

Original Paper

Health Care Trainees' and Professionals' Perceptions of ChatGPT in Improving Medical Knowledge Training: Rapid Survey Study

Je-Ming Hu^{1,2,3}, MD, PhD; Feng-Cheng Liu⁴, MD, PhD; Chi-Ming Chu^{2,5,6,7,8,9}, PhD; Yu-Tien Chang⁵, PhD

¹Division of Colorectal Surgery, Department of Surgery, Tri-service General Hospital, National Defense Medical Center, Taipei, Taiwan

²Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan

³School of Medicine, National Defense Medical Center, Taipei, Taiwan

⁴Division of Rheumatology/Immunology and Allergy, Department of Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

⁵School of Public Health, National Defense Medical Center, Taipei, Taiwan

⁶Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan

⁷Big Data Research Center, College of Medicine, Fu-Jen Catholic University, New Taipei City, Taiwan

⁸Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan

⁹Department of Public Health, China Medical University, Taichung, Taiwan

Corresponding Author:

Yu-Tien Chang, PhD

School of Public Health

National Defense Medical Center

No 161, Sec 6

Minquan E Rd, Neihu Dist

Taipei, 114

Taiwan

Phone: 886 2 8792 3100 ext 18454

Fax: 886 2 87923147

Email: greengarden720925@gmail.com

Abstract

Background: ChatGPT is a powerful pretrained large language model. It has both demonstrated potential and raised concerns related to knowledge translation and knowledge transfer. To apply and improve knowledge transfer in the real world, it is essential to assess the perceptions and acceptance of the users of ChatGPT-assisted training.

Objective: We aimed to investigate the perceptions of health care trainees and professionals on ChatGPT-assisted training, using biomedical informatics as an example.

Methods: We used purposeful sampling to include all health care undergraduate trainees and graduate professionals (n=195) from January to May 2023 in the School of Public Health at the National Defense Medical Center in Taiwan. Subjects were asked to watch a 2-minute video introducing 5 scenarios about ChatGPT-assisted training in biomedical informatics and then answer a self-designed online (web- and mobile-based) questionnaire according to the Kirkpatrick model. The survey responses were used to develop 4 constructs: "perceived knowledge acquisition," "perceived training motivation," "perceived training satisfaction," and "perceived training effectiveness." The study used structural equation modeling (SEM) to evaluate and test the structural model and hypotheses.

Results: The online questionnaire response rate was 152 of 195 (78%); 88 of 152 participants (58%) were undergraduate trainees and 90 of 152 participants (59%) were women. The ages ranged from 18 to 53 years (mean 23.3, SD 6.0 years). There was no statistical difference in perceptions of training evaluation between men and women. Most participants were enthusiastic about the ChatGPT-assisted training, while the graduate professionals were more enthusiastic than undergraduate trainees. Nevertheless, some concerns were raised about potential cheating on training assessment. The average scores for knowledge acquisition, training motivation, training satisfaction, and training effectiveness were 3.84 (SD 0.80), 3.76 (SD 0.93), 3.75 (SD 0.87), and 3.72 (SD 0.91), respectively (Likert scale 1-5: strongly disagree to strongly agree). Knowledge acquisition had the highest score and training effectiveness the lowest. In the SEM results, training effectiveness was influenced predominantly by knowledge acquisition and

partially met the hypotheses in the research framework. Knowledge acquisition had a direct effect on training effectiveness, training satisfaction, and training motivation, with β coefficients of .80, .87, and .97, respectively (all $P < .001$).

Conclusions: Most health care trainees and professionals perceived ChatGPT-assisted training as an aid in knowledge transfer. However, to improve training effectiveness, it should be combined with empirical experts for proper guidance and dual interaction. In a future study, we recommend using a larger sample size for evaluation of internet-connected large language models in medical knowledge transfer.

(*J Med Internet Res* 2023;25:e49385) doi: [10.2196/49385](https://doi.org/10.2196/49385)

KEYWORDS

ChatGPT; large language model; medicine; perception evaluation; internet survey; structural equation modeling; SEM

Introduction

ChatGPT is a powerful, cutting-edge large language model (LLM) launched on Nov 30, 2022, developed by OpenAI, an artificial intelligence (AI) research and deployment company. It is a variant of the GPT-3 model, designed explicitly for chatbot and conversational AI applications [1]. It showcased an impressive advancement in the ability to understand and generate human-like text, opening up new possibilities for language models, and has expanded their capability to perform a wide range of tasks. It reflected a significant step forward in language model technology. Its ability to understand and respond naturally to language inputs makes it useful for various tasks, especially in higher education [2].

ChatGPT was trained on diverse online texts, including books, articles, and websites, on topics ranging from news to Wikipedia to fiction, and it received reinforcement learning from human feedback [3,4]. It attempts to incorporate human-like responses during human-to-human conversations. As a result, humans can fine-tune it to become more accurate and capable of understanding details and narrowly pitched questions [5]. It has been used for a variety of natural language processing tasks, such as language translation, text summarization, and answering questions.

ChatGPT quickly gained popularity for its detailed responses and articulate answers across multiple domains of knowledge [6]. It was the first time a powerful chatbot was made freely available, and it is user-friendly for the general public [3]. The release of ChatGPT made enormous impacts on all walks of life. Concerns and debates that arose immediately following the widespread release of ChatGPT appear understandable owing to the well-described phenomenon of innate resistance to any change of the human mind [4]. However, when users provide insufficient information, ChatGPT tends to make assumptions about what the user wants to hear instead of asking clarifying questions. This tendency can lead to unintended consequences and make ChatGPT, as well as other language models like it, a double-edged sword [7].

Despite ChatGPT's lack of domain-specific training, its knowledge and interpretation ability could pass or nearly pass examinations designed for postgraduate levels of specialization in a wide range of fields, such as medicine, law, and finance [8]. In particular, several articles have discussed its use in improving knowledge interpretation and transfer [9,10]. Regardless, the major concerns of ChatGPT are plagiarism [11]

and generating fraudulent but authentic-looking outcomes [12]. Zuccon and Koopman [13] discovered that relying solely on model knowledge, ChatGPT correctly answered 80% of health care questions. The evidence presented in the prompt can heavily influence the correctness of answers, reducing ChatGPT's accuracy to only 63%. In fact, when evidence is provided, the model often tends to agree with the evidence's stance, even if the model would produce an answer with the opposite stance if the evidence was not provided in the input [14]. The study highlights the importance of carefully considering prompt knowledge and the potential biases it may introduce. Concerns raised by ChatGPT have yet to be thoroughly investigated. It is still being determined whether or not ChatGPT will alleviate or exacerbate the issues raised by previous chatbots [15]. Therefore, it is critical to understand how to leverage the strengths and weaknesses of ChatGPT to ensure safe use in knowledge interpretation and transfer.

