enabled clinical decision support recommendations may diverge from learner clinical reasoning.

https://doi.org/10.2147/AMEP.S592114

Educational Risks of Unstructured Responses to AI-Learner Discordance

When educators do not share a common understanding or language to address the issue of discord between AI and learners, some harmful ways of responding to ambiguity in clinical teaching may emerge with associated negative educational outcomes. The literature on clinical supervision (ie the process by which faculty members assist learners in developing professional competencies) as well as the "hidden curriculum" (ie the unintended messages sent to students through the day-to-day behaviors of educators) suggests that how faculty members respond to ambiguous clinical situations is a critical determinant of the way learners develop.[20,27,28] As shown in Table 1, there are many different patterns of responding to ambiguity in clinical education; each pattern has its own set of characteristic and corresponding potential educational risks.

Implications for education are significant for each of the first three response types. Learners may struggle to determine when, how, and under what circumstances they should question the recommendation of an AI system; how responsibility will be assigned or held for those recommendations; and what constitutes a reasonable exercise of professional judgement while using AI systems that assist in patient care. These issues remain relatively unexamined and could affect the "hidden curriculum", and in turn, learners' perceptions of their accountability, authority, and roles in clinical decision-making.[18,21]

Recent work has underscored the need for developing ethics and governance mechanisms for the deployment of AI systems in healthcare. Ethical firewalls have been proposed as a method of ensuring that decisions made by AI systems are consistent with the values of clinicians, medical standards, and accountable practice, through embedded and technically verifiable ethical constraint mechanisms.[29] The development of these frameworks also supports ongoing clinician involvement in the oversight of AI and in the ethical evaluation of AI system outputs.

The fourth response type (Reflective Integration) sees learner-AI conflicts as opportunities for learning, rather than something to be resolved as soon as possible.

Using the SEED Framework in Developing a Pedagogical Response

Instead of viewing the conflict caused by AI and learners as a disagreement that needs to be resolved quickly, educators may begin to view these disagreements as an opportunity for intentional instruction and reflective dialogue regarding reasoning strategies, uncertainty, and contextual judgment.[30] To operationalize this educational approach, this article proposes the SEED (Surface-Explore-Evaluate-Decide) framework as a method of developing a pedagogical response. The four components of the SEED framework are shown in Figure 2.

Table 1 Example Educator Response Patterns to AI-Learner Discordance and Educational Consequences

Response Pattern	Description	Educational Consequences
Uncritical Deference	AI recommendations accepted by default; algorithmic outputs prioritized over learner reasoning without discussion	• Discourages independent clinical reasoning • Reinforces algorithmic authority at expense of professional judgment • Undermines learner confidence in assessment skills
Reflexive Dismissal	AI outputs dismissed outright to preserve traditional expertise; technology portrayed as unreliable or peripheral	• Limits learner ability to engage critically with AI systems • Fails to prepare learners for AI-augmented practice environment
Silent Avoidance	Discordance not explicitly addressed; AI influence on decision-making remains implicit and unexamined	• Learners uncertain how to reconcile conflicting guidance • Hidden curriculum shapes attitudes without reflection • Ambiguity about responsibility and accountability
Reflective Integration	Discordance made explicit; structured exploration of reasoning processes and contextual factors; human accountability emphasized	• Develops critical judgment and epistemic humility • Reinforces professional accountability and responsibility

