Nursing, AI, and Clinical Research: 2025 Insights

May 1, 2026

Andrew Phillips PhD, RN, FAMIA

Conflict of Interest

Have no real or apparent conflicts of interest to report.

Learning Objectives

- How will AI potentially impact nursing practice

Methods – Scoping Study

Arksey and O'Malley[1]

- Step 1 – Identify the Research Question
- Step 2 – Search for Relevant Literature
- Step 3 – Study Selection
- Step 4 – Charting the Data
- Step 5 – Summarizing, and reporting the results
- Step 6 – Consultation – This is you guys

[1]Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

Step 1: Research Question

- What trends and themes emerge from a survey of the published literature in the area of AI in nursing during the past year

Step 2: Identify Relevant Literature

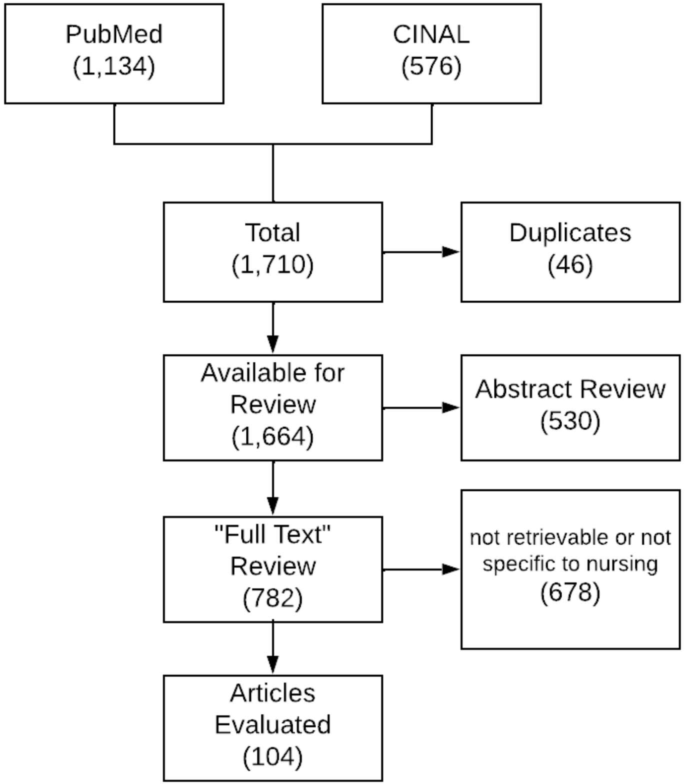
- Search Strategy
  - Databases: PubMed and CINAHL
  - Search terms
- (“artificial intelligence” OR “AI” OR “large language models” OR “machine learning”) AND healthcare AND nurs*
- Publication Dates 3/1/2025 – 3/31/2026

Step 3: Study Selection

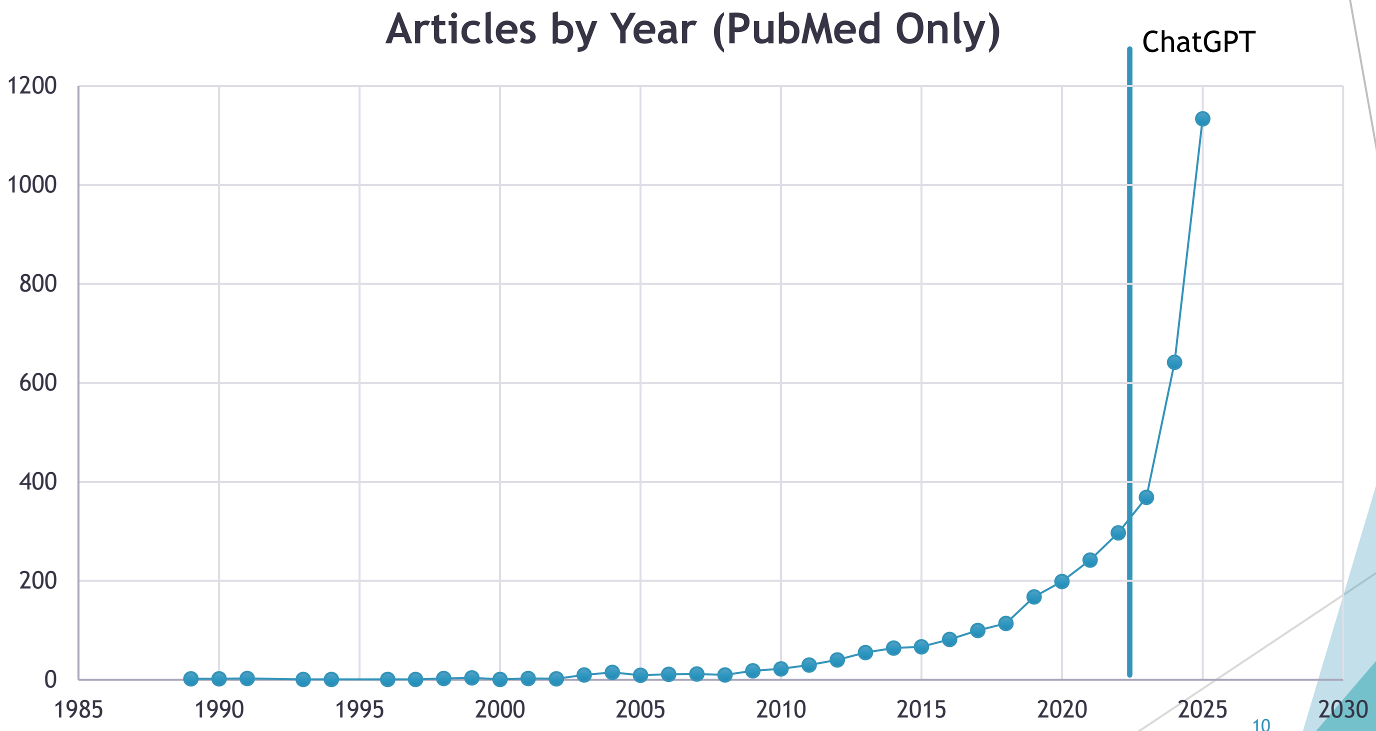
Inclusion and Exclusion Criteria

- Inclusion criteria: healthcare delivery, relevant to nursing
- Exclusions: Articles without an AI informatics focus

