Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2099–2110
https://doi.org/10.54941/ahfe1008436

An Entrustment-Based Human Factors Framework for Integrating Artificial Intelligence Into Healthcare and Medical Education

Bryan Johnston[1,2], Scott Adams[3], Kiven Erique Lukong[4], and Jay Kalra[1,2]

[1]Department of Pathology and Laboratory Medicine, College of Medicine, University of Saskatchewan, Saskatoon, Canada
[2]Royal University Hospital, Saskatchewan Health Authority, 103 Hospital Drive, Saskatoon, Saskatchewan, S7N 0W8, Canada
[3]Department of Medical Imaging, College of Medicine, University of Saskatchewan, Saskatoon, Canada
[4]Department of Biochemistry, Microbiology and Immunology, College of Medicine, University of Saskatchewan, 107 Wiggins Road, Saskatoon, Saskatchewan, S7N 5E5, Canada

ABSTRACT

Artificial intelligence (AI) is increasingly transforming medical education and clinical practice through image interpretation, risk prediction, diagnostic decision support, and automated documentation. As learners encounter these technologies, medical education must prepare them not only to understand and use AI, but also to critically evaluate when to trust, question, or override AI outputs. Existing approaches to AI education commonly emphasize digital literacy, technical understanding, and ethical use, yet provide limited guidance regarding how supervisors should determine whether a learner can use AI responsibly with increasing independence. This work proposes an entrustment-based human factors framework for integrating AI into healthcare and medical education. The framework brings together three elements: human factors principles, which emphasize the interaction among clinicians, learners, technologies, tasks, teams, and clinical environments; the patient-centered CARES framework (Communication, Autonomy, Respect, Equity, and Safety) which defines the responsibilities that should guide AI-supported care; and entrustment, which links these responsibilities to observable learner behaviours, supervision decisions, and progressive independence. Within this framework, Communication requires clinicians to explain the role, limitations, and uncertainty of AI-supported decisions. Autonomy requires preservation of meaningful patient and clinician choice. Respect includes attention to privacy, dignity, and individual context. Equity requires recognition of biased or poorly generalizable performance. Safety requires active monitoring, appropriate human oversight, and thus the prevention of harm. We have observed that AI introduces a second-order entrustment challenge into medical education. Supervisors must determine not only when learners can be trusted to perform a clinical task, but whether they can be trusted to judge when, why, and to what extent an AI-enabled system should itself be trusted. This requires a metacognitive capacity for trust calibration. Learners must recognize

Received July 29, 2026; Revised September 23, 2026; Accepted October 5, 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/

the limitations of both human and algorithmic judgment, communicate uncertainty, and seek guidance when the limits of safe independent action are reached. As learners progress toward independent practice, they increasingly assume a learner-supervisor role in which they are responsible for overseeing the contribution of AI to patient care. Within this framework, entrustment decisions can help supervisors determine when learners are ready for greater independence in using AI-supported tools while maintaining appropriate human oversight, continuous learning, and patient safety. We recommend integrating CARES-informed competencies into entrustment decisions in order to provide a practical foundation for curriculum development, assessment, and the safe incorporation of AI into healthcare and medical education.

Keywords: Artificial intelligence, Human factors, Entrustment, Patient safety, Medical education, Quality care

INTRODUCTION

Artificial intelligence (AI) increasingly contributes to clinical interpretation, decision support, and documentation. These applications may improve efficiency, strengthen pattern recognition, and allow clinicians to devote more attention to the human aspects of care. Their value, however, depends on careful implementation and a clear understanding of the opportunities and limits of each technology (Kalra and Johnston, 2025; Kalra and Seitzinger, 2022). Available review evidence does not establish that machine learning-based decision support consistently improves clinician diagnostic performance (Vasey et al., 2021). Learners are therefore not simply learning about AI. They are increasingly performing clinical activities in which AI shapes how information is interpreted, communicated, and translated into action.

