Accepted Manuscript online: 2024-07-25 Article published online: 2024-10-16

Increasing Generative Artificial Intelligence Competency among Students Enrolled in Doctoral Nursing Research Coursework

Meghan Reading Turchioe[1] Sergey Kisselev[1] Liesbet Van Bulck[2] Suzanne Bakken[1,3,4]

[1] Columbia University School of Nursing, New York, New York, United States
[2] Department of Public Health and Primary Care, KU Leuven - University of Leuven, Leuven, Belgium
[3] Department of Biomedical Informatics, Columbia University, New York, New York, United States
[4] Data Science Institute, Columbia University, New York, New York, United States

Address for correspondence Meghan Reading Turchioe, PhD, MPH, RN, 560 W. 168th Street, New York, NY 10032, United States (e-mail: mr3554@cumc.columbia.edu).

Appl Clin Inform 2024;15:842–851.

Abstract

Background Generative artificial intelligence (AI) tools may soon be integrated into health care practice and research. Nurses in leadership roles, many of whom are doctorally prepared, will need to determine whether and how to integrate them in a safe and useful way.

Objective This study aimed to develop and evaluate a brief intervention to increase PhD nursing students' knowledge of appropriate applications for using generative AI tools in health care.

Methods We created didactic lectures and laboratory-based activities to introduce generative AI to students enrolled in a nursing PhD data science and visualization course. Students were provided with a subscription to Chat Generative Pretrained Transformer (ChatGPT) 4.0, a general-purpose generative AI tool, for use in and outside the class. During the didactic portion, we described generative AI and its current and potential future applications in health care, including examples of appropriate and inappropriate applications. In the laboratory sessions, students were given three tasks representing different use cases of generative AI in health care practice and research (clinical decision support, patient decision support, and scientific communication) and asked to engage with ChatGPT on each. Students ($n = 10$) independently wrote a brief reflection for each task evaluating safety (accuracy, hallucinations) and usability (ease of use, usefulness, and intention to use in the future). Reflections were analyzed using directed content analysis.

Results Students were able to identify the strengths and limitations of ChatGPT in completing all three tasks and developed opinions on whether they would feel comfortable using ChatGPT for similar tasks in the future. All of them reported increasing their self-rated competency in generative AI by one to two points on a five-point rating scale.

Conclusion This brief educational intervention supported doctoral nursing students in understanding the appropriate uses of ChatGPT, which may support their ability to appraise and use these tools in their future work.

Keywords

- artificial intelligence
- nursing informatics
- informatics
- medical informatics
- nursing students
- graduate nursing education
- nursing education research

received
March 1, 2024
accepted after revision
July 24, 2024
accepted manuscript online
July 25, 2024

© 2024. Thieme. All rights reserved.
Georg Thieme Verlag KG,
Rüdigerstraße 14,
70469 Stuttgart, Germany

DOI https://doi.org/10.1055/a-2373-3151.
ISSN 1869-0327.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

Background and Significance

Generative artificial intelligence (AI), which uses generative models such as large language models to create text, images, or other media in response to a prompt,[1] is rapidly becoming integrated into health care practice and research.[2] Although it has existed in various forms for several years, it is receiving new attention with the release of Chat Generative Pretrained Transformer (ChatGPT), an AI chatbot created by OpenAI that provides human-like responses to a wide range of text and image prompts across numerous knowledge domains, in late 2022.[3] A survey conducted by *Nature* found that 80% of scientific researchers who responded to a survey conducted by the journal have used generative AI tools such as ChatGPT to brainstorm and summarize literature, write and debug code, and create presentations and manuscripts.[4] Studies have described how ChatGPT could be used to train medical students,[5] provide clinical decision support,[6–8] summarize clinical guidelines,[9] facilitate data collection, particularly across languages,[9] and write scientific manuscripts.[10] To signify the new importance placed on generative AI within health care, the electronic health record (EHR) vendor, Epic, and Microsoft recently announced a partnership to incorporate generative AI into EHRs.[11] Although in its infancy, generative AI is already being tested in clinical practice to draft replies to patient portal messages, query clinical information within EHRs, and book patient appointments.[12]

Given the growing role of generative AI in health care, nurse leaders need to understand these tools and their risks and benefits. Nurses in leadership roles will need to determine whether and how to integrate generative AI tools into their work in a safe and useful way. PhD-prepared nurse scientists assume a variety of leadership roles after graduation across academia, health care, health policy, and the private sector, including as Chief Nursing Officer or Chief Nursing Informatics Officer.[13] Therefore, they are one of the groups who must become acculturated to generative AI. Statistics courses have long been included in many baccalaureate nursing programs[14] and more recently, leading nurse informaticists[15] and the American Association of Colleges of Nursing Essentials (Domain 8[16]) have advocated for computational thinking[17] and data science[15] competencies to be included in nursing education. There is a need to build on and update these foundational competencies in acknowledgment of the pivotal role generative AI is posed to play in health care research, education, and practice.

Currently, generative AI and its applications in health care are not covered in nursing PhD curricula, creating a knowledge gap. PhD-prepared nurses must have knowledge of generative AI to participate fully in decision-making and thought leadership around its integration into health care, particularly as difficult choices are weighed such as its relative benefits and costs. Given the novelty of ChatGPT and generative AI in health care, few educational interventions have been developed to prepare nurses for this new era.[18] Generative AI has been examined as a learning tool within nursing curricula, for example, to assist nursing students in creating evidence-based care plans during clinical simulations,[19] and the potential threats to academic integrity and other risks have been discussed.[10,20,21] Guidelines on the use of generative AI in nursing education and practice have been proposed[22,23] but there is a need for accompanying empirical research evaluating educational interventions teaching nurses about *generative AI itself* as a potential health care tool they may one day use or govern the use of, similar to the EHR. There remains much to learn about the appropriate and inappropriate use of generative AI tools, making exposure during the formal education process a key strategy for evidence generation.

