

INTEGRATIVE REVIEW

ChatGPT and Large Language Models in Contemporary Nursing

Izabella Uchmanowicz¹  | Heba M. Aldossary²  | Christopher S. Lee^{3,4} | Mariachiara Figura⁵  | Maria Jędrzejczyk¹  | Maddalena De Maria⁶  | Gianluca Pucciarelli^{1,7} | Ewa Molka⁸  | Nicola Straiton⁹  | Aleksandra Kołtuniuk¹ 

¹Department of Nursing, Faculty of Nursing and Midwifery, Wrocław Medical University, Wrocław, Poland | ²Department of Nursing, Prince Sultan Military College of Health Sciences, Dhahran, Saudi Arabia | ³Division of Scientific Research and Innovation in Emergency Medical Service, Department of Emergency Medical Service, Wrocław Medical University, Wrocław, Poland | ⁴Boston College William F. Connell School of Nursing, Chestnut Hill, Massachusetts, USA | ⁵Department of Maternal and Child Health Promotion, Internal Medicine & Specialties of Excellence 'G. D'Alessandro' (PROMISE) Piazza delle Cliniche 2, Palermo, Italy | ⁶Department of Life Science, Health, and Health Professions, Link Campus University, Rome, Italy | ⁷Department of Biomedicine and Prevention, Tor Vergata University, Rome, Italy | ⁸Department of Propedeutics of Nursing, Faculty of Health Sciences, Medical University of Silesia in Katowice, Katowice, Poland | ⁹School of Nursing, Midwifery and Paramedicine, Australian Catholic University, Sydney, New South Wales, Australia

Correspondence: Aleksandra Kołtuniuk (aleksandra.koltuniuk@umw.edu.pl)

Received: 12 April 2026 | **Revised:** 17 July 2026 | **Accepted:** 31 July 2026

Keywords: algorithmic bias | ChatGPT | clinical nursing practice | digital health governance | generative artificial intelligence | hallucination | large language models | nursing documentation | nursing education | patient safety

ABSTRACT

Objectives: This narrative review synthesizes published evidence on the applications, benefits, limitations and governance considerations of ChatGPT and large language models (LLMs) in nursing, across three domains: education, clinical practice and workflow management.

Design: The article was conducted as a narrative review.

Methods: A structured literature search was conducted in Medline (via PubMed), Scopus and arXiv, covering publications from January 2019 to March 2026. Peer-reviewed original studies, systematic reviews, scoping reviews, narrative reviews and expert commentaries addressing LLM applications in nursing education, clinical practice or workflow were eligible for inclusion. Studies limited exclusively to non-nursing medical specialties without transferable nursing implications were excluded. Findings were narratively synthesized across five thematic domains by authors with subject-matter expertise in each area.

Results: In nursing education, ChatGPT demonstrates utility as an adaptive cognitive scaffold, supporting theoretical learning, simulation-based training and virtual patient encounters, though unregulated use poses risks to academic integrity and independent clinical reasoning. In clinical practice, LLMs can assist with preliminary symptom assessment and patient education material generation; however, performance deteriorates markedly in complex or data-sparse clinical scenarios and hallucination rates remain clinically significant. In workflow management, ChatGPT shows promise in reducing documentation burden and supporting administrative tasks, though data privacy obligations under frameworks such as GDPR constrain real-world deployment. Across all domains, concerns persist regarding algorithmic bias, professional accountability and the absence of clear medico-legal frameworks governing AI-related clinical errors.

Conclusion: ChatGPT and related LLMs are best positioned as auxiliary tools that augment rather than replace professional nursing judgement. Safe and ethical integration requires the development of AI literacy curricula, institutionally governed deployment frameworks, mandatory human-in-the-loop verification protocols and longitudinal evaluation of patient safety outcomes. Nurses must play an active role in shaping the responsible adoption of generative AI in healthcare.

Implications for Practice and Research: 1. Nurses must treat AI-generated content as a preliminary draft requiring mandatory human verification before clinical or documentation use. 2. Institutions should prioritize closed-loop, enterprise-grade AI deployments over public platforms to ensure GDPR compliance. 3. AI literacy must be embedded in undergraduate and continuing nursing education curricula. 4. Longitudinal research on patient safety outcomes following real-world LLM deployment in nursing is urgently needed.

Reporting Method: As a narrative review, this article followed established guidance for the conduct and reporting of narrative reviews.

Patient or Public Contribution: There was no patient or public involvement in this narrative review.

Highlights

What Is Already Known?

1. ChatGPT and large language models are increasingly used in healthcare, but evidence in nursing remains fragmented and dominated by single-domain studies.
2. Hallucinations and data privacy risks have been identified as barriers to clinical deployment.
3. Most existing reviews focus on Anglo-American contexts and do not address governance constraints specific to European regulatory frameworks.

What This Paper Adds?

1. A comprehensive, up-to-date synthesis (to March 2026) of ChatGPT applications across nursing education, clinical practice and workflow management within a single review.
2. Consolidation of quantified hallucination rates (up to 83% in clinical vignettes; > 25% in bibliographic reference generation) is directly relevant to nursing safety decision-making.
3. Identification of a medico-legal accountability vacuum when LLM-generated content contributes to clinical errors—a gap not previously articulated in nursing literature.
4. Integration of GDPR constraints and Central European (Polish) healthcare workforce context, extending the evidence base beyond predominantly Anglo-American settings.

LLMs are trained to predict language patterns from large datasets rather than to verify facts against curated clinical databases in real time. Consequently, they may produce hallucinations—confident but inaccurate or unsupported statements—reflecting probabilistic text generation rather than clinical reasoning (Ji et al. 2023). In healthcare contexts, such plausible but incorrect outputs pose patient safety risks, particularly when persuasive language amplifies automation bias and reduces critical appraisal by users (Nguyen 2024). These risks are especially salient in nursing, where documentation has medico-legal implications, communication quality directly influences patient understanding and adherence, and clinical decisions must integrate evolving patient data with professional judgement.

Recent reviews indicate growing interest in ChatGPT within nursing across education, clinical practice and administrative workflow, yet the evidence base remains uneven and dominated by simulations, perception studies and controlled vignettes rather than real-world implementation outcomes (Hobensack et al. 2024; Peng et al. 2025; Zhou et al. 2024). Existing literature highlights opportunities for increased efficiency, improved access to knowledge and enhanced support for education and documentation, while simultaneously emphasizing unresolved concerns regarding accuracy, transparency, data privacy and academic integrity (Abujaber et al. 2023; Hobensack et al. 2024; Peng et al. 2025).

Globally, governance frameworks increasingly emphasize that AI for health must protect autonomy, safety, accountability and equity. The World Health Organization (WHO) calls for human oversight, clear allocation of responsibility and safeguards against harm in AI-enabled care (WHO 2021). Similarly, the OECD principles on trustworthy AI stress transparency, robustness and alignment with human rights and democratic values (OECD 2019). Professional nursing organizations reinforce that AI should augment—not replace—clinical judgement and relational care, highlighting the need for AI literacy and human-in-the-loop use to maintain accountability and patient trust (American Nurses Association 2022; International Council of Nurses 2023).

Against this rapidly evolving global landscape, understanding the implications of ChatGPT for nursing practice, education and workflow is essential. This narrative review synthesizes published evidence on current and emerging applications of ChatGPT in nursing and examines: (1) areas of use across education, clinical practice and administration; (2) tasks where performance appears reliable versus those associated with higher risk; and (3) safeguards—including verification practices,

1 | Introduction

Artificial intelligence (AI), particularly generative AI based on large language models (LLMs) such as ChatGPT, is becoming an influential component of digital transformation in health care. Unlike earlier clinical AI systems focused primarily on prediction or imaging, LLMs generate fluent text and therefore intersect directly with domains central to nursing practice, including clinical documentation, patient education, handovers, care planning and professional learning. At the same time, health systems worldwide face converging pressures—workforce shortages, increasing patient complexity and expanding documentation requirements—creating strong incentives to adopt tools that promise efficiency gains and improved access to information (Wang et al. 2023; Zhou et al. 2024).

governance frameworks and professional competencies—required for safe and ethical integration internationally.

1.1 | Review Design and Search Strategy

This article was conducted as a narrative review aimed at synthesizing published evidence on the applications, benefits, limitations and governance considerations of ChatGPT and large language models (LLMs) in contemporary nursing. The narrative review format was selected given the breadth and heterogeneity of the available literature, which encompasses diverse study designs, clinical contexts and disciplinary perspectives that preclude formal meta-analysis. The review addresses three overarching objectives: (1) to map current and emerging areas of application of ChatGPT across nursing education, clinical practice and workflow management; (2) to appraise the reliability and limitations of LLM performance, distinguishing tasks where evidence supports acceptable accuracy from those associated with elevated clinical or educational risk; and (3) to identify safeguards required for safe and ethical integration, including verification practices, institutional governance frameworks and professional competencies. To achieve these objectives, the review is structured across five thematic domains. The first examines ChatGPT in nursing education, encompassing theoretical learning, simulation-based and virtual patient training and challenges to academic integrity. The second addresses clinical nursing practice, covering clinical reasoning and symptom assessment, emergency nursing and triage and patient communication and education. The third analyses nursing workflow applications, including documentation efficiency, care planning and administrative and organizational functions. The fourth considers ethical, legal and professional dimensions, specifically accuracy and patient safety, data confidentiality and GDPR compliance, algorithmic bias and fairness and professional responsibility and accountability. The fifth section discusses practical implications for nursing practice, education and policy in the context of ongoing digital health transformation.

1.2 | Search Strategy

A structured literature search was performed in Medline (via PubMed), Scopus and arXiv. The search covered publications from January 2019 to March 2026. The start date of 2019 was selected to coincide with the publication of the OECD Principles on Artificial Intelligence and the emergence of transformer-based large language models as clinically relevant technologies, thereby capturing both the foundational governance frameworks and the evolving empirical evidence on their application in healthcare and nursing. The search was limited to publications in English. No restrictions were placed on study design; however, priority was given to peer-reviewed original studies, systematic reviews, scoping reviews, narrative reviews and expert commentaries published in nursing, medical education and health informatics journals.

Search terms were combined using Boolean operators (AND/OR) and included: 'ChatGPT' OR 'large language model' OR 'generative artificial intelligence' OR 'GPT-4' OR 'LLM' AND 'nursing' OR 'nurse' OR 'nursing education' OR 'nursing practice' OR

'nursing documentation' AND 'healthcare' OR 'clinical practice' OR 'patient safety' OR 'patient education'. Additional targeted searches were conducted using terms specific to each domain covered in the review, including 'triage', 'care planning', 'clinical simulation', 'academic integrity', 'hallucination', 'bias' and 'GDPR'.

Reference lists of included studies were manually screened to identify additional relevant publications. All records were imported into Zotero for reference management and deduplication.

1.3 | Inclusion and Exclusion Criteria

Studies were included if they: (1) reported empirical or theoretical findings concerning the application of ChatGPT or other LLMs in nursing education, clinical nursing practice, or nursing workflow; (2) were published in peer-reviewed journals or as preprints on recognized repositories (arXiv); and (3) were available in full text. Studies were excluded if they were limited exclusively to non-nursing medical specialties without transferable implications for nursing, if they addressed AI systems unrelated to natural language generation, or if they were conference abstracts without sufficient methodological detail. Given the rapidly evolving nature of the field, preprint publications were included where relevant and where findings were consistent with peer-reviewed sources.

1.4 | Data Synthesis

Given the heterogeneity of study designs and outcomes across the included literature, a formal quantitative synthesis was not conducted. Instead, findings were narratively synthesized by thematic domain. Each section of this review was developed by a designated author with subject-matter expertise in the relevant area, and all contributions were subsequently harmonized for consistency of terminology and analytical framing. The review followed guidance for narrative reviews as outlined by Ferrari (2015) and adhered to the principles of transparent reporting throughout. Where evidence was conflicting or limited, this is explicitly acknowledged within the relevant sections. Screening proceeded in two stages. Records were screened against the eligibility criteria by title and abstract and then read in full where relevant and reference lists were checked for further sources. In keeping with a narrative review, sources were appraised qualitatively for relevance, design and consistency with the wider literature rather than scored with a formal risk-of-bias tool (Page et al. 2021).

