



Artificial Intelligence–Enabled Educational Tools and Clinical Reasoning in Undergraduate Nursing Education: A Systematic Review

Houyam Jelloul^{1*}, Souad Arhoun², Malika Rmili³ and Youssef Nafidi⁴

¹Laboratory of Education and Social Dynamics, Faculty of Educational Sciences, Mohammed V University, Rabat, 10000, Morocco

²ERIPDS, Higher School of Teachers, Abdelmalek Essaâdi University, Tetouan, 93000, Morocco

Author Designation: ^{1,2,3}Doctoral Researcher, ⁴Professor

*Corresponding author: Houyam Jelloul (e-mail: houyamjelloul@gmail.com).

©2026 the Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)

Abstract Background: Clinical reasoning (CR) is a core competency in undergraduate nursing education, directly influencing patient safety and quality of care. Artificial intelligence (AI)–enabled educational tools are increasingly integrated into nursing curricula; however, their effectiveness in enhancing CR remains uncertain due to heterogeneous technologies, outcome measures and study designs. **Objective:** To systematically evaluate the effectiveness of AI-enabled educational tools in improving clinical reasoning among undergraduate nursing students compared with traditional teaching approaches or no AI intervention. **Methods:** This systematic review was conducted in accordance with PRISMA 2020 guidelines. Five databases were searched for studies published between January 2022 and January 15, 2026. Eligible studies involved undergraduate nursing students and reported explicitly operationalized outcomes related to clinical reasoning, clinical judgment or clinical decision-making. After removal of 332 duplicates, 1,015 records were screened; 33 full texts were assessed and 9 studies met inclusion criteria. Methodological quality was appraised using Joanna Briggs Institute (JBI) checklists. **Results:** Nine studies (n = 9) were included, comprising two randomized controlled trials, quasi-experimental designs and qualitative studies. Interventions included AI-enhanced simulations, rule-based educational chatbots, generative AI/LLM tools and AI-integrated tutoring systems. Improvements in self-reported clinical reasoning were observed in some studies, whereas others reported null or mixed findings. Performance-based or proxy measures yielded heterogeneous results and qualitative studies highlighted perceived benefits in information organization and confidence alongside concerns regarding dependency and reduced critical engagement. Risk-of-bias assessment revealed methodological limitations, particularly in non-randomized designs. **Conclusion:** Current evidence suggests preliminary and context-dependent educational potential for AI-enabled tools in supporting aspects of clinical reasoning. However, findings are heterogeneous, frequently based on self-reported or proxy measures and limited by methodological constraints. AI should therefore be considered a complementary pedagogical resource rather than a substitute for supervised clinical mentorship. Further rigorously designed studies using standardized performance-based CR measures are needed.

Key Words Artificial Intelligence, Clinical Reasoning, Nursing Education, Generative AI, Chatbots, Simulation-Based Learning, Undergraduate Nursing Students, Systematic Review

INTRODUCTION

Nursing education is delivered within increasingly complex healthcare systems where patient safety depends on clinicians' ability to synthesize information rapidly, prioritize ambiguous cues and adapt interventions to evolving clinical situations. Within this context, clinical reasoning (CR) constitutes a foundational professional competency. It underpins assessment, decision-making, care planning and modification

of interventions according to patient responses [1,2]. Nevertheless, CR remains challenging to cultivate in undergraduate nursing students, as expertise develops progressively through repeated exposure to diverse clinical scenarios and guided reflection [3].

Clinical reasoning refers to the analysis and interpretation of clinical data to generate hypotheses, establish priorities and select appropriate care strategies [4,5].

In nursing practice, CR integrates holistic assessment, anticipation of risk, contextual interpretation and tailored intervention planning [1,2]. Conceptual precision is necessary because CR is often used interchangeably with clinical judgment, critical thinking and clinical decision-making, complicating synthesis and comparison across studies [6]. These constructs are related but distinct. Clinical judgment, defined by Tanner as “an interpretation or conclusion...” guiding action, may be viewed as an outcome of the reasoning process (Tanner, 2006, p. 204) [6]. Critical thinking contributes to reasoning quality; however, reasoning errors may reflect deficits in critical thinking skills [7,8]. To reduce ambiguity, this review considers critical thinking measures only when assessed within explicit clinical tasks or directly linked to applied clinical performance [8].

Parallel to these educational challenges, artificial intelligence (AI) has become increasingly embedded in higher education and healthcare environments. AI systems, defined as software capable of supporting logical and informed judgments [9], have evolved from early computer-assisted learning and simulation platforms to adaptive learning environments, virtual tutors and generative large language models (LLMs) [10,11]. In nursing education, AI-enabled tools are increasingly employed to facilitate simulation, knowledge retrieval, feedback provision and self-directed learning, with some reports suggesting enhanced engagement and perceived performance [1,12–15].

However, “AI tools” represent a heterogeneous group of technologies that differ substantially in architecture and pedagogical function. These include: (1) AI-enhanced simulations and virtual patient environments; (2) rule-based or natural language processing (NLP) educational chatbots; (3) generative AI/LLM assistants such as ChatGPT used as learning resources; and (4) intelligent tutoring systems embedded within simulation scenarios. Aggregating these technologies without differentiation risks obscuring meaningful differences in educational mechanisms and outcomes [9].

Previous syntheses examining digital simulation and virtual patient approaches have suggested potential benefits for applied skill development and engagement [16–19]. Chatbot-facilitated learning may support problem-solving processes and immersive environments may enhance clinical performance under certain conditions [16–19]. However, CR is frequently treated as a secondary or indirectly measured outcome and findings remain inconsistent. Furthermore, earlier reviews often examine digital or virtual simulation broadly, without isolating AI-specific mechanisms or distinguishing between performance-based outcomes and self-reported perceptions.

Beyond effectiveness, AI integration raises important ethical and pedagogical concerns. Generative AI systems may produce inaccurate or fabricated outputs (“hallucinations”), embed algorithmic bias or lack transparency in decision pathways [9]. In educational settings, issues of data privacy, academic integrity and potential cognitive dependency, where learners rely

excessively on automated reasoning support, warrant careful scrutiny. These challenges are central to evaluating AI’s role in clinical education rather than peripheral considerations.

Taken together, the literature indicates educational promise but substantial uncertainty regarding the differential impact of distinct AI tool categories, the validity of CR outcome measures (objective performance versus self-reported or proxy indicators) and the contextual conditions required for effective and responsible implementation. Accordingly, this systematic review examines whether AI-enabled educational tools contribute to the development of clinical reasoning among undergraduate nursing students compared with traditional teaching approaches or no AI exposure. Specifically, this review aims to: (1) categorize AI tools used in undergraduate nursing education; (2) synthesize their effects on explicitly operationalized CR outcomes; (3) distinguish primary CR outcomes from secondary measures such as confidence and satisfaction; and (4) identify reported implementation conditions, methodological limitations and ethical considerations associated with AI integration.

By clarifying both the potential benefits and the limitations of AI-enabled interventions, this review seeks to inform evidence-based and ethically responsible decisions regarding the integration of artificial intelligence into nursing curricula.

METHODS

Design

This systematic review was conducted in accordance with the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20] to ensure methodological transparency, reproducibility, and comprehensive reporting. The PRISMA flow diagram is presented in Figure 1. The review protocol was not prospectively registered.