Objective

This study aimed to develop and evaluate a brief intervention to increase PhD nursing students' knowledge of appropriate applications for using generative AI tools in health care.

Methods

Study Design

We created a brief educational intervention composed of didactic lectures and laboratory-based activities to introduce generative AI. The sample consisted of students enrolled in a required PhD-level nursing data science and visualization course; most students were PhD-level nursing students but the course is open to other graduate students across the university, and one student from another discipline audited the course. All students enrolled in the course were invited to participate but they were permitted to opt out of their reflections being used for this research study if they did not consent to participate.

Many PhD programs in nursing worldwide last 3 to 4 years[24]; in some cases, this is a compressed timeline that has been prioritized in part due to the nursing faculty shortage which is a major driver of nursing workforce shortages.[25,26] With many other competencies and dissertation research objectives to achieve, there is limited time in the typical nursing PhD program to integrate novel coursework, making a brief intervention a compelling solution to students' needs compared with a longer lecture series or semester-long course.

In this intervention, students were provided with a subscription to Chat GPT 4.0, a general-purpose generative AI tool, for use in and outside the class. This subscription cost $20 per month at the time of the study. The intervention consisted of didactic and laboratory-based activities. We conducted a mixed-methods evaluation of the intervention comprising quantitative competency self-assessments and qualitative reflections, which provided complementary insights into students' objective improvements in knowledge and experiences with ChatGPT. We defined competency as the "core abilities that are required for fulfilling one's role as a nurse scientist."[27] The Columbia University Institutional Review Board (IRB) approved this study.

Pedagogical Design

The educational intervention was guided by Bloom's Taxonomy of Learning, which describes a hierarchy of cognitive

 © 2024. Thieme. All rights reserved.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

domains associated with learning.[28] The intervention was designed to advance students' familiarity with generative AI beyond understanding to analyzing and evaluating it through laboratory-based activities, with the ultimate goal of supporting students who may apply this knowledge in leadership roles in their future work (►Fig. 1). The intervention was developed by the first author and course director, both nurses with nursing informatics expertise, a lecturer and teaching assistant with expert knowledge of ChatGPT and generative AI, and an associate director and learning designer employed by the Columbia Center for Teaching and Learning.[29]

We introduced the intervention into a PhD-level nursing course focused on research synthesis through data and information visualization, the graphical display of abstract information for sense-making and communication, and the application of data science and visualization tools and techniques. The course design involves both didactic lectures focusing on concepts and case studies describing how to scope and manage complex data science projects, as well as laboratory sessions focused on gaining competency with data science and visualization tools and techniques. It is a required course for all second-year PhD students at the Columbia University School of Nursing and most students are nurses; however, it is also open to other interested graduate students across the university.

To develop the brief intervention, we conducted an informal review of the literature on generative AI broadly and its uses in health care and research in the Summer of 2023. The main applications of generative AI in health care were to perform administrative functions (such as processing medical claims and creating medical records), enhance the interpretation of clinical data, provide clinical decision support, develop patient-facing education, and deliver personalized patient decision support.[5–9] The applications in research were to assist scientific writing, efficiently generate code to analyze datasets, advance drug discovery, and development, conduct literature reviews, and write lay and graphical abstracts.[4,19] The major risks were ethical and legal issues including plagiarism, copyright, transparency, risks of bias including the risks of "infodemics" biasing results, risks of hallucination (i.e., the phenomenon of models providing fabricated or misleading outputs), models having limited and incorrect knowledge, and compromised cybersecurity.[10,30]

Based on these findings, we identified two new competencies relevant to generative AI that the students would need: (1) describe generative AI applications in health care and (2) identify appropriate and inappropriate applications of generative AI in health research. These competencies complement existing data science and visualization competencies and formed the basis of the didactic lectures and laboratory-based activities used in the intervention. The first author drafted the initial concept for these competencies and intervention components and iteratively refined them through discussions with the study team.

Generative Artificial Intelligence Intervention

The brief generative AI intervention included two, face-to-face didactic lectures and laboratory-based activities delivered during one, 3-hour class. During the didactic portion, we described generative AI and its current and potential future applications in health care, including examples of appropriate and inappropriate applications. The lectures lasted for approximately 75 minutes and included a guest lecture from a researcher with expert knowledge of ChatGPT

