Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1296–1305
https://doi.org/10.54941/ahfe1008358

From Recommendation Correctness to Clinical Action: Transparency, Trust, and Verification in AI Decision Support

Sim Dhami, Dennis Vincenzi, and Kat Moran

Embry-Riddle Aeronautical University, USA

ABSTRACT

Artificial intelligence assisted clinical decision support systems are commonly evaluated using aggregate measures such as accuracy, sensitivity, and specificity. These measures do not establish whether a recommendation is dependable for an individual case. Clinical safety also depends on how interfaces communicate reliability and how clinicians interpret, verify, and use the output. This paper examines the pathway connecting AI recommendation correctness to clinical action through transparency, trust calibration, and verification. A qualitative, theory guided integrative framework approach was used to synthesize peer reviewed empirical and theoretical literature identified through Scopus, Web of Science Core Collection, and PubMed. The full multilevel safety framework provides context for a focused conceptual pathway linking recommendation correctness, interface communication, clinician reliance, verification, and action. The reviewed evidence suggests that correct recommendations can improve diagnostic performance, while incorrect recommendations can impair judgment when accepted without sufficient evaluation. Explanations, confidence displays, and uncertainty cues may help clinicians assess recommendations, but can also create false confidence or additional cognitive demands. Their value depends on whether they support interpretation and verification during clinical work. Clinician responses may include questioning, modifying, independently checking, or overriding an output. The proposed pathway identifies opportunities for interface design to support informed reliance and preserve clinician agency. It emphasizes communicating uncertainty and helping clinicians compare recommendations with clinical evidence before acting. These relationships remain conceptual and require empirical evaluation. Future studies should examine how reliability and uncertainty displays affect verification, reliance, diagnostic performance, and workload in realistic clinical workflows.

Keywords: Clinical decision support systems, Recommendation correctness, Interface transparency, Trust calibration, Automation bias, Clinician verification, Diagnostic safety

INTRODUCTION

Artificial intelligence assisted clinical decision support systems are increasingly used to help clinicians interpret patient information and make diagnostic decisions (Abbas et al., 2025). They can identify findings, prioritize cases, and offer recommendations during clinical review (Abbas et al., 2025).

Received October 16, 2026; Revised November 17, 2026; Accepted November 27, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

These systems are commonly evaluated using measures such as accuracy, sensitivity, and specificity. Those measures describe performance across a group of cases, but they cannot establish whether a recommendation is dependable for the patient currently being assessed (Cabitza, 2026). In patient care, the margin for error is non-existent; erroneous decisions can lead to profound, irreversible patient outcomes.

The effect of AI assistance depends partly on whether its recommendation is correct. Correct recommendations can help clinicians reach an accurate diagnosis, while incorrect recommendations can lead them away from it when the output is accepted without sufficient evaluation (Kücking et al., 2026). Clinicians may also use AI output to confirm an initial impression rather than critically evaluate it (Romeo & Conti, 2026). A model may perform well overall without providing equally reliable recommendations for every individual case (Cabitza, 2026). Decisions and recommendations made by AI must be checked by a live clinician, who must agree that they are viable, plausible, and accurate.

The interface influences that judgment. Confidence scores, explanations, and uncertainty cues affect how clinicians interpret a recommendation and whether they feel further checking is necessary (Rezaeian et al., 2026). Clear information about a recommendation's basis and limitations may help clinicians recognize when to investigate further (Antoniadi et al., 2021). However, an explanation can also make an incorrect output appear more convincing, while too much information can make the task harder to manage (González-Alday et al., 2023). Transparency helps when clinicians can use it to understand and evaluate a recommendation during clinical work.

Trust and verification are closely connected to this process. Clinicians need to rely on useful recommendations without treating the system's output as a substitute for their own judgment. Overreliance can allow an incorrect recommendation to influence the final decision, while underreliance can lead clinicians to dismiss support that might have helped them (Romeo & Conti, 2026). Verification provides an opportunity to catch an error before it becomes a diagnostic action (Romeo & Conti, 2026). Clinicians may also question or modify a recommendation rather than simply accept or reject it (Salimparsa et al., 2025).

This paper examines the pathway from AI recommendation correctness to clinical action through transparency, trust calibration, and verification behavior. It draws on a qualitative, theory guided synthesis of empirical and theoretical literature conducted for a broader multilevel study of safety in AI assisted clinical decision support. The paper focuses on how an AI recommendation becomes part of a clinician's diagnostic reasoning. The analysis identifies how interface design can support informed reliance and verification before clinical action.