The review question was structured using the PICO framework to guide eligibility criteria, data extraction and synthesis.

Population (P)

Undergraduate nursing students enrolled in bachelor’s degree programs or equivalent pre-licensure nursing education.

Intervention (I)

Educational interventions explicitly integrating an artificial intelligence tool within a structured learning activity. Eligible AI categories included:

- Generative AI and large language models (LLMs) (e.g., ChatGPT)
- Rule-based or NLP-enabled educational chatbots;
- AI-enhanced virtual patient or simulation environments (including VR-based systems)
- Intelligent tutoring systems embedded within simulation or case-based scenarios
- Expert or decision-support systems used in a training context

Comparator (C)

Traditional teaching approaches (e.g., lectures, standard simulation without AI, textbooks, videos), alternative digital modalities without AI or no comparator. Within-group pre–post comparisons were considered when no parallel control group was available.

Outcomes (O)

Explicitly operationalized measures of clinical reasoning, clinical judgment or clinical decision-making. Outcomes were eligible when they were measured using validated instruments, structured performance tasks or clearly defined assessment grids.

The primary research question was:

How effective are AI-enabled educational tools in improving clinical reasoning skills (and explicitly operationalized components) among undergraduate nursing students compared with standard teaching approaches or no AI intervention?

To reduce conceptual ambiguity, outcomes such as self-efficacy, confidence, satisfaction, perceived competence or engagement were classified as secondary outcomes and analyzed separately from clinical reasoning. Measures of critical thinking were included only when assessed within clearly described clinical tasks (e.g., simulation scenarios or case analyses) and explicitly linked to clinical reasoning or decision-making processes.

Given anticipated heterogeneity in intervention types, comparators and outcome measures, a narrative synthesis approach was planned a priori.

Search Strategy

A systematic literature search was conducted to identify studies published between January 2022 and January 15, 2026. The timeframe was deliberately restricted to reflect the rapid evolution and widespread adoption of generative AI and LLM technologies in educational contexts, rendering earlier studies less comparable in terms of technological capacity and pedagogical mechanisms.

The following electronic databases were searched: PubMed, Scopus, Web of Science Core Collection, ScienceDirect and ERIC.

Search strategies combined three concept blocks:

- **Artificial intelligence:** artificial intelligence, machine learning, generative AI, large language model, LLM, ChatGPT, chatbot;
- **Nursing education:** nursing education, nursing student;
- **Clinical reasoning and related constructs:** clinical reasoning, clinical judgment, clinical decision-making

Controlled vocabularies (e.g., MeSH in PubMed) were used when available and combined with free-text terms to maximize sensitivity. Search syntax was adapted to each database while maintaining conceptual equivalence across

platforms. Full search strategies for all databases are provided in Supplementary Table S1.

Reference lists of included studies were screened manually to identify additional eligible publications.

Eligibility Criteria

Studies were included if they met all of the following criteria:

- Empirical research (quantitative, qualitative or mixed-methods) published in full text;
- Conducted within a nursing education context;
- Involved undergraduate nursing students (bachelor’s degree or equivalent);
- Evaluated an educational intervention incorporating an AI tool as defined above;
- Reported at least one explicitly operationalized outcome related to clinical reasoning, clinical judgment or clinical decision-making;
- Published in English between January 2022 and January 2026

Studies reporting only critical thinking outcomes were included only if critical thinking was assessed within a clearly described clinical task and explicitly linked to reasoning or decision-making processes.

Studies were excluded if they:

- Were non-empirical publications (reviews, editorials, commentaries, protocols, conference abstracts, letters);
- Were conducted outside an educational setting;
- Did not include an AI-enabled intervention;
- Included only practicing professionals without a separately analyzable undergraduate subgroup;
- Failed to report outcomes relevant to clinical reasoning

Selection Process and Data Extraction

All references were imported into Zotero for management. Duplicates were identified and removed prior to screening.

Two reviewers independently screened titles and abstracts against eligibility criteria. Potentially relevant articles underwent full-text assessment. Discrepancies were resolved through discussion and consensus.

Data extraction was performed independently by both reviewers using a predefined Microsoft Excel template. Extracted data included:

- Author(s), year, country and educational setting;
- Study design and sample characteristics;
- AI tool category and technological description;
- Intervention characteristics (duration, frequency, supervision, debriefing, learning context);
- Comparator characteristics;
- Outcome measures (instrument type and validation status);
- Reported results;
- Implementation conditions and contextual factors

Clinical reasoning–related outcomes were coded according to:

- **Measurement type:** performance-based (e.g., structured clinical task, objective scoring grid) versus self-reported scale;
- **Outcome status:** primary (CR/judgment/decision-making) versus secondary (confidence, satisfaction, engagement, AI readiness, self-directed learning);
- **Direction of effect:** positive, null, mixed or negative.

When study data were incomplete or ambiguously reported, the limitation was documented and considered during synthesis.

Qualitative studies were included to capture implementation conditions, learner perceptions and ethical or pedagogical concerns; however, they were not treated as evidence of effectiveness.

Assessment of Methodological Quality and Risk of Bias

Methodological quality and risk of bias were assessed in accordance with PRISMA 2020 recommendations (Item 11) [20].

Design-specific Joanna Briggs Institute (JBI) critical appraisal checklists were used:

- JBI RCT checklist for randomized controlled trials;
- JBI quasi-experimental checklist for non-randomized studies;
- JBI qualitative checklist for qualitative designs

Each item was rated as “Yes,” “No,” “Unclear,” or “Not applicable.” Two reviewers conducted assessments independently, with disagreements resolved through discussion.

Risk-of-bias findings were not used to exclude studies but were integrated into interpretation. In particular, limitations related to allocation concealment, lack of blinding, confounding management or insufficient reflexivity in qualitative research were considered when synthesizing findings and formulating conclusions.

Data Synthesis

Due to heterogeneity in AI tool categories, study designs, comparator types and outcome measures, statistical meta-analysis was not appropriate. Therefore, a structured narrative synthesis was conducted.

Studies were grouped according to:

- AI tool category (AI-enhanced simulation, rule-based chatbot, generative AI/LLM, intelligent tutoring system);
- Type of clinical reasoning outcome (performance-based, self-reported, proxy/composite);
- Direction of effect

This approach allowed systematic comparison across intervention types while preserving methodological transparency. Both positive and null findings were reported to avoid selective emphasis.

Publication bias could not be formally assessed due to the limited number and heterogeneity of studies; this limitation is acknowledged.

RESULTS

The database search identified 1,347 records (Scopus = 190; Web of Science Core Collection = 224; PubMed = 203; ScienceDirect = 716; ERIC = 14). After importing the references into Zotero and removing 332 duplicates, 1,015 unique records remained for title and abstract screening, resulting in the exclusion of 982 records. Of the 33 reports selected for full-text retrieval, 16 could not be obtained because they were not available in full text through the institutional resources accessible at the time of the search. The full texts of 17 articles were subsequently assessed for eligibility; 8 were excluded for documented reasons, including the absence of an eligible clinical reasoning-related outcome ($n = 4$), the absence of a relevant AI intervention ($n = 1$) and an ineligible population (i.e., not undergraduate students [bachelor’s degree or equivalent]) ($n = 3$). These exclusions involved studies conducted with graduate/advanced practice students (FNP programs) and postgraduate students. Finally, 9 studies were included in the synthesis. The study selection process is presented in the PRISMA flow diagram (Figure 1).