Because of the rapid advancements and massive information flux in medical and health technologies, there is an urgent need for trainees and professionals to acquire new knowledge across multiple domains. It is critical to train them efficiently and effectively in new knowledge and techniques in order to ensure their proficiency in real-world medicine and health care applications. In this study, we aimed to evaluate trainees' and professionals' perceptions of ChatGPT-assisted knowledge transfer, using biomedical informatics as an example.

Methods

Ethics Approval

This study was approved by the Tri-Service General Hospital Institutional Review Board in Taiwan (202005069). This board was organized and operates in accordance with the International Conference on Harmonization, World Health Organization Good Clinical Practice, and applicable laws and regulations.

Study Population

In 2023, we used purposeful sampling to include all undergraduate trainees ($n=90$) and graduate professionals ($n=105$) in the School of Public Health at the National Defense Medical Center in Taiwan. There were a total of 195 eligible participants. They have all received biomedical informatics training.

Demonstration Video and Questionnaires

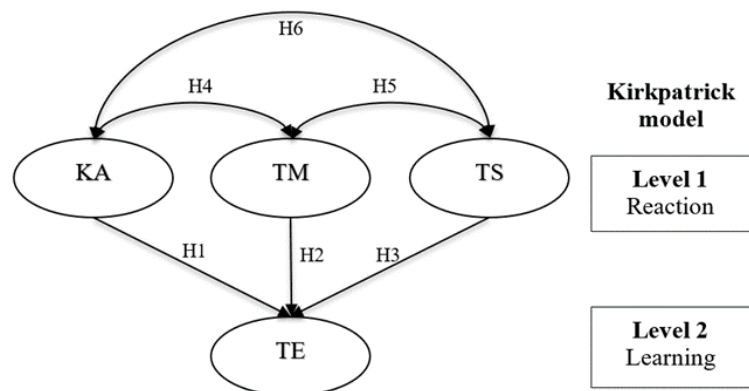
We created a 2-minute demonstration video ([Multimedia Appendix 1](#)) introducing ChatGPT and demonstrating five

ChatGPT-assisted training scenarios in biomedical informatics: (1) query question, (2) code debugging, (3) query coding examples, (4) creating self-evaluation quizzes, and (5) query training resources. To help participants understand a practical use scenario, we recoded the practical process of using ChatGPT in learning biomedical informatics (Multimedia Appendix 2, Table S1). Data were collected using an online (web- and mobile-based) questionnaire (Multimedia Appendix 3). We emailed the questionnaire to all eligible participants. The questionnaire was designed according to the Kirkpatrick model [16], with four dimensions to understand the thoughts of the students: (1) perceived knowledge acquisition (KA), (2) perceived training motivation (TM), (3) perceived training effectiveness (TE), and (4) perceived training satisfaction (TS). Three experts reviewed and edited the questionnaire, which has 12 questions, including 1 open-ended question. A 5-point Likert scale was adopted for all questionnaire items (from 1=strongly disagree to 5=strongly agree). The content of the questionnaire is shown in Multimedia Appendix 2, Table S2. The questionnaire was edited and collected using Google Forms. After signing an informed consent form, all participants were required to answer the questionnaire immediately after watching the video.

Thematic Analysis

The participants provided open-ended responses to question 13, “Do you have any other thoughts or suggestions on the application of ChatGPT to biomedical informatics training? Please give us feedback.” Thematic analysis was used to evaluate these open-text data without the use of a specific theoretical lens [17] via the web tool Taguette (Rémi Rampin). We coded and encapsulated the texts into “benefit” and “concern” themes. Finally, the results were displayed as a sunburst plot using the R package *plotly*.

Figure 1. Proposed research framework using Kirkpatrick model for participants’ perception of learning biomedical informatics using ChatGPT. KA: perceived knowledge acquisition; TE: perceived training effectiveness; TM: perceived training motivation; TS: perceived training satisfaction.



Hypotheses, Pathway Analysis, and Statistics

The Kirkpatrick model is a well-recognized method of evaluating the results of training effectiveness in 4 levels, including reaction, learning, behavior, and results [16]. This study evaluated the levels of reaction and learning. As depicted, we included a total of 6 hypotheses (H1-H6) to dissect how perceived KA, perceived TM, and perceived TS can together affect the perceived TE. In H1, H2, and H3, we tested whether, with a more positive perception of KA, TM, and TS, the participant would be more likely to have a positive perception of TE. In the other 3 hypotheses (H4-H6), we analyzed the relationship between KA, TM, and TS. In H4, we proposed that KA and TM are positively associated. In H5, we tested whether TS and TM are positively associated. Finally, in H6, we tested whether KA and TS are positively associated (Figure 1).

The association of training evaluation scores between undergraduate trainees and graduate professionals and male and female subjects was analyzed with a 2-tailed Student *t* test. The Kirkpatrick model was analyzed with structural equation modeling (SEM) using R Studio (February 3, 2022, build, version 4.2.0; R Core Team) with the package *lavaan*. The goodness-of-fit indices for the hypothetical model were evaluated as follows: chi-squared test, $P > .05$; relative chi-squared (χ^2/df) < 2 ; root mean square error of approximation < 0.06 ; goodness-of-fit index > 0.95 ; and normed fit index > 0.90 [18]. SEM is a multivariate statistical analysis technique used to analyze structural relationships. SEM is a method that combines factor analysis and multiple regression analysis. It examines the structural relationship between measured variables and latent constructs. We preferred this method because it estimates multiple and interrelated dependencies in a single analysis [19].

Results

Descriptive Statistics

From January to May 2023, we included all eligible participants (N=195). Participants were required to watch the demonstration video before filling out the online questionnaire. The demonstration video enables participants to quickly understand how to use ChatGPT to assist in knowledge transfer and solving problems. There are 5 application scenarios. For example, while

previewing the materials, participants can use ChatGPT to ask questions or query difficult codes for a detailed and easy-to-understand explanation. They can use ChatGPT to inquire about specific functional coding, such as how to write a BMI calculation function, or to debug in order to improve their training when writing analysis programming. They can request that ChatGPT create quizzes on specific training topics and provide additional online training resources for self-evaluation and extended training.