Medical education has responded by emphasizing AI literacy, critical appraisal, and ethical use. These capabilities are necessary, but they do not tell a supervisor whether a learner is ready to use AI with increasing independence. A correct outcome is not enough because it may conceal what the learner understood, how the AI influenced the decision, and whether uncertainty was recognized (Abdulnour et al., 2025; Preiksaitis et al., 2026; Pupic et al., 2023). Guidance for AI-assisted practice also remains incomplete, and it is uncertain whether clinicians are adequately prepared to use and monitor these systems (Schuitmaker et al., 2025). The unresolved question is therefore not only what learners should know, but what supervisors should observe and how those observations should shape future supervision.

This is a human factors problem because performance does not arise from the learner or the AI product alone. It emerges from their interaction within a clinical task, workflow, team, and environment (Bienefeld et al., 2024; Holden et al., 2013; Salwei and Carayon, 2022). Workload, interface design, communication, hierarchy, and organizational expectations may support or undermine otherwise capable people and technologies (Goddard et al., 2012). A systems perspective avoids reducing unsafe performance to individual incompetence while preserving the learner's responsibility to exercise clinical judgment, communicate uncertainty, and seek assistance. AI may contribute to team performance, but it should complement rather than replace human

judgment. Accountability and entrustment remain with people (Kalra and Johnston, 2025).

AI therefore introduces a second-order entrustment challenge. Supervisors must determine not only whether a learner can perform the clinical activity, but whether the learner can judge when an AI contribution should be trusted, questioned, verified, overridden, or escalated. This requires trust calibration and recognition of the limits of both human and algorithmic judgment (Sakamoto et al., 2024; Teng et al., 2026). This paper develops a literature-informed, entrustment-based human factors framework that brings together human factors, the patient-centered CARES domains of Communication, Autonomy, Respect, Equity, and Safety, and entrustment. The framework organizes observable learner behaviours and links them to decisions about supervision and progressive independence.

METHODS

We conducted a targeted, literature-informed conceptual synthesis across human factors and human-automation interaction, patient-centered responsibilities for clinical AI, and competency-based medical education and entrustment. Sources were screened for their relevance to observable AI-supported performance, patient-centered responsibilities, and supervision, and the included evidence was mapped to the proposed framework. The synthesis was intended to support framework development rather than estimate effects or produce a systematic evidence map.

Framework development focused on a practical decision about the level of responsibility and supervision appropriate when a learner performs a clinical activity in which AI contributes. We described performance as arising from the interaction among the learner, AI technology, task, patient, team, workflow, supervision, and clinical environment. We then translated the CARES domains into observable learner behaviours and linked those observations to prospective entrustment decisions. Product validation, authorization, data security, and ongoing monitoring remain organizational responsibilities. Learner assessment concerns how AI is used within the professional activity.

RESULTS

The proposed framework is a three-part decision framework for assessing AI-supported learner performance. Human factors identifies the conditions that shaped performance. CARES specifies the patient-centered responsibilities that should be observed. Entrustment uses that evidence to determine the level of supervision and the conditions for future performance. The learner remains the person entrusted to perform the professional activity, while AI remains a bounded contribution within it. Each component serves a different purpose and should not be collapsed into a single score. Figure 1 shows how human factors conditions shape the learner's enactment of CARES responsibilities and generate observable evidence for entrustment and progressive independence.

