



Received: 08-09-2026
Accepted: 18-10-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

AI and Patient Safety - Trust but Verify: Safeguarding Patient Safety in the Age of AI Driven Nursing

¹ Pratik Ravindra Salve, ² Neha Michael-Salve

¹ Assistant Professor, M.Sc., MAEER's VishwaRaj Institute of Nursing, Pune, Maharashtra, India

² Clinical Instructor, M.Sc., Sadhu Vaswani College of Nursing, Pune, Maharashtra, India

DOI: <https://doi.org/10.62225/2583049X.2026.6.5.7089>

Corresponding Author: Pratik Ravindra Salve

Abstract

Artificial intelligence (AI) is reshaping patient safety by increasingly influencing clinical assessment, risk prediction, and nursing decision-making. Although AI-enabled systems offer opportunities to improve early detection, workflow efficiency, and clinical decision support, they also introduce risks associated with algorithmic bias, automation bias, limited contextual understanding, and misplaced confidence in machine-generated recommendations. These developments require a reconsideration of professional accountability and the evolving role of nurses in technology-mediated healthcare. This commentary examines the transition from conventional patient safety approaches towards an emerging paradigm in which safety is increasingly influenced by algorithmic intelligence. It argues that nurses must move beyond the operational use of AI to assume an active role in evaluating, interpreting, and

governing its clinical influence. The concept of Verified Intelligence is proposed as a conceptual framework for integrating algorithmic capabilities with nursing expertise, ethical reasoning, contextual judgement, and person-centred care. Within this perspective, verification is positioned as an essential patient safety competency rather than an optional safeguard. The commentary further highlights implications for nursing education, continuing professional development, clinical leadership, and AI governance. Preparing nurses to critically assess AI-generated recommendations will be essential to maintaining human accountability and preventing avoidable harm. Ultimately, responsible AI integration requires a partnership in which technological intelligence supports, rather than supplants, professional judgement and patient advocacy.

Keywords: Artificial Intelligence, Patient Safety, Nursing Accountability, Clinical Decision Support

Introduction

Patient safety has seen significant changes with the development of healthcare throughout history. The main focus in the early days was on preventing individual clinical mistakes, while the current trend promotes the understanding of systems thinking, interdisciplinary cooperation, standardizing care, and continuous learning. In recent years, healthcare has seen a significant rise in introducing electronic health records and other digital tools aiming to support medical practice [1]. The upcoming stage in the development of postmodern healthcare is artificial intelligence (AI). Unlike traditional digital devices, AI can prioritize, predict future changes, suggest appropriate actions, and make decisions [2, 3].

For nursing practice, such changes are not only another step in the development of healthcare technologies. AI is already integrated into dozens of medical procedures, systems supporting medication use, documentation, and work organization, patient monitoring, analyzing trends, and clinical support systems [4]. Each of these technologies aims to help clinicians in performing their duties more efficiently and effectively and reduce their workload, supporting evidence-based practice [2]. At the same time, they represent a significant change in the way patient safety is approached. By being integrated into the work processes, such systems transform safety into a hybrid concept that includes recommendations from clinicians and robots [5].

This setting raises numerous questions about the role of nurses in patient safety. What does it mean for nursing to trust technology to make decisions for them? The answer lies between the two opposites, suggesting that nurses should promote trustworthiness and use technologies as an aid while being aware of the limitations they impose. Thus, I believe that the professional mindset of "Trust but Verify" is the best way to approach the upcoming changes. While many suggestions made by robots can and should be accepted, nurses should not completely rely on the recommendations, trusting them blindly. The

use of such technology to guide evidence-based practices should be questioned in terms of ethics and professional responsibility to patients^[5].

The suggested editorial review considers that the more AI is integrated into the mechanism of patient safety, the more the nursing practice will change from direct patient contact to a higher level of responsibility. Nurses will have to develop the ability to clinically assess the recommendations of algorithms, verify their safety, and adopt technologies into practice. At the same time, verification of such recommendations should not be seen as skepticism regarding the use of technologies. It is an extension of their mandate to ensure patient safety and health by following the necessary standards and guidelines. Such paradigmatic changes will require adjustments at the educational, professional, leadership, and governance levels.

The introduction of artificial intelligence (AI) into healthcare is changing the decision-making paradigm. Nurses traditionally relied on patient assessments, professional expertise, and interdisciplinary cooperation to make clinical decisions^[6]. Evidence-based practice guidelines and technologies contributed to the process by providing access to relevant research findings, but the responsibility of nursing professionals remained the same – they were to assess and interpret the available evidence^[7]. By suggesting changes to clinical practice, AI implements a new level of decision-making.

Modern technologies can analyze vast amounts of data, both structured and unstructured, identify patterns, correlations, and hidden trends, predict negative events, and suggest ways to avoid them in seconds. The ability of AI to process data is beneficial to nurses in terms of medication safety, patient safety, documentation, work organization, and remote patient monitoring systems. By using such technologies, healthcare providers can ensure safer medication administration, reduce the nursing staff's workload, detect subtle trends preceding negative patient outcomes, and improve the efficiency of care^[2, 8].