2 | ChatGPT in Nursing Education

2.1 | Enhancing Theoretical Learning

Nursing students often use ChatGPT to clarify difficult concepts, summarize course content and generate explanatory materials. Studies show that ChatGPT can assist in understanding pathophysiology, pharmacology and nursing theories, offering accessible explanations at various levels of complexity (Luo et al. 2025; Wang, Wan, et al. 2024). Educators similarly leverage

the tool to develop quizzes, adapt materials for diverse learners and support formative assessment (Abujaber et al. 2023).

ChatGPT is particularly effective for knowledge-based tasks, such as answering quiz-style questions and demonstrates high accuracy in areas aligned with standardized nursing curricula. Some studies found that ChatGPT achieved passing-level performance on nursing examinations or national licensing test items (Jin et al. 2024; Wang, Wan, et al. 2024). However, accuracy decreases when exam questions require multi-step reasoning or knowledge of local guidelines, which is especially relevant in countries like Poland, where practice standards differ from Anglo-American models.

Theoretical learning constitutes the cognitive foundation of nursing education, encompassing not only the acquisition and integration of conceptual knowledge but also the metacognitive regulation of learning processes that structure subsequent clinical reasoning and professional judgement (Guerreiro and Aschidamini 2025). Emerging literature increasingly positions generative AI systems, such as ChatGPT, as interactive cognitive scaffolds that can enhance this domain when embedded within intentional pedagogical design (Gerçek et al. 2025; Tam et al. 2023). These systems extend the learning environment by supporting conceptual clarification, adaptive explanation and iterative reinforcement of theoretical content while activating reflective engagement with knowledge construction rather than replacing instruction (Tran et al. 2025).

Empirical evidence suggests that nursing students employ ChatGPT primarily to clarify complex theoretical material, synthesize information and explore conceptual relationships across biomedical and disciplinary knowledge domains (Abujaber et al. 2023; Gunawan et al. 2024). Such interactions reduce cognitive barriers to comprehension, particularly for novice learners navigating abstract theoretical frameworks (Tam et al. 2023). Performance analyses in standardized examination contexts further demonstrate that large language models can align with theoretical knowledge structures, although variability persists depending on contextual complexity and question design (Moskovich and Rozani 2025; Zhang et al. 2025). These findings indicate that AI systems may support foundational knowledge acquisition while encouraging self-directed engagement with theoretical content and personalized learning trajectories (Gerçek et al. 2025).

Importantly, the pedagogical value of generative AI arises not solely from information delivery but from the metacognitive processes activated during interaction. Iterative prompting, response evaluation and dialogic exploration foster reflective learning and conceptual deepening, aligning with principles of adaptive and personalized education (Liu et al. 2023; Tam et al. 2023). Within this framework, generative AI functions as adaptive scaffolding that stimulates epistemic awareness, prompting students to justify reasoning and monitor understanding through iterative questioning (Srinivasan et al. 2024). However, unregulated reliance on AI risks superficial learning, erosion of independent judgement and loss of epistemic ownership, threats that carry professional implications in a discipline grounded in accountability and patient safety (Abujaber et al. 2023; Tam et al. 2023).

The literature consistently cautions against reducing theoretical learning to rapid information retrieval. Conceptual mastery requires disciplined reasoning and epistemic accountability; technologies that accelerate access to answers must not erode cognitive ownership of understanding (Srinivasan et al. 2024). Qualitative evidence demonstrates that students maintain ambivalent trust toward AI, recognizing its utility while rejecting its substitution for professional judgement (Gunawan et al. 2024). Studies evaluating AI-generated nursing texts reveal variability in accuracy and completeness, reinforcing the necessity of educator oversight, epistemic verification of outputs and critical appraisal against evidence-based sources (Gerçek et al. 2025; Ma et al. 2025).

From a normative perspective, integration of generative AI into theoretical learning must follow explicit pedagogical principles. AI should function as guided scaffolding rather than cognitive outsourcing; its use must remain transparent, ethically governed and aligned with academic integrity; and educators must retain epistemic authority over knowledge validation (American Nurses Association 2022). Responsible AI literacy, therefore, becomes part of professional accountability in nursing education. Institutions are encouraged to adopt blended pedagogical models that combine AI-augmented study with secure in-person assessment to preserve independent competence (Abujaber et al. 2023). Under these conditions, generative AI represents a legitimate amplifier of theoretical learning. The challenge for nursing education is therefore not whether AI belongs in the curriculum, but how to harness its scaffolding potential while preserving the intellectual rigour and humanistic core that define professional formation (Gerçek et al. 2025; Tam et al. 2023).

2.2 | Simulation and Virtual Patient Training

Simulation-based education plays a central role in contemporary nursing training because it allows students to practice clinical assessment, communication and decision-making in a safe environment before interacting with real patients. Traditionally, simulation relies on standardized patients, high-fidelity mannequins, or immersive technologies such as virtual reality. Although these approaches are effective, they require substantial financial investment, specialized infrastructure and trained personnel, which may limit accessibility and scalability across educational institutions. Recent advances in artificial intelligence (AI), particularly large language models (LLMs) such as ChatGPT, offer a complementary approach by enabling dynamic and interactive virtual patient simulations. These systems allow students to engage in conversational clinical encounters that replicate elements of real patient interviews, thereby expanding opportunities for experiential learning in nursing education (Chokkakula et al. 2025; Scherr et al. 2023).

One of the most significant advantages of LLM-based simulations is their capacity for natural language interaction. Unlike traditional scripted simulations, ChatGPT can generate responses in real time based on student questions, allowing learners to conduct history taking, clinical reasoning and patient education within a conversational framework. This interactive format promotes active learning and repeated practice, which are essential for developing communication and assessment

skills. In a study evaluating ChatGPT as a virtual patient in nursing simulation training, students reported high engagement and perceived educational benefits, particularly in practicing patient interviews and therapeutic communication. Such findings suggest that conversational AI can function as an accessible supplementary training tool alongside traditional simulation modalities (Benfatah et al. 2024).

Recent experimental research further supports the potential educational value of AI-driven patient simulations. A cross-over randomized controlled trial conducted by Fung et al. (2025) compared generative AI-based patient simulation with immersive 360° virtual reality simulation among undergraduate nursing students. Both simulation approaches significantly improved students' perceived clinical competence, cultural awareness and readiness to use AI in healthcare. Notably, when generative AI simulation was implemented first, students demonstrated greater initial improvements in clinical competence and AI readiness compared with those who first experienced virtual reality simulation. These results indicate that AI-driven patient simulation may represent an effective and scalable alternative to more resource-intensive digital simulation methods.

In addition to functioning as virtual patients, large language models can support educators in the development of simulation content. ChatGPT can generate clinical case scenarios, patient histories, assessment findings and dialogue prompts that align with specific learning objectives. This capability may reduce the time and effort required to design simulation exercises while also allowing educators to rapidly adapt cases to different clinical contexts. For example, AI-generated scenarios can be used to support training in emergency care, chronic disease management, mental health assessment, or telehealth consultations. A recent pilot study demonstrated that ChatGPT-supported dialogues and simulated scenarios improved nursing students' confidence and readiness for telehealth clinical practice, illustrating how AI-generated simulations can prepare students for emerging models of healthcare delivery (Paul 2025; Peralta Ramirez et al. 2025).

Despite these promising developments, the integration of ChatGPT and other LLMs into nursing simulation raises several important limitations and ethical considerations. Studies have shown that generative AI systems may occasionally produce incomplete, inaccurate, or overly generalized clinical information. Such inaccuracies may negatively influence learning if simulations are used without appropriate supervision. Furthermore, LLM-based systems currently lack nonverbal cues and contextual awareness that are essential components of real clinical encounters. For these reasons, recent literature consistently emphasizes that AI-generated simulations should be used as complementary educational tools rather than replacements for traditional simulation methods. Faculty oversight remains essential to validate clinical accuracy, guide reflection and ensure alignment with evidence-based nursing practice (Chokkakula et al. 2025).

Overall, the literature suggests that ChatGPT and other large language models have considerable potential to enhance simulation-based learning in contemporary nursing education. Their primary strengths include scalability, conversational realism

and the ability to rapidly generate diverse clinical scenarios for repeated practice. At the same time, responsible implementation requires careful integration into curricula, ongoing evaluation of educational outcomes and continued involvement of educators to ensure pedagogical quality and clinical reliability. When applied thoughtfully, AI-supported virtual patient training may complement traditional simulation approaches and contribute to more flexible, accessible and adaptive nursing education in the digital era (Fung et al. 2025; Scherr et al. 2023).

2.3 | Challenges for Academic Integrity

The incorporation of large language models (LLMs), such as ChatGPT, into nursing education presents substantial challenges to academic integrity and scholarly practice. ChatGPT can rapidly generate coherent, well-structured text on complex nursing topics, thereby challenging traditional assessment approaches that rely heavily on essays, case studies and reflective writing as indicators of individual learning (Dempere et al. 2023; Evangelista 2025). Consequently, educators face growing difficulty in distinguishing authentic student work from AI-generated outputs and in verifying that assignments accurately reflect students' knowledge, reasoning abilities and ethical awareness (Ateeq et al. 2024; Sengul et al. 2025).

Plagiarism and ghostwriting are significant risks in this context. Because ChatGPT produces novel text rather than reproducing material from existing sources, conventional plagiarism-detection software often fails to identify AI-generated submissions, thereby obscuring breaches of academic integrity (Eke 2023). Moreover, LLMs sometimes fabricate or improperly modify references, creating citations that appear legitimate but either do not exist or fail to support the associated arguments (Chelli et al. 2024; Walters and Wilder 2023). The uncritical inclusion of such fabricated citations undermines the reliability of scholarly discourse and increases the verification burden placed on instructors.

Beyond issues of originality, the growing reliance on generative AI risks eroding essential cognitive and professional skills. Overdependence on ChatGPT can weaken students' capacity for critical appraisal, ethical deliberation and evidence-based synthesis, which are fundamental to sound clinical judgement (El Arab et al. 2025; Naqvi et al. 2025). Automation bias may also emerge when learners accept AI-generated content as authoritative, given the model's persuasive and self-assured tone (Nguyen 2024; Sun et al. 2025). In a profession grounded in contextual reasoning and empathetic decision-making, such tendencies raise concerns regarding the deception of future practitioners and potential implications for patient safety.

In response, scholars advocate for a shift from prohibition to structured, transparent and pedagogically informed integration of AI within nursing education. Recommended strategies include the establishment of institutional policies governing generative AI, explicit expectations for disclosure of AI use and assessment designs that prioritize process-oriented components—such as drafts, reflective commentaries and viva examinations—alongside final submissions (Dempere et al. 2023; Evangelista 2025). Embedding AI literacy within nursing

curricula—encompassing principles of capability, limitation, bias, hallucination and data privacy—can empower students to employ ChatGPT as a supportive tool for brainstorming, clarification, or linguistic enhancement while maintaining intellectual accountability and ethical standards (Ruksakulpiwat et al. 2024). Academic integrity, therefore, can be preserved not by rejecting technological advancement but by integrating it within a deliberate ethical and pedagogical framework.

Recent qualitative evidence from nursing education adds an empirical anchor to these largely conceptual concerns. In a descriptive phenomenological study of fourteen nursing faculty members, Durmuş Sarıkahya et al. (2025) found that ChatGPT was valued for supporting theoretical learning, time management and the preparation of instructional materials, yet faculty consistently identified the same risks that recur throughout this review, namely uncertain information accuracy, a perceived erosion of students' critical thinking and unresolved ethical concerns. Faculty also judged the tool useful for care planning and patient education but insufficient for developing practical clinical skills, which echoes the performance ceilings reported in the clinical-reasoning and care-planning literature discussed later in this review. The participants framed the appropriate response not as prohibition but as a need for explicit institutional ethical policies and structured faculty training, aligning closely with the AI literacy and governance measures advocated here. This convergence between faculty experience and the wider evidence base reinforces the argument that the hallucination, accountability and integrity risks examined in this review are not merely theoretical but are already shaping how educators adopt the technology.

3 | ChatGPT in Clinical Nursing Practice

3.1 | Clinical Reasoning and Symptom Assessment

Clinical reasoning and symptom assessment represent the interpretive core of nursing practice, requiring the transformation of evolving patient data into safe and accountable decisions. Emerging research suggests that large language models (LLMs) may assist in preliminary reasoning tasks by summarizing symptom descriptions, generating differential considerations and referencing general clinical guidelines (Shin et al. 2024; Yu et al. 2026). In simulation-based and quasi-experimental educational settings, AI-assisted reasoning has been associated with increased analytic confidence and structured reflection among nursing learners, indicating potential value as a cognitive scaffold during early stages of reasoning development (Saad et al. 2025; Shin et al. 2024).