METHOD

This paper presents a focused analysis of evidence gathered for a completed thesis on safety in artificial intelligence assisted clinical decision support systems. The thesis used a qualitative, theory guided integrative framework approach to examine how model behavior, clinician activity, interface design, and wider clinical conditions interact. This paper concentrates on the part of that framework connecting recommendation correctness to clinical action. No new participant data, clinical records, or experimental observations were collected.

The thesis literature search covered Scopus, Web of Science Core Collection, and PubMed. The final literature search was conducted on April 26, 2026. Search terms combined artificial intelligence or clinical decision support with clinical context, human factors concepts, and outcomes such as diagnostic accuracy, reliance, verification, monitoring, or error detection. Searches focused on English language peer reviewed literature published from 2010 to 2026. Earlier foundational sources were retained where they were needed to explain established human factors concepts. Records were organized and deduplicated in Zotero before titles, abstracts, and full texts were screened for relevance.

The focused analysis examined recommendation correctness, the communication of confidence and uncertainty, and clinicians' responses to AI output. Sources were selected for their relevance to recommendation correctness, interface communication, or clinicians' interpretation and use of AI output.

Relevant findings were examined as relationships between the conditions under which a recommendation is presented and the actions a clinician may take in response. For example, the analysis considered whether a confidence cue helped a clinician recognize uncertainty, encouraged reliance without further checking, or made the task more difficult. Evidence from clinical AI research was considered alongside human factors literature on trust, automation bias, and reliance. The relationships were then organized into a conceptual pathway from recommendation correctness, through interface communication and clinician interpretation, to verification and diagnostic action.

The analysis is interpretive rather than statistical. It does not pool effect sizes or test whether transparency or trust causes a particular clinical outcome. Instead, it uses findings from the reviewed literature to identify where an AI error may be caught or carried forward into a decision. The broader framework draws on a multilevel sociotechnical approach to examining interactions across a work system (Brady & Naikar, 2022).

Figure 1 presents the broader multilevel framework that informed this analysis. It connects policy and organizational conditions with local workflow, interface design, clinician activity, and patient outcomes. The framework also includes monitoring and feedback that can inform changes across these levels. The present paper focuses on the recommendation correctness and transparency pathways, examining how clinicians interpret AI output, judge its usefulness, and verify it before acting.

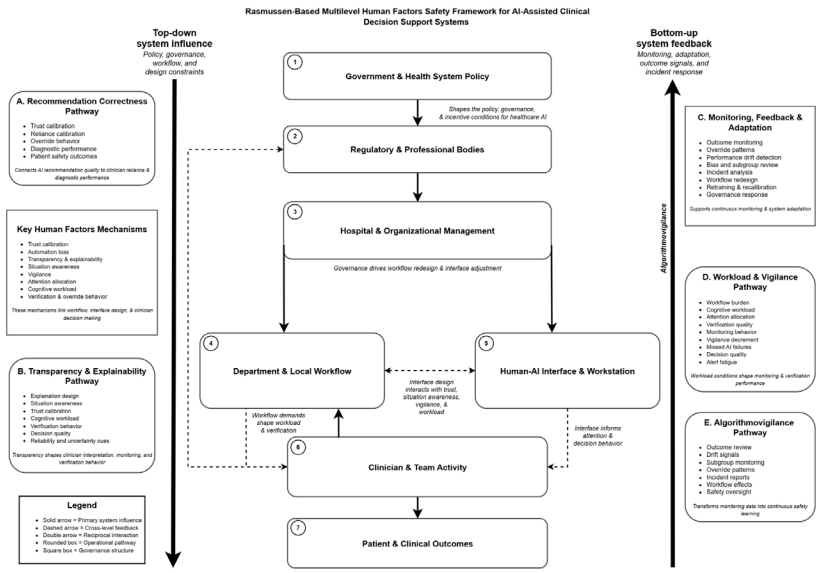


Figure 1: Rasmussen based multilevel human factors safety framework for AI assisted clinical decision support systems.

FINDINGS

Recommendation Correctness and Diagnostic Performance

The evidence showed that the effect of AI support depends in part on whether the recommendation is correct for the case being reviewed. In a simulated diagnostic study involving physicians and nurses assessing wound images, diagnostic accuracy was 92.1% when AI recommendations were correct and 55.6% when they were incorrect (Kücking et al., 2026). These results illustrate why average model performance does not fully describe what happens when a clinician encounters an individual recommendation. Participants were substantially more accurate when the AI recommendation was correct than when it was incorrect (Kücking et al., 2026).

