Journal of Primary Care & Community Health
Volume 16: 1–7

DOI: 10.1177/21501319251326663
journals.sagepub.com/home/jpc
Sage

Augmenting Community Nursing Practice With Generative AI: A Formative Study of Diagnostic Synergies Using Simulation-Based Clinical Cases

Odelyah Saad[1,2], Mor Saban[3], Erika Kerner[3], and Chedva Levin[1,4]

Abstract

Objective: To compare the diagnostic accuracy and clinical decision-making of experienced community nurses versus state-of-the-art generative AI (GenAI) systems for simulated patient case scenarios. **Methods**: In the months of 5 to 6/2024, 114 community Israeli nurses completed a questionnaire including 4 medical case studies. Responses were also collected from 3 GenAI models (ChatGPT-4, Claude 3.0, and Gemini 1.5), analyzed both without word limits and with a 10-word constraint. Responses were scored on accuracy, speed, and comprehensiveness. **Results**: Nurses scored higher on average compared to the shortened GenAI responses. GenAI responses were faster but more verbose, and contained unnecessary information. Gemini (full version) and Claude (full version) achieved the highest accuracy among the GenAI models. **Conclusions**: While GenAI shows potential to support aspects of nursing practice, human clinicians currently exhibit advantages in holistic clinical reasoning abilities, a skill requiring experience, contextual knowledge, and ability to bring concise and practical responses. Further research is needed before GenAI can adequately substitute nursing expertise.

Keywords

artificial intelligence, ChatGPT, Claude, clinical reasoning, community service

Dates received: 11 January 2025; revised: 14 February 2025; accepted: 19 February 2025.

Introduction

The landscape of healthcare problem-solving underwent a significant transformation in November 2022 with the introduction of ChatGPT, the first widely accessible large language model.[1,2] This development marked a turning point in artificial intelligence (AI), particularly in its potential applications within the medical field.

As one of the first widely accessible generative AI (GenAI) systems, ChatGPT highlighted opportunities for these models to augment nursing practice and support patient education.[3,4] For nurses, intelligent conversational agents may help extend limited staff resources by assisting with common patient questions. If carefully designed and validated, such AI could help nurses efficiently communicate important health information to diverse communities while maintaining human oversight of clinical content.[5] Overall, generative models demonstrate potential to bolster frontline providers' efforts to educate and empower individuals to manage their health.[6,7] However, ongoing research and real-world testing are still needed to fully realize these benefits while prioritizing patient safety, privacy, and equitable access to care.

Since then, large language models have been increasingly explored as problem-solving aids across various healthcare domains, from assisting in diagnostic processes to providing quick access to medical information.[8-10] In the

[1]Jerusalem College of Technology, Jerusalem, Israel
[2]Ariel University, Ariel, Israel
[3]Tel Aviv University, Tel Aviv, Israel
[4]The Chaim Sheba Medical Center, Tel Hashomer, Ramat Gan, Tel Aviv, Israel

Corresponding Author:
Odelyah Saad, School of Life and Health Sciences, Nursing Department, Jerusalem College of Technology, Havaad Haleumi 21, Jerusalem 91160, Israel.
Email: Odelyahs@gmail.com

nursing field specifically, language models have shown promise in areas such as patient education, care planning, and clinical decision support. Researchers have continued working to expand the capabilities of these models for use in community healthcare, including exploring how they can assist clinics and organizations in answering patients' questions, providing self-care recommendations, and addressing limitations in access to care.

Nurses working in community settings face unique challenges that differ from those in traditional hospital environments.[11] They work in various settings including primary care clinics, telehealth services, and community health centers. Nurses respond to acute health events at the same time as focusing on prevention, chronic disease management, and long-term patient care, providing crucial continuity within complex healthcare systems.[12]

Community nurses often operate with greater autonomy, manage diverse patient populations, and navigate complex social determinants of health. As the use of language models increases, so does the interest in using them in the clinical field. However, despite their advantages and possible usefulness in performing the nursing process,[13] it is not yet clear whether language models can be effectively used in clinical decision-making.[14] Systematic reviews recommend to continue research on the topic in order to understand the capabilities and limitations of the technology.[13,15] Thus, the effectiveness of language models in supporting problem-solving within these contexts, as compared to the skilled decision-making of experienced community nurses, presents an intriguing area of study.[11]

This study aimed to provide a comprehensive comparison of clinical reasoning capabilities between human nurses and GenAI models in community medicine. Most studies that examined language models focused on nursing education.[15] We chose to focus on community nurses whose work requires a significant amount of clinical reasoning. By incorporating both qualitative and quantitative analyses, the study sought to explore the diagnostic strengths and limitations of GenAI in supporting clinical decision-making within community nursing practice.