However, empirical evidence consistently demonstrates that performance is highly context-dependent. LLMs tend to perform adequately when reasoning about common, well-documented conditions but show marked deterioration in specialized or data-sparse clinical domains (Choi et al. 2024; Geracitano et al. 2025). Studies examining symptom interpretation reveal that AI systems may produce plausible yet incomplete or inaccurate conclusions, particularly when nuanced clinical judgement is required. Diamond et al. (2024) demonstrated that generative AI struggled to reason over nursing flowsheet data, highlighting the gap

between linguistic fluency and genuine clinical understanding. A structural explanation for these failures lies in the phenomenon of hallucinations, confidently generated but factually incorrect outputs arising from probabilistic language modelling rather than grounded clinical reasoning (Ji et al. 2023). In healthcare contexts, such hallucinated statements may appear persuasive while lacking evidentiary support, posing safety risks if accepted uncritically (Meskó and Topol 2023).

A further limitation concerns the dynamic nature of clinical reasoning. Follow-up investigations indicate that LLM recommendations often remain static when new patient information is introduced, failing to recalibrate hypotheses in response to evolving clinical contexts (Geracitano et al. 2025). This rigidity contrasts with the iterative reasoning demanded in real patient care, where symptom trajectories require continuous reassessment. Educational use without explicit supervision therefore risks reinforcing fixed interpretive shortcuts instead of adaptive clinical thinking (Gunawan et al. 2024).

From an educational perspective, the value of AI lies in its capacity to function as a dialogic reasoning partner rather than an authoritative decision-maker. When embedded within supervised simulation and debriefing environments, AI-generated differentials can prompt comparison, critique and metacognitive reflection, strengthening learners' awareness of uncertainty and professional accountability (Ammenwerth and Zwan 2025; Jang et al. 2025). Students themselves report ambivalent trust in AI outputs, recognizing usefulness while rejecting substitution for human judgement (Gunawan et al. 2024). This ambivalence may serve as a productive pedagogical entry point for teaching epistemic vigilance and verification practices.

Normatively, integration of generative AI into symptom assessment education requires explicit safeguards. Learners must be trained to treat AI suggestions as provisional hypotheses requiring validation against clinical evidence and professional standards (Watson et al. 2025). Educators retain responsibility for modelling diagnostic reasoning, supervising AI-assisted exercises and ensuring that independent judgement remains central to decision-making (Hersh 2025). Under these conditions, LLMs may enrich exposure to reasoning scenarios without displacing the ethical and cognitive foundations of nursing practice. The challenge is not technological adoption per se, but disciplined integration that preserves human accountability while leveraging AI as an educational amplifier.

3.2 | Emergency Nursing and Triage

In acute care settings, such as emergency departments (EDs), often defined by high patient volumes, clinical complexity and time-critical decision-making, LLMs such as ChatGPT present a potentially ideal implementation setting. Offering solutions for common challenges in acute care such as real-time documentation, rapid access to up-to-date evidence and the creation of plain-language patient education materials, these tools have the potential to alleviate the substantial cognitive load and stress-related burnout experienced by many acute care nurses and other healthcare professionals (Sarraf and Ghasempour 2025; Szymczyk et al. 2024).

Despite these potential benefits, research on ChatGPT in emergency care settings remains limited, particularly in the nursing domain and studies that do exist report inconsistent performance (Mu and He 2024). Meyer and Meyer (2025) synthesized findings from ten experimental studies evaluating ChatGPT for ED triage—the most extensively examined application to date. Six studies reported diagnostic accuracy comparable to medical residents or closely aligned with gold-standard triage practices, whereas four studies identified significant performance limitations, with only moderate agreement with established triage procedures. Arslan et al. (2025) conducted a single-site comparative study of ChatGPT Plus, Copilot Pro and triage nurses, finding similar overall accuracy across groups (65.2% for nurses, 66.5% for ChatGPT and 61.8% for Copilot). Although triage alignment is similar between LLM and healthcare professionals, findings are emerging suggesting that ChatGPT has a tendency to over-triage patients to higher acuity levels than warranted (Masannek et al. 2024). It remains unclear whether this pattern is intrinsic to LLMs or reflects intentional design choices that favour caution.

Kaboudi et al. (2024) further detailed these discrepancies, noting that ChatGPT occasionally underestimated the urgency of certain emergency scenarios or was unable to provide a diagnosis of rare conditions and at times, produced unsafe triage recommendations. Other studies similarly indicate that the model may overlook red-flag symptoms or misinterpret subtle contextual cues, suggesting that it is not currently suitable for autonomous triage decisions (Wang, Zhang, et al. 2024). Further compounding concerns, ChatGPT can occasionally generate reasonable-sounding but incorrect information (hallucinations) (Antonie et al. 2025). This makes it difficult for clinicians to distinguish accurate content from misinformation. In the fast-paced ED environment, where providers rely on precise, up-to-date evidence for critical decisions, time spent verifying unsupported findings would render the tool more burdensome than beneficial.

Despite these potential benefits, substantial challenges must be addressed before widespread adoption in acute care settings is feasible. ChatGPT's reliance on probabilistic language generation raises concerns about the reliability of its outputs, particularly in complex clinical situations where even minor inaccuracies may lead to inappropriate triage decisions or diagnostic errors. To advance safe and effective integration, further setting-specific research—particularly studies reflecting the multidisciplinary nature of the ED and acute care settings and the central role of nurses—is essential to robustly evaluate ChatGPT's performance.

3.3 | Patient Communication and Education

Large language models (LLMs) can support nurses in patient communication by translating complex medical terminology into accessible language adjusted to different levels of health literacy. Clear and comprehensible communication is central to nursing practice, influencing patient safety, adherence to treatment and self-management outcomes. Tools such as ChatGPT have therefore been evaluated for their ability to generate patient-oriented educational content.

Yeo et al. (2023) assessed ChatGPT's performance in answering patient questions regarding cirrhosis and hepatocellular carcinoma and found that responses were generally accurate and understandable, although variability in completeness was observed. These findings suggest that LLMs may serve as supportive tools in preparing educational materials related to chronic diseases.

Evidence from health communication research indicates that LLMs can reduce the reading level of complex patient education materials. Will et al. (2025) demonstrated that large language models significantly improved the readability of online patient education texts while preserving essential clinical content. Similarly, Swisher et al. (2024) showed that ChatGPT-revised patient handouts achieved improved readability scores compared with original versions, although expert review remained necessary to ensure accuracy and contextual adequacy.

A scoping review by Aydin et al. (2024) synthesizing applications of LLMs in patient education concluded that these tools may enhance accessibility and engagement but identified persistent concerns regarding factual accuracy, bias and variability across medical specialties. These findings are consistent with specialty-specific evaluations, which report improved linguistic simplification but inconsistent depth and completeness of clinical information.

Concerns regarding quality and reliability remain significant. Fahy et al. (2024) identified inconsistencies in content precision and omissions of relevant safety information in AI-generated materials related to platelet-rich plasma therapy. More broadly, studies evaluating AI-generated medical information highlight the risk of incomplete guidance, lack of contextualisation and insufficient emphasis on contraindications or region-specific recommendations.

In nursing practice, such limitations have direct implications for patient safety, informed consent and health outcomes. Therefore, AI-generated materials should be treated as preliminary drafts subject to professional verification, contextual adaptation and alignment with evidence-based guidelines. The responsibility for accuracy, ethical appropriateness and cultural adaptation remains with the nurse.

Overall, ChatGPT and related LLMs demonstrate potential as adjunct tools to enhance patient communication and education within clinical nursing practice; however, their implementation requires structured oversight and critical evaluation to mitigate risks associated with misinformation and variable quality.

Evidence from highly specialized clinical fields further illustrates both the expanding scope and the conditional nature of these applications. In paediatric oncology nursing, Çınar Özbay et al. (2026) explored paediatric nurses' perspectives on the growing role of artificial intelligence in this setting and reported that nurses recognized meaningful potential for AI-supported clinical decision-making and for the preparation of patient and family education materials, while raising substantive ethical concerns and pointing to new competencies that nurses in highly specialized fields will need in order to appraise and supervise AI-generated content. These observations mirror the

pattern documented across the general clinical literature: the more specialized, data-sparse and ethically sensitive the clinical context, the greater the need for rigorous human oversight, domain-specific validation and targeted education before generative AI can safely inform direct patient care.

4 | ChatGPT in Nursing Workflow: Documentation, Care Plans and Administrative Tasks

4.1 | Documentation Efficiency

Nursing documentation is foundational to continuity of care, interprofessional communication, quality assurance and medico-legal accountability, yet it remains one of the principal contributors to workload and professional stress. LLMs such as ChatGPT offer novel opportunities to enhance documentation efficiency by automating routine writing tasks, structuring narrative text and tailoring language to diverse audiences (Zaretsky et al. 2024). Within this context, ChatGPT has been explored as a drafting assistant for clinical notes, discharge summaries, handover reports and patient education materials, thereby real-locating nurses' time from administrative burdens to direct patient care (Bracken et al. 2025).

Emerging evidence indicates that AI-generated documentation can approach the quality of clinician-authored texts under controlled conditions. In a feasibility study using mock patient vignettes, Clough et al. (2024) found that general practitioners rated ChatGPT-generated discharge summaries as acceptable and comparable to those produced by junior doctors. Similarly, other studies have demonstrated that generative AI can organize clinical information coherently, enhance clarity and standardize structure, thereby supporting greater consistency in documentation practices (Ganzinger et al. 2025; Perkins et al. 2024). For nursing practice, potential applications include transforming technical terminology into patient-friendly language, generating preliminary notes from bullet-point data and producing template-based narratives that nurses can adapt and verify (Ju et al. 2025; Zaretsky et al. 2024).

Nevertheless, significant safety and ethical concerns persist. Because ChatGPT relies on probabilistic text generation rather than genuine clinical reasoning, it may introduce hallucinations—fabricated clinical details, erroneous temporal sequences, or recommendations that contradict established guidelines (Asgari et al. 2025). Even minor inaccuracies, if unreviewed, can compromise patient care and legal accountability. Consequently, scholars emphasize a human-in-the-loop model wherein nurses retain ultimate responsibility for reviewing, editing and validating AI-produced content before integration into clinical records (Bakken 2023; Perkins et al. 2024).

Data protection and regulatory compliance further constrain real-world implementation. Entering identifiable health data into publicly accessible AI platforms may breach confidentiality standards and violate data protection laws such as the General Data Protection Regulation (GDPR) within the European Union (Y. Chen and Esmaeilzadeh 2024). To mitigate this risk, researchers propose the adoption of institutionally hosted or

enterprise-grade AI systems equipped with secure infrastructures, traceable audit mechanisms and well-defined governance frameworks that delineate authorized use, roles and accountability (Dennstädt et al. 2025).

Within these governance boundaries, ChatGPT also shows potential in supporting broader administrative functions—such as summarizing meetings, drafting correspondence and preparing policy documentation—which can further ease the documentation burden (Y. Chen and Esmaeilzadeh 2024). Overall, literature positions generative AI not as a replacement for professional documentation but as a complementary instrument capable of enhancing efficiency when embedded within robust ethical, institutional and regulatory safeguards (Y. Chen and Esmaeilzadeh 2024; Clough et al. 2024; Perkins et al. 2024).

4.2 | Care Planning

Recent literature examining ChatGPT's application in nursing care plan development reveals significant limitations that underscore the necessity for human oversight and verification. While large language models demonstrate proficiency in retrieving and organizing clinical knowledge, their capacity to produce clinically sound, individualized nursing care plans remains constrained by several critical factors.

First, ChatGPT exhibits substantial limitations in clinical reasoning depth and contextualization. Studies indicate that AI-generated care plans often lack the nuanced clinical judgement required to prioritize interventions based on individual patient complexity, comorbidities and psychosocial factors (Dağci et al. 2024; Sallam 2023; Thirunavukarasu et al. 2023). The technology tends to produce generic, textbook-like responses that fail to capture the holistic assessment fundamental to nursing practice, particularly when addressing patients with multi-morbidity or atypical presentations (Cascella et al. 2023; Sblendorio et al. 2024).

