ORIGINAL ARTICLE

pISSN 1226-6396, eISSN 2234-4942
J of Oriental Neuropsychiatry 2024;35(4):413-427
https://doi.org/10.7231/jon.2024.35.4.413

JON Journal of Oriental Neuropsychiatry

Development and Application of a ChatGPT-Based Simulation Training Platform for Korean Medicine

Jin-Woo Ryu[#], Chan-Young Kwon[*,#], Jea-Seong Park[†], Seong-Ryong Lim[†], Hye-Li Jeon, Hui-Jun Kim[‡], Sung-Hee Kim[†]

Departments of Artificial Intelligence, *Oriental Neuropsychiatry, [†]Industrial ICT Technology Engineering, [‡]IT Convergence, Dong-eui University

Received: October 30, 2024
Revised: November 22, 2024
Accepted: December 11, 2024

Correspondence to
Sung-Hee Kim
Department of Industrial ICT Technology Engineering, Dong-eui University, 176 Eomgwang-ro, Busanjin-gu, Busan, Korea.
Tel: +82-51-890-2366
Fax: +82-0505-182-6866
E-mail: sh.kim@deu.ac.kr
[#]The first two authors contributed equally to this study.

Acknowledgement
This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MIST) (IITP-2024-RS-2020-II201791).

Objectives: This study aimed to develop a ChatGPT-based stimulation training platform for Korean medicine (KM) education to provide students with virtual patient interactions and enhance their clinical competencies and evaluate its performance.

Methods: A web-based platform integrating ChatGPT API was developed using Node.js and MongoDB with a react-based frontend interface. The platform included functions for instructors to create patient scenarios and for students to conduct simulated clinical consultations. The system was piloted with 47 fourth-year KM students in a KM Neuropsychiatry course, focusing on suicide risk assessment. Student perceptions were evaluated through an anonymous survey measuring satisfaction with simulated consultation experience and AI chatbot usability.

Results: Of a total of 47 students, 42 completed the survey with a response rate of 89.4%. Simulation experience received high satisfaction scores across multiple dimensions, including interest (4.44±0.62), appropriateness (4.71±0.38), knowledge acquisition (4.24±0.79), and skill development (4.14±0.62). The chatbot system demonstrated strong usability scores in usefulness (4.40±0.54), accessibility (4.42±0.56), and overall discoverability (4.43±0.63). However, students expressed some skepticism about the authenticity of virtual patient responses (3.69±0.95).

Conclusions: The platform shows promise as an accessible and standardized tool for clinical education in KM. While it effectively addresses resource constraints in traditional standardized patient programs, its low authenticity ratings suggest that it should be used to complement rather than replace conventional clinical training methods. Future developments should focus on enhancing response authenticity and expanding the range of clinical scenarios while maintaining high levels of accessibility and user satisfaction.

Key Words: Korean medicine, Education, Clinical performance examination, ChatGPT.

Copyright © 2024 by The Korean Society of Oriental Neuropsychiatry. All rights reserved.
This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. INTRODUCTION

Korean medicine (KM) education has received increasing attention, with efforts focused on enhancing educational content, curriculum design, and teaching methodologies[1]. However, a 2020 study highlighted areas of concern, revealing that overall student satisfaction with the KM curriculum was relatively low, averaging 2.83±1.08 out of 5, which indicates a suboptimal outcome[2]. Key sources of dissatisfaction included "difficulty in applying the curriculum to clinical practice after graduation (23.7%)," "ineffective lecture content and methods (20.0%)," and "insufficient integration between subjects (16.9%)," underscoring the need for more structured and practice-oriented clinical education[2].