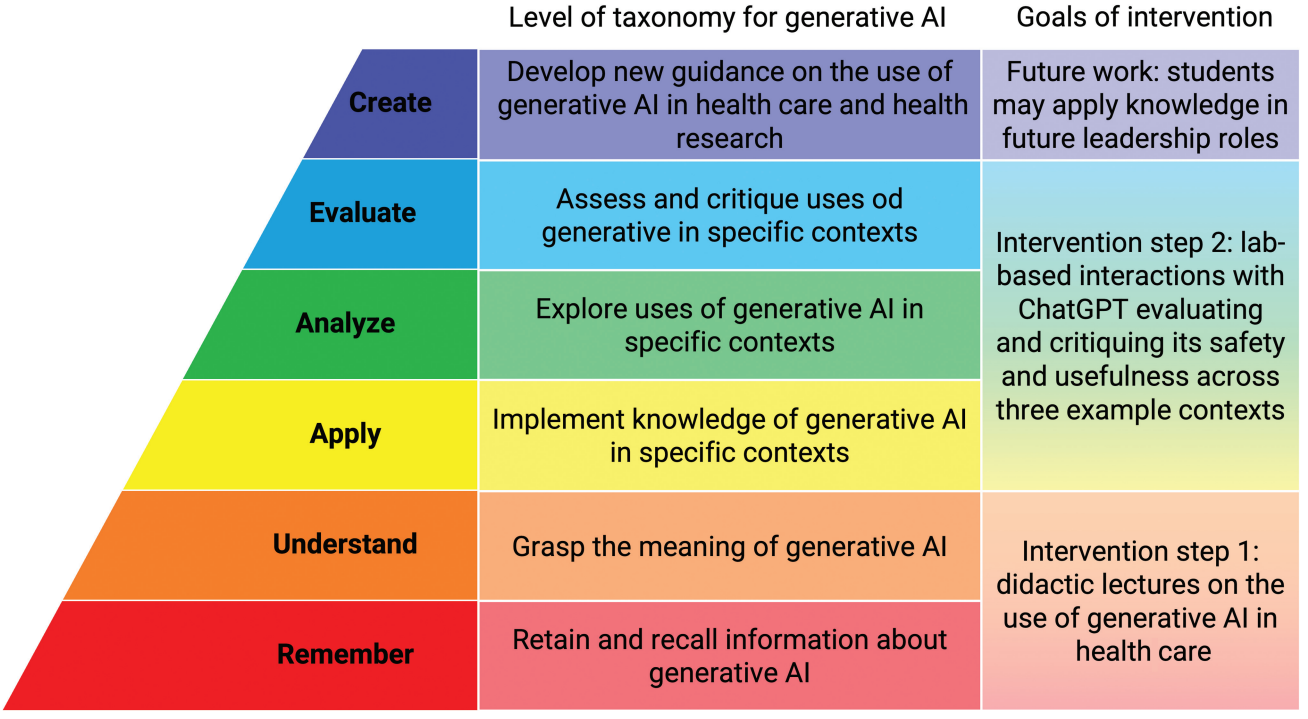

Fig. 1 Application of Bloom's Taxonomy of Learning to the cognitive tasks and components of this brief intervention on generative artificial intelligence. AI, artificial intelligence; ChatGPT, Chat Generative Pretrained Transformer.

 © 2024. Thieme. All rights reserved.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

who has published on its applications in health care,[9,31] and a supplementary mini-lecture reviewing the main applications of generative AI in health care and research and the major risks. The mini-lecture was intended to directly prepare students for the subsequent laboratory-based activities.

In the laboratory sessions, students were given three tasks representing different use cases of generative AI in health care practice and research and asked to engage with ChatGPT on each. The three tasks were created to represent the spectrum of applications of ChatGPT in health care and research discussed at the time of intervention development (mid-2023) in the literature, and to illustrate the major risks; these included clinical decision support, patient empowerment and decision support, and scientific communication to lay audiences (►Table 1). For example, we specifically chose a coronavirus disease 2019 (COVID-19) vaccine use case for patient decision support in an attempt to elicit misinformation that ChatGPT may have learned from certain Internet sources. The goal of the laboratory-based activities was to stimulate students to form their own opinions of ChatGPT's safety and usability. Students were introduced to the assignment during the remainder of the class and had approximately 75 minutes to work independently on the tasks. The professors and teaching assistants provided clarification on the tasks and technical support if needed, but students were encouraged to independently design prompts and interact with ChatGPT.

Table 1 ChatGPT tasks used during laboratory-based activities

Use case	Risks illustrated	Prompt instructions	Reflection instructions
Clinical decision support	Inaccurate information, hallucinations	Imagine that you are a new graduate nurse practitioner (NP) in a family medicine practice. You see a patient with a blood pressure of 139/78 and a history of cardiovascular disease. You want to know whether you should start an antihypertensive agent. Give ChatGPT your patient's information and ask for clinical guidance. Ask follow-up questions that you imagine an NP would need to know to treat the patient. Ask for citations	Write 150 to 200 words comparing the answers from ChatGPT with guidelines published online.[46] Was ChatGPT accurate compared with these guidelines? Was it easier or harder to understand what you should do? Were the citations provided real and relevant to the question? If you were a Chief Nursing Officer or Chief Nursing Informatics Officer, would you be comfortable with your clinicians using ChatGPT for clinical decision support?
Patient decision support	Infodemics, hallucinations	Imagine that you are a patient with the following medical history: a 67-year-old female with a history of Type 2 Diabetes and heart failure. Your physician advises you that you should be vaccinated against COVID-19, but you are unsure if it is a good idea. You have heard on Facebook that the vaccine causes heart problems and even death. You decide to ask ChatGPT what to do. Use ChatGPT as a patient to ask questions using nonmedical terminology and follow-up with questions you imagine patients asking about the vaccine. Ask for citations	Write 150 to 200 words considering ChatGPT's answers in terms of accuracy and misinformation—is any of the information false or misleading? Is a clear answer provided? Were the citations provided real and relevant to the question? Would you recommend that patients use ChatGPT for these types of use cases?
Scientific communication	Ethical and legal issues, limited and incorrect knowledge	Imagine that you are a researcher who is submitting a research paper to a journal. The journal asks you to write a lay abstract. You decide to ask ChatGPT to write a lay abstract based on the scientific abstract that you have already written. Instructions: Find an abstract in PubMed that describes a research study that you are interested in (and ideally, in an area of research you are already familiar with). Copy and paste the abstract into ChatGPT and ask for a lay abstract. Refine it as needed based on your assessment of the quality and clarity of the abstract	Write 150 to 200 words analyzing how ChatGPT performed. How long did it take ChatGPT to produce a lay abstract that you were satisfied with? Do you think the average layperson (without scientific or medical knowledge) would be able to understand this abstract? Where do you think this could be useful in research applications?