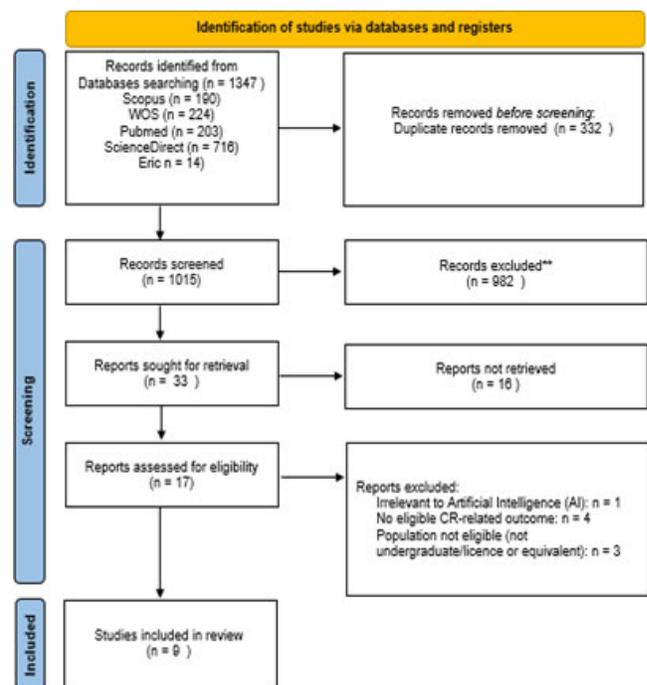


Figure 1: The PRISMA flow diagram illustrates the steps in the systematic review and shows the article selection process

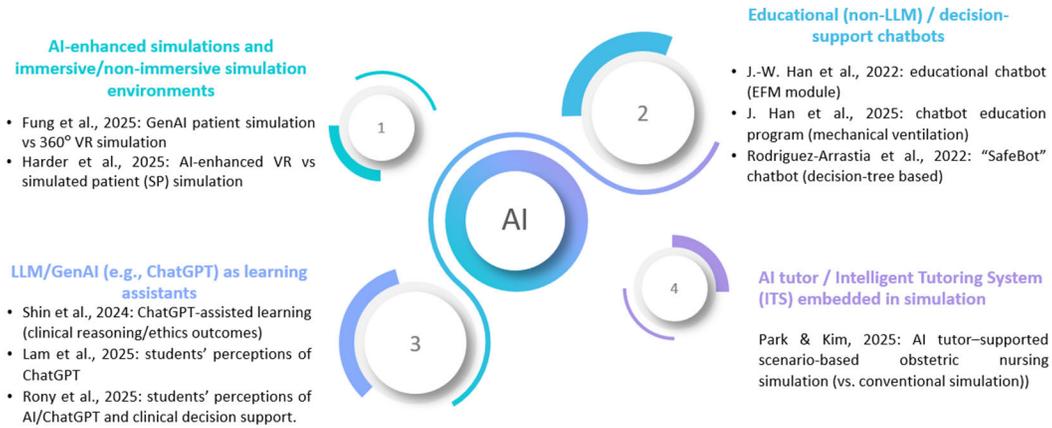


Figure 2: Tool families and dominant pedagogical mechanisms across included studies (n = 9).

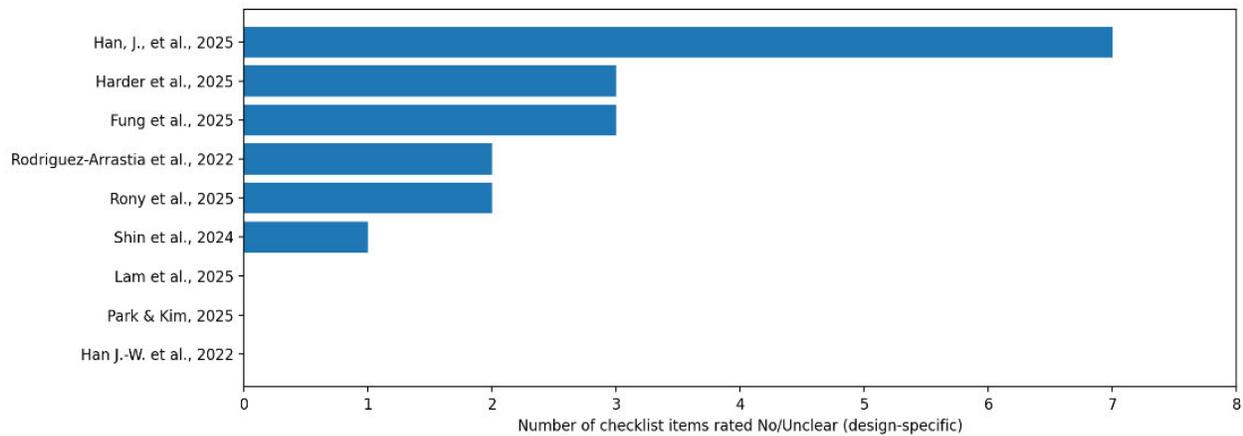


Figure 3: Summary of checklist items rated No/Unclear across studies

Characteristics of the Included Studies

A total of nine studies published between 2022 and 2025 were included. These studies were conducted across several geographical contexts, mainly Asia (South Korea, Hong Kong), Europe (Spain) and North America (Canada), with one study also conducted in Bangladesh, reflecting international interest in integrating AI into nursing education. Methodologically, the corpus comprised two randomized controlled trials (including one crossover trial), quasi- experimental studies and qualitative studies (primarily based on focus groups and/or descriptive/interpretive qualitative analyses). Given the diversity of designs and outcome measures, a narrative synthesis was deemed most appropriate (Table 1).

The evaluated interventions fell into four tool categories: (i) AI-enhanced simulation environments (including GenAI patient simulations and VR/AI-VR formats), (ii) “classic” educational chatbots (non-LLM), (iii) GenAI/LLM assistants (e.g., ChatGPT) used as a learning resource and (iv) an AI tutor integrated into a scripted simulation. Across all studies, participants were exclusively undergraduate nursing students (bachelor’s degree or equivalent), recruited within course and/or simulation activities.

Clinical reasoning-related outcomes were heterogeneous. Some studies used self-reported clinical reasoning/competence scales, whereas others relied on proxies (e.g., subcomponents related to Evidence-Based Practice (EBP) integration into clinical reasoning, critical thinking disposition, perceived performance) or qualitative data addressing perceived mechanisms (realism, psychological safety, communication, confidence/self-efficacy). This heterogeneity in tools, operational definitions and measurement modalities limited the feasibility of quantitatively pooling results.

Furthermore, the included studies employed a range of AI-enabled educational tools (e.g., AI- enhanced simulation environments, non-LLM chatbots and GenAI/LLM assistants). Figure 2 summarizes these interventions by categorizing them according to the dominant techno-pedagogical mechanism reported by the study authors.