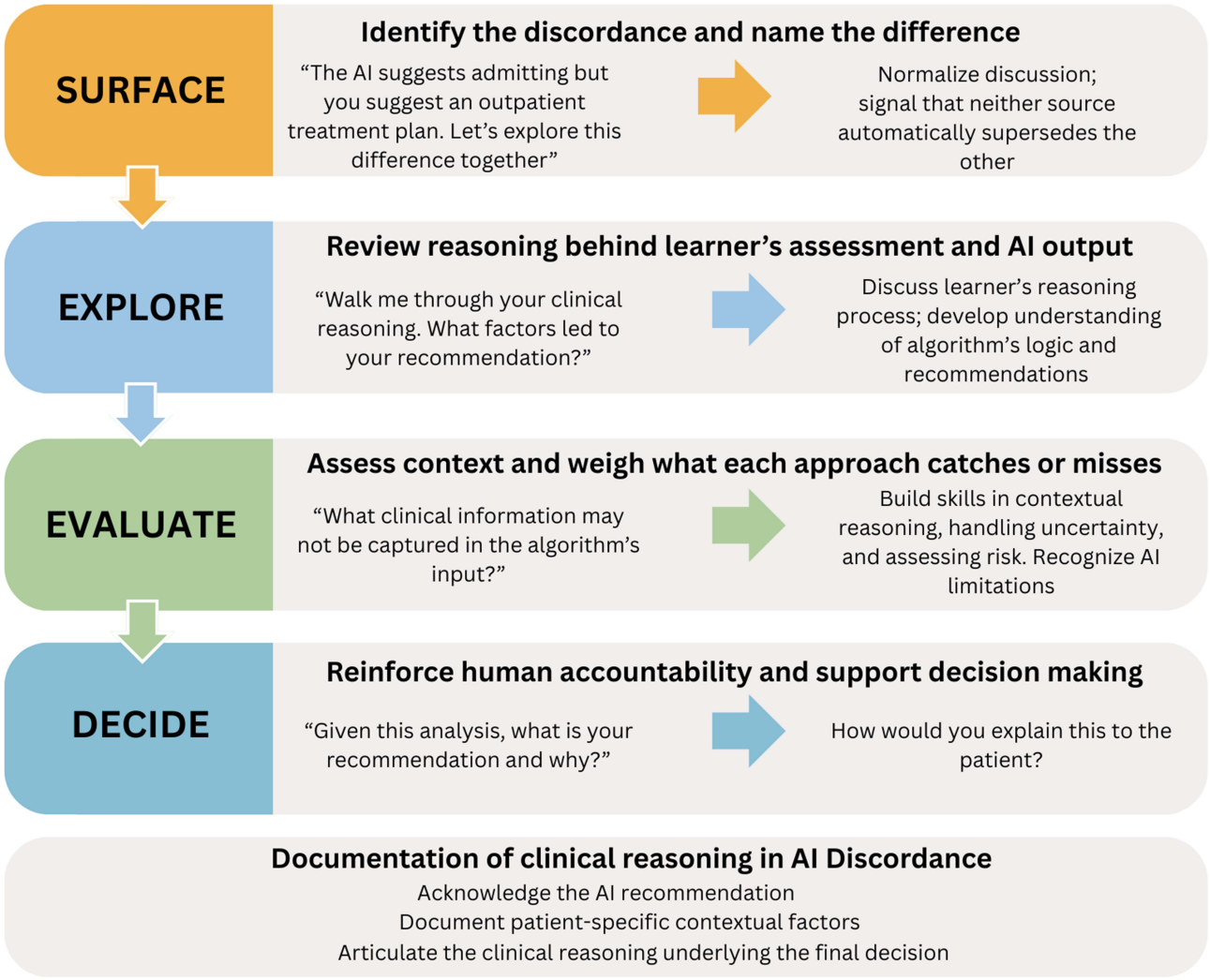

Figure 2 The SEED Framework for Addressing AI-Learner Discordance. The SEED framework (Surface, Explore, Evaluate, Decide) provides a structured approach for educators to guide learners through situations in which AI-enabled CDS recommendations conflict with learner clinical reasoning. The dialogue examples included in this figure are illustrative and were created solely for hypothetical educational purposes.

The SEED framework is based on three complementary theoretical frameworks. First, dual-process theory of clinical reasoning defines two clinical reasoning processes: fast, intuitive pattern recognition (System 1) and slower, analytical deliberation (System 2). This framework has been used to explore how clinicians may use analytical reasoning when they encounter diagnostic uncertainty or cognitive bias.[31–33] Second, trust calibration theory from human factors research suggests that human-automation interactions are optimized when individuals develop appropriately calibrated trust in automated systems. Users may become overly reliant or, conversely, develop unwarranted distrust in automated systems. Therefore, users need to calibrate their levels of trust accordingly.[34–36] This principle is becoming increasingly relevant in the clinical use of AI, which requires that clinicians interpret, contextualize, and critically evaluate the recommendations of AI algorithms. Finally, situated cognition theory states that learning is enhanced when learners engage within the environments in which the knowledge is being applied. The SEED framework incorporates this principle by embedding real-time clinical training in AI trust calibration as opposed to separate didactic education.[37–39] Together, these theoretical perspectives provide a conceptual foundation for the structured instructional approach outlined in the SEED framework.

