



Development of AI-generated medical responses using the ChatGPT for cancer patients

Jae-woo Lee^{a,b,1}, In-Sang Yoo^{c,h,1}, Ji-Hye Kim^c, Won Tae Kim^{d,e}, Hyun Jeong Jeon^{f,g},
Hyo-Sun Yoo^a, Jae Gwang Shin^c, Geun-Hyeong Kim^c, ShinJi Hwang^c, Seung Park^{c,h,2}, Yong-
June Kim^{d,e,2,*}

^a Department of Family Medicine, Chungbuk National University Hospital, Cheongju, Republic of Korea

^b Department of Family Medicine, Chungbuk National University College of Medicine, Cheongju, Republic of Korea

^c Department of Biomedical Engineering, Chungbuk National University Hospital, Cheongju, Republic of Korea

^d Department of Urology, Chungbuk National University Hospital, Cheongju, Republic of Korea

^e Department of Urology, Chungbuk National University College of Medicine, 1 Chungdae-ro, Seowon-gu, Cheongju, Chungcheongbuk-do 28644, Republic of Korea

^f Department of Internal Medicine, Chungbuk National University Hospital, Cheongju, Republic of Korea

^g Department of Internal Medicine, College of Medicine, Chungbuk National University, Cheongju, Republic of Korea

^h Department of Medicine, Chungbuk National University College of Medicine, Cheongju, Republic of Korea

ARTICLE INFO

Keywords:

Chatbot

AI

Cancer patients

ChatGPT

Natural language processing models

ABSTRACT

Background and objective: To develop a healthcare chatbot service (AI-guided bot) that conducts real-time conversations using large language models to provide accurate health information to patients.

Methods: To provide accurate and specialized medical responses, we integrated several cancer practice guidelines. The size of the integrated meta-dataset was 1.17 million tokens. The integrated and classified metadata were extracted, transformed into text, segmented to specific character lengths, and vectorized using the embedding model. The AI-guide bot was implemented using Python 3.9. To enhance the scalability and incorporate the integrated dataset, we combined the AI-guide bot with OpenAI and the LangChain framework. To generate user-friendly conversations, a language model was developed based on Chat-Generative Pretrained Transformer (ChatGPT), an interactive conversational chatbot powered by GPT-3.5. The AI-guide bot was implemented using ChatGPT3.5 from Sep. 2023 to Jan. 2024.

Results: The AI-guide bot allowed users to select their desired cancer type and language for conversational interactions. The AI-guided bot was designed to expand its capabilities to encompass multiple major cancer types. The performance of the AI-guide bot responses was 90.98 ± 4.02 (obtained by summing up the Likert scores).

Conclusions: The AI-guide bot can provide medical information quickly and accurately to patients with cancer who are concerned about their health.

1. Introduction

Cancer is one of the leading causes of death worldwide; it accounted for over 19.3 million new diagnoses and 10 million deaths in 2020 [1]. According to the CONCORD-3 reports, the five-year net survival trends have increased for most cancers, including the cases associated with some of the more lethal cancers [2]. Although there are differences among cancer sites and countries, cancer survival rates are high and continuously increasing [2]. By 2030, approximately 21.6 million

patients per year worldwide are expected to be diagnosed with cancer [3]. All cancer patients need prompt diagnosis and optimal treatment to improve survival. Therefore, the importance of an effective healthcare system for the treatment and care of patients with cancer is increasing. Cancer patients have several concerns about physical, practical, and emotional problems, but they do not receive appropriate information to manage these concerns [4]. In particular, patients often receive misinformation about their health from Internet sources, such as Google or YouTube [5–7]. In previous studies, [6], it was reported that harmful

* Corresponding author.

E-mail address: urokyj@gmail.com (Y.-J. Kim).

¹ These authors contributed equally to the manuscript.

² These authors contributed equally to this work.

misinformation about cancer was being disseminated through YouTube and that misinformation was being indiscriminately uploaded to the Internet not only by nonprofessional groups but also by professional groups.

Conversely, natural language processing (NLP) models can be applied to enhance access to medical information for patients [8]. Large language models (LLMs) are NLP tools that can understand and generate human-like text, such as the Chat-Generative Pretrained Transformer (ChatGPT) [9]. ChatGPT is a conversational chatbot based on GPT-3.5 with 175 billion parameters trained on a wide range of sources, such as books and articles [10]. This chatbot was fine-tuned for conversational tasks using reinforcement learning with human feedback. This training enables it to handle the complexity of user intent and respond effectively [11]. Conventional LLMs are simpler models predicting sequences based on fixed windows of words, whereas ChatGPT utilizes vast datasets to generate human-like conversations [9]. By using ChatGPT to develop chatbots for cancer patients, it can be expected to improve dramatically the accessibility of medical information. In addition, effective health-information delivery can be an alternative solution to the unsatisfied health information needs of patients.