Finally, 152 of 195 (75%) participants completed the online questionnaires and were recruited for analysis, with 88 of 152 (58%) being undergraduate trainees and 90 of 152 (59%) being women. The ages ranged from 18 to 53 years (mean 23.3, SD 6.0 years). There was no difference between women and men in perceptions on ChatGPT-assisted learning (Table 1). For both men and women, graduate professionals scored significantly higher on all questions than undergraduate trainees (Table 1). For undergraduate trainees, men and women scored similarly. For graduate professionals, however, men scored significantly

higher than women. Male undergraduate trainees and graduate professionals had similar thoughts on question 9: "I believe that by using ChatGPT-assisted training rather than traditional training approaches, I will be more confident in my ability to gain the knowledge and skills being taught" (Table 1). On the other hand, female trainees and professionals had similar thoughts on question 12: "I believe that the information provided by ChatGPT-assisted training meets and exceeds my expectations."

Table 1. Evaluation of perceptions on ChatGPT-assisted training in biomedical informatics. Values for items other than age represent scores on a 5-point Likert scale (from 1=strongly agree to 5=strongly disagree).

| Item | Men, mean (SD) | | P value | Women, mean (SD) | | P value | All | P value ^a |
|--|------------------------|------------------|---------|------------------------|------------------|---------|-----------------|----------------------|
| | Under-graduates (n=41) | Graduates (n=21) | | Under-graduates (n=47) | Graduates (n=43) | | | |
| Age (years) | 20.1 (1.27) | 28.4 (6.68) | <.001 | 20.1 (1.42) | 27.5 (7.13) | <.001 | 23.3 (5.98) | .12 |
| Q1: I believe that using ChatGPT to assist with training will allow me to gain more knowledge than traditional training approaches. | 3.66 (0.825) | 4.48 (0.602) | <.001 | 3.74 (0.765) | 4.09 (0.684) | .03 | 3.92 (0.785) | .85 |
| Q2: I believe that using ChatGPT to assist training will help me understand the code content better than traditional training approaches. | 3.61 (0.862) | 4.10 (0.768) | .03 | 3.60 (0.771) | 3.98 (0.707) | .02 | 3.78 (0.799) | .98 |
| Q3: In comparison to traditional training approaches, I believe that using ChatGPT to assist training will make it easier for me to write the analysis programming I want. | 3.61 (0.891) | 4.14 (0.910) | .03 | 3.68 (0.726) | 4.00 (0.756) | .04 | 3.82 (0.825) | .75 |
| Q4: I believe that using ChatGPT-assisted training will increase my desire to do independent training more than traditional training methods. | 3.59 (0.974) | 4.24 (0.995) | .02 | 3.43 (0.827) | 3.95 (1.05) | .009 | 3.73 (0.990) | .43 |
| Q5: I believe that using ChatGPT to assist training will increase my motivation to train when compared to traditional training approaches. | 3.51 (0.925) | 4.14 (0.964) | .02 | 3.43 (0.773) | 4.09 (0.840) | <.001 | 3.74 (0.912) | .90 |
| Q6: I believe that using ChatGPT to assist with training will help me better understand the complex and difficult parts of training material than traditional training approaches. | 3.51 (1.00) | 4.14 (0.910) | .02 | 3.47 (0.776) | 3.81 (0.906) | .05 | 3.67 (0.919) | .54 |
| Q7: I believe that using ChatGPT-assisted training will increase my motivation to practice more when compared to traditional training approaches. | 3.63 (0.915) | 4.33 (0.856) | .005 | 3.57 (0.903) | 3.95 (0.844) | .04 | 3.80 (0.914) | .45 |
| Q8: I believe that using ChatGPT-assisted training will benefit my motivation and comprehension of preview training materials more than traditional training approaches. | 3.59 (0.865) | 4.33 (0.856) | .002 | 3.49 (0.906) | 4.02 (0.801) | .004 | 3.78 (0.906) | .53 |
| Q9: I believe that by using ChatGPT-assisted training rather than traditional training approaches, I will be more confident in my ability to gain the knowledge and skills being taught. | 3.59 (0.999) | 4.38 (0.865) | .003 | 3.57 (0.801) | 3.86 (0.774) | .09 | 3.77 (0.895) | .33 |
| Q10: I believe that by using ChatGPT-assisted training rather than traditional training approaches, I am more confident in my ability to learn programming grammar and data analysis. | 3.54 (0.897) | 4.38 (0.865) | <.001 | 3.47 (0.856) | 3.91 (0.781) | .01 | 3.74 (0.897) | .33 |
| Q11: I believe that using ChatGPT-assisted training is more satisfying than traditional training approaches. | 3.51 (0.925) | 4.10 (0.700) | .02 | 3.51 (0.882) | 3.91 (0.718) | .02 | 3.70 (0.852) | .95 |
| Q12: I believe that the information provided by ChatGPT-assisted training meets and exceeds my expectations. | 3.63 (0.994) | 4.05 (0.805) | 0.1 | 3.64 (0.792) | 4.07 (0.704) | .008 | 3.82 (0.849) | .62 |

^aThis represents a test of the difference in each item between men and women.

When we analyzed the whole group of participants, we found that among the 12 questions the top 2 questions agreed on the

most were question 1, "I believe that using ChatGPT to assist with training will allow me to gain more knowledge than

traditional training approaches,” and question 3, “In comparison to traditional training approaches, I believe that using ChatGPT to assist training will make it easier for me to write the analysis programming I want”; these questions had scores of 3.92 (SD 0.79) and 3.82 (SD 0.83), respectively (Table 1). On the other hand, the top 2 questions least agreed on were question 6, “I believe that using ChatGPT to assist with training will help me better understand the complex and difficult parts of training material than traditional training approaches,” and question 11, “I believe that using ChatGPT-assisted training is more satisfying than traditional training approaches”; these questions had scores of 3.67 (SD 0.91) and 3.70 (SD 0.85), respectively (Table 1).

