jdhs
Journal of Dental Hygiene Science

Performance of ChatGPT on the Korean National Examination for Dental Hygienists

Soo-Myoung Bae[1,2,3], Hye-Rim Jeon[1], Gyoung-Nam Kim[1], Seon-Hui Kwak[1,2], and Hyo-Jin Lee[1,2,3,†]

[1]Department of Dental Hygiene, College of Dentistry, Gangneung−Wonju National University, Gangneung 25457, [2]Research Institute of Dental Hygiene Science, Gangneung−Wonju National University, Gangneung 25457, [3]Research Institute of Oral Science, Gangneung−Wonju National University, Gangneung 25457, Korea

Background: This study aimed to evaluate ChatGPT's performance accuracy in responding to questions from the national dental hygienist examination. Moreover, through an analysis of ChatGPT's incorrect responses, this research intended to pinpoint the predominant types of errors.
Methods: To evaluate ChatGPT−3.5's performance according to the type of national examination questions, the researchers classified 200 questions of the 49th National Dental Hygienist Examination into recall, interpretation, and solving type questions. The researchers strategically modified the questions to counteract potential misunderstandings from implied meanings or technical terminology in Korea. To assess ChatGPT−3.5's problem−solving capabilities in applying previously acquired knowledge, the questions were first converted to subjective type. If ChatGPT−3.5 generated an incorrect response, an original multiple−choice framework was provided again. Two hundred questions were input into ChatGPT−3.5 and the generated responses were analyzed. After using ChatGPT, the accuracy of each response was evaluated by researchers according to the types of questions, and the types of incorrect responses were categorized (logical, information, and statistical errors). Finally, hallucination was evaluated when ChatGPT provided misleading information by answering something that was not true as if it were true.
Results: ChatGPT's responses to the national examination were 45.5% accurate. Accuracy by question type was 60.3% for recall and 13.0% for problem−solving type questions. The accuracy rate for the subjective solving questions was 13.0%, while the accuracy for the objective questions increased to 43.5%. The most common types of incorrect responses were logical errors 65.1% of all. Of the total 102 incorrectly answered questions, 100 were categorized as hallucinations.
Conclusion: ChatGPT−3.5 was found to be limited in its ability to provide evidence−based correct responses to the Korean national dental hygiene examination. Therefore, dental hygienists in the education or clinical fields should be careful to use artificial intelligence−generated materials with a critical view.

Key Words: Artificial intelligence, ChatGPT, Dental hygiene, Large language models, National examination

Introduction

1. Background

OpenAI's ChatGPT (Chat Generative Pre-Trained Transformer) is an interactive artificial intelligence (AI) service that generates responses to users' queries using a pre-trained deep learning algorithm that has been made publicly available. ChatGPT is the latest in a class of large-scale language models known as large language models (LLMs). LLMs are pre-trained language models that can be further trained or fine-tuned thereby expanding the range of applications and scope of LLM[1,2]. Compared to other LLMs, ChatGPT was reported to have a distinct improvement in conversation output[3]. ChatGPT goes beyond responding to user-written questions to write essays, translate languages, or generate program code because

ChatGPT was designed to respond to a wide range of topics[4]. ChatGPT, a generative AI, autonomously creates texts, images, and videos using a self-learning algorithm. It's often referred to as a "Hyperscale AI" due to its capacity to process vast amounts of data[2]. ChatGPT interacts conversationally and provides answers within a natural context, unlike search engines like Google, which deliver a list of websites related to the user's search.

ChatGPT has been used for various tasks, including programming, translation, writing, speech writing, and, more recently, in research contexts[1,2]. Sarraju et al.[5] reported the capability of ChatGPT by evaluating its responses to 25 questions related to the prevention of cardiovascular disease. The aim was to determine how effectively ChatGPT could handle questions concerning basic concepts in cardiovascular disease prevention. The findings revealed that ChatGPT provided appropriate responses to 21 of the 25 questions, achieving an adequacy rate of 84%. These findings emphasized the potential of conversational AI, such as ChatGPT, to enhance clinical workflows, particularly in facilitating patient education and improving communication between patients and physicians regarding the prevention of cardiovascular diseases[5]. Savery et al.[6], assessed ChatGPT's capability in domain-specific conversations as an interactive chatbot for psychotherapy. LLMs have also been explored in the medical fields as tools for direct interaction with patients and educational resources[7]. In addition, the performance of ChatGPT in the United States Medical Licensing Examination (USMLE) has been evaluated, with the model correctly answering over 60% of the questions in the first and second exams. Such a performance level, comparable to that of a third-year medical student, suggested ChatGPT's viability as an educational tool in medical training[2]. Wójcik et al.[8] evaluated the performance of ChatGPT in medical education and suggested that ChatGPT may be used as a valuable assistance tool in medical education. However, they emphasized that the tool cannot completely replace human knowledge and expertise. In Korea, Huh[9] suggested that ChatGPT's knowledge and interpretation ability for parasitology examination were not yet comparable to those of medical students.