In this study, we aimed to develop a healthcare chatbot service (hereafter referred to as the AI-guide bot) that conducts real-time conversations using an LLM model designed by a cancer-related expert group to provide accurate health information to patients. This service is expected to provide assistance to address the unmet needs of patients with cancer.

2. Methods

2.1. Data source

To develop a real-time chatbot in which AI answers medical questions for cancer patients, we first constructed a specialized medical knowledge database related to cancer. To provide accurate and specialized medical responses, cancer experts participated in the construction of the dataset using accredited medical guidelines. For prostate cancer, we integrated prostate cancer practice guidelines from the Korean Prostate Society, [12]. accredited health guidelines for patients, such as the National Comprehensive Cancer Network, American Society of Clinical Oncology, and European Association of Urology Guidelines, [13–18]. urology book, [19,20]. and instructional videos [21]. The total size of the integrated meta-dataset was 1.17 million tokens.

The integrated meta-dataset was divided into the following specific categories: definition, epidemiology, causes, symptoms, diagnosis, staging, treatment, prognosis, and lifestyle. The categorized meta-dataset was organized by considering actual questions posed by patients and nonspecialized medical providers as well as questions raised by cancer experts (Supplementary Fig. 1). The integrated and classified metadata were extracted, transformed into text, segmented to a specific character length, and vectorized using the embedding model provided by OpenAI [22].

This study was approved by the Institutional Review Board of the Chungbuk National University Hospital (CBNUH-04-020) and adhered to the guidelines of the Declaration of Helsinki (1975). The Ethics Committee waived the need for informed consent because the data from this study were anonymized at all stages.

2.2. Development process

The AI-guide bot was implemented using Python 3.9. To enhance the scalability and incorporate the integrated dataset, we combined the AI-guide bot with OpenAI [23] and the LangChain framework [24]. To generate user-friendly conversations, a language model was developed based on ChatGPT, an interactive conversational chatbot powered by GPT-3.5. The AI-guide bot was implemented using ChatGPT3.5 from Sep. 2023 to Jan. 2024. We adopted the GPT 3.5 version as our baseline

rather than GPT 4.0 owing to its cost-effectiveness. The collected data, described in the “Data source” section, was vectorized using the OpenAI embedding model and stored in a database using the FAISS Python library, a vector index (Fig. 1). When a user submits a question, the AI-Chatbot evaluates whether the query contains cancer-related information in the database. We used a fixed prompt for this study and deliberately set the temperature parameter to a relatively low value of 0.1 (default = 0.7). This decision was based on our objective to generate responses that are consistent and closely aligned with the intent of the questions posed. Additionally, while modifying the GPT prompts, we aimed to ensure consistent responses by repeating the same questions 10 times. By adopting this setting, we observed that the model predominantly produced answers that were coherent and relevant to the questions asked. If the question is unrelated to the designated type of cancer, the AI-Chatbot is configured to respond that it could not provide an answer because of the irrelevance of its content as the database contains medically specialized terminologies, the AI-Chatbot interprets the user’s intent and formulates an appropriate keyword. The refined question then undergoes vectorization, and its similarity is assessed in relation to each entry in the database. To ensure that the GPT generated a response using the pertinent information, we incorporated the top four similar entries into the GPT prompt.

To facilitate accurate communication of medical information, when the similarity score of the retrieved response vector index was below 0.4, the query was output as unrelated to the specific cancer type or owing to a lack of information. This prompts users to rephrase their questions more accurately. When the similarity score of the retrieved response vector index was 0.4 or higher, a prompt was constructed using the user’s question, selected vector index with the highest similarity, and the predefined instructions for structuring the output sentence. For questions unrelated to health, we confirmed that the model tended to reject generating answers. This prompt was then delivered to the output server of the AI-guide bot. The fixed directives were designed in a simple and clear format to induce familiar and easy conversations with the user. In cases wherein the responses generated by the AI-guide bot for the output were not very readable, they were corrected and output in user-friendly terms. This entire output process was recorded as part of the conversation history between the user and the AI-guide bot, thus making it accessible for future prompts (Fig. 1, Supplementary Fig. 2).

2.3. Performance evaluation process

To evaluate the performance of the AI-guide bot, we constructed a cancer-related question dataset and assessed whether the AI-guide bot provided correct and accurate medical responses. A total of 100 questions were validated by multiple oncology experts to ensure clear and uncontroversial answers in line with the latest medical guidelines. All questions were selected to represent expert knowledge in oncology (Supplementary Table 1).