Table 2. Scores for perceptions on the 4 constructs.

| Kirkpatrick model/construct | Score, mean (SD) |
|----------------------------------|------------------|
| Reaction | |
| Perceived training satisfaction | 3.75 (0.87) |
| Perceived training motivation | 3.76 (0.93) |
| Learning | |
| Perceived knowledge acquisition | 3.84 (0.80) |
| Perceived training effectiveness | 3.72 (0.91) |

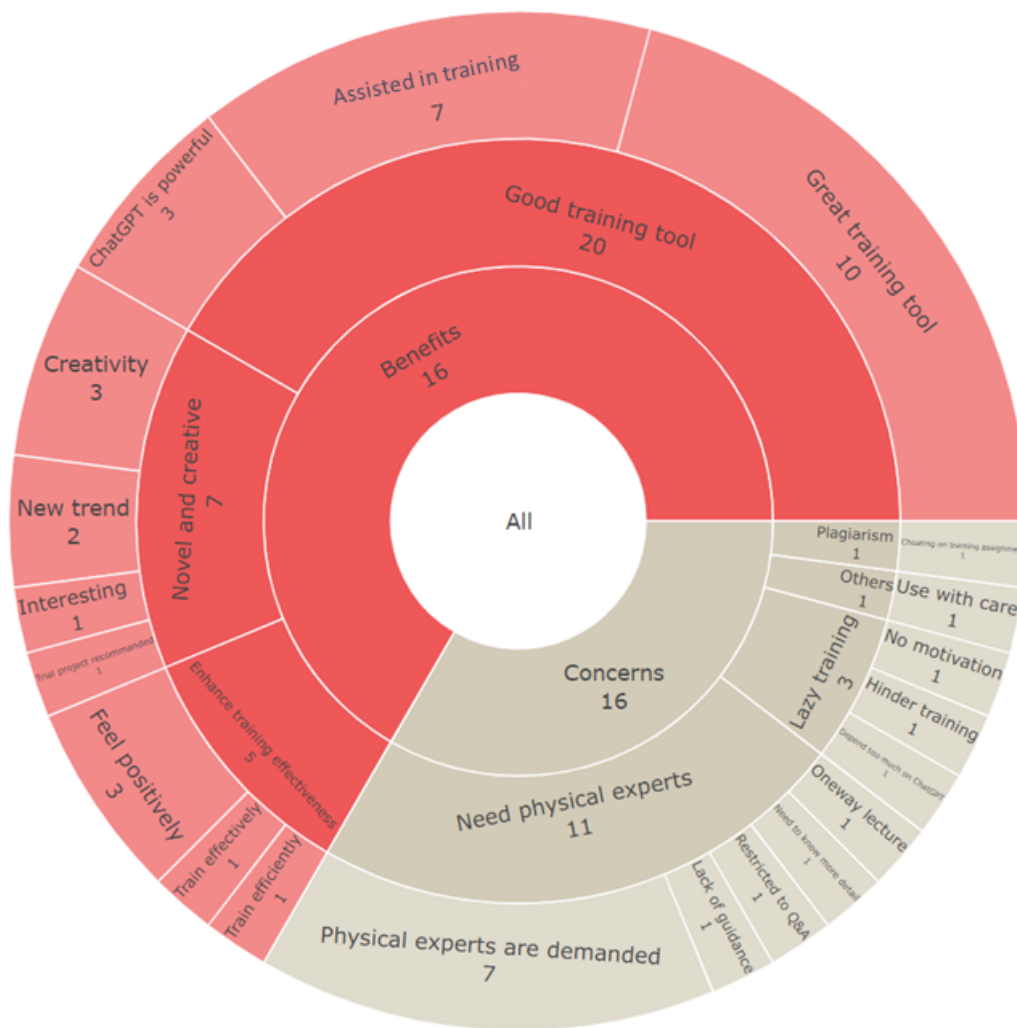
Open-Ended Responses

We included an open-ended question (question 13), “Do you have any other thoughts or suggestions on the application of ChatGPT to biomedical informatics training? Please give us feedback,” in addition to the 12 questions in our online questionnaire. We conducted a thematic analysis on the responses of the 25 participants who answered this question (Multimedia Appendix 2, Table S3). They identified 32 of 48 (67%) potential benefits of ChatGPT-assisted training in biomedical informatics (Figure 2). They believed ChatGPT is novel and creative and a good training tool. For example, one subject said, “Running the trial training program is highly recommended. I eagerly anticipate its implementation. However, it is unfortunate that my graduation is imminent. Nevertheless, this remarkable robot possesses the ability to engage in conversations and seamlessly continue previous discussions, making it feel like interacting with a real interlocutor rather than a mere automaton. Furthermore, ChatGPT can also compose captivating poems and articles upon request. Its impressive capabilities and potential for learning make it an immensely powerful entity deserving of our attention and study.”

In general, trainees and professionals had positive attitudes toward ChatGPT-assisted training in biomedical informatics. The average scores for the 12 questions ranged from 3.67 to 3.92 (SD 0.88; data not shown), higher than the median answer score of 3 points. The average scores for perceived KA, perceived TM, perceived TS, and perceived TE were 3.84 (SD 0.80), 3.76 (SD 0.93), 3.75 (SD 0.87), and 3.72 (SD 0.91), respectively (Table 2). Perceived KA had the highest scores, while perceived TE had the lowest scores. This suggests that ChatGPT-assisted training in biomedical informatics contributed most to KA yet relatively less to TM, TS, and TE.

On the other hand, some participants identified 16 of 48 (33%) potential concerns about ChatGPT-assisted training in biomedical informatics. They believed that physical empirical experts were still required to provide guidance, and thought that dual-track interaction would result in effective knowledge transfer. For example, a subject stated, “Although self-study can be used to gain knowledge, I believe that we still need an expert to lead and guide us so that we can learn the knowledge effectively. ChatGPT can be useful as an auxiliary tool, but it is limited to the question-and-answer mode of communication. To be able to explore a subject in depth, a physical expert’s guidance is still required...” Furthermore, this subject was concerned that relying too heavily on ChatGPT would impair self-reflective ability and lead to cheating on training evaluations and that ChatGPT should be used with caution to avoid fraudulent prompts. Quantitative and qualitative results showed that trainees and professionals were generally positive and excited about using ChatGPT as a training tool. Most of them were excited about the potential of ChatGPT to enhance knowledge transfer.

Figure 2. The perceived benefits of and concerns regarding ChatGPT-assisted in training biomedical informatics. The categories presented were defined by thematic analysis of open-ended answers. The numbers in the plot represent the number of participants who elicited each theme.