To use ChatGPT as an effective educational tool in the domains of dental hygiene and dentistry, the answers it provides must be academically valid and reliable. However, the accuracy of these answers must undergo rigorous evaluation. Few studies are exploring the use of ChatGPT within educational contexts, especially within dental hygiene in Korea[10]. Since 1975, the national examination for dental hygienists in Korea has been conducted by the government, with the responsibility for its administration being transferred in 1998 to the Korea Health Personnel Licensing Examination Institute. The primary objective of this examination is to assess whether dental hygienists possess the necessary clinical practice competencies related to dental hygiene. Thus, the national dental hygienist examination is conducted as a standard to determine if a dental hygienist has the requisite knowledge, attitude, and skills to fulfill their professional responsibilities[11].

2. Objectives

This study aimed to evaluate ChatGPT's performance accurately in responding to questions from the national dental hygienist examination by applying relevant knowledge, including theories and laws pertinent to dental hygiene. Moreover, through an analysis of ChatGPT's incorrect responses, this research intends to pinpoint the predominant types of errors. The goal is to evaluate ChatGPT's potential as a learning tool in the domain of dental hygiene education.

Materials and Methods

1. Study design

This was an exploratory study conducted to examine the applicability of ChatGPT as a learning tool in the field of dental hygiene by entering dental hygienist national examination questions into ChatGPT and analyzing the content it generates.

2. Study participants

The participants of this study were the responses generated by ChatGPT-3.5 to 200 questions of the 49th national dental hygienist examination in Korea. The questions of the national examination were provided from the publicly available data on the website of the Korea Health Personnel Licensing Examination Institute. The institute

provides the national examination questions and the national examination analysis report in the second half of every year. Regarding the study date, the 50th National Dental Hygienist Examination Analysis report had not been published hence the 49th National Dental Hygienist examination was chosen as the latest data for the national examination analysis report. The national dental hygienist examination consisted of 20 questions on the Health and Medical related Law, 80 questions on dental hygiene I (basic dental hygiene, dental hygiene management), and 100 questions on dental hygiene II (clinical dental hygiene).

3. Study tools

In this study, the ChatGPT-3.5 version was selected for its advanced natural language processing capabilities to generate responses to questions from the national dental hygienist examination. To systematically analyze the answers provided by ChatGPT, a comprehensive analysis framework was developed. This framework included several components: the original questions in Korean, questions augmented with English terminology, question types, ChatGPT's responses, categorization into correct and incorrect responses, and classification of the types of inaccuracies observed. This structured approach allowed for an evaluation of ChatGPT-3.5's utility as an educational tool in the field of dental hygiene, focusing on its ability to accurately interpret and respond to examination questions[12].

4. Analysis methods

To systematically analyze the national dental hygienist examination using ChatGPT-3.5, a structured process was followed, as illustrated in Fig. 1.