Search Results

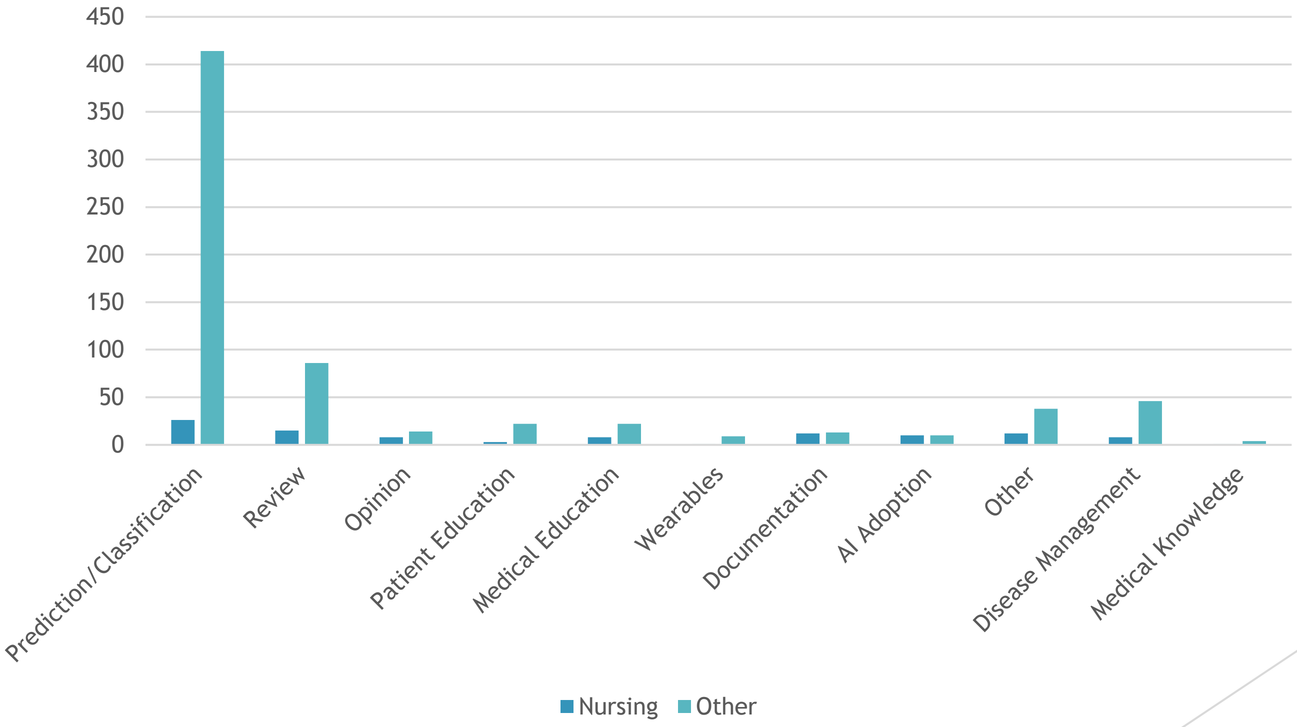


Search Results

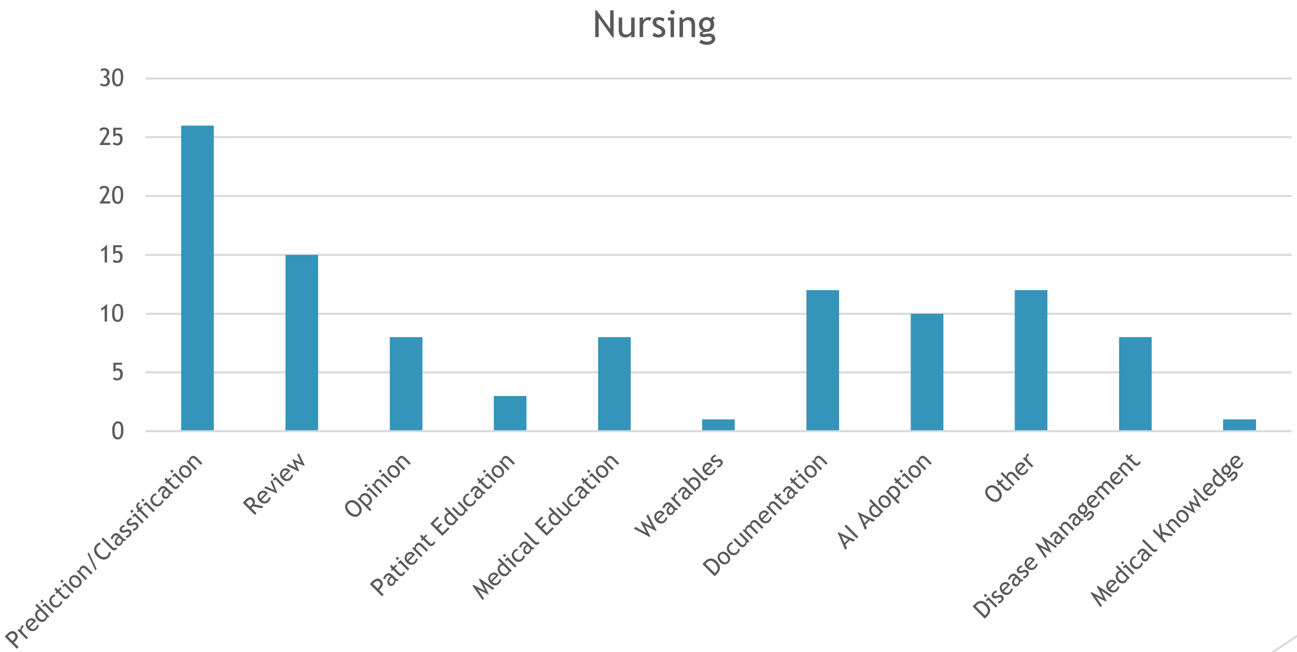


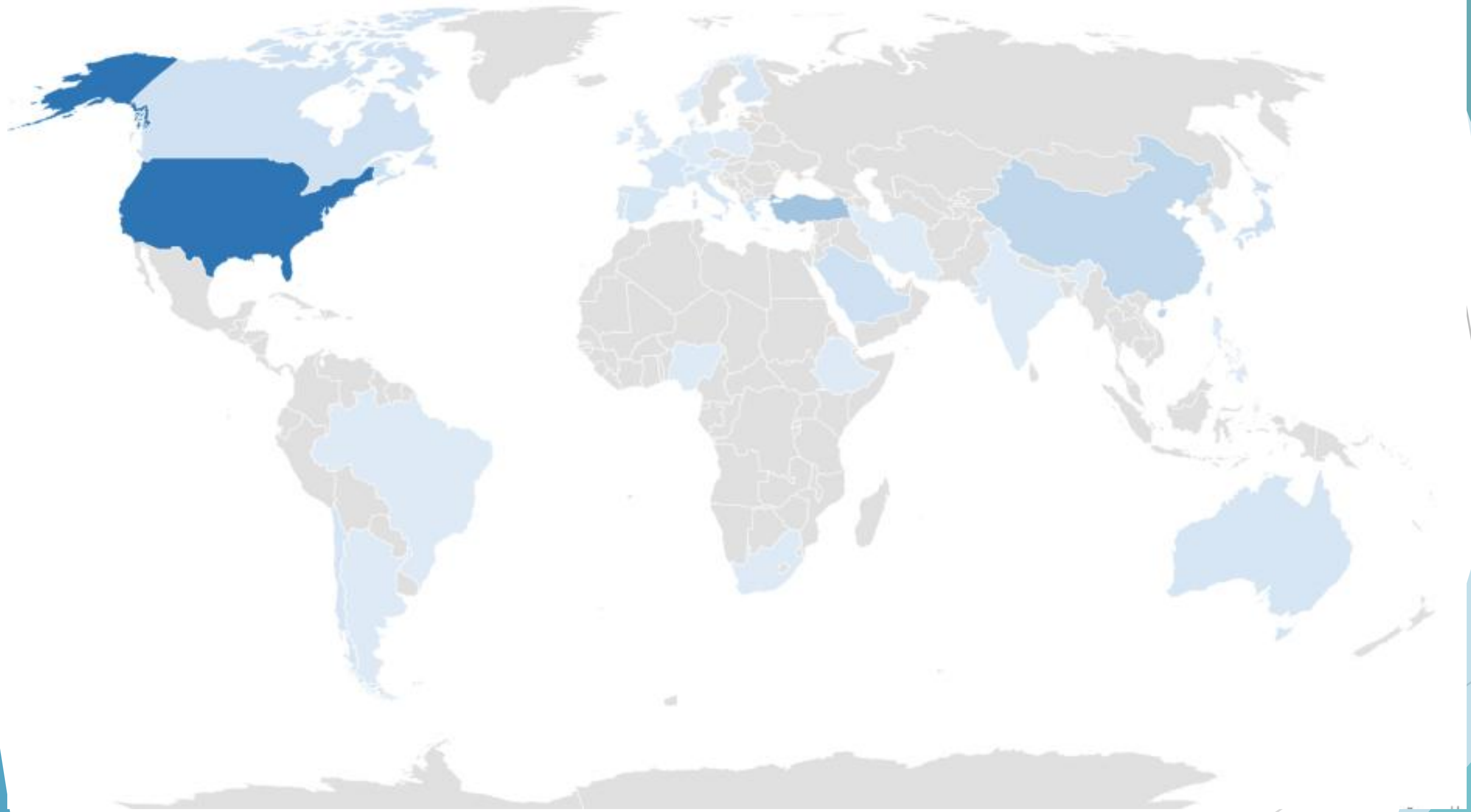
Step 4 – Charting the Data

Article Classifications



Article Classifications





Step 5 – Collating, summarizing, and reporting the results

Themes Identified

1. AI is here. Forgone conclusion that LLMs are coming
2. Education critical for effective and ethical use
3. Potential professional risk of accepting or rejecting AI recommendations
4. Those familiar with AI more likely to find it “exciting”
5. Documentation burden
6. Burnout
7. Prediction/classification models galore
    1. Alert fatigue
8. Much more I did not understand this year
9. Ethics
10. Bias
11. Barriers to adoption
12. Literature reviews
13. Machine Learning vs. LLMs
14. Prompts are important

Uses Identified

- Scheduling
- Clinical Decision Support (CDS)
  - Falls
  - Readmission
  - Pressure injury
  - sepsis
- Nursing Care Plans
- Mental Health Nursing
- Patient Assignments
- Incident reports
- Classification
- Assessing burnout
- Simulation education
- Nursing diagnoses
- Patient education/communication
- Chronic pain management
- Nursing exams and education

Representative/Interesting Citations

Dai, Q., Li, M., Yang, M., Shi, S., Wang, Z., Liao, J., ... & Tang, Y. D. (2025). Attitudes, perceptions, and factors influencing the adoption of AI in health care among medical staff: nationwide cross-sectional survey study. *Journal of Medical Internet Research*, *27*, e75343.

Gokalp, M. G., Yucel, S. C., Cakir, Z., & Kargi, R. (2025). Readability, reliability, and quality of nursing care plan texts generated by ChatGPT. *BMC nursing*.

Gupta, P., Zhang, Z., Song, M., Michalowski, M., Hu, X., Stiglic, G., ... & Topaz, M. (2025). Video-Based Fall Risk Assessment Using Multimodal Large Language Models in Home Health Care: A Proof-of-Concept Feasibility Study. *CIN: Computers, Informatics, Nursing*, 10-1097.