In response to global trends in higher education, which emphasize competency-based and outcome-driven learning models, the KM education system has developed a competency framework for KM doctors[3]. This framework prioritizes competencies such as optimal patient care, effective communication, professional development, social responsibility, and efficient healthcare management[4]. Among these, the ability to deliver optimal patient care is considered the most critical competency for KM doctors[4]. To support this goal, the clinical practice examination (CPX) has become an integral component of KM education, serving as an assessment tool that evaluates students' competencies in history taking, physical examination, and patient education through simulated clinical encounters with standardized patients[5].

Despite the importance of CPX in training KM doctors, its implementation faces challenges, primarily due to the substantial human and material resources required for securing and managing standardized patients[6]. These logistical and financial constraints have hindered the widespread adoption of CPX across KM schools[6]. Moreover, the quality of education can be affected by the variability in recruiting and training standardized patients, prompting the need for alternative solutions[6].

A promising alternative is the integration of large language models (LLMs) such as ChatGPT, a generative AI model that has gained significant attention since its release in 2022. LLMs have evolved rapidly in scale and now exhibit advanced contextual understanding and cognitive reasoning abilities[7]. ChatGPT, with its capacity for natural language processing and generation, has increasingly been applied in various fields, including medical education, where it serves as a cost-effective and accessible tool[8]. By assisting students in understanding complex medical procedures and diagnostic criteria, ChatGPT can generate realistic scenarios that allow students to practice diagnostic and treatment planning[9]. Furthermore, as an interactive AI, ChatGPT can go beyond simply providing information by simulating practical clinical experiences for students[10].

In medical education, training students to assess suicide risk and manage mental health patients is particularly crucial[11,12], considering that South Korea has the highest suicide rate among Organization for Economic Cooperation and Development countries[13]. In KM education, however, there is a notable gap in standardized training for mental health assessment. A recent survey found that KM doctors frequently encounter patients with psychological disorders, including those at risk for suicide, but are not familiar with systematically assessing suicide risk in these patients[14]. While our previous study demonstrated the potential of using generative AI chatbots for suicide risk assessment training[15], developing a comprehensive CPX education platform offers several additional advantages. For example, it provides a standardized framework for creating and managing various clinical scenarios beyond suicide risk assessment. Also, it enables systematic evaluation and feedback

mechanisms that were not available in the previous chatbot-only approach. Most importantly, it allows for scalable implementation across multiple institutions while maintaining consistent educational quality.

In this context, the present study aims to develop a CPX education platform utilizing the ChatGPT API, with the objective of enhancing KM students' clinical competencies of suicide risk assessment through virtual patient interactions.

II. METHODS

1. Development of a platform to implement simulated patients using ChatGPT

1) Development environment

The platform was developed using Node.js (v18.17.1) for the backend and MongoDB (v8.4.4) for data storage (Fig. 1A). The frontend interface was built using React (v18.3.1). To implement the ChatGPT API, the GPT-4o-mini model, optimized for low-cost and fast-response services, was employed to simulate vir-