Methods

Study Design

This cross-sectional study was conducted between May and July 2024 using an online survey designed to evaluate clinical reasoning in community nursing practice. The survey included 4 clinical scenarios that represented common medical challenges encountered in community healthcare settings. The study aimed to compare the clinical reasoning processes of human nurses with those of 3 GenAI models: ChatGPT-4, Claude-3.0, and Gemini-1.5.

Participants

The study included 4 groups of participants: community nurses, ChatGPT-4, Claude-3.0, and Gemini-1.5. The community nurses were drawn from various practice settings, including primary care clinics, home health care services, professional community clinics, and community urgent care centers. Primary clinics provided general outpatient services focused on preventive care and chronic disease management. Home health care nurses offered medical and nursing services to patients in their homes. Professional community clinics specialized in managing specific diseases, such as diabetes or cardiology care. Community urgent care centers operated as walk-in clinics providing immediate but non-emergency care.

The inclusion criteria for nurses were as follows: being a registered nurse, holding a valid nursing license in their country of practice, being actively employed in a community clinic during the data collection period, and providing informed consent to participate. Sociodemographic data collected from the participating nurses included gender, age, total years of professional experience, years of experience in community clinics, and highest level of academic qualification.

Procedure and Data Collection

A clinical reasoning questionnaire was administered to human participants, who were recruited using the snowball sampling technique, via an online survey platform (Qualtrics XM). The questionnaire included 4 clinical scenarios that required participants to assess the presented cases, interpret diagnostic tests, and determine appropriate management strategies. The same scenarios were provided to the 3 GenAI models, which were tasked with generating initial assessments and treatment recommendations. Each GenAI model was prompted twice: once without word limitations (Full Version) and once with a 10-word constraint (Short Version). The rationale for including both versions was to examine the impact of response length on clinical reasoning quality and conciseness.

The AI responses were collected from different platforms. ChatGPT-4's responses were generated using the OpenAI Playground system. Claude-3.0's responses were obtained via the Poe system, an AI chatbot developed by Anthropic that incorporates Constitutional GenAI principles for safe and transparent interactions. Gemini-1.5's responses were generated through its standard user interface.

The clinical scenarios were developed by 2 senior nurses, each with over 30 years of experience and a PhD qualification. These scenarios were designed based on established literature and were intentionally constructed to introduce diagnostic ambiguity, presenting 2 possible diagnoses for each case. The cases included the following community medicine scenarios: a suspected cardiac event, a

diabetic ulcer, an anaphylactic reaction following vaccination, and a urinary tract infection (UTI) in pregnancy. Additional details regarding the scenarios are provided in Supplemental File 1.

To ensure validity and consistency, the clinical cases were reviewed by 2 additional nurses, both of whom held a master's degree and nurse practitioner certification. These reviewers assessed the scenarios for clarity, clinical accuracy, and appropriateness. The scenarios incorporated comprehensive details regarding patient history, comorbidities, and clinical signs. Participants were required to provide an initial evaluation, interpret laboratory and imaging test results, and explain the rationale for their diagnostic and treatment decisions. In the second phase of the questionnaire, additional patient information was provided, requiring participants to adjust their clinical decisions accordingly.

The clinical reasoning assessment was structured around 3 key criteria: accuracy in evaluating the scenarios, including the interpretation of laboratory and imaging tests; accuracy in treatment decision-making following the additional data provided in the second phase; and overall clinical judgment, assessed based on response time and word count for each scenario. A predefined scoring rubric, developed in alignment with clinical guidelines, was used to evaluate responses and ensure consistency across participants. Clinical decision performance for each case scenario was evaluated using scores ranging from 0 to 100, with higher scores indicating better clinical reasoning

Data Analysis

Descriptive and inferential statistical analyses were conducted to compare the performance of human nurses and AI models. The statistical tests used to analyze the data included chi-square tests for categorical variables and *t*-tests or ANOVA for continuous variables, depending on the normality of the distribution. Inter-rater reliability was assessed to ensure consistency in the evaluation of responses across human participants and AI-generated outputs. All statistical analyses were conducted using SPSS Version 28 software. Additional details regarding the statistical approach and specific tests employed are provided in the Results section.