A model's overall accuracy cannot tell a clinician whether its current recommendation is right. Errors may occur in cases involving patient groups, images, or clinical conditions that are poorly represented in the data used to evaluate the model (Cabitza, 2026). This creates a practical problem at the point of care: the clinician must decide how much weight to give a recommendation before its correctness is known. Information about uncertainty and case level reliability may help with that judgment, but it cannot replace review of the patient and the available clinical evidence (Cabitza, 2026).

Transparency and the Interpretation of AI Output

The interface determines what clinicians can see about a recommendation and how easily they can evaluate it. Explanations may help clinicians understand the basis of an output and compare it with the evidence in the current case

(Antoniadi et al., 2021). Confidence and uncertainty cues can also influence whether clinicians accept a recommendation or seek further confirmation (Rezaeian et al., 2026). These features are most useful when they make the recommendation easier to assess during the clinical task.

The benefits of transparency depended on how explanations were designed and used. An explanation may make an output seem convincing even when the recommendation is incomplete or incorrect (González-Alday et al., 2023). Explanations may complicate interpretation when they do not match the task or the user's information needs (González-Alday et al., 2023). The amount of information displayed does not establish whether an interface supports safe use. The more relevant question is whether clinicians can use that information to identify uncertainty, compare the output with clinical evidence, and recognize when further checking is needed.

This distinction is especially important for confidence information. A confidence display can influence reliance, but clinicians still need to understand what the displayed value means for the particular recommendation and what it does not establish (Rezaeian et al., 2026). Case level reliability information may be more useful than a general statement about model accuracy when the immediate task is deciding how to respond to one patient's result (Cabitza, 2026).

Trust, Reliance, and Verification

Trust calibration concerns how closely a clinician's confidence in the system reflects its capabilities and limitations (Romeo & Conti, 2026). Reliance concerns how the clinician uses its recommendations. Clinicians can benefit from correct recommendations while remaining able to question outputs that are uncertain or inconsistent with the clinical evidence (Romeo & Conti, 2026). When reliance is too high, an incorrect recommendation may receive more weight than it deserves. When reliance is too low, a useful recommendation may be dismissed without adequate consideration (Romeo & Conti, 2026).

Verification offers an opportunity to detect an AI error before it influences the final decision. Comparing the recommendation with patient history, clinical presentation, imaging findings, and independent judgment can reveal a disagreement that requires further review (Romeo & Conti, 2026). This does not mean that every recommendation leads to a simple choice between acceptance and rejection. Clinicians may use part of a recommendation, question another part, modify their interpretation, or seek additional evidence before deciding what to do (Salimparsa et al., 2025).

The depth of that review matters. A clinician may appear to have checked an output while still using the AI recommendation mainly to confirm an initial impression (Romeo & Conti, 2026). Active comparison with clinical evidence can help clinicians detect errors in AI recommendations (Romeo & Conti, 2026). The interface should make disagreements visible and practical to investigate. Retaining final decision authority may provide limited protection when clinicians lack a practical opportunity to evaluate the recommendation (Romeo & Conti, 2026).

The Pathway From Recommendation to Action

Figure 2 connects recommendation correctness with the clinician's interpretation, verification, and subsequent action. First, the AI produces a recommendation that may be correct or incorrect for the individual case. The interface then presents that recommendation, along with whatever explanation, confidence, or uncertainty information is available. The clinician interprets this information, decides how much to rely on it, and may verify, question, modify, accept, or override the output. Those actions determine how the recommendation contributes to the diagnostic decision.

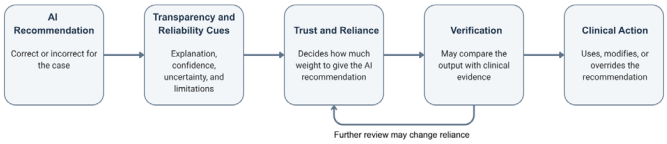


Figure 2: Conceptual pathway from AI recommendation correctness to clinical action.

This pathway is a conceptual synthesis of the reviewed evidence. It does not mean that clinicians move through a fixed sequence or that this study measured the causal effect of each step. Trust can change while a clinician examines the case, and verification may lead them back to the recommendation or to other clinical information. An incorrect recommendation can be caught before it affects a decision, while a correct one can provide useful support when it is understood and applied in context.

DISCUSSION

The Interface as Part of the Safety Pathway

The findings place the clinical interface between a model's output and the clinician's decision. At this point, a recommendation becomes something the clinician must interpret and weigh against the rest of the case. The interface can help by showing information that is relevant to that judgment, including uncertainty and the basis for the recommendation. It can also make evaluation harder if the output appears more certain than the evidence supports or if important information is difficult to find.