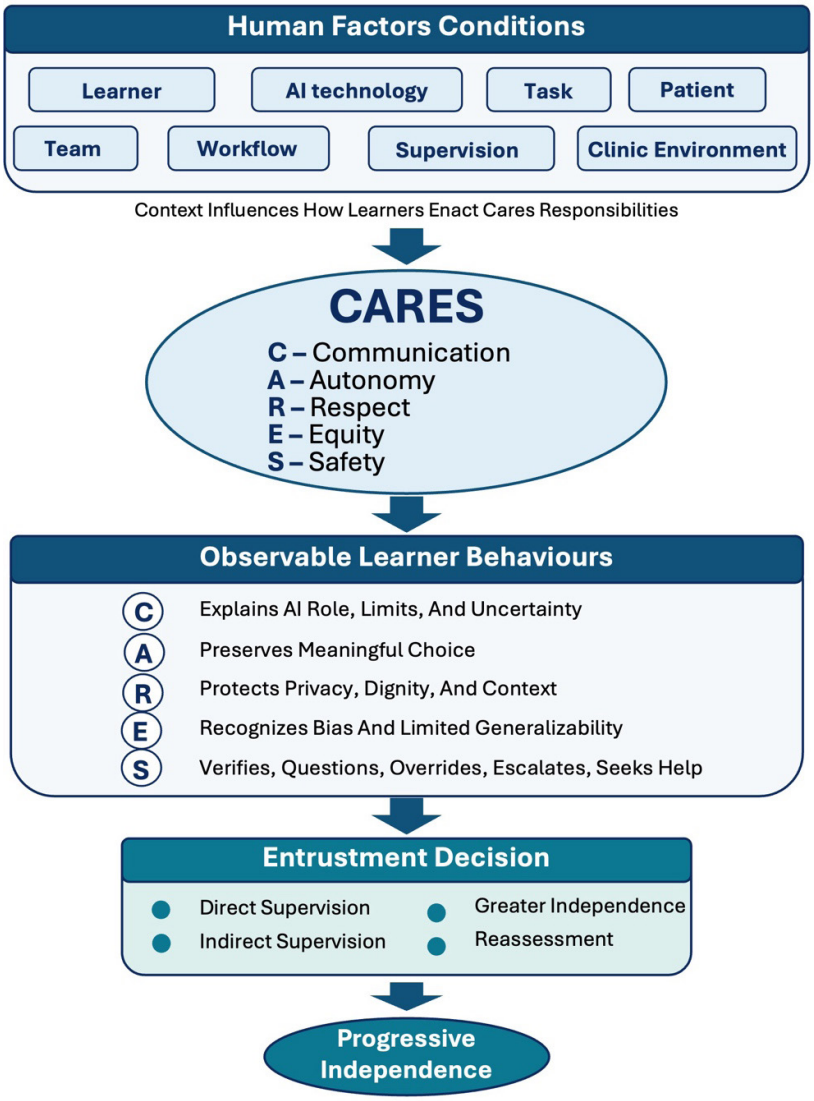


Figure 1: Human factors conditions shape how CARES responsibilities become observable evidence for entrustment and progressive independence.

Human Factors Conditions Shape Learner-AI Performance

A capable learner and an authorized AI product do not by themselves establish safe performance (Vasey et al., 2021). Supervisors also need evidence about how AI entered the learner's reasoning and how the surrounding conditions shaped what occurred. Interaction evidence includes the information supplied to the learner, how the output was presented and interpreted, and whether it was accepted, questioned, verified, rejected, or overridden. Contextual evidence includes the demands of the task, the needs of the patient, the interface, workflow, team, and supervision available at the time (Abdulnour et al., 2025; Bienefeld et al., 2024; Vasey et al., 2022). An unsafe outcome may reflect difficult system conditions rather than learner ability, while a correct outcome may conceal automation bias or overreliance.

Context is therefore explanatory evidence rather than background. Supervisors may draw on direct observation, discussion with the learner, work products, debriefing, selected interaction records, and workflow observation. Evidence should be collected across more than one case because a single encounter cannot establish stable competence. Learners should not be judged negatively when the situation did not provide an opportunity to demonstrate a particular behaviour (Hauer, 2019; Kogan et al., 2009; Ten Cate, 2013).

CARES Defines the Responsibilities to Observe

CARES makes patient-centered responsibilities visible in the learner's use of AI (see Table 1). Communication involves explaining the role of AI, its benefits and limitations, and the uncertainty that remains. Autonomy preserves meaningful choice for both patients and clinicians rather than allowing an AI recommendation to determine the decision (Lorenzini et al., 2023). Respect protects privacy, dignity, cultural values, and the individual circumstances of the patient. Equity requires learners to recognize possible bias, subgroup underperformance, and limits to generalizability without making them responsible for correcting product-level deficiencies (Cross et al., 2024). Safety involves proportionate verification, questioning, override, reporting, escalation, and help-seeking when the limits of safe action are reached (Johnston and Kalra, 2025; Pupic et al., 2023; Vo et al., 2023).