Both authors (OS and CL) collaboratively coded all responses. Each of the reviewers independently scored each response based on predetermined scoring distribution. In cases of scoring discrepancies, the authors reviewed the literature and reached a consensus on the final grade.

Response time (in s) for each system (nurses vs Large Language Models) was recorded from the presentation of case details to final response generation. Mean response times were calculated for all cases. Word counts in written responses for each case were tallied using an automated tool. Mean word counts and standard deviations were calculated for each system across all case responses.

Table 1. Sociodemographic Characteristics of the Study Sample (N = 114).

Variable		Range	Mean (SD)
Age		24-65	43.91 (8.59)
Seniority		1-45	18.44 (10.36)
Seniority in community clinic		1-44	11.42 (9.09)
Variable			**N (%)**
Gender	Male		13 (11.4)
	Female		101 (88.6)
Academic Status	B. A		62 (54.4)
	M.A		52 (45.6)
Post basic course	Yes		30 (26.3)
Type of community clinic	Primary clinic in the community		63 (55.3)
	A professional clinic in the community		17 (14.9)
	Community medical emergency center		3 (2.6)
	Home health care		31 (27.2)

Ethical Considerations

Before the study began, approval was secured from the university's ethics committee. Anonymity was maintained throughout all data collection procedures. Nurses provided informed consent prior to participation and were assured they could withdraw from the study at any time and for any reason.

Results

A total of 114 academic nurses working in community clinics participated in the study, with 52 holding a master's degree (45.6%). The majority (55.3%) were employed in primary community clinics, while slightly more than a quarter (27.2%) worked in home health care. The remaining nurses worked in professional community clinics or community medical emergency centers. Only 30 nurses had completed "post-basic course" training. The mean age of participants was 43.91 ± 8.59 years, ranging from 24 to 65, with 88.6% being women. The average professional seniority was 18.44 ± 10.36 years, varying from 1 to 45 years. The average professional experience in community nursing specifically, was 11.42 ± 9.09 years, ranging from 1 to 44 years, and a median of 10 years. This indicates that most participants, have experience in community nursing. Table 1 provides a comprehensive overview of the study participants' characteristics.

From Table 2, we can observe that there is no consistency, and there are scenarios where the nurses received the highest scores, while at other ones, the language model received the

Table 2. Clinical Decision-making Performance Scores for Nurses Compared to Large Language Models.

CASE mean (SD) (range 0-100)	Nurse	Claude-3.0	Short Claude-3.0	ChatGPT-4.0	Short ChatGPT-4.0	Gemini-1.5	Short Gemini-1.5	*F* score	*P* value
Case 1: Cardiac event	89.55 (11.56)	80.50	63.83	76.60	56.60	83.33	57.00	613.87	**.00**
Case 2: Diabetic ulcer	88.34 (13.79)	93.75	86.75	100	79.25	93.75	100	129.43	**.00**
Case 3: Anaphylactic shock	86.56 (16.41)	83.33	77.66	83.33	58.33	100	66.66	308.70	**.00**
Case 4: Urinary tract infection in pregnancy	85.48 (13.75)	100	68.75	58.25	45.75	100	100	1186.30	**.00**
Average score for all 4 cases	87.52 (9.23)	89.39	74.25	79.54	59.98	94.27	80.91	656.85	**.00**

highest scores. In the first scenario, which dealt with a cardiac event, nurses received the highest scores compared to the large language models. In the second scenario—diabetic ulcer, most types of large language models achieved higher scores than the nurses, except for Short Claude and Short ChatGPT. In the scenario addressing anaphylactic shock, nurses received higher scores compared to large language models, except compared to Gemini. In the fourth scenario, dealing with UTI in pregnancy, nurses scored higher compared to Short Claude and ChatGPT (both full and short versions), while full Claude, Gemini, and Short Gemini achieved 100% accuracy in solving the scenario.