Assessment of Methodological Quality and Risk of Bias

Critical appraisal of the nine included studies (using JBI checklists aligned with each study design) indicated variable risk of bias, mainly related to reporting completeness and the management of selection and confounding biases. Among the

Table 1: Summary of included studies (n=9)

Study	Country/context	Estimate of crossover (crossover RCT)	AI tool category (GenAI, patient simulation (scenarios))	Comparator (Immersive 360° simulation (cross-over, 1-week washout))	Sample (n=44 (1st-3rd years; international cohorts))	Primary outcome (CR) - instrument (QCC/CCO) - proxy for RC	Type of measurement RC (Self-reported (perceived))	Direction of effect on primary outcome (Positive (larger, T1 gains when GenAI delivered first; improvements maintained at T2 after crossover))	Secondary outcomes (instrument) - direction (CAS: Improvement in both sequences; no statistically significant difference between-group differences. MAIRS-MS: Increased after exposure to GenAI; larger gain when GenAI was delivered first (Group B compared with Group A). SEPS: favorable perceptions; comparisons were not reported (descriptive results only).)	Notes/interpretation risks (Primary outcome = perceived clinical competence (self-reported); interpret as proxy for CR, not as objective performance.)
[21]	Hong Kong (University of Hong Kong)	Estimate of crossover (crossover RCT)	GenAI, patient simulation (scenarios)	Immersive 360° simulation (cross-over, 1-week washout)	n=44 (1st-3rd years; international cohorts)	Perceived clinical competence (QCC/CCO) - proxy for RC	Self-reported (perceived)	Positive (larger, T1 gains when GenAI delivered first; improvements maintained at T2 after crossover)	CAS: Improvement in both sequences; no statistically significant difference between-group differences. MAIRS-MS: Increased after exposure to GenAI; larger gain when GenAI was delivered first (Group B compared with Group A). SEPS: favorable perceptions; comparisons were not reported (descriptive results only).	Primary outcome = perceived clinical competence (self-reported); interpret as proxy for CR, not as objective performance.
[22]	Canada (undergraduate nursing)	Qualitative thematic analysis after exposure to both modalities	AI-enhanced virtual simulation (AI-VS/AI-VR)	AI-VR simulation vs. standardized patients (SP)	Exposures n=240 (120/arm); qualitative n=20 (4 focus groups)	No RC instruments; qualitative data (4 focus groups) exploring perceived mechanisms related to clinical competence (realism, psychological safety, reinforcement of skills/practice).	Participants report standardized patients promote interactions that are considered more realistic and greater emotional engagement. AI-VR simulations are perceived as a non-judgmental space, facilitating trial and error and iterative practice; they support communication and build confidence in decision-making.	Not applicable (qualitative)	Secondary themes reported: realism; psychological safety; communication; confidence (perceived mechanisms that can support clinical judgment/decision-making).	n exposed = 240 (120 AI-VR; 120 SP); n qualitative = 20 (4 focus groups, 4-6 students/group).
[23]	Korea (Gumi University)	AI tutor simulation (labor care scenarios)	AI tutor simulation (labor care scenarios)	Conventional high-fidelity simulation	n=72 (38 exp; 34 control), 4th year	Clinical performance (Lee/Choi, 45 items; 45-225), proxy RC	Self-reported (Likert 1-5)	Experimental group showed higher scores than the control group (t = 7.80, p = 0.020).	Obstetric knowledge: significantly higher in the experimental group than in the control group (p < 0.001). Critical thinking disposition (Yoon): no significant difference (p = 0.098). Digital literacy: significantly higher in the experimental group than in the control group (p < 0.001).	Possible selection/confounding bias (quasi-exp) despite efforts to ensure equivalence. CR measured indirectly (performance/CT).
[24]	Korea (university)	Quasi-experimental, non-randomized control group (non-synchronized pre/post) RCT	Educational AI chatbot (online EFM module)	Traditional online course without chatbot	n=61 (30 exp; 31 control), 3rd year	Clinical Reasoning Competency Scale (CRCS, Korean version, 15 items)	Self-reported (RC scale)	None (t=0.75; p=0.455)	Knowledge (RCNRS): NS (t=0.75; p=0.455). Confidence (NS): NS. Feedback learning satisfaction: NS. Self-directed learning (SDLBS): improvement (t=2.72; p=0.006).	Report zero effect on RC and positive effect on self-directed learning.
[25]	Spain (University of Almeria)	Qualitative descriptive	Decision-making chatbot (tree 'SafeBot')	NA (qualitative)	n=114 (final year; bachelor's degree)	No CR instruments; qualitative data (focus groups) on the acceptability/easiness of decision-making chatbot and on clinical safety (perceived).	Students describe SafeBot as useful in complex situations: access to evidence-based information, clarification of doubts "at any time," and perceived support for clinical decision-making and problem-solving with a heightened sense of confidence.	Not applicable (qualitative)	Reported acceptability and ease of use; perceived safety of evidence-based informational support; self-confidence (perceived).	Do not convert perceptions into evidence of effectiveness; use for mechanisms/implement ation.
[26]	Korea (pediatric care course)	Quasi-experimental with control group (post-test only) + reflective trials	ChatGPT-assisted learning	Traditional textbook	n=99 (52 exp; 47 control), 3rd year	Subscores on 2 objectives (ethical standards; care process) + written reflections [tables provided]	Proxy/compo site (post-test): sub-score for "integration of evidence-based knowledge and clinical reasoning" in a case process assessment grid (with other sub-scores: reflection/tin movement).	Mixed results: the control condition outperformed the AI condition on several sub-conditions (p < 0.001) while other sub-conditions showed no significant differences (tables provided)	Objective 1 (ethical standards) Control condition scored higher than the ChatGPT condition for understanding ethical concepts, analyzing challenges/obstacles and applying principles (p < 0.001). No significant difference for knowledge communication (p = 0.260) and integration of professional standards (p = 0.812). Objective 2 (care process) Control condition scored higher than the ChatGPT condition for critical thinking skills, integration of evidence-based practice with clinical reasoning, and reflection/improvement (p < 0.001). No significant difference for process application (p = 0.455).	Study focused on the use of ChatGPT as a resource; interpret with caution (proxies/rank scores). Qualitative: AI perceived as unreliable by the majority.
[27]	Canada (British Columbia)	Qualitative (interpretative description)	ChatGPT (GenAI/LLM)	NA (qualitative)	n=16 (2nd semester; bachelor's degree)	Topics (EBP, ethics, critical thinking)	Qualitative (risks/conditions)	NA (qualitative)	Also report risks (hindrance to critical thinking) as secondary outcomes.	Useful for discussing conditions for responsible use and alignment with critical thinking.
[28]	Bangladesh (5 middle schools)	Descriptive qualitative	ChatGPT diagnostic support systems	NA (qualitative)	n=25	Themes (AI as a "second brain," balance between AI and human expertise)	Qualitative (perceptions)	NA (qualitative)	NA (qualitative)	

randomized trials, Fung *et al.* (2025) received predominantly “Yes” ratings, with a few “Unclear” items, whereas Han *et al.* showed a higher proportion of “Unclear/No” ratings, particularly regarding allocation concealment, blinding, the handling of post-allocation exclusions and/or ITT-related issues and the documentation of certain measures. In the quasi-experimental studies, recurrent limitations included the absence of randomization, limited baseline comparability, and/or insufficient consideration of confounding factors. For the qualitative studies, methodological congruence was generally reported, but reflexivity and the researcher-participant relationship were more frequently incompletely described. Item-level ratings are presented in Table 2 and justifications for “No/Unclear” ratings are provided in the supplementary material.