Liang, A. S., Vedak, S., Dussaq, A., Yao, D. H., Villarreal, J. A., Thomas, S., ... & Morse, K. (2025). Artificial intelligence-generated draft replies to patient messages in pediatrics. *JAMIA open*, *8*(6), ooaf159.

Tuncer, M., & Yalcinkaya, T. (2025). Evaluation of ChatGPT-4, Gemini, Claude, and Copilot in Generating Nursing Diagnoses Based on NANDA-I Taxonomy II: A Comparative Cross-Sectional Study. *International Nursing Review*, *72*(4), e70135.

Turchioe, M. R., Austin, R., & Lytle, K. (2023). Opportunities and challenges for digital health and artificial intelligence to support nurses: results of a survey of nursing informaticists. *CIN: Computers, Informatics, Nursing*, 10-1097.

Gokalp, M. G., Yucel, S. C., Cakir, Z., & Kargi, R. (2025). Readability, reliability, and quality of nursing care plan texts generated by ChatGPT. BMC nursing.

Background

- Nursing care plans require clinical reasoning, prioritization, and patient-centered decision-making, which distinguishes them from more general AI-generated educational texts. As large language models such as ChatGPT are increasingly used to support nursing education and care planning, it is essential to evaluate the readability, reliability, and quality of the nursing care plans they produce.

Purpose

- This study aims to evaluate the readability, reliability, and quality of nursing care plan texts generated by ChatGPT.

Methods

- The study sample consisted of 50 texts generated by **ChatGPT (version 4.0)** based on selected nursing diagnoses from NANDA 2021–2023. These texts were evaluated using a descriptive criteria form, the DISCERN tool, and readability indices including the Flesch Reading Ease Score (FRES), Simple Measure of **Gobbledygook** (SMOG), Gunning Fog Index, and Flesch-Kincaid Grade Level (FKGL).