A useful design goal is to help clinicians distinguish a model's general performance from the reliability of a recommendation for the current case. Aggregate accuracy does not resolve uncertainty about a particular patient or image (Cabitza, 2026). Confidence displays should explain what the value represents and how it relates to the current decision. High confidence may encourage reliance even when independent evaluation is still needed (Rezaeian et al., 2026). Designers should assess whether reliability information helps clinicians identify cases that deserve closer review, rather than assuming that displaying a score is sufficient.

Explanations require the same care. A concise explanation may help a clinician compare an AI output with the patient history or imaging findings (Antoniadi et al., 2021). A persuasive explanation may instead

increase confidence in an incorrect recommendation without helping the clinician detect the error (González-Alday et al., 2023). Design evaluations should establish whether clinicians can use the information to check the recommendation during the task.

Making Verification Practical

The findings suggest that verification should be supported by the system and the workflow, not left entirely to individual vigilance. Clinicians may be better able to detect an incorrect recommendation when they can compare it with relevant clinical evidence before making a final decision (Romeo & Conti, 2026). An interface could make that comparison easier by keeping the underlying evidence accessible, allowing clinicians to inspect the basis for an output, and making it straightforward to record disagreement or override the recommendation. These are design implications of the synthesis and still need to be tested with clinicians in realistic tasks.

Verification also takes time. AI assistance can reduce interpretation time when it helps clinicians identify relevant findings (Chen et al., 2024). That benefit may be weakened if a faster workflow leaves no opportunity to examine uncertain or conflicting recommendations. In a knee MRI reader study, AI assistance improved diagnostic accuracy and reduced reading time (Herpe et al., 2026). Whether these time savings allow more careful review of uncertain recommendations in clinical practice remains to be tested. Studies should examine whether clinicians have the time and access to evidence needed to question a recommendation before acting.

This is particularly relevant when clinicians appear to retain final authority. A requirement to approve an AI recommendation provides limited protection if the interface encourages a quick confirmation without meaningful review. Clinicians may also use an output selectively, revise it, or seek more evidence before acting, so an acceptance rate alone cannot describe the quality of their evaluation (Salimparsa et al., 2025).

Evaluating Transparency Through Clinician Behavior

Transparency should be evaluated by what it helps clinicians do. A display may be understandable or well liked while still failing to help users detect an incorrect recommendation. Future evaluations should compare how different displays affect error detection, the decision to verify, the depth of verification, diagnostic accuracy, and workload. They should include both correct and incorrect recommendations so that increased reliance is not mistaken for improved performance.

The proposed pathway also suggests examining overreliance and underreliance together. A display that encourages clinicians to use more AI recommendations could be helpful when the recommendations are correct and harmful when they are wrong (Romeo & Conti, 2026). A display that encourages caution may help clinicians avoid errors, but it could also lead them to dismiss useful recommendations (Romeo & Conti, 2026). The goal is to help clinicians judge when an AI recommendation is useful and when further checking is needed.

Evaluation should extend beyond a single session. Clinicians' expectations may change as they gain experience with a system, and model performance may shift after deployment as patient populations and clinical practices change (Salwei et al., 2025). Patterns of disagreement, overrides, false positives, and false negatives can provide information about whether the system remains useful and safe in its local setting (Salwei et al., 2025).

Limitations and Future Research

This paper presents a focused synthesis of published literature, not a direct test of the pathway it proposes. The proposed relationships involving trust calibration and verification were not tested experimentally or evaluated as statistical mediators. Studies also differ in their clinical tasks, AI tools, participants, and measures of performance, which limits how broadly any single finding can be applied. The demands of image interpretation, triage, and other forms of decision support may produce different patterns of reliance and verification (Di Fazio et al., 2026).

The next step is to test the pathway in realistic clinical workflows. A study could compare alternative ways of presenting recommendation reliability and uncertainty while observing how clinicians respond to correct and incorrect AI outputs. It should measure diagnostic decisions alongside verification behavior and workload, and examine whether any benefit remains when time pressure and routine workflow demands are present. This would help establish which interface features improve judgment in practice and where additional support is needed.

CONCLUSION

The safety of AI assisted clinical decision support depends on more than how accurately a model performs across many cases. A recommendation must still be interpreted and acted on in the context of an individual patient. Correct recommendations can support diagnostic reasoning, while incorrect ones can lead clinicians toward an error when they are accepted without sufficient checking (Kücking et al., 2026).

This paper brings together three parts of that process: what the AI recommends, how the interface communicates reliability and uncertainty, and how the clinician responds. Explanations and confidence cues may help clinicians evaluate an output, but they can also create misplaced confidence or add work when they are poorly designed (González-Alday et al., 2023). Verification provides an opportunity to identify an error before it becomes part of a diagnostic decision.