The graph in Figure 3 displays, for each study, the number of JBI checklist items rated “No” or “Unclear” (based on the design-appropriate checklist), providing a comparative overview of the main areas of methodological uncertainty.

Effects of AI-Based Interventions on Clinical Reasoning (CR)
Effects on CR Measured by Self-Reported Scales: Two studies assessed clinical reasoning using self-reported scales; accordingly, these findings should be interpreted as perceived

indicators rather than objective clinical performance. In the randomized controlled crossover trial by [21], the primary outcome was perceived clinical competence (PCC), used as a proxy for CR. The authors reported improved scores over time, with an advantage for the sequence in which GenAI was delivered first at T1 and sustained improvements after participants crossed over to the alternate modality. Conversely, in the quasi-experimental study by [12] (EFM course), no statistically significant between-group difference was observed on the clinical reasoning competency scale ($t = 0.75$; $p = 0.455$), although an improvement was reported for self-directed learning (SDLRS) ($t = 2.72$; $p = 0.006$) (Table 1).

Effects on CR-Related Indicators and Secondary Quantitative Outcomes

Several quantitative studies reported effects on dimensions related to clinical reasoning, such as confidence, satisfaction or performance measures treated as proxies (Table 1). In [24], the group receiving the chatbot intervention reported significantly higher self-reported CR scores than the control group ($t = -5.00$; $p < 0.001$). Favorable between-group differences were also reported for confidence ($t = -2.62$; $p = 0.011$) and satisfaction ($t = -3.51$; $p < 0.001$). However, no difference was observed for knowledge ($t = -0.09$; $p = 0.926$).

Table 2: Critical appraisal (JBI) and descriptive summary of risk of bias (9 studies)

Study design JBI tool yes no unclear NA points of attention (from no/unclear items)							
>>Fung <i>et al.</i> , 2025	Randomized>controlled crossover trial (crossover RCT)	>>JBI RCT (13 items)	>>9	>>1	>>2	>>1	>Blinding of participants/clinicians not reported (conditions difficult to blind); outcome assessors not blinded (risk of detection bias).
>>Park & Kim, 2025	Quasi->experimental (pre/post with comparison group)	>JBI Quasi- exp. >(9 items)	>>9	>>0	>>0	>>0	>>No major limitations reported according to the grid
>>J.-w. Han <i>et al.</i> , 2022	Quasi->experimental (non-randomized; pre/post with comparator)	>>JBI Quasi- exp. >(9 items)	>>>9	>>>0	>>>0	>>>0	>>No major limitations reported according to the grid.
>>>J. Han <i>et al.</i> , 2025	>Randomized controlled trial (RCT),>pretest–posttest, parallel groups	>>>>JBI RCT (13 items)	>>>>6	>>>>1	>>>>6	>>>>0	"Concealment of allocation not>described; blinding of participants/clinicians not reported; intention-to-treat (ITT) analysis>not explained (post-allocation exclusions); limited psychometric justification for certain outcomes>(NRS + "knowledge" tool>developed by the authors)."
>>>Shin <i>et al.</i> ,>2024	Quasi->experimental (group comparison) + open-ended questions (descriptive)	>>JBI Quasi- exp.>(9 items)	>>>8	>>>0	>>>1	>>>0	>>Incomplete information on initial comparability/management of confounding factors → "Unclear."
>>Harder <i>et al.</i> , 2025	Qualitative>(focus groups)>– comparison of simulation modalities	>JBI Qualitative (10 items)	>>7	>>0	>>3	>>0	>Reflexivity/positioning of the researcher and researcher- participant relationship insufficiently detailed.
>>Rodríguez-arrastia <i>et al.</i> , 2022	Qualitative>(interviews) – perceptions of a decision-making chatbot	>>JBI Qualitative (10 items)	>>>8	>>>0	>>>2	>>>0	>>Reflexivity/positioning of the researcher insufficiently detailed.
>Rony <i>et al.</i> , 2025	Qualitative>descriptive – perceptions AI/ChatGPT	>JBI Qualitative (10 items)	>>8	>>0	>>2	>>0	>Reflexivity/researcher positioning insufficiently detailed.
>Lam <i>et al.</i> , >2025	>Qualitative	JBI>Qualitative >(10 items)	>8	>0	>2	>0	Reflexivity/researcher positioning>and researcher–participant influence: Unclear

In [23], the primary outcome was questionnaire-based clinical performance (proxy for CR), which was higher in the “AI tutor” group ($p = 0.020$); knowledge also improved ($p < 0.001$). By contrast, critical thinking skills (Yoon) did not differ significantly between groups ($p = 0.098$), suggesting that observed effects varied depending on the domain assessed and the type of outcome used.

Finally, in [25], CR was captured indirectly through sub-scores derived from a care process assessment grid, including a dimension titled “integration of evidence-based knowledge and clinical reasoning.” Findings were heterogeneous: several sub-scores favored the control group ($p < 0.001$), whereas other dimensions were not statistically significant, indicating a variable effect profile across the assessed components.

Qualitative Data

Perceived Mechanisms, Organization of Information and Conditions of Use: Qualitative studies primarily informed perceived mechanisms, acceptability and implementation conditions rather than objective CR effectiveness (Table 1). In Rodriguez-Arrastia *et al.* [2], students described the SafeBot decision-making chatbot as helpful for clarifying doubts, organizing information and supporting perceived decision-making, particularly in situations considered complex, alongside a reported increase in confidence.

In Harder *et al.* [22], participants differentiated perceived contributions by modality: standardized patients were more strongly associated with realism and emotional engagement, whereas AI-VR simulation was more often associated with psychological safety that facilitated trial-and-error learning, iterative practice and perceived gains in communication and confidence in decision-making.

The two qualitative studies focusing on ChatGPT/GenAI [14,26] highlighted use oriented toward generating and organizing information, while also reporting pedagogical and professional tensions. Students expressed concerns about potential dependence, the risk of undermining personal judgment, and, specifically in [14], fear of adverse effects on the caregiver–patient relationship, particularly regarding interaction quality and the empathic dimension of clinical judgment.

Cross-Sectional Synthesis of Observed Patterns

Across the nine studies, findings suggest: (i) favorable effects on self-reported CR under certain conditions [21,24], (ii) favorable effects on related dimensions, particularly confidence and satisfaction [24] and (iii) qualitative contributions described as supporting information organization and perceived decision-making [2,22]. However, null or mixed findings were also reported, notably in [12] (no significant between-group differences on the CR scale), the heterogeneous sub-score pattern in Shin *et al.* (2024) and the perceived risks reported in qualitative studies (dependence, the role of human judgment and the care relationship) [14,26].

DISCUSSION

Key Findings

This systematic review synthesized evidence from nine studies ($n = 9$) examining AI-enabled educational tools in undergraduate nursing education and their relationship to clinical reasoning (CR). Interventions included AI-enhanced simulations, rule-based chatbots, generative AI/LLM systems and AI-integrated tutoring tools. Across these heterogeneous designs and outcome measures, findings indicate preliminary but inconsistent support for AI-assisted learning in relation to CR.