1) Classification of the types of national examination questions

To evaluate ChatGPT-3.5's performance according to the type of national examination questions, five researchers read repeatedly 200 questions of the 49th National Dental Hygienist Examination, and classified them into 66 recall, 112 interpretation, and 22 solving type questions based on
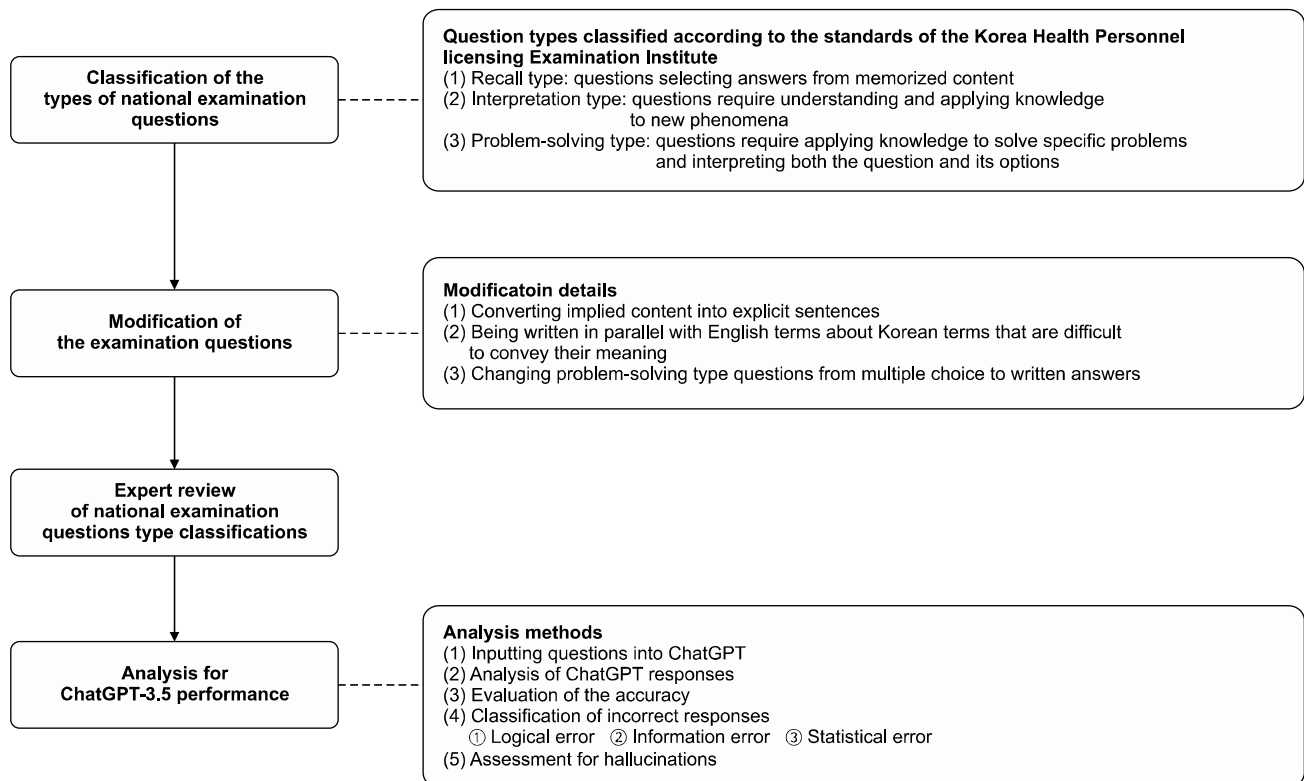

Fig. 1. ChatGPT analysis flow.

the definitions of question types provided by the Korea Health Personnel Licensing Examination Institute[13]. The Korea Health Personnel Licensing Examination Institute defined a recall-type question as a question that can be answered by simply remembering the memorized learning content at the level of memory. An interpretation-type question was defined as a question that can be answered by fully understanding the knowledge and dealing with new phenomena based on the knowledge. A problem-solving type question was defined as a question to solve a specific problem by applying knowledge, which requires the ability to not only interpret the information in the question but also to interpret the meaning of each option.

2) Modification of the examination questions

Given the disparity in ChatGPT's training data volume between English and Korean, which potentially affects its comprehension of technical terms and nuanced Korean sentence constructions[14], we implemented strategic modifications to the national examination questions. To counteract potential misunderstandings arising from implied meanings or technical terminology in Korean, sentences ending with phrases such as "～Which is correct?," "～structure?," or "～tooth?" were rephrased for clarity. For instance, questions were standardized to the format "Which of the following is correct as an explanation for ～?", thereby converting implicit queries into explicit ones to better convey the intended meaning. For example, the following questions were asked in Korean. "Which of the following is a correct description of the operation of a hand instrument when it is used in the patient's mouth?" Furthermore, to bridge the language gap, we added English terms

for each term when the questions and options contained only Korean or Chinese terms.

Moreover, to assess ChatGPT-3.5's problem-solving capabilities in applying previously acquired knowledge, problem-solving questions were first converted to subjective type (Table 1). This modification removed the original multiple-choice framework of these items. If ChatGPT-3.5 generated an incorrect response when inputting a subjective-type solution question, we asked the 'objective problem-solving questions' that provided an original multiple-choice framework again, checked the answer, and checked the accuracy of the answer according to the subjective and objective-type questions. All questions were entered in modified form at the Chat GPT-3.5 prompt verbatim as the examples in Table 1.