Hypothesis Test

SEM was used to investigate the association between the postulated latent constructs of perceptions on ChatGPT-assisted training based on the Kirkpatrick model. We optimized the path model by adding significant paths and deleting insignificant paths. The goodness-of-fit indices are presented in Multimedia Appendix 2, Table S4. The SEM met all the assumption criteria. The total explainable variation of 4 dimensions by the model was 0.84. The latent construct to which the equations should belong was analyzed and defined by SEM. Part of the questions were categorized differently than in the original design.

The results of the research model are shown in Table 3 and Figure 3. Participants with more positive perceptions of KA were more likely to have more positive perceptions of TE (H1 $\beta=.80, P<.001$). However, perceived TM (H2) and perceived TS (H3) did not significantly influence perceived TE. Instead of the expected directed association, perceived TM and perceived TE had a nondirected association (H2 $\beta=.74, P=.01$). Perceived KA influenced both perceived TM (H4 $\beta=.97, P<.001$) and perceived TS (H6 $\beta=.88, P<.001$). Unexpectedly, there was no significant association between perceived TS and perceived TM (H5). These results indicate that the perceived TE of the ChatGPT-assisted training in biomedical informatics was influenced predominantly by perceived KA and partially met the hypotheses in the research framework (Figure 1).

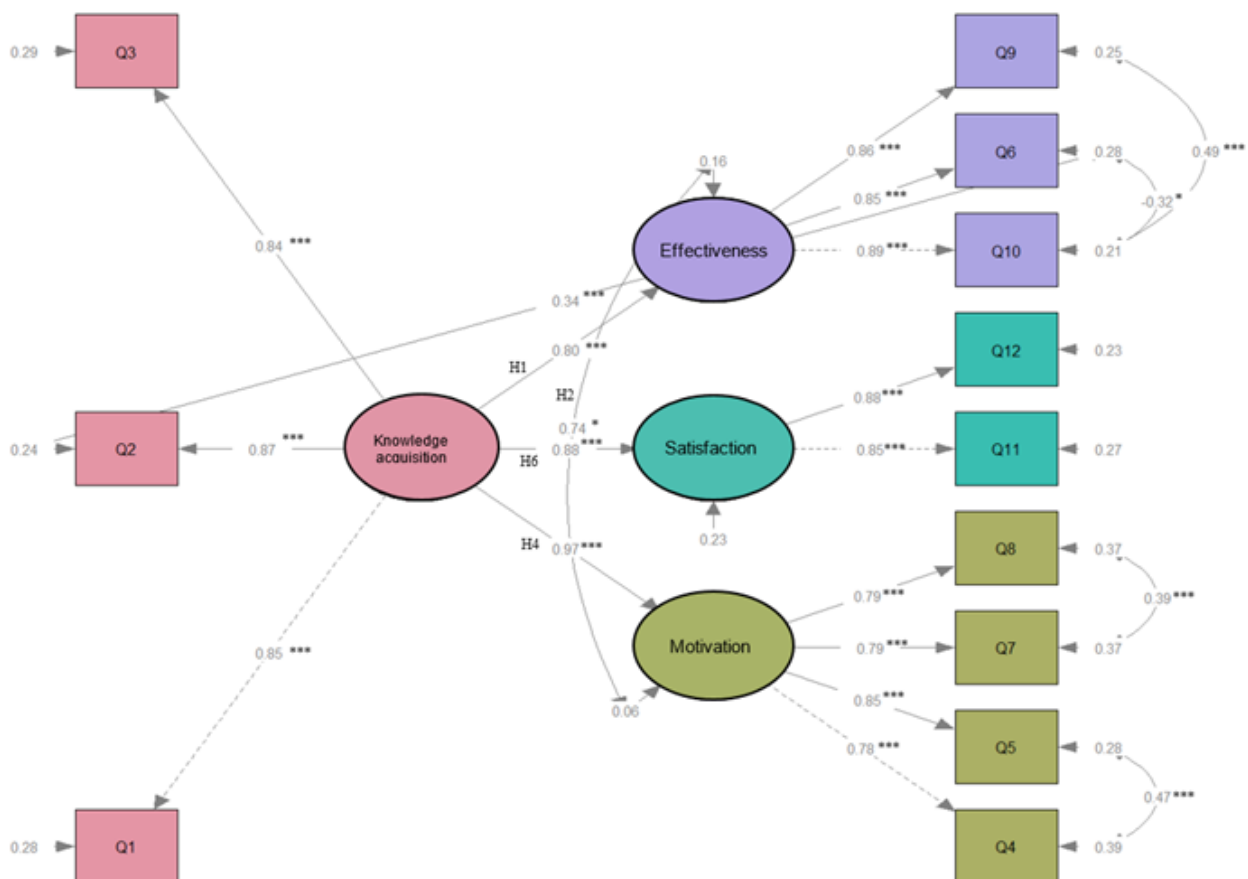
Table 3. The results of research model on ChatGPT-assisted training.

| Model | Hypothesis | Standardized coefficient | P value | Results |
|-------|---|--------------------------|------------------|--|
| H1 | Participants with more positive perceptions of knowledge acquisition are more likely to have more positive perceptions of training effectiveness. | .80 | <.001 | Supported |
| H2 | Participants with more positive perceptions of training motivation are more likely to have more positive perceptions of training effectiveness. | .74 | .01 | Not supported but nondirected association was discovered |
| H3 | Participants with more positive perceptions of training satisfaction are more likely to have more positive perceptions of training effectiveness. | N/A ^{a,b} | N/A ^b | Not supported |
| H4 | Perceptions of knowledge acquisition are positively associated with perceptions of training motivation. | .97 | <.001 | Supported |
| H5 | Perceptions of training motivation are positively associated with perceptions of training satisfaction. | N/A ^b | N/A ^b | Not supported |
| H6 | Perceptions of knowledge acquisition are positively associated with perceptions of training satisfaction. | .88 | <.001 | Supported |

^aN/A: not applicable.

^bDue to the lack of statistical significance, these connections were excluded from the model. Consequently, no statistical data are available.

Figure 3. Results of a path analysis of perceptions on ChatGPT-assisted training in biomedical informatics. The coefficients are standardized. Values on the lines represent β values. * $P < .05$; ** $P < .01$; *** $P < .001$.