The evaluation of the AI-guide bot responses to cancer-related questions was conducted by specialists (Urology, Family Medicine, Endocrinology, and Hematology–Oncology) who were not involved in developing the cancer-related question dataset. The response evaluation focused on whether the AI-guide bot clearly understood the patient’s question (comprehensibility), accuracy of the response content (accuracy), and whether understanding was easy (readability). Comprehensibility was based on how well the AI-guide bot understood the intent of the questions, determining whether the guide bot effectively set the direction for its responses by analyzing the questions entered into the natural language. A Likert score for comprehensibility was assigned to each question on a scale of 10 units (between 10 and 30). Example (10) indicates that the intent of the question and the response are entirely unrelated. Example (20) implies that unrelated content is included; however, the response is relevant to the intent of the question. Example (30) was assigned when a response was considered appropriate for the question’s intent. Content accuracy was assessed based on the accuracy

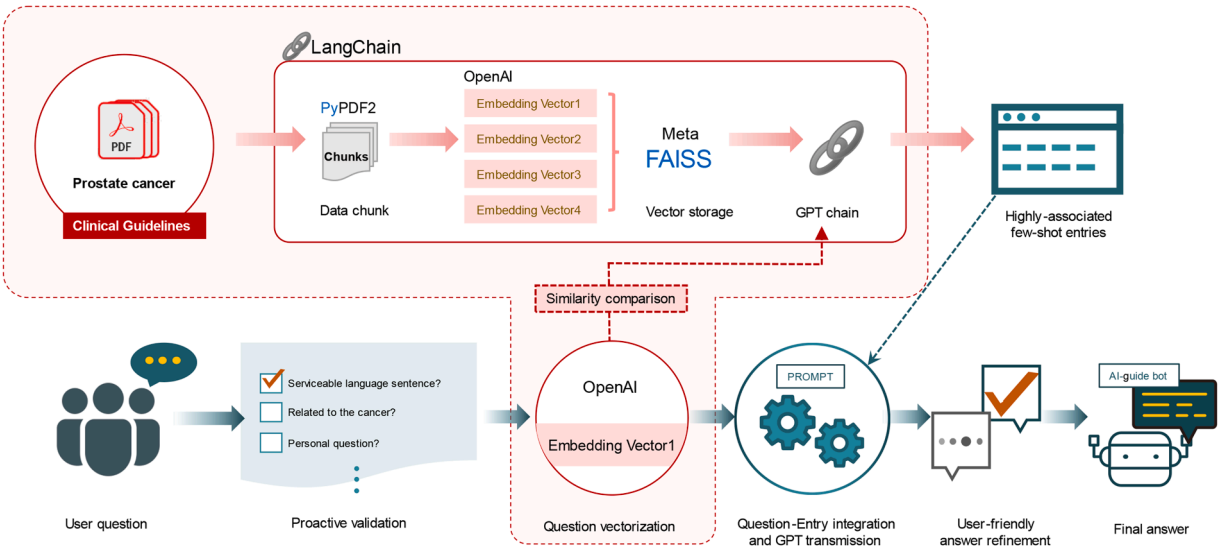


Fig. 1. Development process of artificial intelligence (AI)-guide bot for cancer patients.

of the AI-generated medical responses, regardless of the intent of the question. Content accuracy was established to evaluate the reliability of the responses because a guidebot needs to provide medical-related answers. A Likert score (ranging from 10 to 40 on a scale of 10 units) was assigned to each question for content accuracy. Example (10) indicates that there was no accuracy in the response regardless of the question's intent. Example (20) implies that some facts are included but most information is false. Example (30) is assigned when false hoods are included; however, most of the information is accurate. Example (40) was awarded when the response was considered to provide accurate information. Readability evaluates the comprehension difficulty and naturalness of sentences irrespective of the question's intent and accuracy. Sentence completeness is designed to assess the complexity of sentences and whether they are difficult to understand, as the guide bot is expected to communicate with the general public. Unlike the previous evaluation criteria, the Likert scores were set on a narrower scale because readability is subjective. Additionally, readability solely evaluates sentence completeness, which is unrelated to question relevance or content accuracy. Each question was assigned a Likert score (which ranged from 10 to 30 on a 10-unit scale). Example (10) implies that the use of medical terminology is convoluted, and grammar is incorrect, thus making it difficult to understand. Example (20) suggests that grammar is correct; however, the excessive use of medical terms makes it challenging. Example (30) is assigned when the sentence structure is appropriate and is not considered difficult for adults to understand.

The performance of the AI-guided bot responses was assessed by summing their Likert scores. Responses with a total Likert score of 30 or lower were classified as “inappropriate,” those with total scores of 70 or lower were considered to include some “minor errors or they were incomplete,” and responses with scores above 70 were classified as “appropriate” for the questions [25] (Table 1).

As an additional performance evaluation method, considering that the AI-guide bot user was a cancer patient, performance evaluation was also conducted on a group of 50 non-medical experts. Since this was a non-expert group, accuracy was excluded and a performance evaluation comparison was conducted using ChatGPT in the areas of comprehensibility and readability.

3. Results

The AI-guide bot allows users to select their desired cancer type and language for conversational interactions. Both English and Korean languages provided language support for the study. Users have the

Table 1
Explanation and performance of artificial intelligence (AI)-guide bot for cancer patients.