Abbreviations: ChatGPT, Chat Generative Pretrained Transformer; COVID-19, coronavirus disease 2019.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

Data Collection and Analysis

We collected both quantitative surveys and qualitative written reflections to evaluate the intervention, which students submitted electronically through the Courseworks site for the course. We did not collect or report the demographic characteristics of the students due to the small sample size together with reporting of the specific school and program, which could facilitate unintentional re-identification and privacy violations. The quantitative data consisted of competency self-assessments in generative AI at baseline (prior to the course beginning) and mid-semester (after the intervention had been delivered). The two competencies were assessed: (1) describe generative AI applications in health care and (2) identify appropriate and inappropriate applications of generative AI in health research. On each competency, students self-rated themselves on a five-point scale as novice, advanced beginner, competent, proficient, or expert.

Qualitative data consisted of written reflections on the ChatGPT tasks based on the guidance provided in ►Table 1. The reflections were intended to elicit students' perceptions of ChatGPT's safety (including the accuracy of information presented and hallucinations) and usability (including its helpfulness and real-world applicability). Students also submitted verbatim transcripts of their interactions with ChatGPT to support triangulation with their reflections.[32] The reflections were analyzed using directed content analysis, a method in which a predefined set of concepts guides the initial coding while still allowing new themes to emerge.[33] Following this approach, the first author coded responses according to the task and concept (safety or usability), creating subthemes within these themes as needed, and referencing the transcripts of ChatGPT interactions as needed for further interpretation. A second author reviewed all codes, confirmed interpretations, and provided suggested alternative interpretations in certain cases. Disagreements in coding were discussed and resolved until a final set of themes and illustrative quotes were complete.

Results

Ten students were enrolled in the course and all consented to participate in the research study.

Pre–Post Comparisons of Generative Artificial Intelligence Competencies

The pre–post comparisons of self-reported competencies are illustrated in ►Fig. 2. One student who was auditing the course did not complete the competency self-assessments. For the remaining nine students, on competency 1 (e.g., describe generative AI applications in health care), the majority ($n = 5$ of 9) self-identified as a "novice" at baseline. Postintervention, none self-identified as "novice" and the majority self-identified as "competent" ($n = 5$ of 9). On competency 2 (e.g., identify appropriate and inappropriate applications of generative AI in health), the majority of students ($n = 7$ of 9) self-identified as a "novice" at baseline. Postintervention, none self-identified as "novice" and the majority self-identified as "competent" ($n = 5$ of 9).

Qualitative Findings

All 10 students completed the qualitative reflections. The major findings from the reflections on each task are summarized in ►Table 2. Below we describe the findings in more detail under each task.

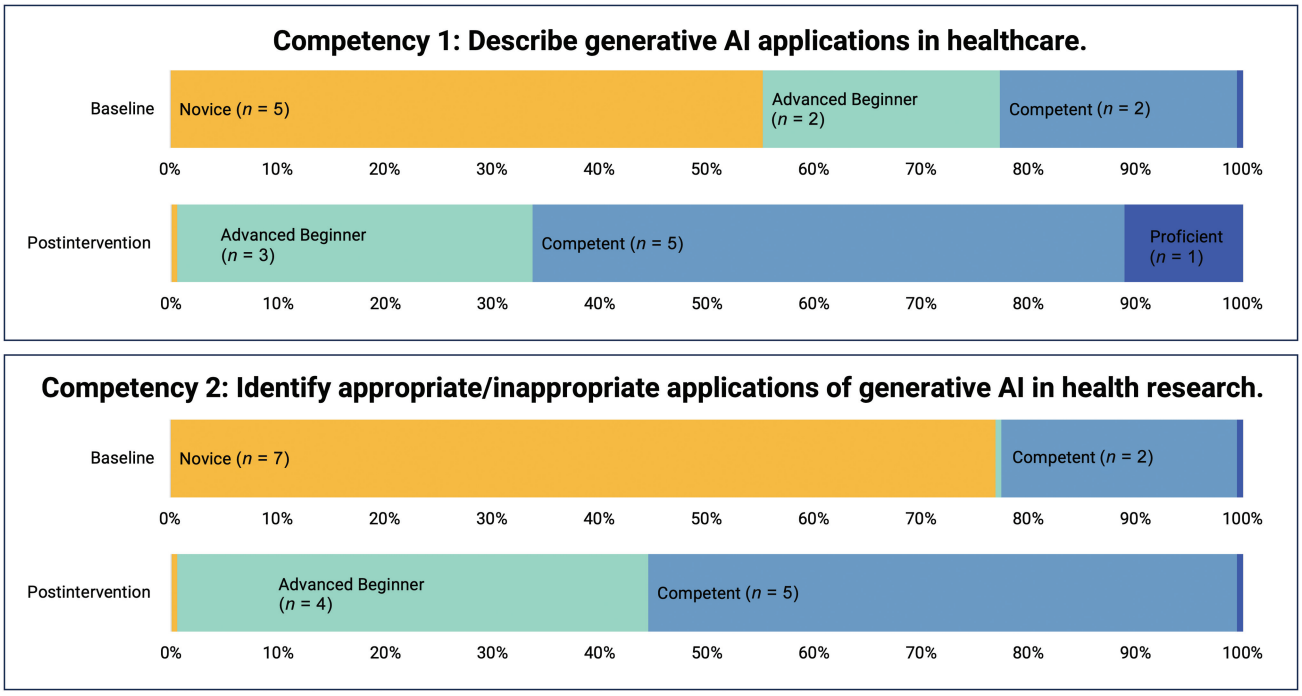

Fig. 2 Self-reported competency in generative AI among students ($n = 9$). AI, artificial intelligence.