Patient-perspective studies support the importance of disclosure, continued human oversight, and privacy when AI contributes to care (Foresman et al., 2025; Park, 2024). The strength of evidence differs across the CARES domains. Communication, Autonomy, Equity, and Safety are supported by empirical and review-level evidence, while Respect remains grounded mainly in broader principles of patient-centered care and has not yet been independently validated as an assessment domain. CARES does not create new professional duties or a separate entrustment decision. It organizes the responsibilities that supervisors should observe when a learner uses AI.

Table 1: CARES-informed observation of AI-supported clinical performance.

CARES Domain	Learner Behaviour	Evidence	Signal for Supervision
Communication	Explains the material role, limitations, uncertainty, and possible error	Presentation, patient discussion, documentation	Presents output as fact or conceals material uncertainty
Autonomy	Preserves patient choice, shared decision-making, and clinical judgment	Observed encounter, case discussion	Treats the recommendation as determinative
Respect	Protects privacy, dignity, cultural values, and individual context	Workflow observation, documentation	Disregards privacy, preferences, dignity, or context
Equity	Recognizes bias, subgroup limitations, and restricted generalizability	Case reasoning, structured questioning	Applies output despite population or context mismatch
Safety	Verifies, questions or overrides, escalates concerns, and seeks help	Observation, work product, audit trail, near-miss review	Fails to verify, escalate, or seek assistance

Entrustment Converts Evidence Into Progressive Independence

An Entrustable Professional Activity (EPA) brings together the capabilities required to perform a defined professional activity. Entrustment uses observed performance to make a prospective decision about what the learner may do next and with what supervision. In AI-supported care, the learner remains the person entrusted to perform the activity. AI contributes to the work but does not become the object of learner entrustment. AI-related behaviours can therefore be incorporated into existing EPAs whenever the technology materially affects performance, rather than creating a separate EPA for every tool (Gin et al., 2025; Jacobs et al., 2023; Ten Cate, 2013; Ten Cate and Scheele, 2007). Other frameworks evaluate whether an AI tool is sufficiently trustworthy for a particular educational task (Gin et al., 2025; Preiksaitis et al., 2026). The present framework is complementary because it asks whether the learner can safely oversee the AI contribution within a clinical activity.

Progressive independence should reflect what the learner demonstrates in practice (Kennedy et al., 2005). Direct supervision remains appropriate when the learner can identify the AI contribution but cannot yet explain, verify, or act on it safely. Indirect supervision becomes more appropriate when the learner consistently recognizes limitations, communicates uncertainty, verifies the output, and seeks help when needed. Evidence from non-AI settings suggests that effective supervision can improve both patient care and educational outcomes (Farnan et al., 2012). Greater independence should still remain within the learner's scope and should reflect the clinical risk, the permitted use of AI, and the availability of assistance. We propose that learners should supervise others only after demonstrating reliable oversight of both the clinical activity and the AI contribution. This learner-supervisor role is a conceptual part of the framework and has not yet been established empirically. The supervision levels are therefore illustrative rather than a validated universal scale. Table 2 shows how authorization of the AI contribution and entrustment of the learner jointly determine whether AI-supported performance may proceed.

Table 2: Relationship between AI authorization and learner entrustment.

	Entrustment Sufficient for the Proposed Level of Supervision	**Entrustment Not Yet Sufficient for the Proposed Level of Supervision**
AI contribution authorized	Proceed with the AI-supported activity within the authorized scope and supervision conditions	Continue direct supervision and training before reducing supervision
AI contribution not authorized	Perform the activity without AI, within the learner's entrusted scope	Do not perform independently; supervised training may continue without the unauthorized AI contribution

Consider a learner using an institutionally authorized AI tool during diagnostic image interpretation. The tool may identify an abnormality, direct attention to a region of interest, or suggest a classification. A correct recommendation would not justify reduced supervision if the learner could

not review the source information, explain how the output informed the decision, recognize important limitations, communicate uncertainty, or seek help when disagreement could not be safely resolved. Conversely, a learner already entrusted to perform the activity could continue without AI if its use was no longer authorized or appropriate. Entrustment therefore applies to the learner's responsibility for the professional activity, not to the product itself.