Criteria		Explanation	Performance (Likert sum average)
Comprehensibility	10	The question's intent and the response are entirely unrelated	28.28 ± 0.38
	20	While some unrelated content is mixed, the response is relevant to the intent of the question	
	30	The response is considered appropriate to the intent of the question	
Accuracy	10	No accurate information in the response	34.17 ± 2.91
	20	Some facts are included, but most of the information is false	
	30	Some falsehoods are included, but most of the information is accurate	
	40	The response is considered to provide entirely accurate information	
Readability	10	The response is difficult owing to the disorganized use of technical terminology and lack of proper grammar	28.53 ± 1.24
	20	The grammar is clean, but an excessive use of medical terms makes it challenging	
	30	Sentence structure is appropriate and it is not considered difficult for adults to understand	
Total	100		90.98 ± 4.02

flexibility to switch to different cancer types or languages during conversations with the AI-guided bot. An AI-generated Chatbot was designed to expand its capabilities to encompass multiple major cancer types. The initial model was developed with a strong focus on prostate cancer, benefiting from established clinical guidelines. Fig. 2 shows a screenshot of the actual output results of the AI-guide bot. When a user asks a question related to cancer, the AI-guide bot understands the question and outputs an appropriate and accurate answer from the medical guideline metadata. An example of a slightly more complex question is given here; “I was diagnosed with stage 1 prostate cancer.

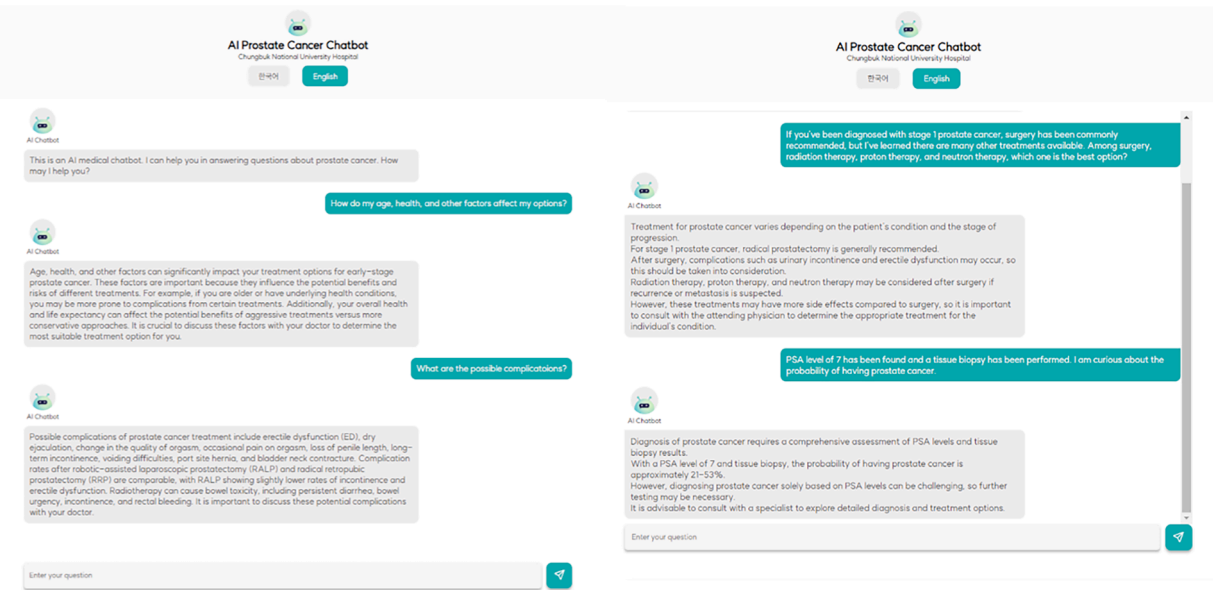


Fig. 2. Output screenshot of AI-Guide bot for cancer patients.

The hospital recommended surgery, but I've heard other treatments are available. What's the best option between surgery, radiation therapy, proton therapy, and neutron therapy?" For this question (supplementary Table 1, #88. Question), the AI-guide bot's automatic medical answer output was as follows; "Treatment for prostate cancer varies depending on the patient's condition and the stage of progression. For stage 1 prostate cancer, radical prostatectomy is generally recommended. After surgery, complications such as urinary incontinence and erectile dysfunction may occur, so this should be taken into consideration. Radiation therapy, proton therapy, and neutron therapy may be considered after surgery if recurrence or metastasis is suspected. However, these treatments may have more side effects compared to surgery, so it is important to consult with the attending physician to determine the appropriate treatment for the individual's condition". In the case of prostate cancer, the PSA level is often checked during health checkups, and patients are curious about the meaning of the results of this test. When the PSA-related questions: "My PSA is seven, and I had a tissue biopsy. I'm wondering about the likelihood of having prostate cancer" (Supplementary Table 1. #89. Question) were entered into the AI-guide bot, and the output was "Diagnosis of prostate cancer requires a comprehensive assessment of PSA levels and tissue biopsy results. With a PSA level of 7 and a tissue biopsy, the probability of having prostate cancer is approximately 21–53 %. However, diagnosing prostate cancer solely based on PSA levels can be challenging, so further testing may be necessary. The answer is generated as "It is advisable to consult with a specialist to explore detailed diagnosis and treatment options." This answer provides information such as the meaning of PSA level and the incidence of prostate cancer that roughly corresponds to this level (Fig. 2).