Results

- The analysis demonstrated that the nursing care plans generated by ChatGPT showed a moderate level of quality and reliability. However, the readability levels were generally higher than what is desirable for clinical and educational use, indicating that the texts may be difficult for some users to understand without adaptation. The findings also suggest that the presence of verifiable references contributes positively to the overall quality and reliability of the generated care plans.

Conclusion

- Evaluating the readability, reliability, and quality of AI-generated nursing care plans is essential for ensuring their safe and meaningful use in nursing education and clinical practice. These findings highlight the importance of guiding and refining AI-supported care planning to better align with professional standards and patient-centered care needs.

<u>PROMPT</u>:

Please write a nursing care plan for the following NANDA-I nursing diagnosis: [DIAGNOSIS NAME]. Include defining characteristics, related factors, expected outcomes, nursing interventions, and rationales. Use in-text citations and provide references at the end of the text.

Table 3 Comparison of the readability scores of all texts according to text content using calculator 1 and 2 with the 6th grade reading level (https://readabilityformulas.com/free-readability-formula-tests.php)

Calculator 1			Calculator 2	
Calculator 1–2 statistics	ChatGPT 4.0	Chat GPT C6thGRL (*P*)*1	ChatGPT 4.0	Chat GPT 4.0 C6thGRL (*P*)*1
FRES	31.35 (9–55)	**< 0.001**	29.73	**< 0.001**
GFOG	17.75 (13.40–25.65)	**< 0.001**	(9.49–51.29)	
FKGL	14.29 (11.74–20.72)	**< 0.001**	16.53	**< 0.001**
CLI	15.17 (12.37–21.42)	**< 0.001**	(13.22–25.16)	**< 0.001**
SMOG	12.85 (9.57–18.09)	**< 0.001**	13.78	**< 0.001**
ARI	16.52 (12.58–23.45)	**< 0.001**	(11.25–20.04)	**< 0.001**
LW	16.27 (12.44–22.38)	**< 0.001**		
Grade level	15.00 (12.00–20.00)	**< 0.001**		

Table 5 Average scores of DISCERN total scale and subscales and evaluation of all texts using DISCERN

DISCERN Bölümler	Min-Max	Ort ± SD	Poor (Score 79%)	Fair (Score 40%-79%)	Good (Score > 79%)
S1 to S8 Reliability of Publication*	21–35	26.62 ± 2.85	0 (0)	50 (100)	0 (0)
S9 ila S15 Quality of information on nursing care**	13–36	27.22 ± 4,65	2 (4)	48 (96)	0 (0)
S16 Overall quality***	3–5	3.56 ± 0.6	10 (20)	37 (74)	3 (6)
Total****	40–69	57,41 ± 5,9			

* Minimum-maximum scores for reliability of publication range from 8 to 40 points.** Minimum-maximum scores for quality of information on nursing care range from 7 to 35 points.*** Minimum-maximum scores for overall quality range from 1 to 5 points.**** dMinimum-maximum scores of DISCERN tool range from 16 to 80 points

Dai, Q., Li, M., Yang, M., Shi, S., Wang, Z., Liao, J., ... & Tang, Y. D. (2025). Attitudes, perceptions, and factors influencing the adoption of AI in health care among medical staff: nationwide cross-sectional survey study. Journal of Medical Internet Research, 27, e75343.

.

- **Background**: Artificial intelligence (AI) has demonstrated transformative potential in the health care field; yet, its clinical adoption faces challenges such as inaccuracy, bias, and data privacy concerns. As the primary operators of AI systems, physicians and nurses play a pivotal role in integrating AI into clinical workflows. Their acceptance and use of AI are essential for bridging the gap between technological innovation and practical implementation. Exploring Chinese medical staff's attitudes and identifying key factors influencing AI adoption are fundamental to developing targeted strategies to facilitate the effective application of AI in clinical settings. **Objective**: **This study aims to investigate attitudes and perceptions regarding medical AI among physicians and nurses in China and identify the factors influencing its adoption.** **Methods**: A nationwide cross-sectional survey was conducted online from December 12 to 26, 2024. Participants were recruited from the Chinese Medical Association and the Chinese Nursing Association. The structured questionnaire assessed demographic characteristics, knowledge and attitudes toward medical AI, experiences and insights regarding using medical AI, and perceptions and factors influencing the adoption of AI based on the unified theory of acceptance and use of technology (UTAUT) model. Multiple linear regression and Karlson-Holm-Breen mediation analysis were used to identify influencing factors. Sample weighting by regional distribution was applied for sensitivity analysis. **Results**: The survey included 991 physicians and 1714 nurses. Among the respondents, 92.4% (916/991) of the physicians and 84.19% (1443/1714) of the nurses reported awareness of medical AI applications, 22.8% (226/991) of the physicians and 17% (291/1714) of the nurses had used AI, and 82.6% (819/991) of the physicians and 80.22% (1375/1714) of the nurses held optimistic views about AI's prospects. After adjusting for covariates, performance expectancy (physicians: B=0.144, 95% CI 0.092-0.197; nurses: B=0.292, 95% CI 0.245-0.338), effort expectancy (physicians: B=0.681, 95% CI 0.562-0.800; nurses: B=0.440, 95% CI 0.342-0.538), social influence (physicians: B=0.264, 95% CI 0.187-0.341; nurses: B=0.098, 95% CI 0.045-0.152), and facilitating conditions (physicians: B=0.098, 95% CI 0.030-0.165; nurses: B=0.158, 95% CI 0.105-0.212) had significant positive impacts on willingness to use AI. **Perceived risk showed no significant effect on physicians' intention to use AI (B=0.012, 95% CI −0.022 to 0.045) but negatively impacted nurses' intention to use AI (B=−0.041, 95% CI −0.066 to −0.015).** Performance expectancy and effort expectancy partially mediated the relationship between facilitating conditions and intention to use. Age, educational level, hospital level, work experience, and personal views also significantly influenced willingness. Weighted and unweighted analyses yielded consistent results, confirming the robustness of the findings. **Conclusions**: Substantial disparities exist between high willingness to adopt medical AI and its low actual use among Chinese medical staff. System optimization focusing on utility enhancement, workflow integration, and risk mitigation for medical staff, along with support for infrastructure and training, could accelerate AI adoption in clinical practice.

Table 2. Knowledge and attitudes toward medical artificial intelligence (AI) of participants.