Second, significant concerns emerge regarding accuracy and potential for clinical errors. Research demonstrates that ChatGPT can generate plausible-sounding but factually incorrect recommendations, a phenomenon known as 'hallucination' (Leclerc et al. 2023). Dağci et al. (2024) specifically evaluated nursing care planning texts generated by ChatGPT and found concerning issues with reliability and quality that necessitate careful verification. The model's inability to access real-time patient data, current clinical guidelines, or institutional protocols further compromises its reliability, with documented instances of outdated practices or misalignment with evidence-based standards (Kung et al. 2023; Sblendorio et al. 2024; Thirunavukarasu et al. 2023).

Third, ChatGPT demonstrates limited capacity for person-centered care planning. Nursing care plans require integration of patient preferences, values, cultural considerations and individual goals—dimensions that AI struggles to incorporate meaningfully (Dağci et al. 2024; Sallam 2023). Effective care planning extends beyond clinical data to encompass therapeutic relationships, patient advocacy and ethical deliberation, competencies that remain distinctly human (Cascella et al. 2023).

Additionally, researchers highlight concerns regarding professional accountability and liability when AI-generated content is used without thorough validation (Thirunavukarasu et al. 2023). Sblendorio et al. (2024) emphasize the necessity of integrating human expertise with automated methods for dynamic evaluation of large language models' feasibility in clinical decision-making.

The emerging consensus advocates for a collaborative human-AI model rather than autonomous AI application. Scholars recommend that ChatGPT be positioned as a supportive tool requiring rigorous nurse validation, adaptation to individual patient contexts and integration with clinical expertise (Dağci et al. 2024; Sallam 2023; Thirunavukarasu et al. 2023). This approach preserves essential clinical judgement, ethical reasoning and therapeutic presence that define professional nursing practice.

4.3 | Administrative and Organizational Benefits

Studies have demonstrated that ChatGPT can support administrative tasks such as drafting emails, summarizing meeting notes and converting complex documents into different formats and in some instances different languages (Jo and Park 2024; Shojaei et al. 2024). A systematic review by Bracken et al. (2025) of 11 studies evaluating AI-driven documentation systems, including ChatGPT and ambient AI technologies showed significant improvements in documentation efficiency, reducing administrative burden for clinicians. Healthcare professionals also reported reduced task load, improved ease of use and enhanced workflow efficiency yet expressed concerns about the reliability and validity of AI-generated documentation.

In addition, some research also suggests that ChatGPT can add an extra layer of support providing unbiased decision-making in certain contexts by avoiding gender or socioeconomic bias. For example, a study compared ChatGPT with 1249 primary care physicians on depression management and found ChatGPT showed no gender or socioeconomic bias in recommending psychotherapy or antidepressants, whereas physicians demonstrated bias in that women were often prescribed fewer antidepressants than men (Levkovich and Elyoseph 2023). Yet the authors of these studies caution that isolated neutrality does not guarantee systemic fairness (Bracken et al. 2025; Shojaei et al. 2024).

At an organizational level, ChatGPT and other LLMs may offer further benefits by strengthening the education and training pipeline for healthcare professionals. In medical education, ChatGPT has been demonstrated to enhance learning for medical students by generating concise, accessible explanations of complex medical concepts, reducing reliance on lengthy traditional texts and enabling more efficient knowledge acquisition (Wu et al. 2024). Additionally, it supports the development of multilingual digital learning resources, such as interactive tutorials and instructional videos, broadening access to high-quality training materials (McGee et al. 2024). Comparable gains have been demonstrated in nursing education. A cross-sectional study of Bachelor of Nursing students in Spain found that learners who used ChatGPT experienced significant improvements in academic grades, with 89.5% reporting enhanced

academic performance (Gonzalez-Garcia et al. 2025). Notably, women nursing students perceived ChatGPT as particularly helpful for completing academic tasks, highlighting its potential to offer equitable, personalized learning support. These educational benefits have broader implications for addressing the global healthcare workforce shortage. By accelerating learning, improving academic outcomes and expanding access to educational materials across languages and regions, ChatGPT may be able to help scale the training of healthcare professionals more efficiently than traditional methods alone. This capability is especially important for countries facing acute shortages of physicians, nurses and allied health workers.

5 | Ethical, Legal and Professional Considerations

5.1 | Accuracy and Safety

The integration of ChatGPT and large language models (LLMs) in contemporary nursing practice raises critical ethical, legal and professional considerations centered on accuracy and safety. Hallucinations—confident but incorrect statements generated by AI systems—pose significant risks to clinical nursing care. Even small inaccuracies can lead to unsafe care if not identified by clinicians, highlighting the potential for cascading errors when flawed AI-generated recommendations enter clinical workflows (Alkaissi and McFarlane 2023; Chelli et al. 2024).

Numerous studies document the presence of hallucinations across critical nursing domains. Siontis et al. (2024) explore the humanlike nature of ChatGPT's hallucinations, noting that the model's ability to generate plausible yet incorrect clinical content creates substantial verification burdens for healthcare professionals. Colakca et al. (2024) evaluated ChatGPT's performance in emergency department triaging and identified concerning inaccuracies that could compromise patient safety during time-sensitive decisions. Dağci et al. (2024) examined nursing care planning texts generated by ChatGPT and found significant reliability and quality issues. Research demonstrates hallucination rates exceeding 25% in systematic review reference generation (Chelli et al. 2024) and up to 83% in clinical vignettes containing fabricated details (Asgari et al. 2025), underscoring the pervasive nature of this threat to clinical accuracy.

These findings create substantial ethical obligations for nurses and healthcare institutions. The principle of non-maleficence demands rigorous validation protocols before implementing AI recommendations in practice (Armitage 2025). Professional accountability standards require that nurses maintain competence and ensure accuracy in all clinical decisions, including those informed by AI tools (American Nurses Association 2022). Legally, the use of LLM-generated content without adequate verification may expose practitioners and institutions to malpractice liability, particularly when errors result in patient harm. However, determining responsibility when an LLM contributes to diagnostic errors or adverse outcomes remains complex and currently lacks clear legal precedent, creating an accountability vacuum that poses significant risks (Jung 2025).

The phenomenon of hallucinations also raises questions about informed consent and patient autonomy. When AI tools

contribute to care planning, patients have the right to understand the role of technology in their care and the potential for computational errors. Professional nursing organizations increasingly emphasize that LLMs should serve only as adjunct tools, with final clinical decisions remaining under human professional judgement (American Nurses Association 2022). The literature converges on a consensus that mitigating AI-related risks requires comprehensive human oversight, mandatory verification protocols, institutional governance frameworks and ongoing professional education. Patient safety, clinical accountability and ethical nursing practice must take precedence over technological convenience (Fareed et al. 2025; Welch et al. 2026).

5.2 | Confidentiality and GDPR

The use of public AI platforms poses risks for patient confidentiality. Uploading any identifiable clinical data into external systems violates privacy standards and may breach GDPR within the EU. Nursing organizations must establish policies for safe AI use, including institutionally hosted models and clear guidelines for staff.

The integration of artificial intelligence into healthcare is increasingly seen as a game-changer that can boost diagnostic accuracy, aid in clinical decision-making and ultimately improve patient outcomes. However, as AI applications, including large language models like ChatGPT, continue to grow rapidly, we need to approach their use with careful ethical and regulatory consideration Almufarreh et al. (2025). Without unified standards and strong oversight, rolling out these technologies could lead to inconsistent clinical practices, legal uncertainties and potential risks to patient safety (Gaddas et al. 2025). Tackling these issues calls for not just updated regulatory frameworks, but also a fresh look at professional accountability, risk management strategies and how we evaluate AI-driven interventions.

The World Health Organization (WHO) has laid out six guiding principles for the ethical use of AI in healthcare (Goodman et al. 2023): protecting autonomy; promoting welfare, safety and the public good; ensuring transparency and explainability; maintaining accountability; fostering inclusivity and equity; and being responsive and sustainable. These principles serve as a solid foundation for responsible innovation and implementation. When it comes to autonomy, AI systems should be designed in line with established ethical standards and fundamental human rights (Pham 2025). Their role should be to support, not replace, clinical judgement and patient involvement in decision-making. Respecting autonomy means not only enabling informed choices but also ensuring strong data protection, confidentiality and the validity of informed consent. These protections need to be backed by clear and enforceable legal frameworks that govern digital health technologies.

Regarding welfare and safety, AI tools must go through thorough validation before they are used in clinical settings, proving their reliability, accuracy and effectiveness (Park et al. 2021). Regulatory compliance should not stop at initial approval; it should also include ongoing monitoring and evaluation to ensure these tools continue to meet safety and efficacy standards.

5.3 | Bias and Fairness

LLMs reflect biases from their training data. Although some studies show neutral outputs in specific tasks, this does not mitigate broader systemic biases. Ensuring equitable performance requires ongoing evaluation and careful deployment in sensitive clinical contexts.

Bias can be understood as a systematic distortion in judgement or decision-making that produces inequitable or unjust outcomes (Cross et al. 2024). Within the field of AI, bias does not emerge from a single source; rather, it may be embedded at multiple stages of the system's lifecycle. It can originate during data collection, be reinforced through algorithmic design choices, or be introduced through the ways in which users interpret and apply AI-generated outputs (Hasanzadeh et al. 2025).

Machine learning models, in particular, are especially sensitive to this issue. Because they identify and generalize patterns from historical datasets, they may inadvertently absorb and reproduce existing social, cultural, or structural inequalities reflected in those data. When deployed in real-world contexts—especially in healthcare—such distortions can translate into discriminatory recommendations, unequal access to services, or systematically different outcomes across population groups.

Bias in AI systems has the potential not only to reproduce existing social inequalities, but also to intensify them (Norori et al. 2021). When embedded in healthcare or other essential services, biased algorithms may disadvantage already marginalized groups, restricting access to care or influencing decisions in ways that systematically penalize certain populations. Beyond reinforcing well-known gender stereotypes, AI-driven processes can generate new discriminatory patterns linked to skin colour, ethnicity, or other physical characteristics, particularly when these attributes are implicitly encoded in training data (Belenguer 2022).

To address these concerns, researchers and practitioners have proposed several technical strategies for bias mitigation (Nazer et al. 2023). Interventions may occur at different stages of the AI lifecycle: during data pre-processing (e.g., through rebalancing or data augmentation), at the modelling stage (through fairness-aware algorithms), or after model deployment (via post-processing adjustments to outputs). Despite these efforts, significant challenges remain. Training datasets are often insufficiently diverse, making it difficult to correct structural imbalances. Moreover, bias itself is not always easy to detect or quantify, as different metrics may capture different—and sometimes conflicting—dimensions of unfairness. In practice, improving fairness can also require trade-offs with other performance indicators, such as predictive accuracy. Ethical questions therefore arise regarding which forms of bias should be prioritized and which social groups should receive particular protection.

The concept of fairness in AI, such as ChatGPT is consequently complex and contested (Ueda et al. 2024). Although it is commonly understood as the absence of unjust discrimination, the operationalization of fairness depends on the criteria adopted and the context in which the system is used.

The literature distinguishes among several approaches. Group fairness focuses on ensuring that outcomes are distributed equitably across defined demographic groups (Van Der Meijden et al. 2025). Within this framework, notions such as demographic parity aim to prevent disproportionate allocation of positive or negative outcomes among groups. Individual fairness, by contrast, is grounded in the principle that similar individuals should be treated in similar ways, irrespective of group membership (Anderson and Visweswaran 2024). This perspective often relies on similarity-based metrics designed to compare individuals with comparable characteristics. More recently, counterfactual fairness has been proposed, emphasizing that outcomes should remain stable even if sensitive attributes—such as gender or ethnicity—were hypothetically altered.

Taken together, these perspectives highlight that fairness in AI cannot be reduced to a single technical adjustment. Rather, it requires ongoing normative reflection, methodological rigour and contextual evaluation, particularly in high-stakes domains such as healthcare.

5.4 | Professional Responsibility

The integration of large language models (LLMs) into clinical nursing practice raises important questions regarding professional responsibility, accountability and autonomy. While AI tools such as ChatGPT may enhance efficiency and support aspects of care delivery, excessive reliance on algorithm-generated content risks diminishing clinicians' independent judgement and critical thinking. Nursing is a profession grounded in autonomous clinical reasoning and ethical accountability, competencies that cannot be delegated to automated systems without compromising safe, person-centered care (Wang et al. 2023).