Table 1 presents the explanation and performance evaluation results of medical experts group for the AI-guide bot. Understanding the user's intention to ask questions was assigned a score of 28.28 ± 0.38 , accuracy of content was assigned to a score of 34.17 ± 2.91 , and readability was assigned to a score of 28.53 ± 1.24 . The performance of the AI-guide bot responses was 90.98 ± 4.02 (obtained by summing up the Likert scores). From the performance evaluation results conducted by a non-expert group, the AI-guide bot achieved statistically higher results than ChatGPT in comprehension and readability (Table 2).

Table 2
Performance comparison with ChatGPT in a non-expert group for artificial intelligence (AI)-guide bot.

	AI-guide bot Mean \pm S.D	ChatGPT Mean \pm S.D	P-value
Comprehensibility	28.4 ± 0.38	26.1 ± 0.81	<0.0001
Readability	28.3 ± 0.39	25.6 ± 0.84	<0.0001

Abbreviations: S.D, Standard deviation

The P-value was calculated using the *t*-test between the two groups.

4. Discussion

4.1. Main findings

In this study, a ChatGPT-based health consultation chatbot was developed for cancer patients using an LLM model. This healthcare chatbot program yielded promising results and can be used for cancer patients to understand the intent of the question, accuracy of the answers, and readability. In addition, through the collaboration of cancer-related experts, the accuracy of the answers was reverified so that false information was not provided to patients. We expect the chatbot program to become a compass for health promotion among patients with cancer.

4.2. AI-generated medical responses

Previously, most NLP models were trained using supervised learning and could be used for trained tasks only [26]. However, in 2017, with the introduction of a self-attention network architecture known as a transformer, [27]. ChatGPT was able to apply pretrained language representations to downstream tasks through a combination of unsupervised pretraining and supervised fine-tuning. The strength of ChatGPT lies in its NLP capabilities, which are based on GPT-3.5. With this feature, ChatGPT can extract information from unstructured medical information data sources, such as medical guidelines and medical textbooks, identify patterns, such as specific symptoms, and generate output results from them [28]. ChatGPT has recently been upgraded to include new features and improvements. The new LLM (GPT-4) can process up to 25,000 words and boost creativity and visual input ahead of its predecessor, GPT-3.5, as well as provide longer contextual

memory. According to the internal evaluation, ChatGPT has now become a safer and more accurate tool for tasks, thus providing 40 % more reliable responses [28].

In this study, the developed AI-guide bot utilized AI and natural language processing, similar to other traditional medical chatbots, [29] to understand user intentions and provide appropriate responses tailored to user requirements by retrieving information from medical datasets. Additionally, AI-guide bot can be further improved through user feedback, enabling the provision of knowledge pertaining to various clinical scenarios of the patients' clinical needs. This ongoing process empowers us to customize the performance of AI-guide bots according to individual patient requirements.

In contrast, ChatGPT raises concerns about the accuracy and reliability of the medical information it provides, as it draws information from the internet [28]. To address this issue, medical experts provide meticulous training to the AI-guide bot, thus preventing the dissemination of inaccurate information to patients. In this study, the performance of the AI-guided bot was evaluated using three criteria: comprehensibility, accuracy, and readability. The average total Likert score was 90.98, which surpasses the acceptable threshold of 70. A previous study examined the information quality and adequacy of the ChatGPT output for urological patients, and only 52 % of all responses reported that it was adequate [30]. The performance results of the AI-guide bot developed in this study demonstrate its potential for ongoing development owing to its adaptability through database modifications and additional AI training by medical experts. This adaptability constitutes a considerable advantage for future programs.

4.3. Unmet needs of cancer patients

The provision of medical information to cancer patients can be expected to yield several benefits, including increased patient participation in decision-making and enhanced satisfaction with medical treatment [31,32]. Moreover, it may improve coping abilities and reduce anxiety and worries related to cancer diagnosis, treatment, and post-treatment phases [33]. This patient information has been recognized as an effective individual means to cope with the quality of life impairment associated with cancer diagnosis and treatment [34]. Furthermore, effective communication between physicians and patients is crucial for optimal health outcomes in the medical field [34]. It is known that it is essential to ensure that patients receive and understand appropriate medical information during their treatment process to ensure the provision of quality cancer treatment [31–34].

Medical chatbots offer an effective communication method with ChatGPT and provide a user-friendly and concise conversational approach to information delivery [9]. Cancer patients sometimes have unmet medical information needs, such as on the Internet or YouTube [5–7]. In this regard, for cancer patients in need of accurate medical information, the AI-guided bot can serve as a suitable gateway for effective communication and as a means to address unmet needs. Because the information quality and adequacy of ChatGPT output are not accurate in urological patients, [30] to compensate for information adequacy, we trained several guidelines to improve the quality and readability of the information.