It is notable that the shortened versions of Claude and ChatGPT, reduced the accuracy of the models compared to the unrestricted versions across all 4 scenarios. For Gemini, accuracy decreased in the shortened version for the cardiac event and anaphylactic shock scenarios. Overall, the 3 shortened versions achieved lower scores compared to nurses in the weighted average across all 4 scenarios combined, and the post hoc test revealed a borderline significant difference of .05 between the Claude and Gemini AI models for the average across all 4 scenarios combined. In the unrestricted word count version, among the models, Gemini demonstrated the best accuracy, followed by Claude, and then ChatGPT with a gap of 10 points or more.

The study found no significant correlation between the nurses' clinical accuracy in responding to all case scenarios and various socio-demographic factors. These factors included gender, age, professional academic status, and both general and professional seniority. Additionally, there was no correlation between the type of clinic where nurses were employed and their level of clinical accuracy across the different scenarios presented.

Figure 1 illustrates significant differences in word count between nurses' responses and those of the 3 full large language model types across all 4 cases (case 1: $F = 186\,978.06$, $P = .00$; case 2: $F = 10\,1623.3$, $P = .00$; case 3: $F = 26\,565.1$, $P = .00$; case 4: $F = 82\,904.8$, $P = .00$). Nurses consistently used the fewest words, while Gemini employed the highest number. For instance, in the cardiac event case, the average word count for a nurse's response was 44.25 ± 19.58, compared to 1207 words for Gemini, 348 for Claude, and 248 for ChatGPT. This pattern of nurses using significantly fewer words than the large language models was consistent across all scenarios. It should be noted that although the models mostly provided correct answers, it was necessary to extract it from the entire text provided.

Models for Each Scenario

The data presented in Figure 2 reveals substantial variations in problem-solving speed across all 4 scenarios when comparing nurses to 6 different types of Large Language Models (LLMs). Statistical analysis confirms these differences are significant ($F = 40.59$, $P = .00$). Notably, nurses took considerably longer time to respond, with their reaction times exceeding those of the short large language models by over 70 times and surpassing the full large language models by more than over 21 times.

Discussion

AI technologies have been integrated into healthcare at an unprecedented pace, driven by advancements in machine learning, natural language processing, and big data analytics. AI applications range from predictive analytics and imaging diagnostics to robotic surgery and virtual health assistants.[16] However, while the potential of GenAI is immense, it is crucial to recognize its current limitations and understand that this evolving technology cannot yet fully replace human healthcare providers.

This study compared the performance of nurses and GenAI in handling clinical case descriptions. The results indicate that while GenAI shows promise, it only outperforms the nurses