Importantly, improvements were more frequently observed in self-reported measures of clinical reasoning than in performance-based or objectively assessed outcomes. This distinction is critical. While some studies reported statistically significant gains in perceived competence or reasoning ability [21,24], other investigations demonstrated null or mixed effects [12,25], particularly when CR was assessed through proxy indicators or structured evaluation grids. Consequently, conclusions regarding effectiveness must remain cautious and sensitive to the type of measurement employed.

The variability in findings appears to depend on several interacting factors: (i) the category of AI tool used, (ii) the pedagogical format (guided scenario-based integration versus independent use as a study aid), (iii) the measurement approach (performance-based versus self-reported) and (iv) methodological rigor and risk-of-bias considerations (Table 2). These dimensions collectively shape interpretation and preclude broad generalizations.

Structuring Clinical Reasoning: Guided and Scenario-Based Applications

A consistent pattern across studies suggests that AI tools may assist learners in organizing clinical information and articulating reasoning steps when embedded within structured scenarios. In the crossover randomized trial by [21], perceived clinical competence (used as a proxy for CR) improved following exposure to generative AI simulation, particularly when delivered as the initial modality. While this finding suggests a potential structuring effect, it is based on self-reported outcomes rather than objective performance.

Similarly, [24] reported improvements in self-reported clinical reasoning following chatbot-assisted instruction. In contrast, [12] found no significant improvement in CR as measured by scale, highlighting that AI integration does not automatically translate into measurable reasoning gains. These differences underscore the importance of instructional design and supervision. AI appears more likely to support reasoning processes when embedded within guided, interactive, scenario-based pedagogies rather than when used independently as a supplemental resource.

These observations align with broader theoretical arguments suggesting that AI systems may scaffold

knowledge organization and simplify complex decision pathways [3,27]. Evidence from educational chatbot research also suggests potential stimulation of analytical thinking and structured problem-solving [16,28]. However, these interpretations must be contextualized within the methodological limitations of the included studies.

Clinical Judgment and Decision-Making

Mixed Quantitative Evidence and Perceived Benefits: Qualitative evidence suggests that students often perceive AI tools as helpful in clarifying information and structuring decision-making processes [2,14,26]. For example, decision-making chatbots and LLM systems were described as facilitating access to evidence-based information and supporting perceived judgment formation. In [22], AI/VR simulations were experienced as psychologically safe environments that encouraged iterative learning, whereas standardized patients were associated with higher realism and emotional engagement.

However, quantitative findings were less consistent. In [23], improvements were reported on a performance-related competence scale, although the measure functioned as a proxy rather than a direct assessment of CR. In contrast, [25] found mixed results, with several sub-scores favoring traditional learning over ChatGPT-assisted approaches. Notably, integration of evidence-based practice within reasoning was stronger in the control group in that study, suggesting that generative AI may not replicate the depth of analytical processing fostered by traditional instructional methods.

These discrepancies emphasize that AI tools may facilitate certain cognitive processes, such as information retrieval or initial hypothesis generation, without necessarily strengthening higher-order clinical reasoning or judgment performance. Measurement heterogeneity further complicates interpretation.

Confidence and Engagement

Distinguishing Perception from Competence: Several studies reported improvements in secondary outcomes such as confidence, satisfaction and self-directed learning [24]. While these outcomes are pedagogically relevant, they should not be equated with demonstrated improvement in clinical reasoning. For instance, [24] observed increased confidence without corresponding knowledge gains, suggesting that AI tools may enhance perceived fluency or comfort rather than cognitive mastery.

Similarly, qualitative findings described AI-enabled environments as supportive of experimentation and iterative practice [2,22]. Conversely, [12] reported no significant improvement in CR or satisfaction with feedback, reinforcing that AI integration alone does not guarantee enhanced learning experiences.

Distinguishing between perceived competence and objectively assessed reasoning performance is essential to prevent overinterpretation of findings.

Limits of AI in Clinical Education

Human Dimensions and Contextual Complexity: Despite potential benefits, important limitations must be acknowledged. Certain aspects of clinical reasoning, particularly those involving tactile assessment, subtle sensory cues and relational judgment, are difficult to replicate through AI systems [33]. Virtual patients and AI-driven simulations may lack the emotional nuance and interpersonal complexity inherent in authentic clinical encounters [34,35].

Within this review, [22] highlighted that standardized patients were associated with greater perceived realism and emotional engagement compared with AI-based modalities. This suggests that traditional simulation approaches may retain advantages in developing affective and relational components of clinical judgment. AI tools should therefore be viewed as complementary rather than substitutive in relation to human clinical mentorship.

Financial and technological barriers also warrant consideration. Implementation of AI-enhanced simulation platforms or institutional LLM access may require substantial infrastructure investment, ongoing maintenance and faculty training. These structural factors were underreported in primary studies but are critical for sustainable adoption.

Ethical and Cognitive Risks

Dependency, Bias and Transparency: Beyond pedagogical limitations, ethical and cognitive concerns emerged across qualitative studies. Students reported apprehension regarding excessive reliance on AI tools, potential weakening of independent judgment and concerns about diminished caregiver–patient relational quality [14,26]. These perceptions align with broader scholarly concerns regarding analytical passivity and cognitive outsourcing [30,31,36,37].

Generative AI systems may produce inaccurate outputs, embed algorithmic bias or obscure reasoning pathways due to limited transparency [9,25,35]. Without explicit verification processes and guided supervision, such risks may undermine the very reasoning skills educational programs aim to cultivate.

Accordingly, AI integration should be accompanied by structured safeguards, including verification exercises, reflective debriefing and explicit discussion of algorithmic limitations. Ethical considerations, including data privacy, academic integrity and responsible use, must be embedded within curricular frameworks rather than treated as peripheral issues.

Pedagogical Implications and Research Directions

Taken together, the findings support a hybrid instructional model in which AI tools are integrated under faculty supervision and paired with explicit reasoning justification activities [25,35]. Three practical implications emerge:

- AI tools should be embedded within structured, scenario-based pedagogies that require students to articulate reasoning steps rather than passively consume generated outputs

- Safeguards against cognitive dependency should be implemented, including mandatory verification of AI-generated content and reflective discussion
- AI literacy, encompassing bias detection, hallucination awareness, transparency limitations and responsible use, should be incorporated as a formal learning objective [38,39]

Future research should prioritize rigorously designed randomized trials, standardized and performance-based CR measures, longitudinal follow-up assessing skill retention and clearer reporting of implementation conditions. Comparative analyses across AI categories would further clarify differential educational effects.

Overall, while AI-enabled educational tools demonstrate promising potential under guided conditions, their contribution to sustained, performance-based clinical reasoning development remains variable and dependent on pedagogical context, supervision and methodological rigor.