3) Expert review of national examination question type classifications

An expert review process was conducted on the categorization of national examination questions, as initially classified by five researchers according to the type of definitions provided by the Korea Health Personnel Licensing Examination Institute. The expert was selected as a professor with more than 10 years of experience in teaching dental hygiene classes and experience in developing dental hygienist national examination questions. Based on the expert's review, the 200 questions on the national dental hygienist examination were finally categorized into 63 recall, 114 interpretation, and 23 problem-solving-type questions.

Table 1. Example of Converting Objective Problem-Solving Type Question into Subjective Question

Objective problem-solving type question	Subjective problem-solving type question
This is the result of an oral examination of a 5-year-old child. What is the appropriate space retainer? • The left mandibular canine is healthy. • The mandibular left first premolar is missing. • Mandibular left second premolar is pulp treated. 　① Lingual arch 　② Distal shoe 　③ Band & loop 　④ Crown & loop 　⑤ Nance holding arch	Below are the results of an oral examination of a 5-year-old child. Which space maintainer is appropriate for this child? • The left mandibular canine is healthy. • The mandibular left first premolar is missing. • Mandibular left second premolar is pulp treated.

4) Analysis of ChatGPT-3.5 performance

This study evaluated the accuracy of ChatGPT-3.5 responses by inputting 200 questions into ChatGPT-3.5. The researchers confirmed the answers generated by ChatGPT and compared them with the contents of reliable materials such as dental hygiene textbooks reading them several times. The accuracy was calculated as the percentage of correct answers among the total answers of ChatGPT. When ChatGPT's responses were incorrect, the types of incorrect answers were classified as logical, information, and statistical errors[1]. A logical error was defined as an error that occurred when ChatGPT properly found information related to a question but did not correctly derive the correct answer to that question. An informational error was defined as when ChatGPT did not properly identify the key information needed to solve the input question. A statistical error was categorized as an arithmetic error in questions that required calculations. Finally, hallucination was defined when ChatGPT provided misleading information by answering something that was not true as if it were true[12,14]. The error classification process for each question was verified twice, and the cross-review within the research team was conducted twice and the results were summarized. The Kappa score for interrater reliability was 0.778.

5. Statistical analysis

Statistical analyses were performed using IBM SPSS statistics ver. 28.0 (IBM Corp., Armonk, NY, USA). Frequency analysis was conducted on the correct and incorrect responses to the 200 questions to calculate the correct and incorrect response rate. A chi-square test was conducted to examine the differences in the accuracy of responses by question type, and a McNemar test was conducted to examine whether there was a significant change in the accuracy of responses between the subjective-type solving questions and objective-type questions. p-values of less than 0.05 were considered significant.

Results

1. Accuracy of ChatGPT responses according to the question type

The accuracy of ChatGPT-3.5 answers to the questions is shown in Table 2. The accuracy of ChatGPT answers for all 200 questions on the national exam was 45.5%. According to the question type, recall-type questions had the highest accuracy at 60.3%, while problem-solving-type questions had the lowest accuracy at 13.0% ($p < 0.001$).

2. Comparison of the accuracy of subjective and objective problem−solving questions

ChatGPT-3.5 had an accuracy of 13.0% for subjective problem-solving questions and 43.5% for objective problem-solution questions, showing that adding a multiple-choice fra-

Table 2. Accuracy of ChatGPT Responses by Dental Hygienist National Board Exam Question Type

	Number of questions	Accuracy	Inaccuracy	p-value[a]
Total	200 (100.0)	91 (45.5)	109 (54.5)	< 0.001
Recall type	63 (31.5)	38 (60.3)	25 (39.7)	
Interpretation type	114 (57.0)	50 (43.9)	64 (56.1)	
Problem-solving type	23 (11.5)	3 (13.0)	20 (87.0)	

Values are presented as n (%).
[a]Calculated by a chi-square test.

Table 3. Accuracy for Subjective and Objective Problem−Solving Questions

	The number of question	Subjective problem-solving question		Objective problem-solving question		p-value[a]
		Accuracy	Inaccuracy	Accuracy	Inaccuracy	
Problem-solving type	23	3 (13.0)	20 (87.0)	10 (43.5)	13 (56.5)	0.016

Values are presented as n (%).
[a]Calculated by a McNemar test.

mework significantly increases accuracy (p=0.016) (Table 3).