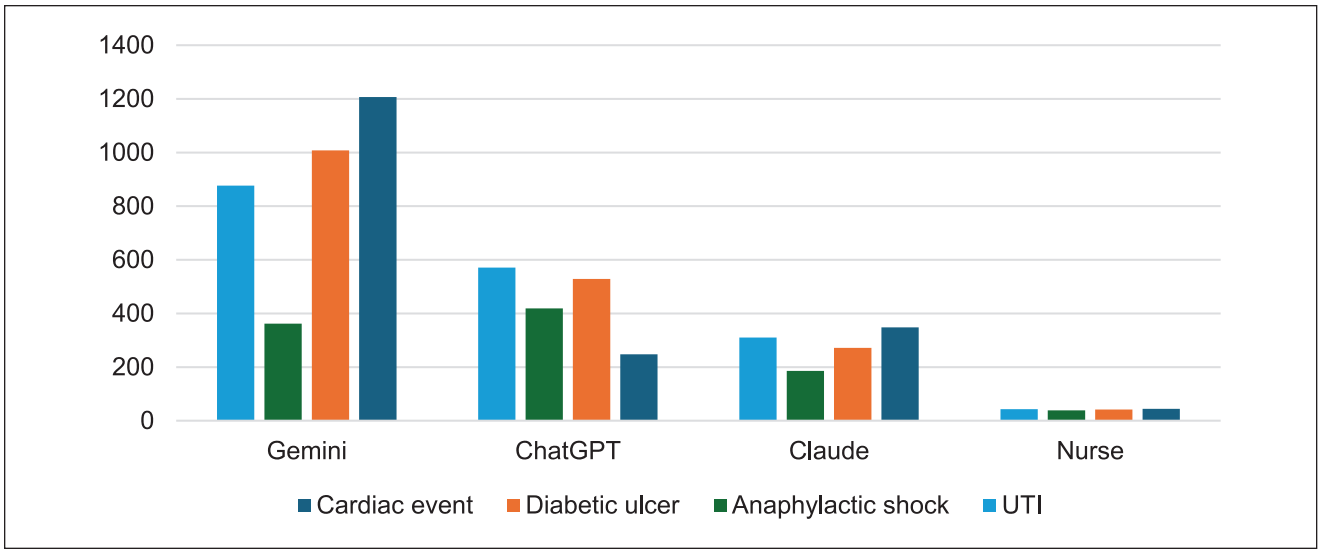

Figure 1. Differences in word counts between nurses and large language.

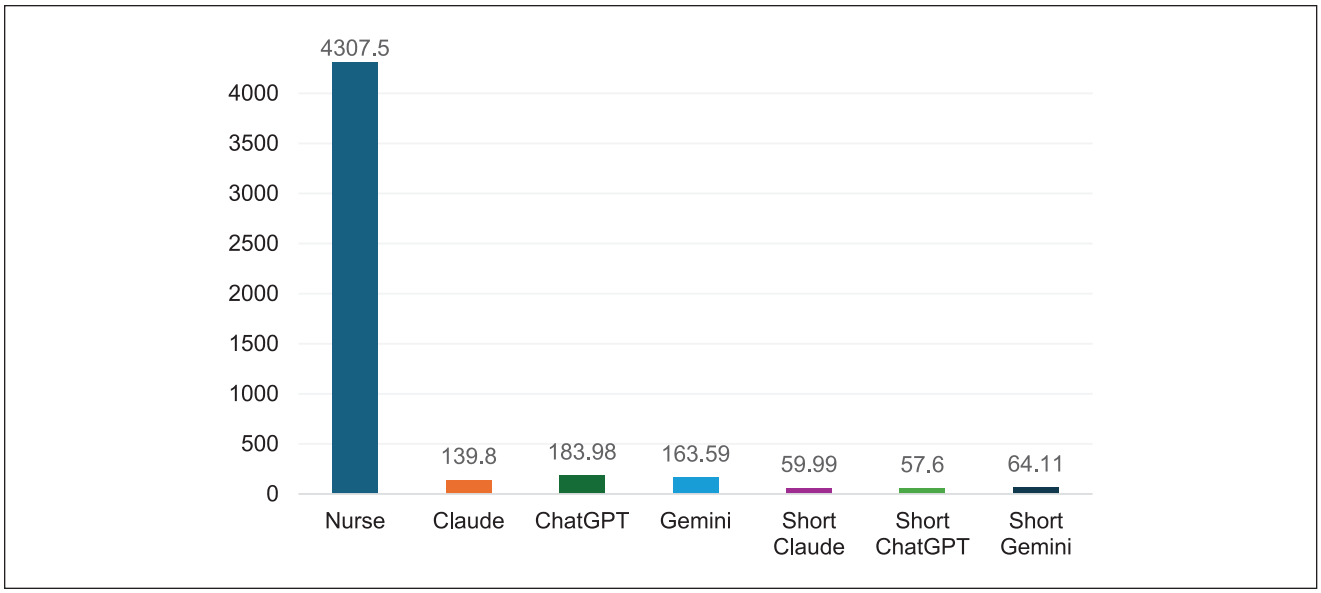

Figure 2. Differences in response time (in s) between nurses and large language models for all 4 case scenarios.

when it is not constrained by word limits for providing solutions. In such cases, GenAI often uses many dozens more words than the nurses used. GenAI models tend to include a lot of unnecessary and irrelevant information, within which the relevant information is hidden. When the models are limited to providing a focused solution of up to 10 words, their accuracy is compromised and falls short of the nurses' expertise. These findings are consistent with previous studies that have demonstrated the superiority of human health professionals in complex clinical decision-making processes.[16,17]

The tendency of GenAI to provide lengthy and convoluted responses, makes them less practical for real-world clinical use.[18] In contrast, nurses provided concise and actionable insights, highlighting the limitation of current GenAI systems in healthcare: the ability to distill complex information into clear, practical guidance[6] that allow for immediate action . This limitation underscores the need for further refinement in GenAI language models to produce more concise and directly applicable outputs.[19,20]