However, the paradigmatic shift caused by the introduction of AI into healthcare is not only technological but also cognitive. Nurses should not only rely on their professional expertise but also consider the recommendations of intelligent systems. Thus, clinical decision-making moves from a human-centered paradigm to a dialectical one, implying that nurses should think not only about their actions but also about the actions suggested by robots^[9].

This shift should not be mistaken for a transfer of responsibility from nurses to technology. The responsibility for the patient's safety and health should remain with the nurses, while technology helps them meet this responsibility. By analyzing large amounts of structured and unstructured data, identifying patterns and correlations invisible to humans, AI allows clinicians to be more accurate in medication administration, predict negative patient outcomes, and prevent them. Thus, such technologies are beneficial in terms of patient safety. However, nurses should not forget about their role as caregivers and rely only on the recommendations of algorithms. The nursing profession requires human interaction, which cannot be replaced by technology. Artificial intelligence can complement the skills of nursing professionals, but nurses should not outsource their responsibility to algorithms^[10]. Understanding the strengths

of both human expertise and technology is paramount to ensuring safe and effective patient care.

Artificial intelligence can significantly affect patients' safety outcomes, which should stimulate nurses and other healthcare professionals to improve their competencies in interacting with new technologies. However, the way this process unfolds should be carefully considered because it raises many ethical concerns. The influence of algorithms on human behavior can lead to several issues affecting patient safety. While AI can be a great source of evidence for diagnosis, predicting post-surgical complications, and suggesting solutions, blindly following the recommendations of algorithms can be dangerous because they are not perfect. Like any instrument, they can only be as accurate as the person who uses them. Thus, even the most sophisticated algorithms will provide incorrect recommendations when they are operated by individuals unprepared to identify these mistakes or when they are trained on incorrect data^[13].

This challenge is especially relevant to nursing because of the complexity of situations arising when interacting with patients. In many cases, proper clinical decisions require considering not only the objective evidence but also the accompanying circumstances, which can be unique for each patient. A nurse must take into account the patient's cultural background, personal beliefs, and expectations, which greatly affects how they will perceive the upcoming interventions. In addition, such factors as the availability of resources, economic status, education, and personal attitudes are important considerations influencing the situation. These issues cannot always be regulated by clinicians, but nurses can suggest possible measures to ensure better patient safety outcomes.

At the same time, healthcare professionals should be aware of automation bias, which consists in an inappropriate reliance on technology and a general tendency to accept its recommendations without critically analyzing them^[14]. When nurses work in conditions of constant automation, they may begin to rely on its recommendations to the detriment of their own expertise, especially in critical situations when the stakes are high^[15]. For example, when faced with highly stressful circumstances, a nurse may choose to follow the recommendations of an automated system rather than spend time on her own, possibly inaccurate, assessment.

Nurses should recognize such biases as a threat to professional judgment and try to prevent them by carefully analyzing the recommendations of technologies. Thus, nurses may find themselves in the position of selecting between the recommendations of algorithms and their professional judgment, which should always be done in the patient's best interest.

A similar dilemma may arise when nurses interact with patients and their families. In such cases, nurses should base their recommendations on the principles of ethical nursing practice, particularly focusing on the patient's best interests. However, the concept of benefits can be different for different people, depending on their personal attitudes, beliefs, and expectations. When communicating with patients, nurses must consider this heterogeneity and individualize their recommendations as much as possible within the framework of their professional nursing practice. This approach will help ensure patient safety by increasing

the likelihood that the recommendations will be followed. Artificial intelligence can support such interaction by providing nurses with objective evidence, but nurses should always remember that patients are not only physical but also psychological beings. Thus, nurses should consider not only the objective evidence provided by AI but also patients' subjective experiences, explaining the necessity of recommended actions and helping them comply with the recommendations. The ability to support patients is one of the core skills of a nurse, which is why nurses can and should rely on their expertise when interacting with patients, even when technology is available.

Nursing practice involves making tough decisions almost every day, and nurses have various ways of doing so. Making ethical decisions in accordance with professional and organizational standards requires critical thinking. Nurses should apply all their professional skills, including analytical skills, to make the right decisions for each individual case. The critical thinking of nurses should be exercised when they work with technologies, suggesting their recommendations. Thus, the maxim of "Trust but verify" makes sense in the context of nurses working with technology.

I assume that trust is a belief that technology can benefit the nurse in terms of saving time, prioritizing decisions, providing evidence, eliminating unnecessary risks. Verification means ensuring that although AI suggests recommendations, they are safe for the patient and do not contradict professional nursing practice. In addition to checking the recommendations of algorithms against available evidence, nurses need to ensure that they are not based on any assumptions and are feasible in a particular setting. Nurses also need to understand when it is necessary to seek advice from other clinicians when verifying the recommendations.