 © 2024. Thieme. All rights reserved.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

Table 2 ChatGPT tasks and qualitative findings from student reflections

Use case	Reflections on safety (accuracy, hallucinations)	Reflections on usability (helpfulness, real-world applicability)
Clinical decision support	ChatGPT was accurate and concordant with guidelines. Some citations were incorrect or fabricated and were only provided when prompted	Recommendations were clear but lacked nuance. Most agreed this should be a complementary, but not primary or sole, tool for clinical decision support
Patient decision support	The information was accurate. Citations were not always provided even when prompted. ChatGPT encouraged caution regarding information from social media	Answers included medical jargon and required a high literacy level. Perceived challenges for low literacy patients limited intention to use or recommend use in the future
Scientific communication	Experiences varied widely with respect to the number of iterations needed to arrive at a satisfactory lay abstract, from one to six	There was a tradeoff between simplicity and inclusion of important details of the study. Most agreed this was useful with researcher oversight

Abbreviation: ChatGPT, Chat Generative Pretrained Transformer.

Clinical Decision Support

Regarding safety, ChatGPT was accurate, concordant with guidelines, did not claim to be an expert, and encouraged consultation of primary sources. One student explained: "*ChatGPT gave clear step-by-step instructions on how to treat hypertension. It was quite easy for me to understand what to do. ChatGPT gave guidance rather than direct advice.*" However, others noted how continued prompting caused ChatGPT to become less concordant with guideline-directed care:

> "[ChatGPT] was able to diagnose my patient with Stage 1 hypertension (HTN) which aligned with the ACC/AHA 2018 Guidelines… and recommended use of an antihypertensive agent alongside lifestyle modifications. When probing the tool about taking more time to assess (e.g. at-home blood pressure monitoring to rule out White Coat hypertension) or using lifestyle modifications only, it pushed for including medications as well. However, the guidelines state that adults with stage 1 hypertension whose estimated 10-year risk of atherosclerotic CVD is less than 10% should be treated with nonpharmacologic interventions and reevaluated in three to six months."

Similarly, one student also noticed that ChatGPT reinforced preferences for specific treatments that the student had expressed in prompts, regardless of the guidelines. In some cases, ChatGPT also fabricated citations or reported them incorrectly. It occasionally deceived the user by providing a combination of accurate and inaccurate citations: "*The citations for the blood pressure information were accurate, but regarding language barriers the hyperlinks were wrong; the literature review on teach-back methods does not exist.*"

Regarding usability, ChatGPT's recommendations were clear and concise but also very general and lacked enough nuance to fully guide a clinician. Some students noted that they needed to prompt ChatGPT to provide adequate detail: "*The first recommendations included a description of antihypertensives and lifestyle modifications. I had to provide a more detailed clinical picture to get direct recommendations, thus I reported my patient also had asthma and angina. This led to the recommendation to place them on an ARB [Angiotensin receptor blocker].*" Another agreed that prompt engineering experience may be needed to elicit guideline recommendations fully: "*ChatGPT was helpful but overall, I still preferred to read the guidelines because it provided recommendations unprompted (e.g. when the patient should follow-up) whereas ChatGPT would only answer what I had specifically asked.*"

Comfort with using ChatGPT in real-world contexts varied but most agreed it would be a helpful complementary, but not a primary or sole, tool for clinical decision support. One student regarded it as "*a starting point to know what kinds of questions to ask their patients and to look for*" but "*not a decision maker.*" Another noted that busy clinicians may fail to verify ChatGPT-provided information:

> "If I were a Chief Nursing Officer/Chief Nursing Informatics Officer, I would be comfortable with clinicians using it for clinical decision support, but not as the only resource that is consulted. However, given the reality of clinical settings being very busy, it may be difficult to expect clinicians to match multiple sources to verify the information. Therefore, I think at its current stage, it should be used in settings that have more time to properly engage with it."

Some noted the potential for another generative AI tool that was trained specifically for clinical decision support and capable of handling protected health information to be more useful: "*I don't think that the responses were specific enough to really help with a clinical decision, however, it is possible that if I had given more clinical information about the patient it would have been more helpful.*"

Patient Decision Support

Regarding safety, students agreed that the information provided by ChatGPT was accurate with no misleading or false information. One student stated that it was able to answer, "*questions about who needs to get the COVID-19 vaccine, whether it's safe in my specific context, what are the side-effects, and then a series of questions specific to my history of*

 © 2024. Thieme. All rights reserved.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

heart failure and the vaccine's potential for heart inflammation." Even when students attempted to elicit misinformation, ChatGPT provided careful and scientific answers. However, they noticed that, due to the extreme caution it exercised, the responses were too vague: "*While the information did not seem misleading or false, ChatGPT was very cautious… I felt as though it was not a clear answer.*" Citations were not always provided, even when they were requested. ChatGPT encouraged caution regarding some sources of information like social media. One student summarized:

> "I introduced some conspiracy theories… The answers acknowledged my concerns while pointing out that I should use credible sources, and rigorous peer-reviewed journals, and that there is a consensus among scientists and public health professionals about the safety and effectiveness of vaccines. It provided citations (after I insisted that I wanted references) relevant to the questions. One of the citations had an author's name that was not included in the actual article."

Regarding usability, all students noticed that ChatGPT's answers included medical jargon and required a high literacy level. For example, one student explained:

> "To imitate the patient's perspective, I used words such as 'heart problems' as a complication of the COVID-19 vaccine. However, ChatGPT named these as myocarditis and pericarditis without much explanation as to what these conditions are…it may be hard for someone to use this program without having the medical background needed to interpret its responses."