Discussion

Principal Findings

This study is an initiative to evaluate the perceptions of health care trainees and professionals on ChatGPT-assisted knowledge transfer concerning biomedical informatics using a hypothesis-driven online questionnaire. We also carried out

further thematic, statistical, and pathway analyses. First, the ChatGPT-assisted training received the highest score in perceived knowledge acquisition but the lowest in perceived TE using the Kirkpatrick model. Secondly, graduate professionals expressed more positive attitudes toward the ChatGPT-assisted course than undergraduate trainees. Finally, there was no statistically significant difference in response scores between men and women. The participants agreed that

ChatGPT could help them gain more knowledge and ease their training, but they thought that physical empirical experts were still required to provide guidance and that a dual-track interaction would result in more effective knowledge transfer.

Although ChatGPT is a powerful tool for dealing with knowledge-based questions, it has limitations. First, it cannot generate content based on events after its most recent internet scan (in 2021). Therefore, some of the references it generates require updates. Secondly, it occasionally generates errors in math problems or word choice [9]. Since it is a brand-new technology, the model still requires further training. Specifically, ChatGPT may provide false answers [20]. From time to time, ChatGPT provides plausible-sounding but incorrect or nonsensical answers, such as citing a scientific study that does not exist [21]. Furthermore, AI can generate completely fabricated scientific articles that appear sophisticated and flawless [12]. Constraints and bias in the training data can have an adverse effect on the output of the model [21,22]. Third, instead of engaging in genuine conversation, it is often seen that ChatGPT passively follows instructions. For instance, when the information the user provides needs to be more comprehensive, ChatGPT frequently makes assumptions about what the user wants to hear instead of asking clarifying questions [7]. As to medical utility, ChatGPT can provide highly consistent but comparable quality medical information with 60% accuracy [23]. ChatGPT-3.5 had high accuracy in responding to common lung cancer questions; however, neither ChatGPT nor Google Bard, Bing, or the Google search engine answered all questions correctly 100% of the time [24].

ChatGPT was not originally designed for health care applications, and one of the major concerns was response reliability. Trust is crucial for users to adopt ChatGPT. Excessive trust in AI-driven chatbots like ChatGPT carries risks. Relying on ChatGPT for health advice could lead to misinformation and health risks. Efforts should focus on improving ChatGPT's ability to distinguish between appropriate and inappropriate queries, redirecting the latter to human experts [25]. However, language models are few-shot learners. Few-shot learners are machine learning models that can learn from a limited amount of labeled data and make accurate predictions based on previously unseen examples [26]. The abovementioned limitations of ChatGPT can be improved by few-shot learning. Besides, other upcoming LLMs and generative AIs, such as GPT-4 and Google Bard, have developed their AI models to access the internet. These features ensure that AIs provide up-to-date information for each question asked and that training data is added via the internet to improve response accuracy.

However, relying solely on ChatGPT-assisted training is insufficient for increasing training effectiveness. The quality of ChatGPT responses is heavily dependent on the user's querying skill. Although ChatGPT can quickly provide the desired information, it lacks the ability to organize and integrate training materials that are best suited to individual aptitudes. Thus, the combination of ChatGPT and empirical experts' guidance would enhance knowledge transfer. Zachary and Pardos [27] compared

the efficacy of ChatGPT hints with hints authored by human tutors among 77 participants across 2 algebra topic areas: elementary algebra and intermediate algebra. Although both human and ChatGPT hints produced positive knowledge transfer, human-created hints produced significantly higher knowledge transfer than ChatGPT hints in both topic areas.

We present a pilot study to assess users' perceptions of ChatGPT-assisted knowledge transfer in biomedical informatics. We started this study with health care students because they are health care trainees and professionals in this field. Although the sample size is insufficient to generalize the findings to other fields of expertise or users, it did provide a paradigm of ChatGPT-assisted training in new knowledge and techniques. We can foresee the flourishing development of AI chatbots being applied in medicine or health care [28-36]. ChatGPT represents a paradigm shift in the field of virtual assistants. Its successors, including ChatGPT-4, hold great potential as valuable tools for both patients and health care providers. The use of ChatGPT emphasizes the pressing need for regulators and health care professionals to actively participate in the development of minimum quality standards and to enhance patient awareness regarding the current limitations of emerging AI assistants [37]. Thus, LLMs still have a considerable journey ahead before they can be effectively used in the medical or health care domains. The demand for regulations concerning patient privacy and legal responsibilities becomes apparent. Additionally, AI-generated responses must undergo rigorous quality control and review processes. Moreover, the establishment of clinical guidelines is essential to ensure the appropriate use and implementation of these assistants in medical practice [38].

Limitations

The main limitation of this study is that we focused on the participants' perceptions of ChatGPT-assisted TE rather than their actual training performance. Furthermore, because the sample size was small, the generalizability of results may be limited. The missing response rate of 22% (43/195) could have resulted in sampling bias. Importance-performance analysis and the Chuchiming index [39] are recommended to determine the demand for and importance of LLM-assisted knowledge transfer for users. Future research should consider actual training performance before and after the intervention, as well as include a larger sample size.

Conclusions

Most trainees and professionals were enthusiastic about the ChatGPT-assisted knowledge transfer. They agreed that ChatGPT could help them gain more knowledge and ease their training. However, they thought that physical empirical experts were still required to provide guidance and that a dual-track interaction would result in more effective knowledge transfer. In a future study, a larger sample size and an evaluation of an internet-connected LLM for medical knowledge transfer are recommended.

Data Availability

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

Authors' Contributions

JMH was responsible for study conceptualization, writing, and editing the manuscript. FCL and CMC were responsible for supervision and editing the manuscript. YTC was responsible for study conceptualization, data curation, formal analysis, and writing and editing the original manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demonstration video of ChatGPT application.

[[MP4 File \(MP4 Video\), 65946 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Tables S1 to S4, including samples and additional statistics.

[[PDF File \(Adobe PDF File\), 446 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Web- and mobile-based questionnaire.