4.4. Role of AI-guide bot in AI-driven infodemic

A potential concern of LLMs, such as ChatGPT, is the submissiveness of users in the following instructions. Despite the restrictions imposed by developers, LLMs can be easily used to generate text on controversial topics, including misinformation [9]. LLMs' ability to generate text similar to human-written texts can be used to create fake news or fabricated or misleading content without the reader realizing that the text was generated by AI [35–37]. The spread of misinformation through fast-disseminating channels, such as mass media and social networks, can lead to a phenomenon known as the infodemic [38]. Infodemic can

affect medical decision-making regarding treatment or preventive measures [39]. Social media platforms and other media outlets can quickly disseminate misinformation and generate an abundant of health-related information inaccurately. Given the nature of social media content proliferation, LLMs can be used to create content for a specific target population and promote the spread of misinformation. With these characteristics of LLM, the concept of infodemic was presented as the concept of "AI-driven infodemic" [9].

Medical information provided to patients must be accurate. Furthermore, it should not infringe on the authority of physicians' medical decisions. It may be prudent not to provide any medical information to patients rather than offering false or unverified medical information. This is particularly true for cancer, which is a pathologically complex clinical condition. Failure to provide cancer patients with accurate information can exacerbate confusion rather than alleviate it. While ChatGPT has passed the United States Medical Licensing Exam, [40] it cannot replace human interactions in the pursuit of more inclusive healthcare [9,41]. Additionally, in AI-driven infodemic situations, there is a risk that ChatGPT may provide incorrect answers. If indiscriminate medical information overflows and AI provides AI-generated answers using incorrect information from the Internet, it can confuse the treatments of cancer patients. In this respect, an AI-guide bot was developed to refine inaccurate medical information overflowing on the Internet and to provide accurate medical information to cancer patients.

4.5. Strengths and limitations

In the course of formulating an AI-driven medical guidance chatbot, designated as an AI-guide bot, via integration of the GPT API and Langchain technology, several limitations need to be considered. First, user-derived feedback is lacking. The AI-guide bot must emphasize the importance of garnering authentic user responses. Augmentation of the bot's precision and functional efficacy mandates a systematic aggregation of such feedback, accompanied by a rigorous analysis. Second, there are inherent limitations to the FAISS similarity-check algorithm. This limitation manifests itself in the obstacles concerning the apprehension and application of medical terminological synonyms. Periodic curation and reinforcement of a medical synonym repository are required to improve the accuracy of the query resolution of the AI-guide bot. Third, we aimed to ensure consistent responses by repeating the same question 10 times while fixing the GPT prompts. This approach yielded consistent answers in the vast majority of cases, with only approximately 2–3 % of the responses deviating from the intended question. Finally, the AI-guided bot may have concerns related to data privacy and infringement. Utilization of the GPT API engenders existing anxieties about data privacy and possible violations thereof. Consequently, it is critical to incorporate protocols for personal data detection and obfuscation in the initial phases of development to safeguard personal information and ensure regulatory compliance. By recognizing and ameliorating these impediments, we can further improve the service quality and operational efficacy of the AI-guide bot.

Despite these limitations, this study had several strengths. The greatest advantage is that the chatbot systems used in this study offer significant advantages in improving the quality of medical care. While the AI-guide bot cannot replace medical professionals, it can assist patients in understanding their medical conditions, the importance of their ongoing diagnostic tests, and disease medical treatments by doctors, thus empowering them with self-awareness. The accurate diagnosis and treatment provided by a physician are essential for treating a patient, but equally important is how well patients adhere to treatment of their disease [34]. The AI-guide bot can serve as a supplementary tool for patient healthcare. In real medical care settings, there are instances in which doctors may lack the time to provide thorough explanations to patients, or even if they do, patients may have difficulty comprehending the medical information. In such situations, an AI-guide bot can play a significant role in a patient's medical journey.

The second strength of this study is that the AI-guided bot can provide precise medical information supported by the involvement of clinicians, and the data can be continually improved through learning. Notably, our AI-guide bot has demonstrated robust outcomes regarding AI hallucinations due to its design to provide information solely from within its database. Furthermore, while currently being developed for prostate cancers, another advantage is the possibility of expanding the scope of provision for cancer patients in the future. Lastly, since there are very few chatbot systems in the world for cancer patients in which AI provides medical information that is user-friendly and appropriate for the patient's unmet medical information needs, the development of an AI-guide bot can be considered valuable.

In conclusion, if the healthcare chatbot program for cancer patients developed in this study is used, it will be possible to provide medical information quickly and accurately to patients who are concerned about their health. More follow-up studies are needed to provide accurate information. However, it is believed that the developed chatbot program could help improve the health of cancer patients.

Funding

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No. HR22C1302).

Data availability

Data from this study may be shared upon reasonable request from the corresponding author.