DISCUSSION

This framework moves AI education beyond knowledge acquisition toward an observable process for deciding when learners may use AI with increasing independence. Human factors explains why performance changes across tasks, technologies, workflows, and environments. CARES identifies the patient-centered responsibilities that should remain visible during AI-supported care. Entrustment connects evidence of those responsibilities to decisions about future supervision. The empirical literature helps explain why this framework is needed. In simulated e-prescribing, senior medical students accepted about half to two-thirds of false-positive alerts, and incorrect decision support increased errors even though correct support reduced them (Lyell et al., 2017). Experienced clinicians have also reversed initially correct decisions after receiving incorrect advice, with reliance shaped by trust, confidence, task difficulty, and experience (Goddard et al., 2014). Review evidence similarly links overreliance to workload, time pressure, task complexity, and the difficulty of verification (Goddard et al., 2012). Among novice medical students, misleading AI explanations reduced diagnostic accuracy while confidence did not reliably distinguish correct from incorrect responses (Teng et al., 2026). Trust calibration is also imperfect. In one study, physicians correctly judged whether an AI-generated differential contained the final diagnosis in 61.5% of cases, and better calibration was associated with greater diagnostic accuracy even though overall accuracy did not improve (Sakamoto et al., 2024). Together, these findings show why supervisors must look beyond the final answer and consider how the learner interacted with AI and the conditions surrounding performance.

The framework can be incorporated into existing EPAs, workplace-based assessments, simulation, case-based learning, feedback, and remediation. CARES-informed behaviours should be attached to the clinical activity whenever AI materially contributes to performance rather than confined to a separate AI curriculum. Simulation can provide safe opportunities to practise verification, disagreement, communication of uncertainty, and escalation. Workplace observation is then needed to determine whether those behaviours transfer to clinical care. Faculty development is important because supervisors need enough familiarity with AI concepts, product limitations, human factors, and entrustment to observe performance and provide useful feedback. Reviews of clinician acceptance similarly identify training as an important condition for safe use (Lambert et al., 2023).

Supervision should also support a just learning culture rather than become a label for blame. Learners need to be able to question AI, report near misses,

and ask for assistance without every concern being treated as evidence of incompetence. Supervisors should explore how the learner interpreted the AI output, what verification occurred, whether uncertainty was communicated, and whether help was sought at the appropriate time. Debriefing and feedback are therefore part of the safety function of supervision.

Healthcare organizations retain responsibilities that cannot be transferred to learners. They must define which AI uses are permitted, communicate known limitations, protect data, monitor performance, and establish clear pathways for escalation. Product monitoring and learner assessment should remain distinct. Product monitoring addresses validity, drift, incidents, and workflow effects, while learner assessment focuses on judgment, verification, communication, help-seeking, and supervision needs. Reassessment may be required when the product, workflow, patient population, learner performance, or available supervision changes (Labkoff et al., 2024; Reddy, 2024; Scott et al., 2024).

Implementation should begin with one clearly defined AI-supported professional activity rather than a comprehensive curriculum or assessment instrument. The activity should identify the permitted AI contribution, the behaviours supervisors should observe, the conditions of supervision, the evidence sources, and the circumstances that should trigger reassessment. A controlled pilot can then determine whether the framework is understandable and feasible before broader adoption. Expansion should depend on evidence that the AI contribution, workflow, and supervision arrangements remain safe, reliable, transparent, and understandable to learners, supervisors, and patients (Kalra and Johnston, 2025). New safeguards should also be examined for unintended consequences such as added documentation burden, workarounds, privacy concerns, or gaps in responsibility.

Research should proceed in stages. Input from experts, learners, supervisors, and patients can first refine the observable behaviours, including through structured consensus methods such as published e-Delphi studies of AI competencies (Çalışkan et al., 2022). Pilot studies can then examine feasibility, acceptability, feedback quality, agreement among assessors, and whether the framework captures developmental progression. Later studies should test whether it improves verification, help-seeking, escalation, supervision decisions, and patient-centered performance. The framework does not yet provide validated thresholds. Its immediate value is to make the decision process visible enough to support implementation, testing, and refinement.