3. Comparison of error classification for incorrect responses of ChatGPT according to the question type

ChatGPT-3.5 responded to 109 out of 200 questions incorrectly. When analyzing the error types for the 109 incorrect responses, the proportion of logical errors was 65.1%, with no statistical errors. When analyzing the types of incorrect responses by question type, logical errors were the most common for recall type among all types of questions (Table 4).

4. Comparison of incorrect response errors in ChatGPT for problem–solving type questions

A total of 13 questions were categorized as incorrect for objective problem-solving questions. The errors for the 13 incorrect responses were logical errors of 76.9%, informational errors of 23.1%, and no statistical errors. When comparing the information errors of the subjective problem-solving questions with the objective problem-solving questions, the information errors decreased in the incorrect responses of the objective problem-solving questions (Table 5). Of the total 102 incorrectly answered questions included as objective problem-solving questions, 100 were categorized as hallucinations.

Discussion

1. Comparison with previous studies and suggestions

In this study, 200 questions from the 49th National Dental Hygienist examination were input into ChatGPT-3.5 one by one to evaluate if ChatGPT-3.5 can accurately derive the correct answers with appropriate dental hygiene knowledge, and then the answers generated were confirmed for accuracy and the types of incorrect answers were analyzed. As a result, ChatGPT-3.5 showed an accuracy rate of 45.5% for 200 questions. Conversely, Kung et al.[1] showed that ChatGPT answered Steps 1 and 2 of more than 60% of all USMLE correctly. The results of a previous study by Kim et al.[12], showed that users could get answers closer to the desired content when inputting questions in English than when inputting questions in Korean in the ChatGPT-3.5 system. It is expected that the correct answer rate will be higher if inputting the questions in English in this study.

The accuracy rate for each type of question on the national dental hygienist examination was 60.3% for recall-type questions, 43.9% for interpretation-type questions, and 13.0% for problem-resolving type questions, with the highest accuracy rate for recall-type questions. Recall-type questions are those that can be answered by recalling the

Table 4. Error Classification for Incorrect Responses of ChatGPT According to the Question Type

Question types	Incorrect number of question	Logical error	Information error	Statistical error	p-value[a]
Total	109	71 (65.1)	38 (34.9)	0 (0.0)	0.848
Recall type	25	17 (68.0)	8 (32.0)	0 (0.0)	
Interpretation type	64	42 (65.6)	22 (34.4)	0 (0.0)	
Problem-solving type	20	12 (60.0)	8 (40.0)	0 (0.0)	

Values are presented as n (%).
[a]Calculated by a chi-square test.

Table 5. Errors of Incorrect Responses in ChatGPT According to Problem–Solving Type Questions

Error for subjective problem-resolving questions				Error for objective problem-resolving questions			
Incorrect question	Logical error	Information error	Statistical error	Incorrect question	Logical error	Information error	Statistical error
20	12 (60.0)	8 (40.0)	0 (0.0)	13	10 (76.9)	3 (23.1)	0 (0.0)

Values are presented as n (%).

memorized learning content, while interpretation-type questions are those that check the ability to fully understand the learned knowledge and cope with new situations based on existing knowledge. From these results, it can be concluded that ChatGPT-3.5's ability to apply knowledge to judge the clinical data given is weaker than the ability to simply answer the memorized content.

To get a more accurate interpretation of ChatGPT's problem-solving performance, we first evaluated its accuracy on subjective problem-solving questions with the multiple-choice items removed. As a result, we found that ChatGPT-3.5's accuracy for subjective problem-solving questions was 13.0%. For questions classified as incorrect, the accuracy of ChatGPT-3.5 was increased to 43.5% by giving the correct answer to the previously classified incorrect questions after providing additional options. In addition, the accuracy of ChatGPT-3.5 for problem-solving questions in this study was found to be higher with multiple-choice items than without multiple-choice items. Gilson et al.[2] found that ChatGPT increased the percentage of correct answers when additional information was provided during the problem-solving process, which is consistent with the results of this study. Therefore, it appeared that ChatGPT-3.5 utilized more information for problem-solving questions when options were present or more information was provided by the user, resulting in more accurate answers.