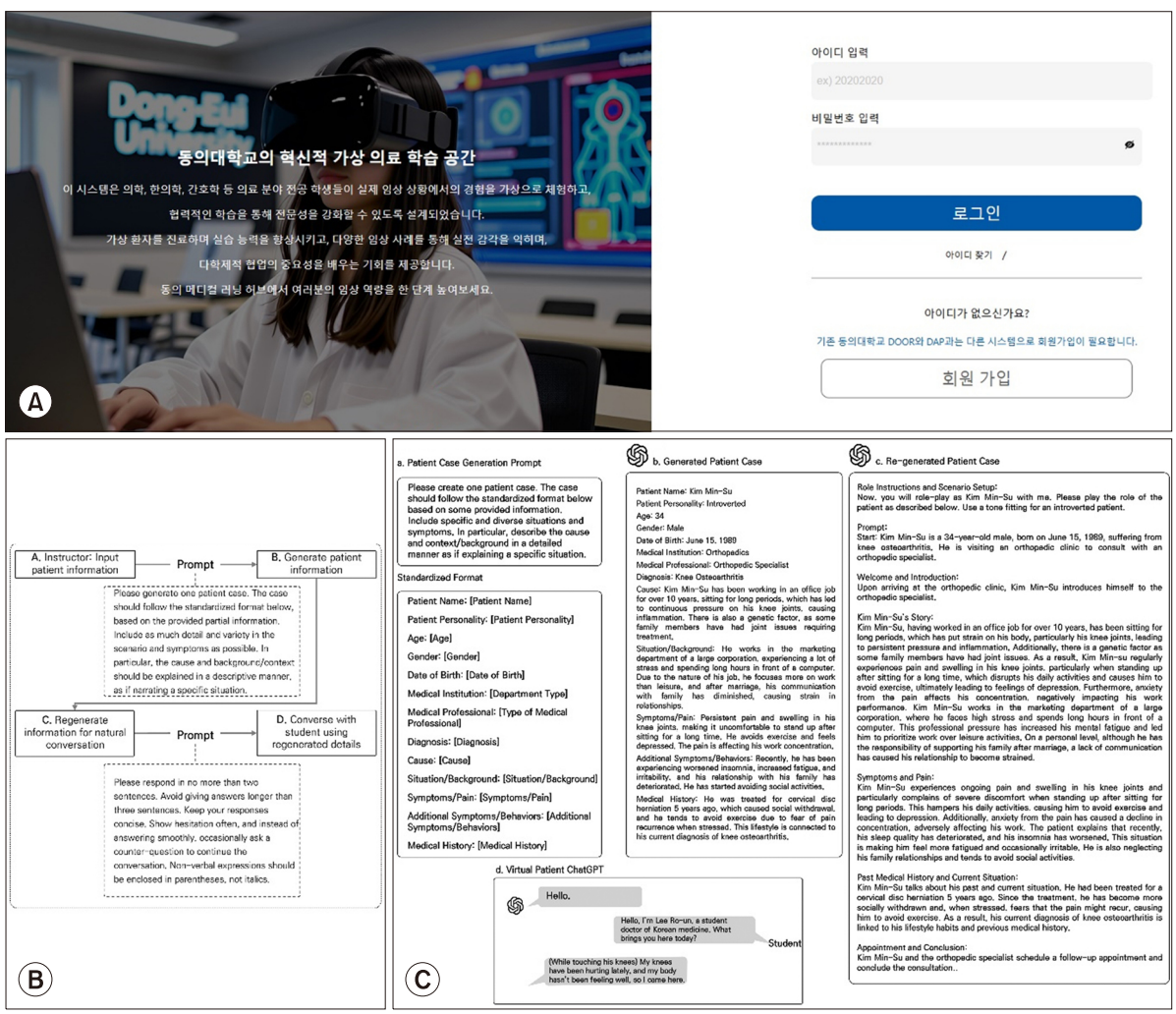


Fig. 1. Overview of the platform to implement simulated patients using ChatGPT. (A) shows the main interface of this chatbot system. (B) illustrates the flow of virtual patient case generation. (C) presents an example of a generated virtual patient case.

tual patients.

2) Prompt engineering

This system utilized prompt engineering techniques exclusively, without fine-tuning or hyperparameter tuning, to generate virtual patients using ChatGPT and facilitate natural, contextually accurate interactions between virtual patients and KM students. Firstly, the initial patient information is generated according to the "A: Patient Case Generation Prompt". Any missing patient details are automatically supplemented in "B: Generated Patient Case". In "C: Refined Patient Case", a refinement prompt is applied to further enhance the case for a more natural conversation. The final result is a virtual patient capable of engaging students in realistic, interactive dialogues (Fig. 1B). The virtual patient retains ChatGPT's conversational style but is modified to simulate a real patient's behavior, with actions such as "(touching his knees)" enclosed in parentheses to enrich the immersion of the experience. This method of prompt engineering allowed the creation of a chatbot capable of realistic patient-student interactions (Fig. 1C).