One student also noted how this could be detrimental to a person who was still deciding whether to receive a vaccine: "*It provided a lot of information very quickly, and I wonder whether someone would take the time to read all this text if they were feeling anxious about the vaccine.*" The lack of direct advice on whether or not to be vaccinated was perceived as unhelpful by some students but favored by others: "*I view it as a positive thing as it still leaves room for human thought and engagement in decision-making and therefore does not impact autonomy.*" The use of medical jargon, which many felt created challenges for low health literacy patients, limited students' perceptions of ChatGPT's real-world applicability for patient decision-support. This was compounded by the knowledge that its training dataset was outdated, as one student explained:

> "ChatGPT provides recommendations from information obtained online prior to September 2021. While it appears ChatGPT has been able to mostly filter 'fake news' out of its recommendations, it still could not provide advice based on up-to-date studies. Similarly, any new information on the COVID-19 vaccine or subsequent boosters will not be reflected in ChatGPT's answers. Therefore, these conversations should stay between the clinician and the patient at this time."

Public Scientific Communication

The number of iterations needed to arrive at a lay abstract perceived as satisfactory to the student varied widely from one to six or more. This was inherently tied to usability because there was a tradeoff between simplicity enabling comprehension at a sixth-grade reading level, and inclusion of relevant and important details of the study; ChatGPT could not achieve both simultaneously. Rather, many reported that abstracts were either overly complex and included medical jargon, or oversimplified and missing critical information:

> "The simplification led to the omission of key details about the study [that] misrepresented the study's findings. For instance, the research focused on the smoking habits among mothers with opioid addiction during pregnancy and postpartum… ChatGPT left out the crucial detail that these mothers were struggling with opioid use disorders while attempting to quit smoking."

Students noted that both the clarity of their prompts and the amount of medical jargon in the abstract could affect the performance. For example, one student said:

> "I intentionally picked a study that used natural language processing (NLP) methods, as this is something I would consider harder for a layperson to understand… In the first draft, NLP was completely removed and instead called a 'special tool', and stigmatizing language was referred to as 'negative words'. I felt neither of these terms adequately described the methods of the study… To overcome this, I specifically asked ChatGPT to describe what stigmatizing language is and to also be more specific about NLP methods. After six rounds of revisions, I felt ChatGPT was able to produce an accurate abstract that the average layperson should be able to understand."

All students were able to generate a lay abstract that they were satisfied with, and noted time-saving benefits even if multiple rounds of iteration were required: "*The process took me less than 10 minutes whereas it would have taken me much longer if doing this on my own.*" Overall, ChatGPT was considered to be a useful tool for generating lay abstracts and other scientific materials for lay audiences (such as informed consent documents). One student reportedly verified that the information was easily understood by those without a health care background:

> "I work in the oncology subspecialty of bone marrow and stem cell transplantation. Describing the process of cell infusion and engraftment can be very abstract, even for clinicians… I found ChatGPT to be very helpful when writing this for the general population to understand. I

 © 2024. Thieme. All rights reserved.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

even shared it with my spouse and parents – all of them were able to understand the abstract."

Discussion

The brief intervention increased PhD nursing students' knowledge of appropriate applications for using generative AI tools in health care. Students were able to identify the strengths and limitations of ChatGPT in completing all three tasks and developed opinions on whether they would feel comfortable using ChatGPT for similar tasks in the future. In particular, students felt ChatGPT was useful in all three use cases but identified important guardrails they would seek to enforce if choosing to implement it in health care practice or research in their future roles as nurse leaders and scholars. All of them also reported increasing their self-rated competency in generative AI by one to two points. This suggests that the brief intervention involving both didactic lectures and interactive work with ChatGPT was successful in supporting the goal of acclimating PhD nursing students to generative AI.

There is increasing recognition that nurses must become acculturated to AI.[34] This builds on the recognition that big data and data science are of high relevance to nursing science and practice.[35] Much has been written about how to leverage AI as a teaching aid, for example, to teach nurses how to document the nursing process[36] and create patient backstories for nursing simulation courses.[37] Specific to generative AI, O'Connor and colleagues discuss integrating prompt engineering into nursing coursework as an educational tool.[19] Less has been written about how to equip nurses to use AI in their future work.[38] *AI literacy*—in which nurses can understand and effectively use AI in their work[39]—may be a useful construct to guide the refinement of coursework on AI in future work.

Generative AI poses unique challenges for education because the field is rapidly advancing, and models themselves are continually becoming more intelligent. Russell and colleagues proposed six competencies for the use of AI among all clinicians[40] that can serve as a useful foundation for developing coursework about generative AI. For example, two competencies are the ability to appraise quality and safety and to understand ethical considerations.[40] In an educational context, creating laboratory-based activities that will predictably illustrate specific quality and safety risks and ethical issues is a challenge because ChatGPT might generate different responses to a series of prompts.[19] Nonetheless, most students demonstrated in their reflections that they grasped the high-level ethical and social challenges and potential safety risks despite ChatGPT's somewhat varied responses.

Providing more training on prompt engineering may facilitate more predictable interactions,[31] although educators should caution against providing too much guidance on prompting to avoid undermining the learning that organically occurs by interacting creatively with ChatGPT. The Problem, AI, Interaction, Reflection framework[41] proposed by O'Connor and colleagues[19] may guide prompt engineering education. It specifies four steps: (1) formulate a clear problem within a specific context, (2) explore different AI tools, (3) experiment with prompts and critically evaluate AI output, and (4) reflect on using generative AI tools and the impacts of their outputs.[19] This framework aligns with the approach taken in our brief intervention and may be useful in updating or creating new educational content related to generative AI in future work.