In this study, we analyzed the errors according to the question type and found that logical errors were higher than informational errors for the recall, interpretation, and problem-solving types of questions, and no statistical errors were found. Logical errors were when ChatGPT properly found information about a problem but failed to properly translate that information into a response, while informational errors were when they failed to properly identify key information needed to solve a given problem. From these results, it could be interpreted that ChatGPT-3.5 was relatively good at identifying key information about the input question, but was limited in its ability to properly translate it into a correct response. A study evaluating the accuracy of ChatGPT on the USMLE[2] reported that inaccurate responses were more likely to contain logical errors than information errors. Because ChatGPT

is constantly changing the way it learns depending on the model, and its functionality depends on the amount of information covered and accumulated, it is difficult to draw clear conclusions about its accuracy and types of incorrect answers at this point, but we believe that the results of this study and the studies reported to date show similar patterns.

In addition, ChatGPT-3.5 responded to 100 of the 102 questions categorized as incorrect by stating untrue statements as if they were true. This suggested that ChatGPT's ability to properly find relevant information about the question and derive the right answer was less than its ability to get to the point of the matter. The previous study[2] suggested that ChatGPT users should be aware that the answers given by ChatGPT may be well-founded but inaccurate. Other previous studies have suggested that ChatGPT could be utilized as a valuable tool in medical education, but emphasized that ChatGPT cannot fully replace human expertise[8,9]. Therefore, the users should continue to attempt to assess the accuracy of ChatGPT's responses independently of their perception of ChatGPT's ability to answer any question. Based on this study and previous reports, it was not yet possible to fully trust the responses of ChatGPT-3.5. Therefore, to use ChatGPT-3.5 as a tool for educational and clinical applications in dental hygiene, it is recommended that dental hygiene majors and dental hygienists with dental hygiene knowledge finally confirm the accuracy of ChatGPT's answers rather than blindly trusting the chatbot's answers. In other words, dental hygienists in the education and clinical fields will need to have the right knowledge and the ability to utilize AI-generated data with a critical view. When dental hygienists use the ChatGPT in the education and clinical fields, dental hygienists should be able to respond to the error by comprehensively utilizing their knowledge and expertise as a way to deal with each error type. Therefore, based on the results of this study, we should consider the direction of dental hygiene education in the era of the Fourth Industrial Revolution. There have been various attempts to utilize AI, ChatGPT, etc. in the medical and educational fields, but it would not be possible to completely replace human intelligence. In the current era where a myriad of diverse information is poured out, it is necessary to iden-

tify and distinguish the fields that could not be handled only by AI, not humans, and the competencies that cannot be replaced, and to train human resources who could perform such competencies in these fields. In addition, it was necessary to train human resources who could fully utilize and apply AI, which was developing in various forms, rather than distinguishing it regardless of AI.

2. Limitations

This study has several limitations. When this study was first conducted, the ChatGPT-4 model had not been released, so it was not possible to confirm the functional differences with ChatGPT-4 as the newer model. Therefore, it is necessary to evaluate whether the ChatGPT-4 model could utilize the correct dental hygiene knowledge base to generate appropriate answers in future studies. In addition, since ChatGPT developed by OpenAI was significantly lacking in Korean information compared to English, caution seems necessary when interpreting the accuracy of ChatGPT-3.5 based on the results of this study. It is necessary to conduct a study to fully evaluate the functionality of ChatGPT by inputting and utilizing English information. Since ChatGPT is constantly being improved and supplemented, it should be kept in mind that if the dental hygienist national examination questions used in this study are input into ChatGPT again in the future, they might generate different answers from this study and might show higher correct answer rate[15]. Nevertheless, the significance of this study was that it was the first study to evaluate the accuracy and usefulness of ChatGPT in Korean dental hygiene education. Based on the results of this study, dental hygiene majors or dental hygienists should be able to independently determine the accuracy of their responses when using ChatGPT as a tool for dental hygiene education and dental hygienist practice.

3. Conclusion

As a result of evaluating the performance of ChatGPT-3.5 on the Korean national dental hygienist examination questions, the overall accuracy of the responses was 45.5%. According to the question type, the percentage of correct responses was significantly lowest for the problem-solving questions and the highest for the recall-type que-

stions. There were also found that logical errors were higher than information errors in the types of incorrect responses. It was evaluated that ChatGPT-3.5 was relatively unable to provide accurate answers in the problem-solving process for the Korean national dental hygienist examination, especially in the case of interpretation or problem-solving type that utilized and applies knowledge to solve compared to recall-type questions that need to be solved with simple knowledge. Therefore, dental hygiene majors or dental hygienists who want to use ChatGPT in dental hygiene should be able to independently confirm the accuracy of their responses when using ChatGPT.