As of today, the nuanced understanding and contextual interpretation that experienced nurses bring to patient care

remain challenging to replicate in GenAI systems whom excel in other medical areas such as image analysis and predicting at-risk populations.[21] The lower scores of GenAI in critical clinical thinking suggest that current GenAI models may lack the depth of clinical reasoning that nurses develop through education, critical thinking, and hands-on experience. This gap is significant in healthcare, where decisions can have life-altering consequences.[22] In situations that require immediate clinical reasoning, large language models are still not good enough and should be used in conjunction with human clinical judgment.[23,24]

Community Health Implications

While our study reveals current limitations of GenAI in nursing tasks, it's important to note that GenAI technology is rapidly evolving. The potential for GenAI to augment rather than replace nursing expertise remains a promising avenue for future developments.[25] Over time, it seems that GenAI can serve as an assistant to medical professionals in considering differential diagnoses and treatment options, especially in situations where the clinical response is not urgent.[24]

However, the results emphasize the irreplaceable value of human nurses in patient care. The ability to synthesize information, draw from experience, and provide empathetic care continues to set human healthcare providers apart from GenAI systems.[26]

While AI shows potential in healthcare applications, our study demonstrates that it currently falls short of matching nursing expertise in critical areas of patient care. The verbose and sometimes impractical nature of GenAI responses highlights the ongoing need for human judgment and experience in clinical settings. As GenAI technology continues to advance, its role in healthcare should be viewed as complementary to, rather than a replacement for, the invaluable skills and intuition of human nurses.

Limitations

Several limitations of this study must be acknowledged. First, only 4 clinical scenarios were used to evaluate clinical reasoning, representing a small sample that does not fully capture the breadth and complexity of real-world nursing practice. Larger and more diverse scenarios may provide different results.

Second, the scenario-based methodology presented static cases without the dynamic evolution of patient conditions over time. Nursing care usually involves iterative adjustment of decisions based on fluctuating clinical factors. Real life situations may emphasize the superiority of nurses over GenAI.

Finally, the GenAI models evaluated in this study represent specific generations that will likely be surpassed by continually advancing natural language processing capabilities. Repeating this comparison longitudinally could show a diminishing performance gap with human experts.

In summary, while providing novel insights, generalizability is constrained by these recognized limitations in study design and scope. Further research addressing these gaps would serve to validate and expand understanding of relative capabilities.

Conclusion

In conclusion, this study provided a first comparison of clinical reasoning performance between experienced community nurses and several state-of-the-art GenAI systems. While GenAI models show promise for supporting administrative and low-complexity nursing functions, human nurses currently demonstrate superiority in diagnostic accuracy, treatment planning, and contextual and concise application of knowledge to patient care—core skills demanding experience and intuition. As GenAI and nursing each continue advancing respectively through technology and education, ongoing evaluation will be essential to define their most effective integration and ensure the preservation of human touch in healthcare.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical Approval

The study was approved by the Institutional ethical committee (#008_20). The committee examines all research proposals in light of acceptable ethical principles.

Data Availability

The data that support the findings of this study are available from the corresponding author, OS, upon reasonable request.

ORCID iD

Odelyah Saad https://orcid.org/0000-0002-7291-9142
Mor Saban https://orcid.org/0000-0001-6869-0907
Chedva Levin https://orcid.org/0000-0002-3336-4279

Supplemental Material

Supplemental material for this article is available online.