Additionally, another AI competency proposed by Russell and colleagues is continued learning and training in AI.[40] The fast pace at which the field of generative AI is moving makes creating educational content that is up-to-date a challenge. For example, while we were developing the intervention in the several months leading up to the course, the literature on applications of ChatGPT in health care and research and the associated risks grew on an almost daily basis. Furthermore, the body of research on the safety mechanisms and interpretability techniques to align generative AI models is also growing rapidly. Generative AI is, by nature, more challenging to align and interpret than supervised machine learning models for which interpretability techniques, such as SHapley Additive exPlanations,[42] have been developed. Rather, generative AI models may behave in unpredictable or even deceptive ways—this problem will become increasingly pronounced as models become increasingly intelligent.[43] OpenAI, who created ChatGPT, has publicly shared its alignment efforts in an effort to invite public discourse and scientific collaboration on this problem.[44] Future work should explore structures that can support continued professional learning on generative AI as it evolves. This should include interpretability and alignment techniques, such as the "AI Lie Detector" which uses representation engineering to represent the way generative AI systems work.[45]

This intervention can easily be adapted in other doctoral nursing and informatics programs, as well as other types of nursing programs, as the need for nurses and others to acculturate to generative AI grows. The task descriptions and reflection instructions, provided in ►Table 1, provide a basis for others to update and adapt to their particular student needs. In future courses, we plan to continue offering the generative AI module described here, updating it as needed based on emerging use cases in health care and research. Additionally, we plan to provide some additional guidance on prompt engineering for students in both the didactic portion and the instructions for the laboratory-based activities.

One limitation of this study is the small sample size. While PhD nursing courses typically have small class sizes, this limited our ability to conduct inferential statistics. In the future, larger cohorts should be recruited, ideally from multiple programs to increase generalizability and enable hypothesis testing. Additionally, we used ChatGPT in the exercises in this intervention. ChatGPT is a general-purpose tool and differs from generative AI tools developed for specific clinical and research tasks, especially those used in health care settings. Finally, the need to prepare materials for the intervention in advance of obtaining IRB approval and beginning the course precluded us from including more updated literature on use cases in health care as it emerged.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

Conclusion

As applications of AI in health care and research continue to expand, AI competencies will be essential for PhD-prepared nurses. This is particularly true for generative AI, which should be considered a core component of AI literacy but also has unique learning challenges and competencies. In this study, we found that a brief intervention in a PhD nursing program increased competency scores and facilitated the appraisal of generative AI tools, such as ChatGPT. This intervention can easily be adapted in other doctoral nursing and informatics programs to prepare students encountering generative AI in their future work.

Clinical Relevance Statement

There is increasing recognition that nurses who are providing direct patient care and functioning in leadership positions must become acculturated to AI. There is limited understanding about how to equip nurses to use AI in their future work. This brief educational intervention provided nurses with generative AI competency and may be adapted to other nursing or informatics coursework.

Multiple-Choice Questions

1. What is one risk of generative AI when used for clinical decision support?
   a. Accuracy
   b. Claims of being an expert
   c. Hallucinations
   d. Lack of concordance with published recommendations

   Correct Answer: The correct answer is option c. Hallucinations are a known risk of generative AI models, especially ChatGPT. In this study, ChatGPT occasionally fabricated citations or reported them incorrectly. In many cases, it deceived the user by providing a combination of accurate and inaccurate citations.

2. Which of the following is not a relevant component of AI literacy for nurses?
   a. The ability to appraise quality and safety
   b. Understanding ethical considerations
   c. Continued learning and education
   d. Programming generative AI models

   Correct Answer: The correct answer is option d. Nurses may use AI in their future work but will not be programming AI models. The other aspects of AI literacy are relevant, however, to be able to use AI effectively and safely in patient care, nursing leadership, and research.

Protection of Human and Animal Subjects

The Columbia University IRB approved this study.

Funding

This project was funded through a grant from the Columbia Center for Teaching and Learning. M.R.T. is also funded by National Institute of Nursing Research (NINR) of the National Institutes of Health (NIH). (grant no.: R00NR019124).

Conflict of Interest

M.R.T. has the following conflicts: Boston Scientific (consulting), and Iris OB Health (co-founder, equity).

References

1 Jo A. What ChatGPT and generative AI mean for science. Nature 2023;614(1). Accessed July 30, 2024 at: https://www.nature.com/articles/d41586-023-00340-6
2 Carroll WM. Generative AI in clinical practice and operations. Nurs Manage 2023;54(10):56
3 Open AI. Introducing ChatGPT. 2022. Accessed April 6, 2023 at: https://openai.com/blog/chatgpt
4 Owens B. How Nature readers are using ChatGPT. Nature 2023; 615(7950):20
5 Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ 2023;9(01):e46885
6 Liu S, Wright AP, Patterson BL, et al. Assessing the value of ChatGPT for clinical decision support optimization. medRxiv (e-pub ahead of print). Published online February 23, 2023 Doi: 10.1101/2023.02.21.23286254
7 Rao A, Kim J, Kamineni M, Pang M, Lie W, Prasad AK, Landman A, Dreyer K, Succi MD. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. J Med Internet Res 2023;25:e48659
8 Rao A, Kim J, Kamineni M, Pang M, Lie W, Dreyer KJ, Succi MD. Evaluating GPT as an Adjunct for Radiologic Decision Making: GPT-4 versus GPT-3.5 in a breast imaging pilot. J Am Coll Radiol 2023;20(10):990–997
9 Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. Eur J Cardiovasc Nurs 2023;22(07):e55–e59
10 O'Connor S. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? Nurse Educ Pract 2023;66:103537
11 Eddy N. Epic, Microsoft partner to use generative AI for better EHRs. Healthcare IT News Published April 18, 2023. Accessed April 19, 2023 at: https://www.healthcareitnews.com/news/epic-microsoft-partner-use-generative-ai-better-ehrs?mkt_tok=NDIwLVlOQS0yOTIAAAGLNZIswFNUMzJqPehB6oRzDwlL–b5VSFOyFTrEsy9JOHwqG1T2_zyPjj1MY51LgwY7D73j-X8bvMI2itbo5fy9hRwFH8BuAKLCeYQ_z26kg
12 Raza MM, Venkatesh KP, Kvedar JC. Generative AI and large language models in health care: pathways to implementation. NPJ Digit Med 2024;7(01):62
13 Morris NS, Wassef ME, Sullivan-Bolyai S, Bova C, Kane AT. Making explicit the development of PhD-prepared nurses to steward the discipline. Nurs Outlook 2021;69(01):50–56
14 Grace JT, D'Aoust R. Evidence-based program requirements: evaluation of statistics as a required course. Nurs Educ Perspect 2006;27(01):28–33
15 Brennan PF, Bakken S. Nursing needs big data and big data needs nursing. J Nurs Scholarsh 2015;47(05):477–484
16 American Association of Colleges of Nursing. The Essentials: Core Competencies for Professional Nursing Education. 2021
17 Wing JM. Computational thinking. Commun ACM 2006;49(03): 33–35
18 Liu J, Liu F, Fang J, Liu S. The application of Chat Generative Pre-trained Transformer in nursing education. Nurs Outlook 2023;71 (06):102064
19 O'Connor S, Peltonen LM, Topaz M, et al. Prompt engineering when using generative AI in nursing education. Nurse Educ Pract 2024;74:103825