The development process included consultations with a diverse group of experts and researchers: a professor of KM Neuropsychiatry, a professor of Industrial ICT Technology Engineering specializing in AI, and five researchers from the Departments of Artificial Intelligence, Industrial ICT Technology Engineering, and IT Convergence. These experts provided input on scenario design, clinical authenticity, and educational effectiveness. Additionally, ten KM students participated in preliminary testing of the chatbot, providing user-perspective feedback. Consequently, the prompt engineering was iteratively refined based on this feedback to ensure both clinical accuracy and educational value.

3) Development of the instructor-student service platform

The platform was designed to resemble a lecture support system, familiar to students, and offers functionalities for both instructors and students. Instructors can create lectures, design patient scenarios, assign tasks, grade assignments, and provide feedback. Students complete the clinical simulations in a chat-based format and can view feedback to further develop their clinical competencies (Fig. 2).

(1) Scenario creation by instructor

Instructors can either input patient details manually or use the platform to auto-generate the following patient data: (1) patient name, (2) personality, (3) age, (4) gender, (5) date of birth, (6) medical institution, (7) healthcare provider, (8) diagnosis, (9) cause, (10) patient situation, (11) primary symptoms, (12) additional symptoms, and (13) medical history. The generated patient data can be edited to allow for flexible creation of diverse patient cases (Fig. 1C). Once the scenario is finalized, instructors can save it by clicking the 'Save' button, and they can optionally upload a patient image to enhance the scenario.

(2) Simulated clinical examination for students

After the scenario is created, students begin the clinical simulation, where the virtual patient initiates the conversation with a greeting. Acting as KM doctors, students conduct a diagnosis by asking relevant questions and collecting information. If a virtual patient's response does not align with the flow of the consultation, students can request the answer to be regenerated to maintain coherence (Fig. 3).

(3) Instructor's assignment evaluation

Instructors review each student's submission, which includes the consultation time, number of questions asked, responses provided, regenerated answers, diagnostic details, and diagnostic outcomes. The entire conversation between the student and the

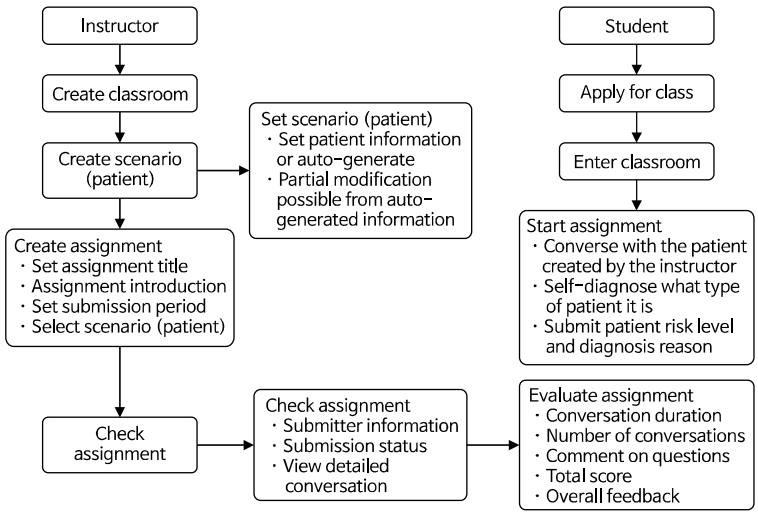


Fig. 2. Instructor-student educational platform flow. The instructor creates a course where students register and complete assignments after approval. The instructor sets up a virtual patient scenario with details like duration, score, and description. Students, acting as Korean medicine doctors, consult the virtual patient by asking questions and reviewing symptoms. They can regenerate irrelevant responses during the consultation. Upon submission, details like consultation duration, questions asked, and regenerated responses are recorded. The instructor reviews the session, provides feedback, and assigns a final grade to complete the evaluation.