The pathway identifies where interface design and verification may influence the use of AI recommendations. It remains a conceptual synthesis and requires testing in realistic clinical workflows. Future studies should assess whether different ways of presenting uncertainty help clinicians use correct recommendations, detect incorrect ones, and retain the ability to question or override AI output.

ACKNOWLEDGMENT

AI writing and editing tools were used to support drafting, language refinement, and organization. The research and conceptual framework were developed in the first author's thesis. The authors remain responsible for the interpretation of the evidence and the accuracy of the paper.

REFERENCES

Abbas, Q., Jeong, W. and Lee, S. W. (2025). Explainable AI in clinical decision support systems: A meta-analysis of methods, applications, and usability challenges. *Healthcare*, 13(17), article 2154. https://doi.org/10.3390/healthcare13172154

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A. and Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review. *Applied Sciences*, 11(11), article 5088. https://doi.org/10.3390/app11115088

Brady, A. and Naikar, N. (2022). Development of Rasmussen's risk management framework for analysing multi-level sociotechnical influences in the design of envisioned work systems. *Ergonomics*, 65(3), pp. 485–518. https://doi.org/10.1080/00140139.2021.2005823

Cabitza, F. (2026). Calibration-informed metrics for instance-level predictive reliability in medical AI. *Artificial Intelligence in Medicine*, 174, article 103366. https://doi.org/10.1016/j.artmed.2026.103366

Chen, M., Wang, Y., Wang, Q., Shi, J., Wang, H., Ye, Z., Xue, P. and Qiao, Y. (2024). Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. *npj Digital Medicine*, 7(1), article 349. https://doi.org/10.1038/s41746-024-01328-w

Di Fazio, N., Zanza, C., Longhitano, Y., Voza, A., Balagna, R., Mosca, S., Balagna, P., Rossignoli, R., Cerenzia, S., Bertozzi, G., Maiese, A., Frati, P. and La Russa, R. (2026). Artificial intelligence for early diagnosis in emergency department. *Journal of Anesthesia, Analgesia and Critical Care*, 6(1), article 7. https://doi.org/10.1186/s44158-025-00334-y

González-Alday, R., García-Cuesta, E., Kulikowski, C. A. and Maojo, V. (2023). A scoping review on the progress, applicability, and future of explainable artificial intelligence in medicine. *Applied Sciences*, 13(19), article 10778. https://doi.org/10.3390/app131910778

Herpe, G., Vesoul, T., Zille, P., Pluot, E., Guillin, R., Rizk, B., Ardon, R., Adam, C., d'Assignies, G. and Gondim Teixeira, P. A. (2026). Impact of AI assistance on radiologist interpretation of knee MRI. *European Radiology*, 36(2), pp. 1294–1305. https://doi.org/10.1007/s00330-025-11820-w

Kücking, F., Busch, D. A., Przysucha, M., Kutza, J.-O., Hannemann, N., Hüsers, J., Babitsch, B. and Hübner, U. (2026). Impact of AI recommendation correctness on diagnostic accuracy in clinical decision-making. *International Journal of Medical Informatics*, 207, article 106223. https://doi.org/10.1016/j.ijmedinf.2025.106223

Rezaeian, O., Bayrak, A. E. and Asan, O. (2026). Explainability and AI confidence in clinical decision support systems: Effects on trust, diagnostic performance, and cognitive load in breast cancer care. *International Journal of Human-Computer Interaction*, 42(6), pp. 4477–4497. https://doi.org/10.1080/10447318.2025.2539458

Romeo, G. and Conti, D. (2026). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & Society*, 41(1), pp. 259–278. https://doi.org/10.1007/s00146-025-02422-7

Salimparsa, M., Sedig, K., Lizotte, D. J., Abdullah, S. S., Chalabianloo, N. and Muanda, F. T. (2025). Explainable AI for clinical decision support systems: Literature review, key gaps, and research synthesis. *Informatics*, 12(4), article 119. https://doi.org/10.3390/informatics12040119

Salwei, M. E., Davis, S. E., Reale, C., Novak, L. L., Walsh, C. G., Beebe, R., Nelson, S., Sundrani, S., Rose, S., Wright, A., Ripperger, M., Shave, P. and Embí, P. (2025). Human-centered design of an artificial intelligence monitoring system: The Vanderbilt Algorithmovigilance Monitoring and Operations System. *JAMIA Open*, 8(5), article ooaf136. https://doi.org/10.1093/jamiaopen/ooaf136