Items	Total (n=2705), n (%)	Physicians (n=991), n (%)	Nurses (n=1714), n (%)	*P* value
Have you ever heard of medical AI?				<.001
No	346 (12.8)	75 (7.6)	271 (15.8)	
Yes	2359 (87.2)	916 (92.4)	1443 (84.2)	
Have you ever used medical AI?				<.001
No	2188 (80.9)	765 (77.2)	1423 (83)	
Yes	517 (19.1)	226 (22.8)	291 (17)	
What is the level of attention your hospital pays to medical AI?				.68
Low	756 (27.9)	281 (28.4)	475 (27.7)	
General	1080 (39.9)	385 (38.8)	695 (40.5)	
High	869 (32.1)	325 (32.8)	544 (31.7)	
What is your view on the prospects of medical AI?				.12
Pessimistic	511 (18.9)	172 (17.4)	339 (19.8)	
Optimistic	2194 (81.1)	819 (82.6)	1375 (80.2)	

Figure 1. Factors associated with intention to use medical artificial intelligence (AI).

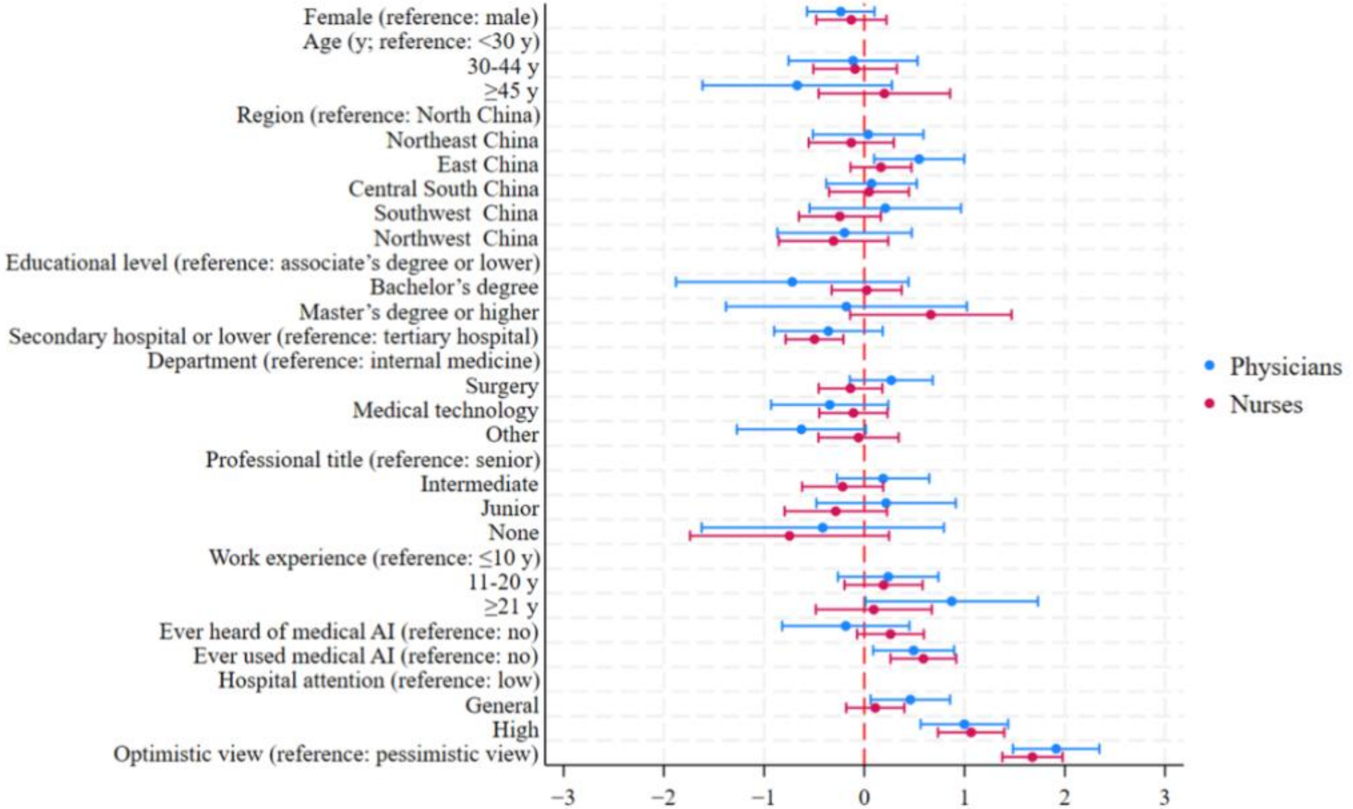
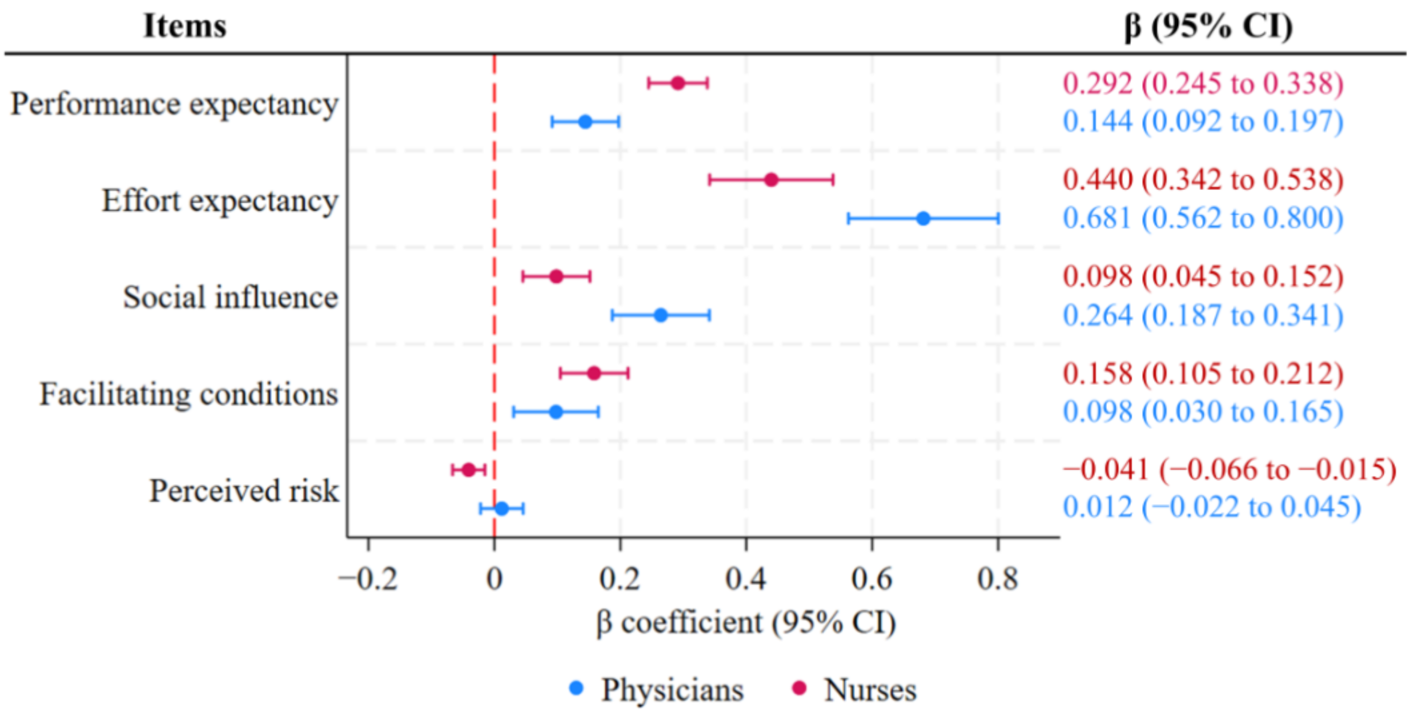
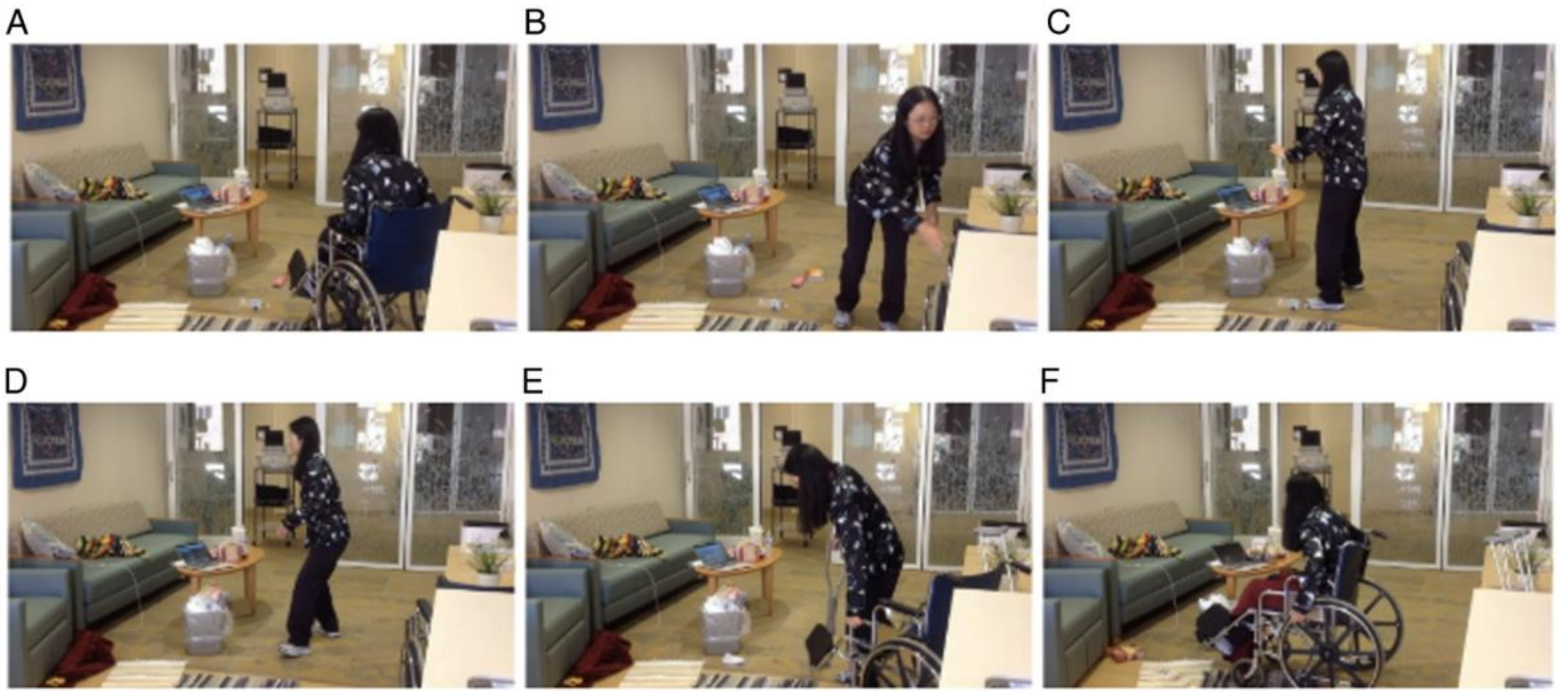