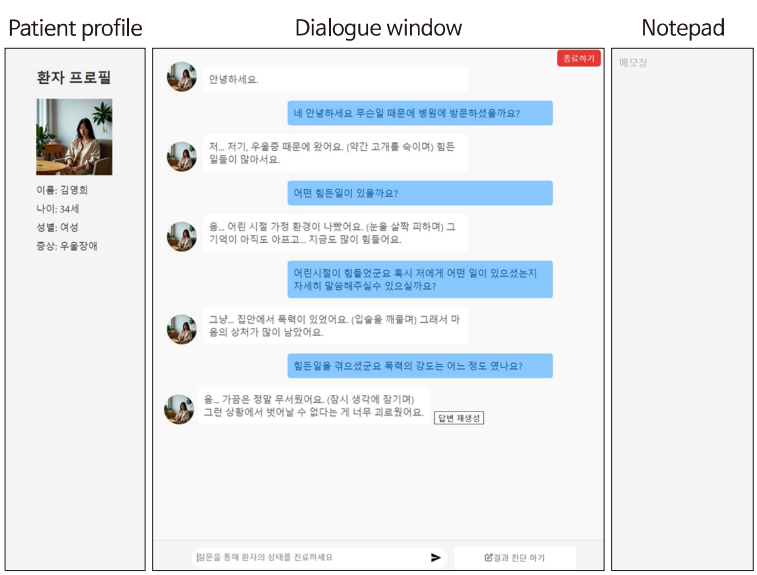


Fig. 3. Student screen view of the interaction with the virtual patient. The left side of the screen displays the patient's profile, including (1) image, (2) name, (3) age, (4) gender, and (5) symptoms, enhancing the immersive experience. On the right side is a notepad where the student can record symptoms during the consultation, which can be referred to when making the final diagnosis.

virtual patient can be examined, and instructors can offer real-time feedback by commenting directly on specific student queries. The evaluation process concludes with the instructor assigning a final grade and providing comprehensive feedback based on the student's diagnostic performance (Fig. 4).

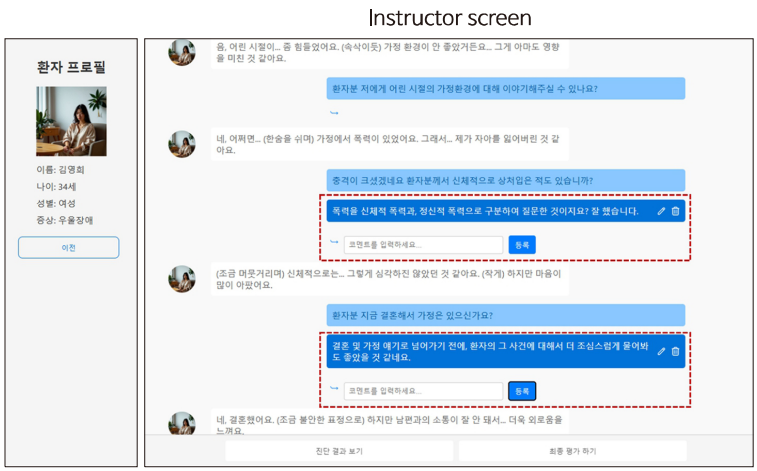


Fig. 4. Instructor screen view for evaluating student submitted dialogues. The dotted lines on the right side of the screen indicate the comments provided by the instructor on individual students' dialogues.

4) Quality control of AI responses

To minimize the risk of AI providing incorrect or inappropriate responses to students, several preventive measures were implemented. First, the prompt engineering was carefully designed to constrain the AI's responses within clinically appropriate boundaries. The patient's core characteristics, medical history, and expected responses were clearly defined in the initial prompt, reducing the likelihood of inconsistent or incorrect information. Second, the system included a 'regenerate response' feature that allows students to flag and regenerate responses they perceive as inappropriate or inconsistent. This feature helped maintain the quality of the conversation while providing valuable feedback for system improvement. Lastly, all student-AI conversations were recorded and reviewed by instructors, allowing for the identification and correction of any incorrect information during the feedback process.