Several authors analysing the role of AI in health professions caution that the fluent and polished language characteristic of LLM outputs may create a false sense of accuracy, which in turn can erode critical thinking among students and staff if uncritically accepted (Abujaber et al. 2023; O'Connor and ChatGPT 2023). Abujaber et al. (2023) highlighted the potential threat that ChatGPT and similar tools pose to the development of autonomous clinical judgement in nursing education, emphasizing that learners must be trained to critically appraise AI-generated information rather than accept it at face value. O'Connor and ChatGPT (2023) similarly argued that generative AI platforms can function as tools for academic progress or as mechanisms that encourage superficial reasoning and misuse, depending on the presence of oversight and guidelines.

From an ethical perspective, overreliance on AI also risks undermining the humanistic and relational dimensions of care. Analyses of ethical concerns in healthcare settings emphasize that uncritical adoption of AI can weaken compassion, disrupt professional patient relationships and erode trust if the clinician's role becomes secondary to algorithmic suggestion (Wang et al. 2023). Ensuring transparency about AI use, preserving the clinician's authority in final decision-making and maintaining active professional engagement with patients are therefore critical considerations.

Legal and regulatory frameworks further underscore the non-delegable nature of professional responsibility. Current guidelines for AI deployment in healthcare emphasize that clinicians must retain ultimate authority over diagnosis and treatment decisions and that users of high-risk AI systems should possess the competencies necessary to evaluate and intervene in automated recommendations. In nursing practice, this translates into the obligation to interpret, verify and contextualize AI outputs rather than treating them as definitive clinical guidance.

Moreover, broader ethical principles such as accountability, transparency and explainability are being incorporated into international frameworks for trustworthy AI in healthcare, reinforcing the imperative that professional judgement remains central to clinical practice (Lekadir et al. 2025). These frameworks advocate for ongoing ethical scrutiny, including human oversight and clear allocation of responsibility, to ensure that the adoption of AI does not compromise professional standards.

In sum, while ChatGPT and other LLMs may augment specific tasks in nursing, maintaining professional responsibility requires that nurses use these tools as supportive aids, not substitutes for clinical judgement. Clear institutional policies, education in digital and AI literacy and reinforcement of reflective practice are essential to preserve autonomy, safeguard patient trust and uphold the ethical foundations of the nursing profession (Lekadir et al. 2025; Wang et al. 2023).

5.5 | Beyond ChatGPT: Comparative Evidence Across Contemporary LLMs

Although this review, consistent with the published literature, draws predominantly on studies of ChatGPT, the conclusions should be read against the broader and rapidly diversifying landscape of large language models, including Claude (Anthropic), Gemini (Google), Copilot (Microsoft) and open-weight models such as Llama (Meta). Direct comparative evidence in nursing remains scarce but is beginning to emerge. In a 100-question nursing competency examination, Dilan S. Hiwa et al. (2024) found that Microsoft Copilot achieved the highest accuracy (84%), followed by ChatGPT-3.5 (77%), Gemini (75%) and Llama 2 (68%), with no single system answering all items correctly and each model uniquely answering at least one question that the others missed. In emergency triage, Arslan et al. (2025) reported broadly comparable accuracy between ChatGPT Plus, Copilot Pro and triage nurses, as discussed earlier in this review. Outside nursing, a Polish comparative study of medical and dental licensing examinations found that Claude achieved the highest accuracy of the models tested and, importantly, that performance varied significantly with the language of the prompt and the specialty examined, with Gemini performing markedly worse in Polish than in English (Wójcik et al. 2025).

Taken together, this emerging comparative literature indicates that the strengths and failure modes described in this review, including hallucination, inconsistency across repeated queries and degradation in specialized domains, are properties of the current generation of LLMs as a class rather than of ChatGPT

alone, while the magnitude of these problems is conditional on the specific model, its version, the task and the language of use. Two implications follow. First, findings obtained with ChatGPT cannot be assumed to transfer to other models and vice versa, so institutional procurement and governance decisions should require local, task-specific and language-specific validation of whichever system is deployed. Second, the sensitivity of model performance to prompt language is directly relevant to non-English-speaking nursing workforces, including the Central European context discussed in this review and argues for multilingual evaluation as a standard component of future nursing research on generative AI.

5.6 | Critical Synthesis Across Domains

Read together, the five domains describe a single underlying pattern rather than five separate literatures. Across education, clinical reasoning, triage, communication, documentation and care planning, the usefulness of ChatGPT tracks a common gradient defined by two properties of the task: how far it depends on language as opposed to dynamic clinical judgement and how readily its output can be verified before it reaches a patient. Utility is highest and most consistent for language-centric, lower-stakes, verifiable tasks such as simplifying patient-education materials, drafting and restructuring documentation and summarizing administrative content and it declines steadily as tasks demand the integration of evolving patient data, individualized reasoning and accountable decision-making, as in symptom assessment, triage and care planning. The same failure mode recurs at every point on this gradient, namely fluent but unsupported output, or hallucination, amplified by automation bias and the same safeguard recurs in every domain, namely human-in-the-loop verification supported by institutional governance and AI literacy. This task-by-verifiability framework links the educational, clinical, workflow and ethical sections into one account and explains why the literature converges on augmentation rather than substitution.

Several genuine tensions in the evidence should not be smoothed over. Reported performance in emergency triage is openly contradictory, with some studies finding accuracy comparable to residents or to gold-standard practice and others finding only moderate agreement, alongside opposing error tendencies of over-triage in some samples and unsafe under-triage in others. The evidence on bias is similarly divided, since demonstrations of apparent neutrality on isolated tasks, such as the absence of gender bias in one depression-management comparison, sit beside well-documented systemic bias arising from training data and the former cannot be generalized to the latter. In education, the same tool that improved grades, engagement and satisfaction is also identified as a threat to the critical thinking and epistemic ownership on which safe practice depends and in documentation, output rated comparable to that of junior clinicians coexists with fabricated clinical detail in other reports. These contradictions are not noise to be averaged away but signals that performance is conditional on task type, model version, clinical complexity and study design, which is precisely why uniform claims about ChatGPT in nursing are untenable and why setting-specific evaluation and human verification remain essential.

6 | Implications for Nursing

The Polish health-care system faces challenges such as workforce shortages, administrative overload and accelerating digital transformation. Artificial intelligence technology that uses LLM like ChatGPT has the potential to fundamentally change nursing education, clinical practice and clinical leadership (Srinivasan et al. 2024; Zhou et al. 2024). For nursing educators, effective use of ChatGPT can be helpful in the creation of exam questions that are in alignment with regulatory test structure (Hargett et al. 2026). ChatGPT can assist nursing educators in generating clinical simulation scenarios with greater efficiency (Reid 2025). Moreover, the use of ChatGPT by nurse educators enhances student engagement and satisfaction with learning (Gonzalez-Garcia et al. 2025). With respect to clinical practice, the use of ChatGPT has been shown to significantly reduce the time spent documenting care while maintaining sufficient documentation quality (C.-J. Chen et al. 2024). ChatGPT also can be used to generate and modify quality education materials used in patient and family teaching (Tian Tran et al. 2024). Finally, ChatGPT is used by the majority of nurses working in clinical care and especially by nursing supervisors who use it to help make their administrative tasks more completed more efficiently (Lin et al. 2025).

Despite numerous existing and emerging prospects for the use of ChatGPT in nursing education and practice, there remain important ethical and professional alignment concerns (Abou Hashish et al. 2025; Cho Kwan et al. 2025). As such, safe adoption of ChatGPT in nursing requires the consideration of several guide rails. First, a nurse is ethically and, in most instances, legally responsible for every action taken even if ChatGPT was used to assist in clinical decision-making (American Nurses Association 2022). As such, ChatGPT and related LLMs should be considered as co-pilots to nurses, never as a replacement to expert nursing practice. Nurses have an ethical duty to provide human-centered care and as such any automation cannot result in displacement of a nurse in the caring process (Abou Hashish et al. 2025). Second, nursing education lags behind on ensuring that the nursing workforce is prepared for the integration of ChatGPT and related LLMs. As a result, most nursing graduates lack the training and experience to use ChatGPT and related LLMs safely and effectively (Cho Kwan et al. 2025). Third, data security remains a major barrier to implementation of ChatGPT and other LLMs into clinical care, largely due to uncertainties about access, integration and protection of protected health information (Abou Hashish et al. 2025; Cho Kwan et al. 2025). National guidance and institution-specific policies on the use of closed-loop versus public versions of artificial intelligence technology that uses LLM like ChatGPT would help with their implementation. Finally, numerous sources of bias and hallucinations plague even the most up-to-date LLMs (Haltaufderheide and Ranisch 2024). Therefore, nurses and other end users of ChatGPT and related LLMs must advocate for less error-ridden technologies before they can be safely integrated into clinical practice.

Taken together, these considerations can be condensed into a practical framework for the safe implementation of LLMs in nursing education and clinical practice, built on four mutually reinforcing

pillars. The first pillar is workforce preparation, comprising structured faculty development and AI literacy education that address model capabilities, limitations, bias, hallucination and data privacy for students, educators and practising nurses alike. The second pillar is institutional policy, meaning explicit, written rules that specify which platforms may be used (favouring closed-loop, enterprise-grade deployments over public tools), require transparent disclosure of AI use and ensure compliance with data protection obligations such as GDPR. The third pillar is human oversight, understood as mandatory human-in-the-loop verification in which every AI-generated output, whether an examination question, a patient education leaflet, or a documentation draft, is treated as a preliminary draft requiring professional review before use. The fourth pillar is ethical governance, comprising standing oversight structures that monitor model performance over time, audit outputs for bias, define professional accountability for AI-assisted decisions and review incidents in which AI-generated content contributed to error. Adopted together, these four pillars translate the evidence synthesized in this review into an actionable pathway that allows nursing to capture the efficiency and educational benefits of generative AI without compromising patient safety or professional standards.

7 | Strengths and Limitations

This review has several strengths, including a synthesis that is current to March 2026, coverage of education, clinical practice and workflow within a single analytical framework and explicit attention to European regulatory constraints that are under-represented in the predominantly Anglo-American literature. Nevertheless, important limitations must be acknowledged. First, as a narrative review, the synthesis did not include a formal risk-of-bias assessment or quantitative pooling and the selection and weighting of evidence necessarily involved expert judgement. Second, the field is evolving at an exceptional pace: large language models are updated continuously, often without public documentation of the changes, so performance figures reported here are tied to specific model versions and may not reflect the systems available at the time of reading. Third, the evidence base is dominated by simulation studies, controlled vignettes and perception surveys, whereas real-world implementation research with patient-level outcomes remains scarce, which limits the strength of any inference about safety and effectiveness in routine care. Fourth, publication bias is likely, since novel and positive findings about generative AI are more readily published than null or unfavourable results, and much of the literature has appeared in rapid-turnaround venues with variable peer review. Finally, the search was restricted to English-language publications and, reflecting the state of the literature, most included studies examined ChatGPT, so the findings may not generalize fully to other models or linguistic contexts.

8 | Future Research Directions

The research agenda that follows from this synthesis extends well beyond the general call for patient safety evidence. Longitudinal implementation studies are needed that follow real-world deployments of LLMs in clinical nursing environments over time and measure effects on documentation quality, workload, care

processes and patient outcomes, rather than relying on single-time-point simulations. In education, adequately powered practicing controlled trials of AI-supported teaching interventions should compare structured, supervised use of LLMs against conventional instruction, with clinical reasoning, knowledge retention and academic integrity as endpoints. Parallel work is required on measurement: validated instruments for assessing AI competency among nursing students, educators and practicing nurses do not yet exist and are a precondition both for curriculum evaluation and for workforce planning. Research should also support the development and empirical evaluation of nursing-specific AI governance frameworks, translating high-level regulatory instruments into auditable processes for verification, accountability and incident review at the level of wards and faculties. Finally, future studies should routinely benchmark multiple models and multiple languages rather than ChatGPT alone and should examine equity effects, including whether model performance differs across patient populations and linguistic groups.

9 | Conclusions

ChatGPT offers promising opportunities to support nursing education, documentation, communication and workflow efficiency. Evidence from published research suggests substantial benefits in generating educational materials, improving access to information, enabling interactive simulations and reducing administrative burden.