Within the SEED framework, the initial phase of the process entails making the conflict between the learner's assessment and the AI-generated assessment visible by explicitly acknowledging a discrepancy between them. As

https://doi.org/10.2147/AMEP.S592114

a result, it legitimizes having an open and honest discussion about the two assessments as well as reducing the perception that one must be superior to the other. Upon identifying why the learner's assessment differed from the AI-generated assessment through facilitated reflection, and once the underlying assumptions associated with each assessment are identified, the framework emphasizes that ultimately the responsibility for clinical decision-making rests with the human clinician, regardless of whether they used an AI tool. The emphasis on the responsibility for clinical decision-making placed on the clinician supports the significance of professional accountability and recognizes the capability of AI to serve as a support mechanism for clinician judgment rather than a replacement. The process assists educators in modeling how to critically evaluate AI-generated assessments and gives learners the opportunity to develop the cognitive skills required for critical evaluation of AI assessments, such as situational awareness and the ethical implications. Future work should evaluate the feasibility, acceptability, and educational impact of the SEED framework across diverse clinical training environments.

Practical Strategies for Implementing the SEED Framework

The SEED framework will need to be implemented with practical tools and institutional support to encourage educators to work toward consistent and effective navigation of AI-learner discordance.

Structured Debriefing Questions

To provide educators with a way to support reflection and engage learners critically, educators can ask structured debriefing questions. These do not have to be long discussions as short exchanges can reinforce the critical engagement. Examples of debriefing questions that could be used based upon the SEED framework stages are:

Surface Stage

- Can you walk me through the clinical reasoning that led to your recommendation?
- What does the AI recommend, and where does the AI recommendation vary from your assessment?

Explore Stage

- What clinical data or patient factors were you weighing the heaviest in your decision?
- What inputs does the AI algorithm use? What factors might it be putting a greater emphasis on or missing?
- What assumptions might the AI algorithm be making regarding the patient's risk or prognosis?

Evaluate Stage

- What clinical context might not be included in the inputs to the AI algorithm (eg social factors, patient preference, functional status)?
- How confident are you in your clinical assessment? What would increase/decrease your confidence?
- What are the potential consequences if we decide to pursue your recommendation vs. the AI recommendation?
- Under what circumstances would you change your mind?

Decide Stage

- If you disagree with the AI recommendation, how would you document and justify your decision?
- How would you explain your decision to the patient and incorporate their preferences?
- What type of follow-up or safety-netting would be warranted considering the uncertainty?

In addition to the structured questions listed above, educators can use the AI system as a means of understanding the discordance by questioning the AI tool regarding the factors that inform its recommendations, reviewing its confidence scores, or asking why the output generated differs from the learner's assessment. All of these could provide a learning opportunity related to the importance of explainability and transparency principles in clinical AI systems.[40–42] Instructors should facilitate learners' viewing of the AI not simply as a tool that is either accepted or rejected, but as a third partner whose reasoning can be questioned, similar to how one might have a dialogue with a colleague who has a differing clinical opinion.

Faculty Development Recommendations

A course on faculty development should contain a section specifically focused on the way faculty can teach learners in an environment with AI-enhanced learning tools. Faculty development program recommendations are:

- *Learning about AI in Clinical Care*: Understanding basic concepts of clinical AI; gaining familiarity with decision support systems that are available at your educational facility; and identifying their capabilities and limitations.
- *Recognizing AI vs. Learner Decision-Making*: Recognizing when the AI recommendation is different than the clinical decision-making process of the learner and why this is important as a teaching opportunity.
- *Using the SEED Framework to Teach Learners When to Trust AI*: Case studies using simulation that allow faculty to practice identifying and addressing AI-learner discordance.
- *Managing Deference to AI-Recommendations*: Methods of being aware of and minimizing the tendency to uncritically accept recommendations made by AI tools.
- *Finding Balance Between Increasing Efficiency and Including Educational Opportunities*: Methods for creating time for brief reflective discussions during the clinical workflow.