By this definition, verification is the stage of critical thinking that allows nurses to ensure that the recommendations of technology are appropriate to the patient, evidence-based, and ethically justified within the framework of professional nursing practice. Such critical review should become a routine activity for nurses when interacting with technology, replacing the blind trust that they may have in the recommendations of algorithms. Nurses should constantly remember that technology is only a tool that is capable of making decisions but does not bear moral responsibility. This moral responsibility remains with nurses who must ensure that the recommendations of AI contribute to patient safety.

At the same time, verification does not require nurses to perform any additional functions since they already follow similar recommendations when working with conventional technologies. For example, they check the identity of patients before administering medications, ensure the dose of a substance in accordance with the prescribed dose, follow the indicators during procedures, and monitor patient responses to various interventions.

As healthcare turns to technology to improve care, verification will become even more important to patient safety. It will be the responsibility of nurses to ensure that the recommendations of AI benefit the patient and support the highest standards of safety. Thus, nurses will face the dilemma of choosing between trusting their professional judgment and relying on the recommendations of technology. Nurses will rely on their knowledge,

professional skills, and critical thinking to find the optimal solution for each case.

References

1. Ratwani R. Electronic Health Records and Improved Patient Care: Opportunities for Applied Psychology. *Current Directories in Psychological Science*, Aug 2017; 26(4):359-365. Doi: 10.1177/0963721417700691
2. Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances, and future perspectives. *European Journal of Medical Research*, Sep 23, 2025; 30:848. Doi: 10.1186/s40001-025-03196-w
3. Salve P, Ray S, Ray SP, Geddugol BJ, Nadaf H, Jabade M, *et al.* Trust and Reliance on Artificial Intelligence-Based Clinical Decision Support Systems. Implications for Nursing Practice. *Investigacion y Educacion en Enfermeria*, Jun 9, 2026; 44(2). Doi: 10.17533/udea.ice.v44n2e02
4. Ahmed MM, Okesanya OJ, Othman ZK, Adebayo UO, Oso TA, Lucero-Prisno DE. Artificial intelligence in modern healthcare: Applications, innovations, and ethical dimensions in personalized, predictive, and inclusive care. *Informatics and Health*, Sep 2026; 3(2):243-256. Doi: 10.1016/j.infoh.2026.03.002
5. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, Feb 29, 2024; 10(4):e26297. Doi: 10.1016/j.heliyon.2024.e26297
6. Wei Q, Pan S, Liu X, Hong M, Nong C, Zhang W. The integration of AI in nursing: addressing current applications, challenges, and future directions. *Frontiers in Medicine*, Feb 11, 2025; 12:1545420. Doi: 10.3389/fmed.2025.1545420
7. Li D, Jiang N, Huang K, Tu R, Ouyang S, Yu H, *et al.* Streamlining evidence based clinical recommendations with large language models. *Npj Digital Medicine*, Dec 22, 2025; 8(1):793. Doi: 10.1038/s41746-025-02273-y
8. De Micco F, Palma GD, Seveso G, Giacomobono F, Scendoni R, Tambone V. AI-Driven Approaches for Adverse Event Detection: A Systematic Review of Current Evidence. *Safety*, Apr 2026; 12(2):52. Doi: 10.3390/safety12020052
9. Abdelwanis M, Simsekler MCE, Gabor AF, Sleptchenko A, Omar M. Artificial intelligence adoption challenges from healthcare providers' perspectives: A comprehensive review and future directions. *Safety Science*, Jan 1, 2026; 193:107028. Doi: 10.1016/j.ssci.2025.107028
10. Khosravi M, Zare Z, Mojtavaeian SM, Izadi R. Artificial Intelligence and Decision-Making in Healthcare: A Thematic Analysis of a Systematic Review of Reviews. *Health Services Research and Managerial Epidemiology*, Mar 5, 2024; 11:23333928241234863. Doi: 10.1177/23333928241234863
11. Pepito JA, Locsin R. Can nurses remain relevant in a technologically advanced future? *International Journal of Nursing Sciences*, Oct 4, 2018; 6(1):106-110. Doi: 10.1016/j.ijnss.2018.09.013
12. Fernandes F, Shinnars L, Mota M, Santos P, Sá L. Contributions of Artificial Intelligence to Decision Making in Nursing: A Scoping Review. *Nursing &*

- Health Sciences, Mar 2026; 28(1):e70308. Doi: 10.1111/nhs.70308
13. Joseph J. Algorithmic bias in public health AI: A silent threat to equity in low-resource settings. *Frontiers in Public Health*, Jul 23, 2025; 13:1643180. Doi: 10.3389/fpubh.2025.1643180
 14. Goddard K, Roudsari A, Wyatt JC. Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*. 2012; 19(1):121-127. Doi: 10.1136/amiajnl-2011-000089
 15. Sokol K, Fackler J, Vogt JE. Artificial intelligence should genuinely support clinical reasoning and decision making to bridge the translational gap. *Npj Digital Medicine*, Jun 10, 2025; 8:345. Doi: 10.1038/s41746-025-01725-9