Limitations

Several limitations must be acknowledged when interpreting the findings of this review. The search period (January 2022–January 15, 2026) was intentionally restricted to capture contemporary AI technologies, particularly generative systems, but may have excluded relevant earlier studies and did not include research published after the final search date. Although major databases were searched, specialized sources such as CINAHL were not included and the English-language restriction may have led to missed studies; additionally, sixteen full-text articles could not be retrieved, introducing potential availability bias. Considerable heterogeneity across studies, including differences in design, sample size, geographic setting, AI tool type, pedagogical integration, exposure duration and outcome measures, limited comparability and precluded meta-analysis. Clinical reasoning was frequently assessed using self-reported or proxy measures rather than objective performance-based instruments and the limited number of randomized controlled trials, along with incomplete reporting of key methodological safeguards, reduced confidence in causal inference. Geographic concentration of studies and the exclusive focus on undergraduate students further constrain generalizability. Finally, the rapid evolution of AI technologies and inconsistent reporting of vendor or proprietary influences may affect the durability and transparency of conclusions. Overall, findings should be interpreted carefully, with clear distinction between perceived improvements and objectively demonstrated gains in clinical reasoning performance.

Implications for Practice and Research

Implications for Educational Practice: The evidence supports a cautious and structured integration of AI tools within undergraduate nursing education. AI should not be

implemented as a replacement for human mentorship or clinical supervision but rather as a complementary resource embedded within guided pedagogical frameworks.

Three practical considerations emerge:

- **Supervised and Hybrid Implementation:** AI tools should be integrated within structured learning activities that include instructor guidance, debriefing and peer-based justification. Passive or unsupervised use may increase the risk of superficial engagement or cognitive dependency
- **Explicit AI Literacy Training:** Students should receive formal instruction in evaluating AI-generated outputs, recognizing hallucinations and bias, understanding limitations of algorithmic systems and maintaining academic integrity. Data privacy and confidentiality safeguards must be clearly articulated
- **Alignment With Learning Objectives:** AI integration should be directly aligned with defined educational outcomes. When the goal is development of clinical reasoning, instructional designs should require articulation of reasoning processes rather than reliance on automated answers

Faculty preparedness also warrants attention. Effective implementation requires educator training in AI literacy, scenario design and ethical oversight. Institutional infrastructure, technical support and cost considerations must be addressed to ensure equitable access and sustainability.

Implications for Research

Future research should prioritize methodological rigor and conceptual clarity.

Key priorities include:

- Larger, multicenter randomized controlled trials with transparent reporting;
- Standardized and validated performance-based clinical reasoning measures;
- Clear differentiation between primary CR outcomes and secondary perceptions;
- Longitudinal studies assessing skill retention and transfer to clinical practice;
- Comparative analyses across distinct AI tool categories;
- Examination of psychological and cognitive effects, including potential dependency;
- Evaluation of financial feasibility and faculty readiness.

Research should also explore hybrid AI–human training models to determine optimal balances between technological scaffolding and human mentorship.

CONCLUSIONS

This systematic review synthesized evidence from nine studies examining AI-enabled educational tools in undergraduate nursing education. The findings suggest that AI may support certain aspects of learning related to clinical reasoning,

particularly when embedded within guided and scenario-based pedagogies. However, reported improvements were frequently based on self-reported or proxy measures rather than standardized performance-based assessments.

Heterogeneity in tool categories, pedagogical approaches, comparators and outcome measures precludes definitive conclusions regarding overall effectiveness. Several studies reported null or mixed findings and methodological limitations further constrain causal interpretation. Consequently, AI should not be viewed as an inherently transformative solution for clinical reasoning development.

Rather, AI appears to function most effectively as a structured cognitive support within supervised educational contexts. Its contribution to complex, context-sensitive reasoning processes remains variable and contingent upon pedagogical design, faculty oversight and learner engagement.

In practice, AI integration in nursing education should proceed cautiously, emphasizing verification, justification, ethical safeguards and preservation of the human dimensions of care. Continued rigorous research is required before firm claims can be made regarding sustained improvements in clinical reasoning performance.

Ethical Statement

As this review synthesizes findings from previously published studies and does not involve direct human participation or identifiable personal data, formal ethical approval was not required.

REFERENCES

- [1] Georgieva-Tsaneva, G. *et al.* “Application of virtual reality, artificial intelligence and other innovative technologies in healthcare education (nursing and midwifery specialties): challenges and strategies.” *Education Sciences*, vol. 15, no. 1, 2025, p. 1. <https://doi.org/10.3390/educsci15010011>.
- [2] Rodriguez-Arrastia, M. *et al.* “Experiences and perceptions of final-year nursing students of using a chatbot in a simulated emergency situation: a qualitative study.” *Journal of Nursing Management*, vol. 30, no. 8, 2022, pp. 3874–3884. <https://doi.org/10.1111/jonm.13630>.
- [3] Saban, M. and I. Dubovi. “A comparative vignette study: evaluating the potential role of a generative AI model in enhancing clinical decision-making in nursing.” *Journal of Advanced Nursing*, vol. 81, no. 11, 2025, pp. 7489–7499. <https://doi.org/10.1111/jan.16101>.
- [4] Baloyi, O.B. and N.G. Mtshali. “A middle-range theory for developing clinical reasoning skills in undergraduate midwifery students.” *International Journal of Africa Nursing Sciences*, vol. 9, 2018, pp. 92–104. <https://doi.org/10.1016/j.ijans.2018.10.004>.
- [5] Antikchi, M. *et al.* “The effect of game-based scenario writing on the clinical reasoning of internship nursing students in cardiovascular emergencies and critical care units.” *BMC Medical Education*, vol. 25, no. 1, 2025, p. 1. <https://doi.org/10.1186/s12909-025-07079-w>.
- [6] Doyon, O. and L. Raymond. “Clinical reasoning and clinical judgment in nursing research: a bibliometric analysis.” *International Journal of Nursing Knowledge*, vol. 36, no. 3, 2025, pp. 339–350. <https://doi.org/10.1111/2047-3095.12484>.
- [7] Audétat, M.C. *et al.* “Clinical reasoning: where do we stand on identifying and remediating difficulties?” *Creative Education*, vol. 4, no. 6, 2013, pp. 42–48. <https://doi.org/10.4236/ce.2013.46A008>.
- [8] Mwale, O.G. *et al.* “Acquisition of clinical reasoning skills by undergraduate nursing students in Malawi: towards the development of a middle-range theory, a qualitative study.” *BMC Nursing*, vol. 24, no. 1, 2025, p. 416. <https://doi.org/10.1186/s12912-025-03064-2>.
- [9] Bozkurt, S.A. *et al.* “A systematic review and sequential explanatory synthesis: artificial intelligence in healthcare education, a case of nursing.” *International Nursing Review*, vol. 72, no. 2, 2025, <https://doi.org/10.1111/inr.70018>.
- [10] Chen, Y. *et al.* “Need assessment for history-taking instruction program using chatbot for nursing students: a qualitative study using focus group interviews.” *Digital Health*, vol. 9, 2023. <https://doi.org/10.1177/20552076231185435>.
- [11] Gagne, D. and C.J. “The state of artificial intelligence in nursing education: past, present and future directions.” *International Journal of Environmental Research and Public Health*, vol. 20, no. 6, 2023, p. 6. <https://doi.org/10.3390/ijerph20064884>.
- [12] Han, J.W. *et al.* “Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: a quasi-experimental study.” *BMC Medical Education*, vol. 22, no. 1, 2022. <https://doi.org/10.1186/s12909-022-03898-3>.
- [13] Owoc, M.L. *et al.* “Artificial intelligence technologies in education: benefits, challenges and strategies of implementation.” *Artificial Intelligence for Knowledge Management*, Springer International Publishing, 2021, pp. 37–58. https://doi.org/10.1007/978-3-030-85001-2_4.
- [14] Rony, M.K.K. *et al.* “Nursing students’ perspectives on integrating artificial intelligence into clinical practice and training: a qualitative descriptive study.” *Health Science Reports*, vol. 8, no. 4, 2025. <https://doi.org/10.1002/hsr.70728>.
- [15] Rony, M.K.K. *et al.* “Nursing educators’ perspectives on the integration of artificial intelligence into academic settings.” *SAGE Open Nursing*, vol. 11, 2025. <https://doi.org/10.1177/23779608251342931>.
- [16] Labrague, L.J. and S.A. Sabei. “Integration of AI-powered chatbots in nursing education: a scoping review of their utilization, outcomes and challenges.” *Teaching and Learning in Nursing*, November 2024. <https://doi.org/10.1016/j.teln.2024.11.010>.
- [17] Pérez-Perdomo, A. and A. Zabalegui. “Teaching strategies for developing clinical reasoning skills in nursing students: a systematic review of randomized controlled trials.” *Healthcare*, vol. 12, no. 1, 2024, p. 1. <https://doi.org/10.3390/healthcare12010090>.
- [18] García-Torres, D. *et al.* “Enhancing clinical reasoning with virtual patients: a hybrid systematic review combining human reviewers and ChatGPT.” *Healthcare*, vol. 12, no. 22, 2024, p. 2241. <https://doi.org/10.3390/healthcare1222241>.
- [19] Sim, J.J.M. *et al.* “Virtual simulation to enhance clinical reasoning in nursing: a systematic review and meta-analysis.” *Clinical Simulation in Nursing*, vol. 69, 2022, pp. 26–39. <https://doi.org/10.1016/j.ecns.2022.05.006>.
- [20] Page, M.J. *et al.* “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.” *BMJ*, vol. 372, 2021, p. n71. <https://doi.org/10.1136/bmj.n71>.
- [21] Fung, T.C.J. *et al.* “Effects of generative artificial intelligence (GenAI) patient simulation on clinical competency among global nursing undergraduates: a cross-over randomised controlled trial.” *Research Square*, May 2025. <https://doi.org/10.21203/rs.3.rs-6250414/v1>.