LIMITATIONS

This framework is limited by the maturity of the available evidence. Much of the supporting literature concerns simulated tasks, clinician rather than learner performance, or individual AI applications. Few studies examine how learners use AI during workplace performance or how supervisors should translate that evidence into prospective entrustment decisions. The evidence is also heterogeneous across technologies, specialties, clinical risks, and outcomes. It therefore does not yet establish which observable behaviours or supervision thresholds are most predictive of safe performance.

The framework has not been empirically validated or refined through consultation with patients, learners, and supervisors. The CARES domains may overlap in practice, and Respect remains grounded mainly in broader principles of patient-centered care. Observable behaviours and supervision decisions will require adaptation to the activity, specialty, AI contribution, and clinical context.

The framework does not establish the validity or safety of a particular AI product, nor has it been shown to improve learner competence, patient outcomes, or safety. The targeted searches were not exhaustive, so relevant evidence may also have been missed. The framework should therefore be treated as a testable organizing model for curriculum and assessment rather than as a validated instrument.

CONCLUSION

AI literacy alone does not show that a learner is ready to use AI with increasing independence. Human factors explains the conditions that shape learner-AI performance, while CARES identifies the patient-centered responsibilities that supervisors should observe. Entrustment then connects that evidence to decisions about future supervision. The learner remains the person entrusted to perform the professional activity, and AI remains a bounded contribution within it. Incorporating CARES-informed behaviours into existing clinical activities and entrustment processes provides a practical foundation for curriculum development, workplace assessment, and further testing of safe AI-supported practice.

ACKNOWLEDGMENT

The authors would like to acknowledge that the material in this paper forms part of a thesis to be submitted by Bryan Johnston to the University of Saskatchewan in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

REFERENCES

Abdulnour, R.-E.E., Gin, B., Boscardin, C.K., 2025. Educational Strategies for Clinical Supervision of Artificial Intelligence Use. N. Engl. J. Med. 393, 786–797. https://doi.org/10.1056/NEJMra2503232

Bienefeld, N., Keller, E., Grote, G., 2024. Human-AI Teaming in Critical Care: A Comparative Analysis of Data Scientists' and Clinicians' Perspectives on AI Augmentation and Automation. J. Med. Internet Res. 26, e50130. https://doi.org/10.2196/50130

Çalışkan, S.A., Demir, K., Karaca, O., 2022. Artificial intelligence in medical education curriculum: An e-Delphi study for competencies. PLOS ONE 17, e0271872. https://doi.org/10.1371/journal.pone.0271872

Cross, J.L., Choma, M.A., Onofrey, J.A., 2024. Bias in medical AI: Implications for clinical decision-making. PLOS Digit. Health 3, e0000651. https://doi.org/10.1371/journal.pdig.0000651

Farnan, J.M., Petty, L.A., Georgitis, E., Martin, S., Chiu, E., Prochaska, M., Arora, V.M., 2012. A Systematic Review: The Effect of Clinical Supervision on Patient and Residency Education Outcomes. Acad. Med. 87, 428–442. https://doi.org/10.1097/ACM.0b013e31824822cc

Foresman, G., Biro, J., Tran, A., MacRae, K., Kazi, S., Schubel, L., Visconti, A., Gallagher, W., Smith, K.M., Giardina, T., Haskell, H., Miller, K., 2025. Patient Perspectives on Artificial Intelligence in Health Care: Focus Group Study for Diagnostic Communication and Tool Implementation. J. Particip. Med. 17, e69564. https://doi.org/10.2196/69564

Gin, B.C., O'Sullivan, P.S., Hauer, K.E., Abdulnour, R.-E., Mackenzie, M., Ten Cate, O., Boscardin, C.K., 2025. Entrustment and EPAs for Artificial Intelligence (AI): A Framework to Safeguard the Use of AI in Health Professions Education. Acad. Med. 100, 264–272. https://doi.org/10.1097/ACM.0000000000005930