CRediT authorship contribution statement

Jae-woo Lee: Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation. **In-Sang Yoo:** Methodology, Formal analysis, Data curation. **Ji-Hye Kim:** Software, Resources, Methodology, Formal analysis, Data curation. **Won Tae Kim:** Validation, Supervision. **Hyun Jeong Jeon:** Writing – review & editing, Validation, Supervision. **Hyo-Sun Yoo:** Validation, Supervision, Project administration, Investigation. **Jae Gwang Shin:** Project administration, Methodology, Formal analysis, Data curation. **Geun-Hyeong Kim:** Methodology, Formal analysis, Data curation. **ShinJi Hwang:** Software, Project administration, Methodology, Formal analysis, Data curation. **Seung Park:** Writing – review & editing, Validation, Supervision, Software, Data curation, Conceptualization. **Yong-June Kim:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have no potential conflicts of interest to disclose.

Acknowledgment

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No. HR22C1302).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.cmpb.2024.108302](https://doi.org/10.1016/j.cmpb.2024.108302).

References

- [1] B.S. Chhikara, K. Parang, Global Cancer Statistics 2022: the trends projection analysis, *Chem. Biol. Lett.* 10 (2023) 451.
- [2] C. Allemani, T. Matsuda, V. Di Carlo, R. Harewood, M. Matz, M. Nikšić, A. Bonaventure, M. Valkov, C.J. Johnson, J. Estève, O.J. Ogunbiyi, E.S.G. Azevedo, W.Q. Chen, S. Eser, G. Engholm, C.A. Stiller, A. Monnereau, R.R. Woods, O. Visser, G.H. Lim, J. Aitken, H.K. Weir, M.P. Coleman, Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries, *Lancet* 391 (2018) 1023–1075.
- [3] World Health Organization, Cancer Prevention and Control in the Context of an Integrated approach: Report By the Secretariat [Internet], WHO, Geneva, 2016 [cited 2019 Apr 29].
- [4] M.I. Fitch, I. Nicoll, G. Lockwood, Cancer survivor's perspectives on the major challenge in the transition to survivorship, *Patient Educ. Couns.* 103 (2020) 2361–2367.
- [5] M. Wasserman, N.N. Baxter, B. Rosen, M. Burnstein, A.L. Halverson, Systematic review of internet patient information on colorectal cancer surgery, *Dis. Colon Rectum* 57 (2014) 64–69.
- [6] H. Park, E. Kang, Y. Kim, H. Ju, Analysis of the spread of misinformation about Lung Cancer on youtube: based on source of information, *Korean J. Fam. Pract.* 13 (2023) 152–158.
- [7] A.N. Sahin, A.S. Sahin, F. Schwenter, H. Sebang, Youtube videos as a source of information on colorectal cancer: what do our patients learn? *J. Cancer Educ.* 34 (2019) 1160–1166.
- [8] S. Locke, A. Bashall, S. Al-Adely, J. Moore, A. Wilson, G.B. Kitchen, Natural language processing in medicine: a review, *Trends Anaesth. Crit. Care* 38 (2021) 4–9.
- [9] L. De Angelis, F. Baglivo, G. Arzilli, G.P. Privitera, P. Ferragina, A.E. Tozzi, C. Rizzo, ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health, *Front. Public Health* 11 (2023) 1166120.
- [10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, Language models are few-shot learners, *Adv. Neural Inf. Process. Syst.* 33 (2020) 1877–1901.
- [11] R.S. Goodman, J.R. Patrinely, C.A. Stone Jr., E. Zimmerman, R.R. Donald, S. S. Chang, S.T. Berkowitz, A.P. Finn, E. Jahangir, E.A. Scoville, T.S. Reese, D. L. Friedman, J.A. Bastarache, Y.F. van der Heijden, J.J. Wright, F. Ye, N. Carter, M. R. Alexander, J.H. Choe, C.A. Chastain, J.A. Zic, S.N. Horst, I. Turker, R. Agarwal, E. Osmundson, K. Idrees, C.M. Kiernan, C. Padmanabhan, C.E. Bailey, C. E. Schlegel, L.B. Chambless, M.K. Gibson, T.J. Osterman, L.E. Wheelless, D. B. Johnson, Accuracy and reliability of chatbot responses to physician questions, *JAMA Netw. Open* 6 (2023) e2336483.
- [12] The Korean prostate society. Prostate Cancer Guidelines, 2021.
- [13] The Korean urological oncology society. Guidance for Prostate Cancer Patients. 2019.
- [14] The Korean Urological Association. The medical treatment of prostate hyperplasia. 2020.
- [15] The Korean Cancer Association. Nutrition and physical activity guidelines for cancer patients. 2019.
- [16] K.A. Moses, P.C. Sprenkle, C. Bahler, G. Box, S.V. Carlsson, W.J. Catalona, D. M. Dahl, M. Dall'Era, J.W. Davis, B.F. Drake, NCCN guidelines® insights: prostate cancer early detection, version 1.2023: featured updates to the NCCN guidelines, *J. Natl. Compr. Cancer Netw.* 21 (2023) 236–246.
- [17] K.S. Virgo, R.B. Rumble, R. de Wit, D.S. Mendelson, T.J. Smith, M.-E. Taplin, J. L. Wade III, C.L. Bennett, H.I. Scher, P.L. Nguyen, Initial management of noncastrate advanced, recurrent, or metastatic prostate cancer: ASCO guideline update, *J. Clin. Oncol.* 39 (2021) 1274–1305.
- [18] N. Mottet, R.C.N. van den Bergh, E. Briers, T. Van den Broeck, M.G. Cumberbatch, M. De Santis, S. Fanti, N. Fossati, G. Gandaglia, S. Gillesen, N. Grivas, J. Grummet, A.M. Henry, T.H. van der Kwast, T.B. Lam, M. Lardas, M. Liew, M.D. Mason, L. Moris, D.E. Oprea-Lager, H.G. van der Poel, O. Rouvière, I.G. Schoots, D. Tilki, T. Wiegel, P.M. Willemsse, P. Cornford, EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. part 1: screening, diagnosis, and local treatment with curative intent, *Eur Urol* 79 (2021) 243–262.
- [19] The Korean Prostate Society, Getting to Know the Prostate, 4th ed., Ilchokak, 2023.
- [20] The Korean Urological Association, Urology 6th Edition, Ilchokak, 2019.
- [21] HealthBreeze, Medical Animation <https://www.healthbreeze.com/>.
- [22] M. Schwarz, K. Chapman, B. Häussler, Multilingual medical entity recognition and cross-lingual zero-shot linking with facebook AI similarity search, (2022).
- [23] OpenAI, Introducing ChatGPT. 2024 <https://openai.com/blog/chatgpt/>.
- [24] LangChain 2024 https://python.langchain.com/docs/get_started/introduction.
- [25] J. Rudolph, S. Tan, S. Tan, War of the chatbots: bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education, *J. Appl. Learn. Teach.* 6 (2023).
- [26] R. Shams, Semi-supervised classification for natural language processing, arXiv preprint arXiv:1409.7612, (2014).
- [27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017) 30.
- [28] J.C. Chow, L. Sanders, K. Li, Impact of ChatGPT on medical chatbots as a disruptive technology, *Front. Artif. Intell.* 6 (2023) 1166014.
- [29] J.C.L. Chow, L. Sanders, K. Li, Design of an educational chatbot using artificial intelligence in radiotherapy, *AI* 4 (2023) 319–332.
- [30] A. Cocci, M. Pezzoli, M. Lo Re, G.I. Russo, M.G. Asmundo, M. Fode, G. Cacciamani, S. Cimino, A. Minervini, E. Durukan, Quality of information and appropriateness of ChatGPT outputs for urology patients, *Prostate Cancer Prostatic Dis.* (2023).