However, ChatGPT's limitations—including hallucinations, inaccurate triage recommendations, shallow reasoning, fabricated references and privacy risks—make it clear that the tool must be used cautiously. ChatGPT is best positioned as an *auxiliary support tool*, never a substitute for professional nursing judgement.

As digital health continues to expand, nurses will play a critical role in shaping ethical, safe and effective integration of AI into practice. Developing AI literacy, robust governance frameworks and long-term evaluation strategies is essential for ensuring that generative AI enhances, rather than undermines, the foundations of nursing care.

Funding

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

- Abou Hashish, E. A., S. A. Alsayed, and N. M. F. Abdel Razek. 2025. "Embracing AI in Academia: A Mixed Methods Study of Nursing Students' and Educators' Perspectives on Using ChatGPT." *PLoS One* 20, no. 7: e0327981. <https://doi.org/10.1371/journal.pone.0327981>.

- Abujaber, A. A., A. Abd-alrazaq, A. R. Al-Qudimat, and A. J. Nashwan. 2023. "A Strengths, Weaknesses, Opportunities, and Threats (SWOT) Analysis of ChatGPT Integration in Nursing Education: A Narrative Review." *Cureus* 15: e48643. <https://doi.org/10.7759/cureus.48643>.
- Alkaiissi, H., and S. I. McFarlane. 2023. "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing." *Cureus* 15: e35179. <https://doi.org/10.7759/cureus.35179>.
- Almufarreh, A., A. Ahmad, M. Arshad, C. W. Onn, and R. Elechi. 2025. "Ethical Implications of ChatGPT and Other Large Language Models in Academia." *Frontiers in Artificial Intelligence* 8: 1615761. <https://doi.org/10.3389/frai.2025.1615761>.
- American Nurses Association. 2022. *The Ethical Use of Artificial Intelligence in Nursing Practice*. American Nurses Association.
- Ammenwerth, E., and L. Zwan. 2025. "Integration of Digital Tools Into Clinical Reasoning Education: A Rapid Review." In *Studies in Health Technology and Informatics*, edited by M. Baumgartner, D. Hayn, B. Pfeifer, and G. Schreier. IOS Press. <https://doi.org/10.3233/SHTI250159>.
- Anderson, J. W., and S. Visweswaran. 2024. "Algorithmic Individual Fairness and Healthcare: A Scoping Review." *JAMIA Open* 8, no. 1: ooae149. <https://doi.org/10.1093/jamiaopen/ooae149>.
- Antonie, N. I., G. Gheorghe, V. A. Ionescu, L.-C. Tiucă, and C. C. Diaconu. 2025. "The Role of ChatGPT and AI Chatbots in Optimizing Antibiotic Therapy: A Comprehensive Narrative Review." *Antibiotics* 14, no. 1: 60. <https://doi.org/10.3390/antibiotics14010060>.
- Armitage, R. C. 2025. "Implications of Large Language Models for Clinical Practice: Ethical Analysis Through the Principlism Framework." *Journal of Evaluation in Clinical Practice* 31, no. 1: e14250. <https://doi.org/10.1111/jep.14250>.
- Arslan, B., C. Nuhoglu, M. O. Satici, and E. Altinbilek. 2025. "Evaluating LLM-Based Generative AI Tools in Emergency Triage: A Comparative Study of ChatGPT Plus, Copilot Pro, and Triage Nurses." *American Journal of Emergency Medicine* 89: 174–181. <https://doi.org/10.1016/j.ajem.2024.12.024>.
- Asgari, E., N. Montaña-Brown, M. Dubois, et al. 2025. "A Framework to Assess Clinical Safety and Hallucination Rates of LLMs for Medical Text Summarisation." *NPJ Digital Medicine* 8, no. 1: 274. <https://doi.org/10.1038/s41746-025-01670-7>.
- Ateeq, A., M. Alzoraiki, M. Milhem, and R. A. Ateeq. 2024. "Artificial Intelligence in Education: Implications for Academic Integrity and the Shift Toward Holistic Assessment." *Frontiers in Education* 9: 1470979. <https://doi.org/10.3389/educ.2024.1470979>.
- Aydin, S., M. Karabacak, V. Vlachos, and K. Margetis. 2024. "Large Language Models in Patient Education: A Scoping Review of Applications in Medicine." *Frontiers in Medicine* 11: 1477898. <https://doi.org/10.3389/fmed.2024.1477898>.
- Bakken, S. 2023. "AI in Health: Keeping the Human in the Loop." *Journal of the American Medical Informatics Association* 30, no. 7: 1225–1226. <https://doi.org/10.1093/jamia/ocad091>.
- Belenguer, L. 2022. "AI Bias: Exploring Discriminatory Algorithmic Decision-Making Models and the Application of Possible Machine-Centric Solutions Adapted From the Pharmaceutical Industry." *AI and Ethics* 2, no. 4: 771–787. <https://doi.org/10.1007/s43681-022-00138-8>.
- Benfatah, M., A. Marfak, E. Saad, A. Hilali, C. Nejari, and I. Youlyouz-Marfak. 2024. "Assessing the Efficacy of ChatGPT as a Virtual Patient in Nursing Simulation Training: A Study on Nursing Students' Experience." *Teaching and Learning in Nursing* 19, no. 3: e486–e493. <https://doi.org/10.1016/j.teln.2024.02.005>.
- Bracken, A., C. Reilly, A. Feeley, E. Sheehan, K. Merghani, and I. Feeley. 2025. "Artificial Intelligence (AI)—Powered Documentation Systems in Healthcare: A Systematic Review." *Journal of Medical Systems* 49, no. 1: 28. <https://doi.org/10.1007/s10916-025-02157-4>.
- Cascella, M., J. Montomoli, V. Bellini, and E. Bignami. 2023. "Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios." *Journal of Medical Systems* 47, no. 1: 33. <https://doi.org/10.1007/s10916-023-01925-4>.
- Chelli, M., J. Descamps, V. Lavoué, et al. 2024. "Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis." *Journal of Medical Internet Research* 26: e53164. <https://doi.org/10.2196/53164>.
- Chen, C.-J., C.-T. Liao, Y.-C. Tung, and C.-F. Liu. 2024. "Enhancing Healthcare Efficiency: Integrating ChatGPT in Nursing Documentation." In *Studies in Health Technology and Informatics*, edited by J. Mantas, A. Hasman, G. Demiris, et al. IOS Press. <https://doi.org/10.3233/SHTI240545>.
- Chen, Y., and P. Esmaeilzadeh. 2024. "Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges." *Journal of Medical Internet Research* 26: e53008. <https://doi.org/10.2196/53008>.
- Cho Kwan, R. Y., A. C. Yan Tang, J. Y. Ha Wong, et al. 2025. "Navigating the Integration of Artificial Intelligence in Nursing: Opportunities, Challenges, and Strategic Actions." *International Journal of Nursing Sciences* 12, no. 3: 241–245. <https://doi.org/10.1016/j.ijnss.2025.04.009>.
- Choi, Y. K., S.-Y. Lin, D. M. Fick, et al. 2024. "Optimizing Chatgpt's Interpretation and Reporting of Delirium Assessment Outcomes: Exploratory Study." *JMIR Formative Research* 8: e51383. <https://doi.org/10.2196/51383>.
- Chokkakula, S., S. Chong, B. Yang, et al. 2025. "Quantum Leap in Medical Mentorship: Exploring Chatgpt's Transition From Textbooks to Terabytes." *Frontiers in Medicine* 12: 1517981. <https://doi.org/10.3389/fmed.2025.1517981>.
- Çınar Özbay, S., D. Gelin, and S. Durmuş Sarıkahya. 2026. "Artificial Intelligence in Pediatric Oncology Care: Pediatric Nurses' Perspectives and Future Implications." *Seminars in Oncology Nursing* 42, no. 1: 152082. <https://doi.org/10.1016/j.soncn.2025.152082>.
- Clough, R. A. J., W. A. Sparkes, O. T. Clough, J. T. Sykes, A. T. Steventon, and K. King. 2024. "Transforming Healthcare Documentation: Harnessing the Potential of AI to Generate Discharge Summaries." *BJGP Open* 8, no. 1: BJGPO.2023.0116. <https://doi.org/10.3399/BJGPO.2023.0116>.
- Colakca, C., M. Ergin, H. S. Ozensoy, A. Sener, S. Guru, and A. Ozhasenekler. 2024. "Emergency Department Triage Using ChatGPT Based on Emergency Severity Index Principles: A Cross-Sectional Study." *Scientific Reports* 14, no. 1: 22106. <https://doi.org/10.1038/s41598-024-73229-7>.
- Cross, J. L., M. A. Choma, and J. A. Onofrey. 2024. "Bias in Medical AI: Implications for Clinical Decision-Making." *PLOS Digital Health* 3, no. 11: e0000651. <https://doi.org/10.1371/journal.pdig.0000651>.
- Dağci, M., F. Çam, and A. Dost. 2024. "Reliability and Quality of the Nursing Care Planning Texts Generated by ChatGPT." *Nurse Educator* 49, no. 3: E109–E114. <https://doi.org/10.1097/NNE.0000000000001566>.
- Dempere, J., K. Modugu, A. Hesham, and L. K. Ramasamy. 2023. "The Impact of ChatGPT on Higher Education." *Frontiers in Education* 8: 1206936. <https://doi.org/10.3389/educ.2023.1206936>.
- Dennstädt, F., J. Hastings, P. M. Putora, M. Schmerder, and N. Cihoric. 2025. "Implementing Large Language Models in Healthcare While Balancing Control, Collaboration, Costs and Security." *NPJ Digital Medicine* 8, no. 1: 143. <https://doi.org/10.1038/s41746-025-01476-7>.
- Diamond, C. J., J. Thate, J. B. Withall, R. Y. Lee, K. Cato, and S. C. Rossetti. 2024. "Generative AI Demonstrated Difficulty Reasoning on Nursing Flowsheet Data." *AMIA... Annual Symposium Proceedings. AMIA Symposium 2024*: 349–358.
- Durmuş Sarıkahya, S., Ö. Özbay, K. Torpuş, G. Usta, and S. Çınar Özbay. 2025. "The Impact of ChatGPT on Nursing Education: A