Figure 2. Relationship between perception of and intention to use medical artificial intelligence (AI). The multiple linear regression analysis wa adjusted for demographic information, knowledge of AI, experience of using AI, hospital attention to AI, and personal views on the prospects of medica AI.



Gupta, P., Zhang, Z., Song, M., Michalowski, M., Hu, X., Stiglic, G., ... & Topaz, M. (2025). Video-Based Fall Risk Assessment Using Multimodal Large Language Models in Home Health Care: A Proof-of-Concept Feasibility Study. CIN: Computers, Informatics, Nursing, 10-1097.

- Falls cause millions of injuries and deaths annually, making prevention a key priority in home health care (HHC). Traditional fall risk assessments often overlook the complex interaction of personal, environmental, and behavioral factors. **This study addresses these limitations by introducing a novel approach that leverages multimodal data, specifically visual frames and structured prompts, to assess fall risk in in-home patients.** Using the multimodal large language model (MLLM), LLaVA-NeXTVideo-7B-hf, we analyze simulated in-home patients' video data to enable a more comprehensive and dynamic evaluation of fall risk, paving the way for intelligent, video-based fall prevention in home health care. Preliminary validation using simulated video data demonstrates the feasibility of using MLLMs for such tasks. Simulated in-home patient video data were processed into 24 equally spaced frames. Twelve visually observable fall risk factors extracted from the literature search, categorized as intrinsic, extrinsic, or behavioral, guided the creation of prompts for the MLLM. Standardized prompts were developed by testing the model with concise prompts for simple inferences and elaborated prompts for complex ones. Each prompt was run 3 times, and consensus results were compared with expert evaluations. The model achieved 85.71% accuracy with concise prompts on 7 simple risk factors and 100% accuracy with elaborated prompts on two complex ones. However, the model consistently failed for 2 risk factors that required clinical judgment or had limited visual data. MLLMs like LLaVA-NeXTVideo 7B-hf show strong potential for augmenting fall risk assessment in HHC when guided by well-structured prompts. The approach focuses on visually inferable factors and is intended to complement, rather than replace, clinical evaluation. **This proof-of-concept feasibility study shows that MLLMs can support preliminary fall risk analysis using simulated home health care video data and lays the groundwork for future video-based research in this setting, where existing work remains limited.** To our knowledge, this is the first study to evaluate the feasibility of MLLM-based video analysis for fall risk assessment in home health care.