[[PDF File \(Adobe PDF File\), 282 KB-Multimedia Appendix 3](#)]

References

1. Prusinkiewicz K. How can we improve language models using reinforcement learning? ChatGPT case study. Deepsense. URL: <https://deepsense.ai/how-can-we-improve-language-models-using-reinforcement-learning-chatgpt-case-study/> [accessed 2023-10-05]
2. Atlas S. ChatGPT for higher education and professional development: a guide to conversational AI. DigitalCommons@URI. URL: https://digitalcommons.uri.edu/cba_facpubs/548 [accessed 2023-10-05]
3. Rudolph J, Tan S, Tan S. The brilliance and weirdness of ChatGPT. New York Times. URL: <https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html> [accessed 2023-10-05]
4. Sallam M. The utility of ChatGPT as an example of large language models in healthcare education, research and practice: systematic review on the future perspectives and potential limitations. medRxiv Preprint posted online February 21, 2023. [FREE Full text] [doi: [10.1101/2023.02.19.23286155](https://doi.org/10.1101/2023.02.19.23286155)]
5. Schwartz E. OpenAI promises customizable ChatGPT after bias complaints. Voicebot. URL: <https://voicebot.ai/2023/02/20/openai-promises-customizable-chatgpt-after-bias-complaints/> [accessed 2023-10-05]
6. Rudolph J, Tan S, Tan S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? J Appl Learn Teach 2023 Jan 25;6(1):22 [FREE Full text] [doi: [10.37074/jalt.2023.6.1.9](https://doi.org/10.37074/jalt.2023.6.1.9)]
7. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. Radiology 2023 Apr;307(2):e230163 [doi: [10.1148/radiol.230163](https://doi.org/10.1148/radiol.230163)] [Medline: [36700838](https://pubmed.ncbi.nlm.nih.gov/36700838/)]
8. Varanasi L. OpenAI just announced GPT-4, which can pass everything from a bar exam to AP Biology with flying colors. Here's a list of difficult exams both AI models have passed. Business Insider. URL: <https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1#the-uniform-bar-exam-2> [accessed 2023-10-05]
9. Lieberman M. What Is ChatGPT and how is it used in education? EducationWeek. URL: <https://www.edweek.org/technology/what-is-chatgpt-and-how-is-it-used-in-education/2023/01> [accessed 2023-10-05]
10. Ofgang E. What is ChatGPT and How can you teach with it? Tips and tricks. Tech & Learning. URL: <https://www.techlearning.com/how-to/what-is-chatgpt-and-how-to-teach-with-it-tips-and-tricks> [accessed 2023-10-05]
11. Bagshaw J, Knight C, Kernohan D. ChatGPT, assessment and cheating – have we tried trusting students? Wonkhe. URL: <https://wonkhe.com/blogs/chatgpt-assessment-and-cheating-have-we-tried-trusting-students/> [accessed 2023-10-05]
12. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. J Med Internet Res 2023 May 31;25:e46924 [FREE Full text] [doi: [10.2196/46924](https://doi.org/10.2196/46924)] [Medline: [37256685](https://pubmed.ncbi.nlm.nih.gov/37256685/)]
13. Zuccon G, Koopman B. Dr ChatGPT, tell me what I want to hear: How prompt knowledge impacts health answer correctness. arXiv Preprint posted online Feb 23, 2023. [FREE Full text] [doi: [10.48550/arXiv.2302.13793](https://doi.org/10.48550/arXiv.2302.13793)]
14. White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H. A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv Preprint posted online Feb 21, 2023. [FREE Full text]

15. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn Environ* 2023 Feb 22;10(1):15 [FREE Full text] [doi: [10.1186/s40561-023-00237-x](https://doi.org/10.1186/s40561-023-00237-x)]
16. Smidt A, Balandin S, Sigafos J, Reed VA. The Kirkpatrick model: A useful tool for evaluating training outcomes. *J Intellect Dev Disabil* 2009 Sep;34(3):266-274 [doi: [10.1080/13668250903093125](https://doi.org/10.1080/13668250903093125)] [Medline: [19681007](https://pubmed.ncbi.nlm.nih.gov/19681007/)]
17. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. *NPJ Digit Med* 2019;2:53 [FREE Full text] [doi: [10.1038/s41746-019-0132-y](https://doi.org/10.1038/s41746-019-0132-y)] [Medline: [31304399](https://pubmed.ncbi.nlm.nih.gov/31304399/)]
18. Hooper D, Coughlan J, Mullen M. Structural equation modelling: Guidelines for determining model fit. *Electron J Bus Res Methods* 2008;6(1):8 [doi: [10.1093/hesc/9780192896599.003.0012](https://doi.org/10.1093/hesc/9780192896599.003.0012)]
19. Structural equation modeling. Statistics Solutions. URL: <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/structural-equation-modeling/> [accessed 2023-10-05]
20. What is ChatGPT? Top capabilities and limitations you must know. Emeritus. URL: <https://emeritus.org/in/learn/ai-ml-what-is-chatgpt/> [accessed 2023-10-05]
21. Thorp HH. ChatGPT is fun, but not an author. *Science* 2023 Jan 27;379(6630):313 [doi: [10.1126/science.adg7879](https://doi.org/10.1126/science.adg7879)] [Medline: [36701446](https://pubmed.ncbi.nlm.nih.gov/36701446/)]
22. Baidoo-Anu D, Owusu Ansah L. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of chatgpt in promoting teaching and learning. *SSRN Journal* 2023;52-62 [FREE Full text] [doi: [10.2139/ssrn.4337484](https://doi.org/10.2139/ssrn.4337484)]
23. Walker HL, Ghani S, Kuemmerli C, Nebiker CA, Müller BP, Raptis DA, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. *J Med Internet Res* 2023 Jun 30;25:e47479 [FREE Full text] [doi: [10.2196/47479](https://doi.org/10.2196/47479)] [Medline: [37389908](https://pubmed.ncbi.nlm.nih.gov/37389908/)]
24. Rahsepar AA, Tavakoli N, Kim GHJ, Hassani C, Abtin F, Bedayat A. How AI responds to common lung cancer questions: ChatGPT vs Google Bard. *Radiology* 2023 Jun;307(5):e230922 [doi: [10.1148/radiol.230922](https://doi.org/10.1148/radiol.230922)] [Medline: [37310252](https://pubmed.ncbi.nlm.nih.gov/37310252/)]
25. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *J Med Internet Res* 2023 Jun 14;25:e47184 [FREE Full text] [doi: [10.2196/47184](https://doi.org/10.2196/47184)] [Medline: [37314848](https://pubmed.ncbi.nlm.nih.gov/37314848/)]
26. Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P. Language models are few-shot learners. In: *Advances in Neural Information Processing Systems*. 2020 Presented at: NeurIPS 2020; Dec 6-12, 2020; Virtual p. 33-901 URL: <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>
27. Zachary A, Pardos S. Learning gain differences between ChatGPT and human tutor generated algebra hints. *arXiv Preprint* posted online May 2023. [FREE Full text] [doi: [10.48550/arXiv.2302.06871](https://doi.org/10.48550/arXiv.2302.06871)]
28. Calleja-López JRT, Rivera-Rosas CN, Ruibal-Tavares E. Impact of ChatGPT and artificial intelligence in the contemporary medical landscape. *Arch Med Res* 2023 Jul;54(5):102835 [doi: [10.1016/j.arcmed.2023.05.003](https://doi.org/10.1016/j.arcmed.2023.05.003)] [Medline: [37248157](https://pubmed.ncbi.nlm.nih.gov/37248157/)]
29. Janamala V, Ram IS, Daram SB. Realization of green 5G cellular network role in medical applications: use of ChatGPT-AI. *Ann Biomed Eng* 2023 Jun 01:1-3 [FREE Full text] [doi: [10.1007/s10439-023-03257-3](https://doi.org/10.1007/s10439-023-03257-3)] [Medline: [37261590](https://pubmed.ncbi.nlm.nih.gov/37261590/)]
30. Jin Q, Leaman R, Lu Z. Retrieve, summarize, and verify: How will ChatGPT affect information seeking from the medical literature? *J Am Soc Nephrol* 2023 Aug 01;34(8):1302-1304 [doi: [10.1681/ASN.000000000000166](https://doi.org/10.1681/ASN.000000000000166)] [Medline: [37254254](https://pubmed.ncbi.nlm.nih.gov/37254254/)]
31. Kleebayoon A, Wiwanitkit V. ChatGPT in medical practice, education and research: malpractice and plagiarism. *Clin Med (Lond)* 2023 May;23(3):280 [doi: [10.7861/clinmed.Let.23.3.2](https://doi.org/10.7861/clinmed.Let.23.3.2)] [Medline: [37236804](https://pubmed.ncbi.nlm.nih.gov/37236804/)]
32. Koga S. The potential of ChatGPT in medical education: Focusing on USMLE preparation. *Ann Biomed Eng* 2023 Oct;51(10):2123-2124 [doi: [10.1007/s10439-023-03253-7](https://doi.org/10.1007/s10439-023-03253-7)] [Medline: [37248408](https://pubmed.ncbi.nlm.nih.gov/37248408/)]
33. Li S. Exploring the clinical capabilities and limitations of ChatGPT: a cautionary tale for medical applications. *Int J Surg* 2023 Sep 01;109(9):2865-2867 [FREE Full text] [doi: [10.1097/JS9.0000000000000523](https://doi.org/10.1097/JS9.0000000000000523)] [Medline: [37222684](https://pubmed.ncbi.nlm.nih.gov/37222684/)]
34. Shay D, Kumar B, Bellamy D, Palepu A, Dershwitz M, Walz JM, et al. Assessment of ChatGPT success with specialty medical knowledge using anaesthesiology board examination practice questions. *Br J Anaesth* 2023 Aug;131(2):e31-e34 [doi: [10.1016/j.bja.2023.04.017](https://doi.org/10.1016/j.bja.2023.04.017)] [Medline: [37210278](https://pubmed.ncbi.nlm.nih.gov/37210278/)]
35. Tsang R. Practical applications of ChatGPT in undergraduate medical education. *J Med Educ Curric Dev* 2023;10:23821205231178449 [FREE Full text] [doi: [10.1177/23821205231178449](https://doi.org/10.1177/23821205231178449)] [Medline: [37255525](https://pubmed.ncbi.nlm.nih.gov/37255525/)]
36. Zhou J, Jia Y, Qiu Y, Lin L. The potential of applying ChatGPT to extract keywords of medical literature in plastic surgery. *Aesthet Surg J* 2023 Aug 17;43(9):NP720-NP723 [doi: [10.1093/asj/sjad158](https://doi.org/10.1093/asj/sjad158)] [Medline: [37210418](https://pubmed.ncbi.nlm.nih.gov/37210418/)]
37. Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. *JNCI Cancer Spectr* 2023 Mar 01;7(2):pkad010 [FREE Full text] [doi: [10.1093/jncics/pkad010](https://doi.org/10.1093/jncics/pkad010)] [Medline: [36808255](https://pubmed.ncbi.nlm.nih.gov/36808255/)]
38. Mesko B. The ChatGPT (generative artificial intelligence) revolution has made artificial intelligence approachable for medical professionals. *J Med Internet Res* 2023 Jun 22;25:e48392 [FREE Full text] [doi: [10.2196/48392](https://doi.org/10.2196/48392)] [Medline: [37347508](https://pubmed.ncbi.nlm.nih.gov/37347508/)]
39. Cheng C, Huang Y, Sun C, An C, Chang Y, Chu C, et al. Gender-specific determinants of eHealth literacy: Results from an adolescent internet behavior survey in Taiwan. *Int J Environ Res Public Health* 2022 Jan 07;19(2):664 [FREE Full text] [doi: [10.3390/ijerph19020664](https://doi.org/10.3390/ijerph19020664)] [Medline: [35055487](https://pubmed.ncbi.nlm.nih.gov/35055487/)]

Abbreviations

AI: artificial intelligence
KA: knowledge acquisition
LLM: large language model
TE: training effectiveness
TM: training motivation
TS: training satisfaction

Edited by G Eysenbach, T de Azevedo Cardoso; submitted 29.05.23; peer-reviewed by YH Lee, K Fultz Hollis; comments to author 28.06.23; revised version received 13.07.23; accepted 29.09.23; published 18.10.23

Please cite as:

Hu JM, Liu FC, Chu CM, Chang YT

Health Care Trainees' and Professionals' Perceptions of ChatGPT in Improving Medical Knowledge Training: Rapid Survey Study
J Med Internet Res 2023;25:e49385

URL: <https://www.jmir.org/2023/1/e49385>

doi: [10.2196/49385](https://doi.org/10.2196/49385)

PMID: [37851495](https://pubmed.ncbi.nlm.nih.gov/37851495/)

©Je-Ming Hu, Feng-Cheng Liu, Chi-Ming Chu, Yu-Tien Chang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 18.10.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.