References

1. The Lancet Digital Health. ChatGPT: friend or foe? *Lancet Digit Health*. 2023;5(3):e102. doi:10.1016/S2589-7500(23)00023-7
2. Sejnowski TJ. Large language models and the reverse turing test. *Neural Comput*. 2023;35(3):309-342. doi:10.1162/neco_a_01563
3. Taira K, Itaya T, Hanada A. Performance of the large language model ChatGPT on the national nurse examinations in Japan: evaluation study. *JMIR Nurs*. 2023;6(1):e47305. doi:10.2196/47305
4. Gunawan J. Exploring the future of nursing: insights from the ChatGPT model. *Belitung Nurs J*. 2023;9(1):1. doi:10.33546/BNJ.2551
5. Nolin-Lapalme A, Theriault-Lauzier P, Corbin D, et al. Maximising large language model utility in cardiovascular care: a practical guide. *Can J Cardiol*. 2024;40:1774-1787. doi:10.1016/J.CJCA.2024.05.024
6. Sai S, Gaur A, Sai R, Chamola V, Guizani M, Rodrigues JJPC. Generative AI for transformative healthcare: a comprehensive study of emerging models, applications, case studies, and limitations. *IEEE Access*. 2024;12:31078-31106. doi:10.1109/ACCESS.2024.3367715
7. Blease C, Torous J, McMillan B, Hägglund M, Mandl KD. Generative language models and open notes: exploring the promise and limitations. *JMIR Med Educ*. 2024;10(1):e51183. doi:10.2196/51183
8. Rosen S, Saban M. Evaluating the reliability of ChatGPT as a tool for imaging test referral: a comparative study with a clinical decision support system. *Eur Radiol*. 2023;1:1-12. doi:10.1007/S00330-023-10230-0/FIGURES/2
9. Rosen S, Saban M. Can ChatGPT assist with the initial triage? A case study of stroke in young females. *Int Emerg Nurs*. 2023;70:101340. doi:10.1016/J.IENJ.2023.101340
10. Saban M, Dubovi I. A comparative Vignette study: evaluating the potential role of a generative AI model in enhancing clinical decision-making in nursing. *J Adv Nurs*. Published online February 12, 2024. doi:10.1111/jan.16101
11. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Trans Benchmarks Stand Eval*. 2023;3(1):100105. doi:10.1016/J.TBENCH.2023.100105
12. Reiss-Brennan B, Hayes R, McCauley L. An overlooked strategic powerhouse: how nurses can rise to the challenge of integrating public health and primary care. *Am J Public Health*. 2022;112:S253-S255. doi:10.2105/AJPH.2022.306861
13. Bohn B, Anselmann V. Artificial intelligence in nursing practice: a Delphi study with ChatGPT. *Appl Nurs Res*. 2024;80:151867. doi:10.1016/j.apnr.2024.151867
14. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning. *JAMA Netw Open*. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969
15. Kleib M, Darko EM, Akingbade O, et al. Current trends and future implications in the utilization of ChatGPT in nursing: a rapid review. *Int J Nurs Stud Adv*. 2024;7:100252. doi:10.1016/j.ijnsa.2024.100252
16. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
17. Levin C, Suliman M, Naimi E, Saban M. Augmenting intensive care unit nursing practice with generative AI: a formative study of diagnostic synergies using simulation-based clinical cases. *J Clin Nurs*. Published online August 5, 2024. doi:10.1111/JOCN.17384
18. Shah YB, Ghosh A, Hochberg AR, et al. Comparison of ChatGPT and traditional patient education materials for men's health. *Urol Pract*. 2024;11(1):87-94. doi:10.1097/UPJ.0000000000000490
19. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Reza Afrash M. Generative artificial intelligence in healthcare: a scoping review on benefits, challenges and applications. *Int J Med Inform*. 2024;188:105474. doi:10.1016/J.IJMEDINF.2024.105474
20. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inform Decis Mak*. 2021;21(1):1-23. doi:10.1186/S12911-021-01488-9/FIGURES/11
21. Thomas LB, Mastorides SM, Viswanadhan NA, Jakey CE, Borkowski AA. Artificial intelligence: review of current and future applications in medicine. *Fed Pract*. 2021;38(11):527. doi:10.12788/FP.0174
22. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA*. 2018;320(21):2199-2200. doi:10.1001/jama.2018.17163
23. Kostopoulou O, Delaney B. AI for medical diagnosis: does a single negative trial mean it is ineffective? *Lancet Digit Health*. 2025;7(2):e108-e109. doi:10.1016/j.landig.2025.01.005
24. Ranji SR. Large language models—misdiagnosing diagnostic excellence? *JAMA Netw Open*. 2024;7(10):e2440901. doi:10.1001/jamanetworkopen.2024.40901
25. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. *J R Soc Med*. 2019;112(1):22-28. doi:10.1177/0141076818815510
26. Bekbolatova M, Mayer J, Ong CW, Toma M. Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. *Healthcare*. 2024;12(2):125. doi:10.3390/healthcare12020125