- [22] Harder, N. *et al.* “Comparing artificial intelligence-enhanced virtual reality and simulated patient simulations in undergraduate nursing education.” *Clinical Simulation in Nursing*, vol. 105, 2025, p. 101780. <https://doi.org/10.1016/j.cens.2025.101780>.
- [23] Park, S.A. and H.Y. Kim. “Development and effects of a scenario-based labor nursing simulation education program using an artificial intelligence tutor: a quasi-experimental study.” *Women’s Health Nursing*, vol. 31, no. 2, 2025, pp. 143–154. <https://doi.org/10.4069/whn.2025.06.18>.
- [24] Han, J.W. *et al.* “Development and effects of a chatbot education program for self-directed learning in nursing students.” *BMC Medical Education*, vol. 25, no. 1, 2025. <https://doi.org/10.1186/s12909-025-07316-2>.
- [25] Shin, H. *et al.* “The impact of artificial intelligence-assisted learning on nursing students’ ethical decision-making and clinical reasoning in pediatric care.” *Computers, Informatics, Nursing*, vol. 42, no. 10, 2024, pp. 704–711. <https://doi.org/10.1097/CIN.0000000000001177>.
- [26] Lam, M. *et al.* “Undergraduate nursing students’ perspectives on artificial intelligence in academia.” *Canadian Journal of Nursing Research*, June 2025. <https://doi.org/10.1177/08445621251347025>.
- [27] Shi, W. *et al.* “Retrieval-augmented large language models for adolescent idiopathic scoliosis patients in shared decision-making.” *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Association for Computing Machinery, 2023, pp. 1–10. <https://doi.org/10.1145/3584371.3612956>.
- [28] Chang, C.Y. *et al.* “Chatbot-facilitated nursing education: incorporating a knowledge-based chatbot system into a nursing training program.” *Educational Technology & Society*, vol. 25, no. 1, 2022, pp. 15–27.
- [29] Simsek-Cetinkaya, S. and Cakir, S.K. “Evaluation of the effectiveness of artificial intelligence-assisted interactive screen-based simulation in breast self-examination: an innovative approach in nursing students.” *Nurse Education Today*, vol. 127, 2023. <https://doi.org/10.1016/j.nedt.2023.105857>.
- [30] Abdulai, A. and Hung, L. “Will ChatGPT undermine ethical values in nursing education, research and practice?” *Nursing Inquiry*, vol. 30, no. 3, 2023. <https://doi.org/10.1111/nin.12556>.
- [31] Cho Kwan, R.Y. *et al.* “Navigating the integration of artificial intelligence in nursing: opportunities, challenges and strategic actions.” *International Journal of Nursing Sciences*, vol. 12, no. 3, 2025, pp. 241–245. <https://doi.org/10.1016/j.ijnss.2025.04.009>.
- [32] Ramírez-Baraldes, E. *et al.* “Artificial intelligence in nursing: new opportunities and challenges.” *European Journal of Education*, vol. 60, no. 1, 2025. <https://doi.org/10.1111/ejed.70033>.
- [33] Anthamatten, A. *et al.* “Developing clinical competence through case presentations with artificial intelligence-driven simulation.” *Journal of Nurse Practitioners*, vol. 21, no. 7, 2025. <https://doi.org/10.1016/j.nurpra.2025.105415>.
- [34] Liaw, S.Y. *et al.* “Artificial intelligence versus human-controlled doctor in virtual reality simulation for sepsis team training: randomized controlled study.” *Journal of Medical Internet Research*, vol. 25, no. 1, 2023, p. e47748. <https://doi.org/10.2196/47748>.
- [35] Ma, J. *et al.* “The role of artificial intelligence in shaping nursing education: a comprehensive systematic review.” *Nurse Education in Practice*, vol. 84, 2025, p. 104345. <https://doi.org/10.1016/j.nepr.2025.104345>.
- [36] Ramirez, K.B.C. *et al.* “Analysis of architect profession career pursuance intention among Filipinos: a TPB-SDT approach under public and private educational administration.” *Acta Psychologica*, vol. 258, 2025, p. 105286. <https://doi.org/10.1016/j.actpsy.2025.105286>.
- [37] Shen, M. *et al.* “Prompts, privacy and personalized learning: integrating AI into nursing education, a qualitative study.” *BMC Nursing*, vol. 24, no. 1, 2025. <https://doi.org/10.1186/s12912-025-03115-8>.
- [38] Labrague, L.J. *et al.* “Artificial intelligence in nursing education: a review of AI-based teaching pedagogies.” *Teaching and Learning in Nursing*, vol. 20, no. 3, 2025, pp. 210–221. <https://doi.org/10.1016/j.teln.2025.01.019>.
- [39] Porsdam Mann, S. *et al.* “Guidelines for ethical use and acknowledgement of large language models in academic writing.” *Nature Machine Intelligence*, vol. 6, no. 11, 2024, pp. 1272–1274. <https://doi.org/10.1038/s42256-024-00922-7>.