Notes

Conflict of interest

No potential conflict of interest relevant to this article was reported.

Ethical approval

Because this study was not research involving humans, Institutional Review Board approval was not required.

Author contributions

ORCID

Soo-Myoung Bae, *https://orcid.org/0000-0002-1802-4129*
Hye-Rim Jeon, *https://orcid.org/0009-0009-6528-4095*
Gyoung-Nam Kim, *https://orcid.org/0009-0001-2311-9050*
Seon-Hui Kwak, *https://orcid.org/0000-0001-6407-0890*
Hyo-Jin Lee, *https://orcid.org/0000-0002-3465-1747*

Acknowledgements

Data availability

Raw data is provided at the request of the corresponding author for reasonable reason.

References

1. Kung TH, Cheatham M, Medenilla A, et al.: Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2: e0000198, 2023.
https://doi.org/10.1371/journal.pdig.0000198

2. Gilson A, Safranek CW, Huang T, et al.: How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 9: e45312, 2023. https://doi.org/10.2196/45312 Erratum in: JMIR Med Educ 10: e57594, 2024.
https://doi.org/10.2196/57594

3. Brown TB, Mann B, Ryder N, et al.: Language models are few-shot learners. Paper presented at: 34th International Conference on Neural Information Processing Systems; 2020 Dec 6-12; Vancouver, Canada. Red Hook: Curran Associates Inc., 2020.

4. JAGRAN Josh: What is ChatGPT: definition, how to use it. Retrieved January 31, 2024, from: https://www.jagranjosh.com/general-knowledge/what-is-chat-gpt-1676870553-1 (2023, February 20).

5. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L: Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. JAMA 329: 842-844, 2023.
https://doi.org/10.1001/jama.2023.1044

6. Savery M, Abacha AB, Gayen S, Demner-Fushman D: Question-driven summarization of answers to consumer health questions. Sci Data 7: 322, 2020.
https://doi.org/10.1038/s41597-020-00667-z

7. Das A, Selek S, Warner AR, et al.: Conversational bots for psychotherapy: a study of generative transformer models using domain-specific dialogues. In: Demner-Fushman D, Cohen KB, Ananiadou S, Tsujii J, eds. Proceedings of the 21st Workshop on Biomedical Language Processing. Association for Computational Linguistics, Dublin, pp.285-297, 2022.
https://doi.org/10.18653/v1/2022.bionlp-1.27

8. Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz J: Reshaping medical education: performance of ChatGPT on a PES medical examination. Cardiol J 2023. doi: 10.5603/cj.97517. [Epub ahead of print]

9. Huh S: Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. J Educ Eval Health Prof 20: 1, 2023.
https://doi.org/10.3352/jeehp.2023.20.1

10. Shin DK, Jung HK, Lee YS: Exploring the potential of using ChatGPT as a content-based English learning and teaching tool. J Korea Eng Educ Soc 22: 171-192, 2023.
https://doi.org/10.18649/jkees.2023.22.1.171

11. Kim SH, Song KH, Kwun HS, et al.: Dental hygienist job analysis for item development for the Korean dental hygienists' licensing examination. J Educ Eval Health Prof 2: 59-74, 2005.
https://doi.org/10.3352/jeehp.2005.2.1.59

12. Kim SM, Kim JH, Choi MJ, Jeong SH: Evaluation of the applicability of ChatGPT in biological nursing science education. J Korean Biol Nurs Sci 25: 183-204, 2023.
https://doi.org/10.7586/jkbns.23.0013

13. Kim JG: Analysis report of the 49th national dental hygienist examination in 2021. Korea Health Personnel Licensing Examination Institute, Seoul, pp.1-33, 2022.

14. Kim SW, Suh JY, Park JY, Kim DG: Evaluating Korean language capability of ChatGPT and discussion. Korean Inst Inf Sci Eng Proc Korea Comput Congr 2023: 286-288, 2023.

15. Sun GH, Hoelscher SH: The ChatGPT storm and what faculty can do. Nurse Educ 48: 119-124, 2023.
https://doi.org/10.1097/NNE.0000000000001390