- [31] K.A. Luker, K. Beaver, S.J. Leinster, R.G. Owens, L.F. Degner, J.A. Sloan, The information needs of women newly diagnosed with breast cancer, *J. Adv. Nurs.* 22 (1995) 134–141.
- [32] M. Cawley, J. Kostic, C. Cappello, Informational and psychosocial needs of women choosing conservative surgery/primary radiation for early stage breast cancer, *Cancer Nurs.* 13 (1990) 90–94.
- [33] D.E. Harrison-Woermke, J.E. Graydon, Perceived informational needs of breast cancer patients receiving radiation therapy after excisional biopsy and axillary node dissection, *Cancer Nurs.* 16 (1993) 449–455.
- [34] L.J.F. Rutten, N.K. Arora, A.D. Bakos, N. Aziz, J. Rowland, Information needs and sources of information among cancer patients: a systematic review of research (1980–2003), *Patient. Educ. Couns.* 57 (2005) 250–261.
- [35] R.F. Sear, R. Leahy, N.J. Restrepo, Y. Lupu, N.F. Johnson, Machine learning language models: achilles heel for social media platforms and a possible solution, *Adv. Artif. Intell. Mach. Learn.* 1 (2021) 191–202.
- [36] A. Gupta, A. Singhal, A. Mahajan, A. Jolly, S. Kumar, Empirical framework for automatic detection of neural and human authored fake news, in: *Proceedings of the 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2022, pp. 1625–1633.
- [37] K. McGuffie, A. Newhouse, The radicalization risks of GPT-3 and advanced neural language models, *arXiv preprint arXiv:2009.06807*, (2020).
- [38] The lancet infectious D. The COVID-19 infodemic, *Lancet Infect. Dis.* 20 (2020) 875.
- [39] S.C. Briand, M. Cinelli, T. Nguyen, R. Lewis, D. Prybylski, C.M. Valensise, V. Colizza, A.E. Tozzi, N. Perra, A. Baronchelli, Infodemics: a new challenge for public health, *Cell* 184 (2021) 6010–6014.
- [40] T.H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, V. Tseng, Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models, *PLOS Digit. Health* 2 (2023) e0000198.
- [41] J.E.H. Brown, J. Halpern, AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare, *SSM Mental Health* 1 (2021) 100017.