Shared Institutional Policies and Expectations

Developing shared institutional expectations about the use of AI in clinical reasoning can assist in creating consistency in how supervisors respond to AI-learner discordance and reduce variability in supervisory response. The following are examples of institutional policy options that can create a shared expectation of the role of AI in clinical reasoning among educational programs:

- *Statements Regarding Human Accountability*: Clearly defined institutional policies stating that clinical decisions are the responsibility of the treating clinician, not the AI system.
- *Documenting Discrepancies Between AI and Clinician Recommendation*: Standardized methods for documenting instances when a clinician deviates from an AI recommendation with a focus on documenting clinical reasoning.
- *Supervision Competency Criteria Including Discussions of AI-Learner Discordance*: Integrating AI-learner discordance into the competency criteria and expectations for clinical evaluators and supervisors.
- *AI Decision Support System Requirements for Transparency*: Requiring AI decision support systems to provide clear information regarding their input variables, logic, and confidence level to support critical analysis.

Evaluation and Assessment Options

Assessment techniques to evaluate learners' application of critical thinking processes responding to recommendations generated by AI systems include:

- *Work-Based Evaluations*: Observation forms specifically designed to assess how well a learner responds to recommendations from an AI system, particularly when such recommendations conflict with a learner's clinical judgment.
- *Clinical Reasoning Sessions Using Cases with AI-Learner Discordance*: Structured clinical reasoning sessions in which learners discuss specific case studies involving AI recommendations that conflict with their own clinical judgments and provide a rationale for their decisions.
- *Learners' Reflections of Experiences Documented in Their Professional Development Portfolios*: Learners write about their experiences in which AI recommendations conflicted with their own clinical judgments and document what they have learned.
- *AI Recommendations – Simulation Scenarios*: A controlled environment in which learners experience an AI recommendation that conflicts with their own clinical judgment and receive immediate feedback.

Potential Barriers to Implementing the SEED Framework and Possible Solutions

Although the SEED framework provides a systematic method for addressing AI-learner discordance, several barriers to implementing this framework can exist (Table 2). It is crucial to anticipate these obstacles and to develop proactive solutions to overcome them to successfully implement the SEED framework.

Proactive measures to address these barriers will be important to the success of implementing AI in clinical supervision. Medical institutions should provide faculty members with training to better understand how to utilize AI in their daily practice, protected time to reflect on how they are utilizing AI in their own practice, and systems to support educators in using technology.

Longer Term Implications for Medical Education

AI-learner discordance extends beyond individual teaching moments to foundational issues concerning the objectives and methodologies of medical education. As AI tools grow, there may be more instances of discordance between the outputs of an algorithm and human reasoning. Therefore, medical education must begin to move beyond educating students how to apply AI in their own practice and toward training educators how to supervise, contextualize, and critically analyze the application of AI in practice.

Conclusion and Call to Action

Unless educators are provided formalized instructional support to help navigate discrepancies between clinical AI suggestions, clinical reasoning, and decision-making, these challenges will continue to be an obstacle for medical education. Educator response to AI-learner discordance can unintentionally impede students from developing their ability to reason clinically and to form their professional identities. Therefore, this article presented a practical way for educators to integrate AI into clinical education by seeing AI-student discordance as an opportunity to reflect on how they teach and on their role in the educational process.

Educators can use the SEED framework to systematically convert AI-learner discordance into intentional instructional opportunities. By making the conflict between what the student thinks and what the AI suggests visible, examining the reasoning processes that influenced both perspectives, and emphasizing educator accountability, educators can help students develop the critical judgment and professional responsibility to make decisions when using AI in clinical practice.