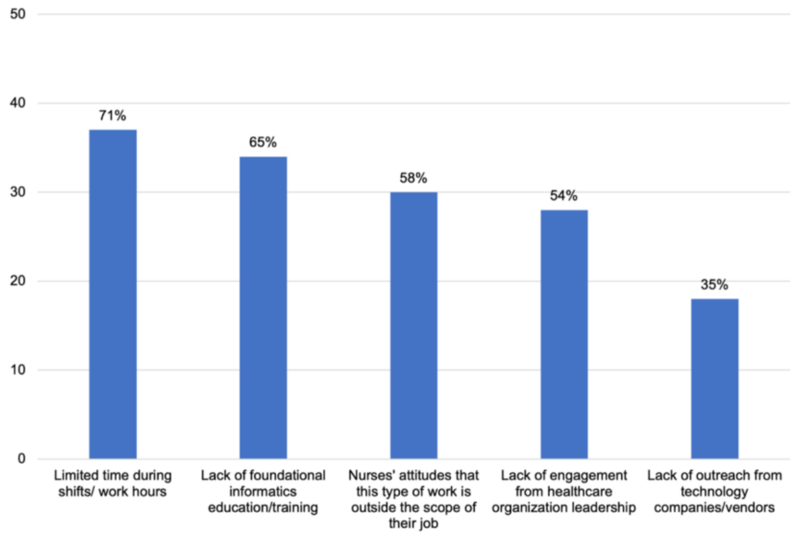
Six representative sequential frames (out of 24 total) extracted from video to illustrate the visuals. A, The patient sits in a wheelchair. B, The patient is a woman wearing glasses. C and D, The patient is exercising. E, The patient sits back in a wheelchair.

Table 2 - Risk Factors, Standardized Prompts, and Responses (MLLM vs. Human Evaluation)

Fall Risk Factors	Standardized Prompts	MLLM Results	Expert Evaluation
Category 1: Direct prompts for easy inferences			
Age	Based on visible characteristics in the video, does the patient seem to be an older adult? Respond 'yes, 'no,' or 'insufficient information.'	Yes	No
Sex	Based on visible features in the video, does the patient seem to be a female or a male? Answer 'male' or 'female.'	Female	Female
Difficulty walking	Based on the video, does the patient display any visible signs of difficulty walking that could increase their risk of falling? Respond 'yes, 'no,' or 'insufficient information.'	Yes	Yes
Impaired mobility	Based on the video, does the patient show any visible signs of impaired mobility that could increase their risk of falling? Respond 'yes, 'no,' or 'insufficient information.'	Yes	Yes
Uneven surfaces	Based on the video, is there any visible indication that the floor is uneven in a way that could increase the risk of falling? Respond 'yes, 'no,' or 'insufficient information.'	Yes	Yes
Category 3: Insufficient Information			
Poor vision(with direct prompt)	Based on observable signs in the video, does the patient seem to have impaired vision? Respond 'yes, 'no,' or 'insufficient information.'	Yes	Insufficient information
Poor vision (with elaborated prompt)	Observe the person in the video and assess their visual ability based on their movements. Look for signs of impaired vision, such as hesitation while walking, difficulty detecting obstacles, excessive squinting, or frequent misjudgment of distances. Describe signs and estimate the likelihood of visual impairment if signs are present. If no such behavior is observed, state that the person seems to have normal vision. Respond 'yes, 'no,' or 'insufficient information.'	Yes	Insufficient information
12. Slippery floor (with direct prompt)	Based on the video, does the floor appear slippery in a way that could increase the risk of falling? Respond 'yes, 'no,' or 'insufficient information.'	Yes	Insufficient information
13. Slippery floor (with elaborated prompt)	Observe the person's walking pattern and body movements to determine if the floor is slippery. Look for signs such as sudden slips, loss of balance, hesitant or cautious walking, quick posture adjustments, or near-falls. If any of these behaviors are present, describe them and assess whether they indicate a slippery surface. If no such signs are observed, state that the floor appears stable. Respond 'yes, 'no,' or 'insufficient information.'	Yes	Insufficient information

Turchioe, M. R., Austin, R., & Lytle, K. (2023). Opportunities and challenges for digital health and artificial intelligence to support nurses: results of a survey of nursing informaticists. CIN: Computers, Informatics, Nursing, 10-1097.

- Artificial intelligence and other digital health technologies may optimize nurses' work. Therefore, we aimed to examine the roles of nurses in facilitating the adoption of digital health technologies and identify opportunities for these technologies to reduce burnout. We conducted a cross-sectional survey study focused on nurses' use of digital health and artificial intelligence technology with nursing informaticists. Data collection was guided by the implementation science framework, Non-Adoption, Abandonment, Scale-up, Spread, and Sustainability. Participants were recruited electronically through professional nursing informatics organizations. Survey data were analyzed using basic descriptive statistics. Fifty-two participants from across the United States completed the survey. Telehealth (73%), patient portals (71%), and medical-grade devices (69%) were most frequently used, whereas artificial intelligence was frequently used by only 38%. Staffing shortages (88%), low staff retention (81%), and inadequate support when adopting new technologies (52%) were among the key drivers of nursing burnout. Participants endorsed most nursing tasks as being supported by digital health, especially patient assessment and evaluating outcomes, and especially artificial intelligence. Engaging nurses early in the process of developing and deploying digital health, especially artificial intelligence, may help address burnout by producing more nursing-centered technologies and providing technology-enabled nursing work alternatives to bedside care.


50
40
30
20
10
0
71%
65%
58%
54%
35%
Limited time during shifts/ work hours
Lack of foundational informatics education/training
Nurses' attitudes that this type of work is outside the scope of their job
Lack of engagement from healthcare organization leadership
Lack of outreach from technology companies/vendors

Tuncer, M., & Yalcinkaya, T. (2025). Evaluation of ChatGPT-4, Gemini, Claude, and Copilot in Generating Nursing Diagnoses Based on NANDA-I Taxonomy II: A Comparative Cross-Sectional Study. International Nursing Review, 72(4), e70135.