- Qualitative Study Based on the Experiences of Faculty Members.” *Nurse Education Today* 152: 106755. <https://doi.org/10.1016/j.nedt.2025.106755>.
- Eke, D. O. 2023. “ChatGPT and the Rise of Generative AI: Threat to Academic Integrity?” *Journal of Responsible Technology* 13: 100060. <https://doi.org/10.1016/j.jrt.2023.100060>.
- El Arab, R. A., O. A. Al Moosa, F. H. Abuadas, and J. Somerville. 2025. “The Role of AI in Nursing Education and Practice: Umbrella Review.” *Journal of Medical Internet Research* 27: e69881. <https://doi.org/10.2196/69881>.
- Evangelista, E. D. L. 2025. “Ensuring Academic Integrity in the Age of ChatGPT: Rethinking Exam Design, Assessment Strategies, and Ethical AI Policies in Higher Education.” *Contemporary Educational Technology* 17, no. 1: ep559. <https://doi.org/10.30935/cedtech/15775>.
- Fahy, S., M. Niemann, P. Böhm, T. Winkler, and S. Oehme. 2024. “Assessment of the Quality and Readability of Information Provided by ChatGPT in Relation to the Use of Platelet-Rich Plasma Therapy for Osteoarthritis.” *Journal of Personalized Medicine* 14, no. 5: 495. <https://doi.org/10.3390/jpm14050495>.
- Fareed, M., M. Fatima, J. Uddin, A. Ahmed, and M. A. Sattar. 2025. “A Systematic Review of Ethical Considerations of Large Language Models in Healthcare and Medicine.” *Frontiers in Digital Health* 7: 1653631. <https://doi.org/10.3389/fdgh.2025.1653631>.
- Ferrari, R. 2015. “Writing Narrative Style Literature Reviews.” *Medical Writing* 24, no. 4: 230–235. <https://doi.org/10.1179/2047480615Z.000000000329>.
- Fung, T. C. J., S. L. Chan, C. F. M. Lam, et al. 2025. “Effects of Generative Artificial Intelligence (GenAI) Patient Simulation on Perceived Clinical Competency Among Global Nursing Undergraduates: A Cross-Over Randomised Controlled Trial.” *BMC Nursing* 24, no. 1: 934. <https://doi.org/10.1186/s12912-025-03492-0>.
- Gaddas, M., M. Ben Dhiab, I. Ben Saida, and H. Ben Saad. 2025. “Artificial Intelligence in Hospitals: Legal Uncertainties and Emerging Risks for Patient Safety.” *EXCLI Journal* 24: 824–827. <https://doi.org/10.17179/EXCLI2025-8679>.
- Ganzinger, M., N. Kunz, P. Fuchs, et al. 2025. “Automated Generation of Discharge Summaries: Leveraging Large Language Models With Clinical Data.” *Scientific Reports* 15, no. 1: 16466. <https://doi.org/10.1038/s41598-025-01618-7>.
- Geracitano, J., B. Anderson, M. Coffel, et al. 2025. “The Accuracy of ChatGPT in Answering FAQs, Making Clinical Recommendations, and Categorizing Patient Symptoms: A Literature Review.” *Advances in Health Information Science and Practice* 1, no. 1: VXUL2925. <https://doi.org/10.63116/VXUL2925>.
- Gerçek, A., N. Çiftci, M. Durmuş, A. Sarman, Ö. Taşçı, and M. Yıldız. 2025. “ChatGPT in Nursing: Applications, Advantages, and Challenges in Education, Research, and Clinical Practice.” *Annals of Biomedical Engineering* 53, no. 11: 3202–3207. <https://doi.org/10.1007/s10439-025-03832-w>.
- Gonzalez-Garcia, A., D. Bermejo-Martinez, A. I. Lopez-Alonso, B. Trevisson-Redondo, C. Martín-Vázquez, and S. Perez-Gonzalez. 2025. “Impact of ChatGPT Usage on Nursing Students Education: A Cross-Sectional Study.” *Heliyon* 11, no. 1: e41559. <https://doi.org/10.1016/j.heliyon.2024.e41559>.
- Goodman, K. W., S. Litewka, R. Malpani, S. Pujari, and A. A. Reis. 2023. “Global Health and Big Data: The WHO’s Artificial Intelligence Guidance.” *South African Journal of Science* 119, no. 5/6: 14725. <https://doi.org/10.17159/sajs.2023/14725>.
- Guerreiro, K. T. M., and C. Aschidamini. 2025. “O conceito de meta-cognição aplicada à formação do enfermeiro: Uma revisão de escopo.” *International Journal of Education and Health* 9: e6298. <https://doi.org/10.17267/2594-7907ijeh.2025.e6298>.
- Gunawan, J., Y. Aunguroch, C. Marzilli, E. Chaerani, and J. Montayre. 2024. “Artificial Intelligence Chatbot as Perceived by Nursing Students: A Qualitative Study.” *SAGE Open* 14, no. 4: 21582440241303453. <https://doi.org/10.1177/21582440241303453>.
- Haltaufderheide, J., and R. Ranisch. 2024. “The Ethics of ChatGPT in Medicine and Healthcare: A Systematic Review on Large Language Models (LLMs).” *NPJ Digital Medicine* 7, no. 1: 183. <https://doi.org/10.1038/s41746-024-01157-x>.
- Hargett, J. L., A. L. Princiotta, P. F. Lowry, L. K. Hosey, and M. N. White. 2026. “ChatGPT as a Tool in Nursing Exam Design: Opportunities and Limitations.” *Teaching and Learning in Nursing* 21, no. 1: e269–e273. <https://doi.org/10.1016/j.teln.2025.09.003>.
- Hasanzadeh, F., C. B. Josephson, G. Waters, D. Adedinsewo, Z. Azizi, and J. A. White. 2025. “Bias Recognition and Mitigation Strategies in Artificial Intelligence Healthcare Applications.” *NPJ Digital Medicine* 8, no. 1: 154. <https://doi.org/10.1038/s41746-025-01503-7>.
- Hersh, W. 2025. “Generative Artificial Intelligence: Implications for Biomedical and Health Professions Education.” *Annual Review of Biomedical Data Science* 8, no. 1: 355–380. <https://doi.org/10.1146/annurev-biodatasci-103123-094756>.
- Hiwa, D. S., S. S. Abdalla, A. S. Muhialdeen, H. M. Hamasalih, and S. O. Karim. 2024. “Assessment of Nursing Skill and Knowledge of ChatGPT, Gemini, Microsoft Copilot, and Llama: A Comparative Study.” *Barw Medical Journal* 2, no. 3: 3–6. <https://doi.org/10.58742/bmj.v2i2.87>.
- Hobensack, M., H. Von Gerich, P. Vyas, et al. 2024. “A Rapid Review on Current and Potential Uses of Large Language Models in Nursing.” *International Journal of Nursing Studies* 154: 104753. <https://doi.org/10.1016/j.ijnurstu.2024.104753>.
- International Council of Nurses. 2023. “Digital Health Transformation and Nursing Practice”.
- Jang, A., M. Oh, and M. O. Song. 2025. “Question-Centered Prompt Engineering for Nursing Simulation Debriefing: Model Development and Validation.” *Nurse Education in Practice* 88: 104541. <https://doi.org/10.1016/j.nepr.2025.104541>.
- Ji, Z., N. Lee, R. Frieske, et al. 2023. “Survey of Hallucination in Natural Language Generation.” *ACM Computing Surveys* 55, no. 12: 1–38. <https://doi.org/10.1145/3571730>.
- Jin, H. K., H. E. Lee, and E. Kim. 2024. “Performance of ChatGPT-3.5 and GPT-4 in National Licensing Examinations for Medicine, Pharmacy, Dentistry, and Nursing: A Systematic Review and Meta-Analysis.” *BMC Medical Education* 24, no. 1: 1013. <https://doi.org/10.1186/s12909-024-05944-8>.
- Jo, H., and D.-H. Park. 2024. “Effects of Chatgpt’s AI Capabilities and Human-Like Traits on Spreading Information in Work Environments.” *Scientific Reports* 14, no. 1: 7806. <https://doi.org/10.1038/s41598-024-57977-0>.
- Ju, H., M. Park, H. Jeong, et al. 2025. “Generative AI-Based Nursing Diagnosis and Documentation Recommendation Using Virtual Patient Electronic Nursing Record Data.” *Healthcare Informatics Research* 31, no. 2: 156–165. <https://doi.org/10.4258/hir.2025.31.2.156>.
- Jung, K.-H. 2025. “Large Language Models in Medicine: Clinical Applications, Technical Challenges, and Ethical Considerations.” *Healthcare Informatics Research* 31, no. 2: 114–124. <https://doi.org/10.4258/hir.2025.31.2.114>.
- Kaboudi, N., S. Firouzbakht, M. Shahr Eftekhari, et al. 2024. “Diagnostic Accuracy of ChatGPT for Patients’ Triage: a Systematic Review and Meta-Analysis.” *Archives of Academic Emergency Medicine* 12, no. 1: e60. <https://doi.org/10.22037/aaem.v12i1.2384>.
- Kung, T. H., M. Cheatham, A. Medenilla, et al. 2023. “Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models.” *PLOS Digital Health* 2, no. 2: e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.

- Lecler, A., L. Duron, and P. Soyer. 2023. "Revolutionizing Radiology With GPT-Based Models: Current Applications, Future Possibilities and Limitations of ChatGPT." *Diagnostic and Interventional Imaging* 104, no. 6: 269–274. <https://doi.org/10.1016/j.diii.2023.02.003>.
- Lekadir, K., A. F. Frangi, A. R. Porras, et al. 2025. "FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare." *BMJ* 388: e081554. <https://doi.org/10.1136/bmj-2024-081554>.
- Levkovich, I., and Z. Elyoseph. 2023. "Identifying Depression and Its Determinants Upon Initiating Treatment: ChatGPT Versus Primary Care Physicians." *Family Medicine and Community Health* 11, no. 4: e002391. <https://doi.org/10.1136/fmch-2023-002391>.
- Lin, H., L. Liao, Y. Wang, and L. Chang. 2025. "Attitude and Utilization of ChatGPT Among Registered Nurses: A Cross-Sectional Study." *International Nursing Review* 72, no. 2: e13012. <https://doi.org/10.1111/inr.13012>.
- Liu, J., F. Liu, J. Fang, and S. Liu. 2023. "The Application of Chat Generative Pre-Trained Transformer in Nursing Education." *Nursing Outlook* 71, no. 6: 102064. <https://doi.org/10.1016/j.outlook.2023.102064>.
- Luo, Y., Y. Miao, Y. Zhao, J. Li, and Y. Wu. 2025. "Exploring the Current Applications and Effectiveness of ChatGPT in Nursing: An Integrative Review." *Journal of Advanced Nursing* 81, no. 7: 3473–3484. <https://doi.org/10.1111/jan.16628>.
- Ma, J., J. Wen, Y. Qiu, et al. 2025. "The Role of Artificial Intelligence in Shaping Nursing Education: A Comprehensive Systematic Review." *Nurse Education in Practice* 84: 104345. <https://doi.org/10.1016/j.nepr.2025.104345>.
- Masanneck, L., L. Schmidt, A. Seifert, et al. 2024. "Triage Performance Across Large Language Models, ChatGPT, and Untrained Doctors in Emergency Medicine: Comparative Study." *Journal of Medical Internet Research* 26: e53297. <https://doi.org/10.2196/53297>.
- McGee, R. G., S. Wark, F. Mwangi, et al. 2024. "Digital Learning of Clinical Skills and Its Impact on Medical Students' Academic Performance: A Systematic Review." *BMC Medical Education* 24, no. 1: 1477. <https://doi.org/10.1186/s12909-024-06471-2>.
- Meskó, B., and E. J. Topol. 2023. "The Imperative for Regulatory Oversight of Large Language Models (Or Generative AI) in Healthcare." *NPJ Digital Medicine* 6, no. 1: 120. <https://doi.org/10.1038/s41746-023-00873-0>.
- Meyer, N. S., and J. W. Meyer. 2025. "A Practical Guide to the Utilization of ChatGPT in the Emergency Department: A Systematic Review of Current Applications, Future Directions, and Limitations." *Cureus* 17: e81802. <https://doi.org/10.7759/cureus.81802>.
- Moskovich, L., and V. Rozani. 2025. "Health Profession Students' Perceptions of ChatGPT in Healthcare and Education: Insights From a Mixed-Methods Study." *BMC Medical Education* 25, no. 1: 98. <https://doi.org/10.1186/s12909-025-06702-0>.
- Mu, Y., and D. He. 2024. "The Potential Applications and Challenges of ChatGPT in the Medical Field." *International Journal of General Medicine* 17: 817–826. <https://doi.org/10.2147/IJGM.S456659>.
- Naqvi, W. M., R. Ganjoo, M. Rowe, A. A. Pashine, and G. V. Mishra. 2025. "Critical Thinking in the Age of Generative AI: Implications for Health Sciences Education." *Frontiers in Artificial Intelligence* 8: 1571527. <https://doi.org/10.3389/frai.2025.1571527>.
- Nazer, L. H., R. Zatarah, S. Waldrip, et al. 2023. "Bias in Artificial Intelligence Algorithms and Recommendations for Mitigation." *PLOS Digital Health* 2, no. 6: e0000278. <https://doi.org/10.1371/journal.pdig.0000278>.
- Nguyen, T. 2024. "ChatGPT in Medical Education: A Precursor for Automation Bias?" *JMIR Medical Education* 10: e50174. <https://doi.org/10.2196/50174>.
- Norori, N., Q. Hu, F. M. Aellen, F. D. Faraci, and A. Tzovara. 2021. "Addressing Bias in Big Data and AI for Health Care: A Call for Open Science." *Patterns* 2, no. 10: 100347. <https://doi.org/10.1016/j.patter.2021.100347>.
- O'Connor, S., and ChatGPT. 2023. "Open Artificial Intelligence Platforms in Nursing Education: Tools for Academic Progress or Abuse?" *Nurse Education in Practice* 66: 103537. <https://doi.org/10.1016/j.nepr.2022.103537>.
- OECD. 2019. "Recommendation of the Council on OECD Legal Instruments Artificial Intelligence. OECD/LEGAL/0449".
- Page, M. J., J. E. McKenzie, P. M. Bossuyt, et al. 2021. "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews." *Bmj*, n71. <https://doi.org/10.1136/bmj.n71>.
- Park, S. H., J. Choi, and J. S. Byeon. 2021. "Key Principles of Clinical Validation, Device Approval, and Insurance Coverage Decisions of Artificial Intelligence." *Korean Journal of Radiology* 22, no. 3: 442–453. <https://doi.org/10.3348/kjr.2021.0048>.
- Paul, L. 2025. "Evaluation of ChatGPT Artificial Intelligence Dialogues and Simulated Scenarios to Prepare Nursing Students for Telehealth Practicum." *Medical Research Archives* 13, no. 7: 1–9. <https://doi.org/10.18103/mra.v13i7.6786>.
- Peng, W., X. Cheng, J. Deng, and X. Zhang. 2025. "ChatGPT Applications in Nursing: Current Status and Future Perspectives." *Nursing Open* 12, no. 6: e70253. <https://doi.org/10.1002/nop2.70253>.
- Peralta Ramirez, A. A., S. Trujillo López, G. A. Navarro Armendariz, S. A. De La Torre Othón, M. R. Sierra Cervantes, and J. A. Medina Aguirre. 2025. "Clinical Simulation With ChatGpt: A Revolution in Medical Education?" *Journal of CME* 14, no. 1: 2525615. <https://doi.org/10.1080/28338073.2025.2525615>.
- Perkins, S. W., J. C. Muste, T. Alam, and R. P. Singh. 2024. "Improving Clinical Documentation With Artificial Intelligence: A Systematic Review." *Perspectives in Health Information Management* 21, no. 2: 1d.
- Pham, T. 2025. "Ethical and Legal Considerations in Healthcare AI: Innovation and Policy for Safe and Fair Use." *Royal Society Open Science* 12, no. 5: 241873. <https://doi.org/10.1098/rsos.241873>.
- Reid, J. A. 2025. "Building Clinical Simulations With ChatGPT in Nursing Education." *Journal of Nursing Education* 64, no. 5: e6–e7. <https://doi.org/10.3928/01484834-20240424-05>.
- Ruksakulpiwat, S., S. Thorngthip, A. Niyomyart, et al. 2024. "A Systematic Review of the Application of Artificial Intelligence in Nursing Care: Where Are We, and What's Next?" *Journal of Multidisciplinary Healthcare* 17: 1603–1616. <https://doi.org/10.2147/JMDH.S459946>.
- Saad, O., M. Saban, E. Kerner, and C. Levin. 2025. "Augmenting Community Nursing Practice With Generative AI: A Formative Study of Diagnostic Synergies Using Simulation-Based Clinical Cases." *Journal of Primary Care & Community Health* 16: 21501319251326663. <https://doi.org/10.1177/21501319251326663>.
- Sallam, M. 2023. "ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns." *Health* 11, no. 6: 887. <https://doi.org/10.3390/health11060887>.
- Sarraf, B., and A. Ghasempour. 2025. "Impact of Artificial Intelligence on Electronic Health Record-Related Burnouts Among Healthcare Professionals: Systematic Review." *Frontiers in Public Health* 13: 1628831. <https://doi.org/10.3389/fpubh.2025.1628831>.
- Sblendorio, E., V. Dentamaro, A. Lo Cascio, F. Germini, M. Piredda, and G. Cicolini. 2024. "Integrating Human Expertise & Automated Methods for a Dynamic and Multi-Parametric Evaluation of Large Language Models' Feasibility in Clinical Decision-Making." *International Journal of Medical Informatics* 188: 105501. <https://doi.org/10.1016/j.ijmedinf.2024.105501>.