 © 2024. Thieme. All rights reserved.

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.

20 Topaz M, Peltonen LM, Michalowski M, et al. The ChatGPT Effect: nursing education and generative artificial intelligence. J Nurs Educ 2024:1–4
21 Dwivedi YK, Kshetri N, Hughes L, et al. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy Int J Inf Manage 2023;71:102642
22 O'Connor S, Leonowicz E, Allen B, Denis-Lalonde D. Artificial intelligence in nursing education 1: strengths and weaknesses. Nursing Times September 11, 2023. Accessed July 30, 2024 at: https://www.nursingtimes.net/roles/nurse-educators/artificial-intelligence-in-nursing-education-1-strengths-and-weaknesses-11-09-2023/
23 O'Connor S, Permana AF, Neville S, Denis-Lalonde D. Artificial intelligence in nursing education 2: opportunities and threats. Nursing Times October 9, 2023. Accessed June 4, 2024 at: https://www.nursingtimes.net/roles/nurse-educators/artificial-intelligence-in-nursing-education-2-opportunities-and-threats-09-10-2023/
24 Nurse.org. Top PhD in nursing programs. 2024. Accessed February 20, 2024 at: https://nurse.org/education/phd-in-nursing-degree/
25 Bakewell-Sachs R, Trautman D, Rosseter R. Addressing the nurse faculty shortage. American Nurse August 3, 2022. Accessed February 20, 2024 at: https://www.myamericannurse.com/addressing-the-nurse-faculty-shortage-2/
26 American Association of Colleges of Nursing. Nursing Faculty Shortage Fact Sheet. Accessed February 20, 2024 at: https://www.aacnnursing.org/news-data/fact-sheets/nursing-faculty-shortage
27 Mrayyan MT, Abunab HY, Abu Khait A, et al. Competency in nursing practice: a concept analysis. BMJ Open 2023;13(06):e067352
28 Bloom BS. Taxonomy of Educational Objectives: The Classification of Educational Goals. Longmans, Green & Co.; 1956
29 Columbia Center for Teaching and Learning at Columbia University. Accessed February 20, 2024 at: https://ctl.columbia.edu/
30 Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11(06):887
31 Durieux B, Davis J, Moons P, Van Bulck L. How to get the most out of ChatGPT? Tips and tricks on prompting. Eur J Cardiovasc Nurs 2024:zvae016
32 Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999;34(5 Pt 2):1189–1208
33 Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(09):1277–1288
34 Ronquillo CE, Peltonen LM, Pruinelli L, et al. Artificial intelligence in nursing: Priorities and opportunities from an international invitational think-tank of the Nursing and Artificial Intelligence Leadership Collaborative. J Adv Nurs 2021;77(09):3707–3717
35 Douthit BJ, Walden RL, Cato K, et al. Data science trends relevant to nursing practice: a rapid review of the 2020 literature. Appl Clin Inform 2022;13(01):161–179
36 Gosak L, Pruinelli L, Topaz M, Štiglic G. The ChatGPT effect and transforming nursing education with generative AI: discussion paper. Nurse Educ Pract 2024;75:103888
37 Reed JM, Dodson TM. Generative AI backstories for simulation preparation. Nurse Educ 2024;49(04):184–188
38 De Gagne JC. The state of artificial intelligence in nursing education: past, present, and future directions. Int J Environ Res Public Health 2023;20(06):4884
39 Long D, Magerko B. What is AI Literacy? Competencies and Design Considerations. Paper presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. CHI '20. Association for Computing Machinery; 2020:1–16
40 Russell RG, Lovett Novak L, Patel M, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med 2023;98(03):348–356
41 Acar OA. Are Your Students Ready for AI?: A Four-Step Framework to Prepare Learners for a ChatGPT World Harvard Business School Publishing; 2023. Accessed February 21, 2024 at: https://kclpure.kcl.ac.uk/portal/en/publications/are-your-students-ready-for-ai-a-four-step-framework-to-prepare-l
42 Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Accessed February 21, 2024 at: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
43 Open AI. Governance of superintelligence. 2023. Accessed February 21, 2024 at: https://openai.com/blog/governance-of-superintelligence
44 Open AI. Our approach to alignment research. 2022. Accessed February 21, 2024 at: https://openai.com/blog/our-approach-to-alignment-research
45 Zou A, Phan L, Chen S, et al. Representation engineering: A top-down approach to AI transparency. arXiv [csLG]. Published online October 2, 2023. Accessed February 21, 2024 at: https://www.ai-transparency.org/
46 Armstrong C. High blood pressure: ACC/AHA Releases Updated Guideline. Am Fam Physician 2018;97(06):413–415

Downloaded by: University of Pennsylvania Libraries. Copyrighted material.