Goddard, K., Roudsari, A., Wyatt, J.C., 2012. Automation bias: A systematic review of frequency, effect mediators, and mitigators. J. Am. Med. Inform. Assoc. JAMIA 19, 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Goddard, K., Roudsari, A., Wyatt, J.C., 2014. Automation bias: Empirical results assessing influencing factors. Int. J. Med. Inf. 83, 368–375. https://doi.org/10.1016/j.ijmedinf.2014.01.001

Hauer, K.E., 2019. Seeking trust in entrustment: Shifting from the planning of entrustable professional activities to implementation. Med. Educ. 53, 752–754. https://doi.org/10.1111/medu.13920

Holden, R.J., Carayon, P., Gurses, A.P., Hoonakker, P., Hundt, A.S., Ozok, A.A., Rivera-Rodriguez, A.J., 2013. SEIPS 2.0: A human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 56, 1669–1686. https://doi.org/10.1080/00140139.2013.838643

Jacobs, S.M., Lundy, N.N., Issenberg, S.B., Chandran, L., 2023. Reimagining Core Entrustable Professional Activities for Undergraduate Medical Education in the Era of Artificial Intelligence. JMIR Med. Educ. 9, e50903. https://doi.org/10.2196/50903

Johnston, B., Kalra, J., 2025. Introducing the CARES Model: Integrating Artificial Intelligence, Medical Education, and Patient-Centered Care. Presented at the AHFE HAWAII International Conference (AHFE HAWAII 2025). https://doi.org/10.54941/ahfe1006960

Kalra, J., Seitzinger, P., 2022. Implications and Consequences of Artificial Intelligence in Healthcare Quality and Medical Training. Presented at the 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022). https://doi.org/10.54941/ahfe1002109

Kalra, J., Johnston, B., 2025. Artificial Intelligence Revolution in Healthcare: Enhancing Clinical Practice with a New Member of the Team. Presented at the 16th International Conference on Applied Human Factors and Ergonomics (AHFE 2025). https://doi.org/10.54941/ahfe1006206

Kennedy, T.J.T., Regehr, G., Baker, G.R., Lingard, L.A., 2005. Progressive Independence in Clinical Training: A Tradition Worth Defending?: Acad. Med. 80, S106–S111. https://doi.org/10.1097/00001888-200510001-00028

Kogan, J.R., Holmboe, E.S., Hauer, K.E., 2009. Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review. JAMA 302, 1316–1326. https://doi.org/10.1001/jama.2009.1365

Labkoff, S., Oladimeji, B., Kannry, J., Solomonides, A., Leftwich, R., Koski, E., Joseph, A.L., Lopez-Gonzalez, M., Fleisher, L.A., Nolen, K., Dutta, S., Levy, D.R., Price, A., Barr, P.J., Hron, J.D., Lin, B., Srivastava, G., Pastor, N., Luque, U.S., Bui, T.T.T., Singh, R., Williams, T., Weiner, M.G., Naumann, T., Sittig, D.F., Jackson, G.P., Quintana, Y., 2024. Toward a responsible future: Recommendations for AI-enabled clinical decision support. J. Am. Med. Inform. Assoc. 31, 2730–2739. https://doi.org/10.1093/jamia/ocae209

Lambert, S.I., Madi, M., Sopka, S., Lenes, A., Stange, H., Buszello, C.-P., Stephan, A., 2023. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. Npj Digit. Med. 6, 111. https://doi.org/10.1038/s41746-023-00852-5

Lorenzini, G., Arbelaez Ossa, L., Shaw, D.M., Elger, B.S., 2023. Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making. Bioethics 37, 424–429. https://doi.org/10.1111/bioe.13158

Lyell, D., Magrabi, F., Raban, M.Z., Pont, L.G., Baysari, M.T., Day, R.O., Coiera, E., 2017. Automation bias in electronic prescribing. BMC Med. Inform. Decis. Mak. 17, 28. https://doi.org/10.1186/s12911-017-0425-5