- Aim: To evaluate the capability of large language models to generate nursing diagnoses based on NANDA-I Taxonomy II and assess their performance across domains and overall.
- Background: Large language models are emerging tools in nursing, showing potential to aid in diagnosis generation and education. However, their accuracy and applicability in clinical and educational settings remain underexplored.
- Methods: This cross-sectional comparative study used 10 realistic patient scenarios based on NANDA-I Taxonomy II, covering 12 domains. The study aimed to evaluate the capability of four models to generate nursing diagnoses based on patient scenarios. The responses were assessed by five nursing experts for accuracy and alignment with NANDA-I Taxonomy II in a single-blind evaluation process.
- Results: All models demonstrated similar performance across different domains and overall, with Claude attaining the highest overall performance score. Expert evaluations indicated moderate interrater reliability.
- Discussion: Small variations between models and occasional omissions suggest that expert review is still required before clinical use.
- Conclusions: Large language models are not yet sufficiently reliable for independent use in clinical settings and nursing education. Their application as supportive tools necessitates a cautious approach. Moreover, the development of specialized models designed to address the unique demands of the nursing field would be advantageous.
- Implications for nursing: When large language models are used in nursing practice, their limitations should be considered, and the outputs they produce should be verified by nurses. Implications for nursing policy: Ensuring the safe integration of artificial intelligence tools into nursing necessitates the establishment of robust regulatory policies to safeguard patient safety, the deployment of effective systems to monitor models' performance, and the development of comprehensive guidelines and training programs.

TABLE 2 | LLMs' performance by domain and overall.

Domain (*)	ChatGPT Mean ± SD (mean rank)	Claude Mean ± SD (mean rank)	Copilot Mean ± SD (mean rank)	Gemini Mean ± SD (mean rank)	Test statistic and *p*-value (**) Chi-square	*p*-value
Domain 1	4.06 ± 0.55 (20.15)	3.92 ± 0.71 (18.8)	3.86 ± 0.79 (17.65)	4.28 ± 0.40 (25.4)	2.626	0.453
Domain 2	3.54 ± 0.89 (20.3)	3.78 ± 0.51 (23.5)	3.76 ± 0.97 (24.15)	3.14 ± 0.71 (14.05)	4.776	0.189
Domain 3	4.20 ± 1.05 (19.4)	4.36 ± 0.62 (19.3)	4.22 ± 0.98 (20.45)	4.48 ± 0.75 (22.85)	0.622	0.891
Domain 4	4.00 ± 0.71 (21.25)	4.38 ± 0.51 (27.9)	3.70 ± 0.63 (15.95)	3.74 ± 0.80 (16.9)	6.581	0.87
Domain 5	4.32 ± 0.55 (22.9)	4.18 ± 0.60 (19.15)	4.24 ± 0.56 (20.65)	4.18 ± 0.58 (19.3)	0.675	0.879
Domain 6	3.68 ± 0.44 (20.0)	3.62 ± 0.72 (19.1)	3.98 ± 0.51 (25.7)	3.46 ± 0.77 (17.2)	3.004	0.391
Domain 7	4.26 ± 0.82 (21.6)	4.36 ± 0.48 (21.5)	4.40 ± 0.48 (22.3)	4.04 ± 0.72 (16.6)	1.567	0.667
Domain 8	4.28 ± 0.55 (20.4)	4.36 ± 0.37 (21.25)	4.14 ± 0.57 (17.5)	4.40 ± 0.43 (22.85)	1.145	0.766
Domain 9	4.70 ± 0.38 (18.65)	4.60 ± 0.78 (20.85)	4.68 ± 0.56 (20.95)	4.76 ± 0.43 (21.55)	0.491	0.921
Domain 10	4.42 ± 0.63 (19.85)	4.62 ± 0.62 (24.95)	4.58 ± 0.45 (22.35)	4.12 ± 0.80 (14.85)	4.244	0.236
Domain 11	4.62 ± 0.57 (19.6)	4.92 ± 0.25 (26.15)	4.22 ± 1.09 (18.7)	4.24 ± 1.04 (17.55)	4.327	0.228
Domain 12	4.42 ± 0.39 (26.45)	4.00 ± 0.58 (17.65)	4.10 ± 0.69 (19.85)	3.98 ± 0.63 (18.05)	3.698	0.296
Overall Scores	4.20 ± 0.71 (246.27)	4.25 ± 0.66 (253.89)	4.15 ± 0.74 (238.55)	4.06 ± 0.78 (223.28)	3.257	0.354

*The domains have been written in accordance with NANDA-I Taxonomy II (Herdman et al. 2024). Domain 1. Health promotion, Domain 2. Nutrition, Domain 3. Elimination and exchange, Domain 4. Activity/rest, Domain 5. Perception/cognition, Domain 6. Self-perception, Domain 7. Role relationship, Domain 8. Sexuality, Domain 9. Coping/stress tolerance, Domain 10. Life principles, Domain 11. Safety/protection, Domain 12. Comfort (Herdman et al. 2024).

**Performance scores are presented as mean ± SD. Statistical comparisons were performed using the Kruskal–Wallis test, and *p*-values ($p < 0.05$) indicate the significance level of observed differences among models.

Liang, A. S., Vedak, S., Dussaq, A., Yao, D. H., Villarreal, J. A., Thomas, S., ... & Morse, K. (2025). Artificial intelligence-generated draft replies to patient messages in pediatrics. JAMIA open, 8(6), ooaf159.

Objectives

- This study describes the utilization and experiences of artificial intelligence (AI)-generated draft responses to patient messages in pediatric ambulatory clinicians and contextualizes their experiences in relation to those of adult specialty clinicians.

Materials and Methods

- A prospective pilot was conducted from September 2023 to August 2024 in 2 pediatric clinics (General Pediatric and Adolescent Medicine) and 2 obstetric clinics (Reproductive Endocrinology and Infertility and General Obstetrics) within an academic health system in Northern California. Participants included physician, nurse, and medical assistant volunteers. The intervention involved a feature utilizing large language models embedded in the electronic health record to generate draft responses. Proportion of AI-generated draft used was collected, as were prepilot and follow-up surveys.

Results

- A total of 61 clinicians (26 pediatric, 35 obstetric) enrolled, with 46 (75%) completing both surveys. Pediatric clinicians utilized 13.3% (95% CI, 12.3%-14.4%) of AI-generated drafts, and usage rates when responding to patients vs their proxies was similar (15% vs 12.9%, $P = .24$). Despite using AI-generated drafts significantly less than obstetric clinicians (18.3% [17.2%-19.5%], $P < .0001$), pediatric clinicians reported a significant reduction in perceived task load (NASA Task Load Index: 59.9–50.9, $P = .04$) and were more likely to recommend the tool (LTR: 7.0 vs 5.2, $P = .04$).

Discussion and Conclusion

- Pediatric clinicians used AI-generated drafts at a rate within previously reported ranges in adult specialties and experienced utility. These findings suggest this tool has potential for enhancing efficiency and reducing task load in pediatric care.

Step 6 – Consultation

Feedback and Professional Input…