- Scherr, R., F. F. Halaseh, A. Spina, S. Andalib, and R. Rivera. 2023. "ChatGPT Interactive Medical Simulations for Early Clinical Education: Case Study." *JMIR Medical Education* 9: e49877. <https://doi.org/10.2196/49877>.
- Sengul, T., S. Bilgic, B. Macit, F. Sevim, S. Alik, and H. Kirkland-Kyhn. 2025. "Evaluation of Nursing Students' Ethical Decision-Making Biases and Attitudes Toward Artificial Intelligence in Nursing Education." *Nurse Education in Practice* 86: 104432. <https://doi.org/10.1016/j.nepr.2025.104432>.
- Shin, H., J. C. De Gagne, S. S. Kim, and M. Hong. 2024. "The Impact of Artificial Intelligence-Assisted Learning on Nursing Students' Ethical Decision-Making and Clinical Reasoning in Pediatric Care: A Quasi-Experimental Study." *CIN: Computers, Informatics, Nursing* 42, no. 10: 704–711. <https://doi.org/10.1097/CIN.0000000000001177>.
- Shojaei, P., M. Khosravi, Y. Jafari, A. H. Mahmoudi, and H. Hassanipourmahani. 2024. "ChatGPT Utilization Within the Building Blocks of the Healthcare Services: A Mixed-Methods Study." *Digital Health* 10: 20552076241297059. <https://doi.org/10.1177/20552076241297059>.
- Siontis, K. C., Z. I. Attia, S. J. Asirvatham, and P. A. Friedman. 2024. "ChatGPT Hallucinating: Can It Get Any More Humanlike?" *European Heart Journal* 45, no. 5: 321–323. <https://doi.org/10.1093/eurheartj/ehad766>.
- Srinivasan, M., A. Venugopal, L. Venkatesan, and R. Kumar. 2024. "Navigating the Pedagogical Landscape: Exploring the Implications of AI and Chatbots in Nursing Education." *JMIR Nursing* 7: e52105. <https://doi.org/10.2196/52105>.
- Sun, F., N. Li, K. Wang, and L. Goette. 2025. "Large Language Models Are Overconfident and Amplify Human Bias (Version 2). arXiv." <https://doi.org/10.48550/ARXIV.2505.02151>.
- Swisher, A. R., A. W. Wu, G. C. Liu, M. K. Lee, T. R. Carle, and D. M. Tang. 2024. "Enhancing Health Literacy: Evaluating the Readability of Patient Handouts Revised by Chatgpt's Large Language Model." *Otolaryngology and Head and Neck Surgery* 171, no. 6: 1751–1757. <https://doi.org/10.1002/ohn.927>.
- Szymczyk, A., R. Krion, K. Krzyzaniak, D. Lubian, and M. Sieminski. 2024. "Artificial Intelligence in Optimizing the Functioning of Emergency Departments; a Systematic Review of Current Solutions." *Archives of Academic Emergency Medicine* 12, no. 1: e22. <https://doi.org/10.22037/aaem.v12i1.2110>.
- Tam, W., T. Huynh, A. Tang, S. Luong, Y. Khatri, and W. Zhou. 2023. "Nursing Education in the Age of Artificial Intelligence Powered Chatbots (AI-Chatbots): Are We Ready Yet?" *Nurse Education Today* 129: 105917. <https://doi.org/10.1016/j.nedt.2023.105917>.
- Thirunavukarasu, A. J., D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting. 2023. "Large Language Models in Medicine." *Nature Medicine* 29, no. 8: 1930–1940. <https://doi.org/10.1038/s41591-023-02448-8>.
- Tian Tran, J., A. Burghall, T. Blydt-Hansen, et al. 2024. "Exploring the Ability of ChatGPT to Create Quality Patient Education Resources About Kidney Transplant." *Patient Education and Counseling* 129: 108400. <https://doi.org/10.1016/j.pec.2024.108400>.
- Tran, M., C. Balasooriya, C. Semmler, and J. Rhee. 2025. "Generative Artificial Intelligence: The 'More Knowledgeable Other' in a Social Constructivist Framework of Medical Education." *NPJ Digital Medicine* 8, no. 1: 430. <https://doi.org/10.1038/s41746-025-01823-8>.
- Ueda, D., T. Kakinuma, S. Fujita, et al. 2024. "Fairness of Artificial Intelligence in Healthcare: Review and Recommendations." *Japanese Journal of Radiology* 42, no. 1: 3–15. <https://doi.org/10.1007/s11604-023-01474-3>.
- Van Der Meijden, S. L., Y. Wang, M. S. Arbous, B. F. Geerts, E. W. Steyerberg, and T. Hernandez-Boussard. 2025. "Navigating Fairness in AI-Based Prediction Models: Theoretical Constructs and Practical Applications." *Preprint, MedRxiv* March 24. <https://doi.org/10.1101/2025.03.24.25324500>.
- Walters, W. H., and E. I. Wilder. 2023. "Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT." *Scientific Reports* 13, no. 1: 14045. <https://doi.org/10.1038/s41598-023-41032-5>.
- Wang, C., S. Liu, H. Yang, J. Guo, Y. Wu, and J. Liu. 2023. "Ethical Considerations of Using ChatGPT in Health Care." *Journal of Medical Internet Research* 25: e48009. <https://doi.org/10.2196/48009>.
- Wang, L., Z. Wan, C. Ni, et al. 2024. "Applications and Concerns of ChatGPT and Other Conversational Large Language Models in Health Care: Systematic Review." *Journal of Medical Internet Research* 26: e22769. <https://doi.org/10.2196/22769>.
- Wang, P., Q. Zhang, W. Zhang, and J. Sun. 2024. "The Application of ChatGPT in Nursing: A Bibliometric and Visualized Analysis." *Frontiers in Medicine* 11: 1521712. <https://doi.org/10.3389/fmed.2024.1521712>.
- Watson, A. L., C. Bond, H. Aveyard, G. D. Smith, and D. Jackson. 2025. "Generative AI at the Bedside: An Integrative Review of Applications and Implications in Clinical Nursing Practice." *Journal of Clinical Nursing* 1–16. <https://doi.org/10.1111/jocn.70151>.
- Welch, M. L., B. Grant, C. Deutschman, et al. 2026. "A Practical Framework for Operationalising Responsible and Equitable Artificial Intelligence in Health Care: Tackling Bias, Inequity, and Implementation Challenges." *Lancet Digital Health* 8: 100957. <https://doi.org/10.1016/j.landig.2025.100957>.
- WHO. 2021. *Ethics and Governance of Artificial Intelligence for Health*. WHO.
- Will, J., M. Gupta, J. Zaretsky, A. Dowlath, P. Testa, and J. Feldman. 2025. "Enhancing the Readability of Online Patient Education Materials Using Large Language Models: Cross-Sectional Study." *Journal of Medical Internet Research* 27: e69955. <https://doi.org/10.2196/69955>.
- Wójcik, D., O. Adamiak, G. Czerepak, O. Tokarczuk, and L. Szalewski. 2025. "Comparing the Performance of ChatGPT, Gemini, and Claude in English and Polish on Medical Examinations." *Scientific Reports* 15, no. 1: 33083. <https://doi.org/10.1038/s41598-025-17030-0>.
- Wu, Y., Y. Zheng, B. Feng, Y. Yang, K. Kang, and A. Zhao. 2024. "Embracing ChatGPT for Medical Education: Exploring Its Impact on Doctors and Medical Students." *JMIR Medical Education* 10: e52483. <https://doi.org/10.2196/52483>.
- Yeo, Y. H., J. S. Samaan, W. H. Ng, et al. 2023. "Assessing the Performance of ChatGPT in Answering Questions Regarding Cirrhosis and Hepatocellular Carcinoma." *Clinical and Molecular Hepatology* 29, no. 3: 721–732. <https://doi.org/10.3350/cmh.2023.0089>.
- Yu, S.-Y., H.-P. Lin, Y.-M. Kung, and S.-C. Chen. 2026. "Artificial Intelligence-Enhanced Clinical Reasoning in Nurse Practitioners: A Systematic Review." *Nurse Education in Practice* 91: 104724. <https://doi.org/10.1016/j.nepr.2026.104724>.
- Zaretsky, J., J. M. Kim, S. Baskharoun, et al. 2024. "Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format." *JAMA Network Open* 7, no. 3: e240357. <https://doi.org/10.1001/jamanetworkopen.2024.0357>.
- Zhang, S., X. Yang, L. Zhang, et al. 2025. "Exploring the Role and Potential of Chatbots in Learning From the Perspective of Nursing Students: A Systematic Review of Qualitative Studies." *International Nursing Review* 72, no. 3: e70060. <https://doi.org/10.1111/inr.70060>.
- Zhou, Y., S.-J. Li, X.-Y. Tang, et al. 2024. "Using ChatGPT in Nursing: Scoping Review of Current Opinions." *JMIR Medical Education* 10: e54297. <https://doi.org/10.2196/54297>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** PRISMA 2020 checklists.