Park, H.J., 2024. Patient perspectives on informed consent for medical AI: A web-based experiment. Digit. Health 10, 20552076241247938. https://doi.org/10.1177/20552076241247938

Preiksaitis, C., Nord, G., Davis, J., Sangal, R., Mackey, C., Oh, L., Rose, C., 2026. SCALES-AI: A Supervision- and Context-Aligned Entrustment Framework for Integrating Artificial Intelligence Into Emergency Medicine Education. AEM Educ. Train. 10, e70195. https://doi.org/10.1002/aet2.70195

Pupic, N., Ghaffari-zadeh, A., Hu, R., Singla, R., Darras, K., Karwowska, A., Forster, B.B., 2023. An evidence-based approach to artificial intelligence education for medical students: A systematic review. PLOS Digit. Health 2, e0000255. https://doi.org/10.1371/journal.pdig.0000255

Reddy, S., 2024. Generative AI in healthcare: An implementation science informed translational path on application, integration and governance. Implement. Sci. 19, 27. https://doi.org/10.1186/s13012-024-01357-9

Sakamoto, T., Harada, Y., Shimizu, T., 2024. Facilitating Trust Calibration in Artificial Intelligence Driven Diagnostic Decision Support Systems for Determining Physicians' Diagnostic Accuracy: Quasi-Experimental Study. JMIR Form. Res. 8, e58666. https://doi.org/10.2196/58666

Salwei, M.E., Carayon, P., 2022. A Sociotechnical Systems Framework for the Application of Artificial Intelligence in Health Care Delivery. J. Cogn. Eng. Decis. Mak. 16, 194–206. https://doi.org/10.1177/15553434221097357

Schuitmaker, L., Drogt, J., Benders, M., Jongsma, K., 2025. Physicians' required competencies in AI-assisted clinical settings: A systematic review. Br. Med. Bull. 153, ldae025. https://doi.org/10.1093/bmb/ldae025

Scott, I.A., Van Der Vegt, A., Lane, P., McPhail, S., Magrabi, F., 2024. Achieving large-scale clinician adoption of AI-enabled decision support. BMJ Health Care Inform. 31, e100971. https://doi.org/10.1136/bmjhci-2023-100971

Ten Cate, O., 2013. Nuts and Bolts of Entrustable Professional Activities. J. Grad. Med. Educ. 5, 157–158. https://doi.org/10.4300/JGME-D-12-00380.1

Ten Cate, O., Scheele, F., 2007. Viewpoint: Competency-Based Postgraduate Training: Can We Bridge the Gap between Theory and Clinical Practice?: Acad. Med. 82, 542–547. https://doi.org/10.1097/ACM.0b013e31805559c7

Teng, D., Tan, L., Cao, Q., Xia, Y., Zhang, N., Li, J., Zhao, D., 2026. Impact of AI misinformation on diagnostic accuracy and confidence calibration in novice medical students. Npj Digit. Med. 9, 356. https://doi.org/10.1038/s41746-026-02547-z

Vasey, B., Ursprung, S., Beddoe, B., Taylor, E.H., Marlow, N., Bilbro, N., Watkinson, P., McCulloch, P., 2021. Association of Clinician Diagnostic Performance With Machine Learning-Based Decision Support Systems: A Systematic Review. JAMA Netw. Open 4, e211276. https://doi.org/10.1001/jamanetworkopen.2021.1276

Vasey, B., Nagendran, M., Campbell, B., Clifton, D.A., Collins, G.S., Denaxas, S., Denniston, A.K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B.A., Mathur, P., McCradden, M.D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D.S.W., Watkinson, P., Weber, W., Wheatstone, P., McCulloch, P., 2022. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 377, e070904. https://doi.org/10.1136/bmj-2022-070904

Vo, V., Chen, G., Aquino, Y.S.J., Carter, S.M., Do, Q.N., Woode, M.E., 2023. Multi-stakeholder preferences for the use of artificial intelligence in healthcare: A systematic review and thematic analysis. Soc. Sci. Med. 338, 116357. https://doi.org/10.1016/j.socscimed.2023.116357