Table 2 Implementation Barriers and Proposed Solutions

Implementation Disruption	Description	Proposed Solutions
Time Constraints in Clinical Settings	Fast-paced clinical environments limit opportunities for extended reflective discussions	• Develop "micro-debriefing" protocols (2–3 minutes) using key SEED questions • Schedule brief end-of-shift reflections on notable cases
Faculty Unfamiliarity with AI Systems	Many clinical educators lack technical understanding of AI decision support tools	• Mandatory faculty development modules on institutional AI systems • Provide accessible summaries of AI tool capabilities
Institutional Variability in AI Tools	Clinical environments use different AI systems with varying capabilities	• Emphasize SEED framework as tool-agnostic approach • Maintain institutional inventory of AI tools with educational guidance
Lack of Assessment Tools	No validated instruments exist for evaluating learner critical engagement with AI	• Develop AI-specific competency frameworks and assessment rubrics • Integrate AI reasoning into existing clinical assessment tools
Organizational Culture and Resistance	Some institutions or specialties may resist changing traditional supervisory approaches	• Engage clinicians as champions • Frame initiative as enhancing, not replacing, traditional teaching

While the SEED framework provides a structural pedagogical approach for addressing discordance at the point of care, its impact may be strengthened when placed within larger educational structures. Standardized curricula, faculty development, and formalized competency expectations should be created to safely and effectively incorporate AI into the clinical education process. These can act as guidelines for how learners interact with, evaluate, and utilize AI-generated recommendations. Additionally, structured oversight mechanisms are needed to assure that human clinical judgment is integral to the decision-making process and that the use of AI aligns with professional, ethical, and educational expectations.[43] Incorporating these elements into the design of education programs will shift the focus from isolated teaching moments to long-term preparation of clinicians to use AI in their practice.

The use of AI in medicine is inevitable; however, how we educate future generations of clinicians to work with these technologies is not. By converting AI-learner discordance into an educational opportunity, we can ensure that clinicians in the future are educated to function with both technological competence and professional responsibility.

Data Sharing Statement

Not applicable. This article does not involve primary data.

Ethics Approval

This article does not involve human participants. The clinical scenarios presented are hypothetical and created solely for illustrative educational purposes; therefore, ethical approval and informed consent were not required.

Funding

No funding support was acquired.

Disclosure

The author reports no conflicts of interest in this work.

References

1. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New Engl J Med*. 2019;380(14):1347–1358. doi:10.1056/NEJMra1814259
2. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318. doi:10.1001/jama.2017.18391
3. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *Npj Digital Med*. 2020;3(1):17. doi:10.1038/s41746-020-0221-y
4. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. 2018;320(21):2199–2200. doi:10.1001/jama.2018.17163
5. Haug CJ, Drazen JM, Drazen JM, Kohane IS, Leong T-Y. Artificial intelligence and machine learning in clinical medicine, 2023. *New Engl J Med*. 2023;388(13):1201–1208. doi:10.1056/NEJMra2302038
6. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? *Npj Digital Med*. 2020;3(1):86. doi:10.1038/s41746-020-0294-7
7. Wartman SA, Combs CD. Reimagining medical education in the age of AI. *AMA J Ethics*. 2019;21(2):E146–E152. doi:10.1001/amajethics.2019.146
8. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR med educ*. 2019;5(2):e16048. doi:10.2196/16048
9. Masters K. Artificial intelligence in medical education. *Med Teach*. 2019;41(9):976–980. doi:10.1080/0142159X.2019.1595557
10. Lee Y-M, Kim S, Lee Y-H, et al. Defining medical AI competencies for medical school graduates: outcomes of a delphi survey and medical student/educator questionnaire of South Korean medical schools. *Acad Med*. 2024;99(5):524–533. doi:10.1097/ACM.0000000000005618
11. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J Ame Med Informat Assoc*. 2006;13(2):138–147. doi:10.1197/jamia.M1809
12. Nanji KC, Seger DL, Slight SP, et al. Medication-related clinical decision support alert overrides in inpatients. *J Ame Med Informat Assoc*. 2018;25(5):476–481. doi:10.1093/jamia/ocx115
13. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med*. 2021;181(8):1065–1070. doi:10.1001/jamainternmed.2021.2626
14. Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: a systematic review. *J Hospital Med*. 2015;10(6):396–402. doi:10.1002/jhm.2347
15. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean J Radiol*. 2019;20(3):405–410. doi:10.3348/kjr.2019.0025
16. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nature Med*. 2022;28(1):31–38. doi:10.1038/s41591-021-01614-0
17. Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*. 2023;307(4):e222176. doi:10.1148/radiol.222176

18. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Ame Med Informat Assoc*. 2012;19(1):121–127. doi:10.1136/amiajnl-2011-000089
19. Nguyen T. ChatGPT in medical education: a precursor for automation bias? *JMIR Med Educ*. 2024;10:e50174. doi:10.2196/50174
20. Cruess SR, Cruess RL, Steinert Y. Supporting the development of a professional identity: general principles. *Med Teach*. 2019;41(6):641–649. doi:10.1080/0142159X.2018.1536260
21. Cruess RL, Cruess SR, Steinert Y. Amending Miller's pyramid to include professional identity formation. *Acad Med*. 2016;91(2):180–185. doi:10.1097/ACM.0000000000000913
22. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *J Med Ethics*. 2021;medethics–2020–106820. doi:10.1136/medethics-2020-106820
23. London A. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Center Report*. 2019;49(1):15–21. doi:10.1002/hast.973
24. Ghassemi M, Oakden-Rayner L, Beam A. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digital Health*. 2021;3(11):e745–e750. doi:10.1016/S2589-7500(21)00208-9
25. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. doi:10.1126/science.aax2342
26. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature Med*. 2024;30(9):2613–2622. doi:10.1038/s41591-024-03097-1
27. Billett S. Learning through health care work: premises, contributions and practices. *Med Educ*. 2016;50(1):124–131. doi:10.1111/medu.12848
28. Kennedy TJT, Regehr G, Baker GR, Lingard L. Point-of-care assessment of medical trainee competence for independent clinical work. *Acad Med*. 2008;83(10 Suppl):S89–S92. doi:10.1097/ACM.0b013e318183c8b7
29. Thurzo A. Provable AI ethics and explainability in medical and educational AI agents: trustworthy ethical firewall. *Electronics*. 2025;14(7):1294. doi:10.3390/electronics14071294
30. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ*. 2008;42(5):468–475. doi:10.1111/j.1365-2923.2008.03030.x
31. Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. *Acad Med*. 2017;92(1):23–30. doi:10.1097/ACM.0000000000001421
32. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. 2009;84(8):1022–1028. doi:10.1097/ACM.0b013e3181ace703
33. Norman G, Pelaccia T, Wyer P, Sherbino J. Dual process models of clinical reasoning: the central role of knowledge in diagnostic expertise. *J Evalu Clin Prac*. 2024;30(5):788–796. doi:10.1111/jep.13998
34. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*. 2015;57(3):407–434. doi:10.1177/0018720814547570
35. Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Human Factors*. 2004;46(1):50–80. doi:10.1518/hfes.46.1.50.30392
36. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *Npj Digital Med*. 2021;4(1):31. doi:10.1038/s41746-021-00385-9
37. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. *Med Teach*. 2011;33(3):188–199. doi:10.3109/0142159X.2011.550965
38. Lave J, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press; 1991.
39. Penner JC, Schuwirth L, Durning SJ. From noise to music: reframing the role of context in clinical reasoning. *J General Internal Med*. 2024;39(5):851–857. doi:10.1007/s11606-024-08612-1
40. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inf Decis Making*. 2020;20(1):310. doi:10.1186/s12911-020-01332-6
41. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Network Open*. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969
42. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. presented at: Proceedings of the 4th Machine Learning for Healthcare Conference; 2019; Proceedings of Machine Learning Research. https://proceedings.mlr.press/v106/tonekaboni19a.html.
43. Thurzo A How is AI transforming medical research, education and practice?. *Bratisl Lek Listy*. 2025;126(3):243–248. doi:10.1007/s44411-025-00063-2