2. Application of the system to KM education

1) Case development

This system was applied in the second semester of the KM Neuropsychiatry Practical Course for fourth-year students at the KM College of D University. A patient case was developed using the scenario creation feature of the system (Table 1). Forty-seven students enrolled in this practicum were tasked with assessing the suicide risk of a virtual patient. As outlined in our previous work[15], the students were instructed to follow a structured conversational flow with the virtual patient: greeting, obtaining the patient's name, date of birth, purpose of care, and underlying conditions, followed by suicide risk assessment, and concluding the session. Each student was required to conduct a minimum of 50 interactions with the virtual patient and assess suicide risk using the Ask Suicide-Screening Questions Toolkit (ASQ)[16], a standardized risk assessment tool. After completing their consultations, the students classified the suicide risk as low, high, or very high, and submitted a rationale for their classification.

Table 1. Case of Virtual Patient Development

Items	Content entered by instructor
The patient's name	Kyung-Yeon Kim
Personality	Introverted and does not express emotions well, but is considerate of others and tends to endure inner pain alone.
Age	76
Gender	Female
Date of birth	March 14, 1948
Medical institution	Dong-eui Medical Center
Healthcare provider	Korean medicine doctor
Diagnosis	Depression with cognitive impairment
Cause	Her depression with cognitive impairment was largely influenced by the loss of her spouse. She experiences social isolation due to a lack of contact with her children and the severed relationships with old friends, which exacerbates her psychological stress. Furthermore, the decline in cognitive function associated with aging, combined with depression, has had a negative impact on her overall cognitive abilities.
Patient situation	In recent months, she has been experiencing severe depression and confusion. She suffers from memory loss and disorganized thinking, which have led her to lose the will to live. Suicidal impulses, coupled with a past suicide attempt, have worsened her current psychological distress. Her loneliness and sense of loss have been intensified by the lack of contact with friends and family.
Main symptoms	She frequently complains of persistent feelings of depression and lethargy, along with sleep disturbances and loss of appetite. Her cognitive impairment symptoms include memory decline and decreased reasoning ability, which have affected her daily life. These physical and mental symptoms further isolate her socially and diminish her motivation for daily activities.
Additional symptoms	She fears being alone, especially due to her lack of contact with her children, which deepens her feelings of loneliness and loss. At times, she experiences hallucinations or delusional thoughts, which harm her overall mental health. Her psychological distress is becoming more severe due to her state of social isolation.
Medical history	In the past, she suffered from severe depression following the death of her spouse and had attempted suicide. The mild cognitive impairment symptoms that began a few years ago have recently worsened, making daily life more difficult for her. She suffers from hypertension and diabetes, for which she takes regular medication, and these chronic conditions further aggravate her mental suffering.

2) Evaluation of virtual consultation

Upon submission of the students' conversations and suicide risk assessments, the instructor conducted a comprehensive evaluation. This evaluation included both a quantitative assessment, graded on a scale of 0~100, and a qualitative review of each student's performance. Individual feedback on specific interactions was also provided (Fig. 4).

3. Survey of students' perceptions and experiences

All students who participated in the class were invited to complete an anonymous online survey via Moaform (URL: https://www.moaform.com/) (Qoom Networks Inc., Seoul, Republic of Korea), a free platform for anonymous surveys. This survey was conducted from October 24 to October 30, 2024. The purpose of the survey was to gather students' perceptions and experiences of using the chatbot system in the context of their education. Only students who were informed that the survey data could be used for research purposes, and who voluntarily consented to participate, were included in the study.

The survey consisted of 13 questions evaluating students' experiences with mock clinical consultations[17] and 17 questions assessing the usability of the AI chatbot[18]. The survey items were adapted from a prior study[15] and modified appropriately for use in an KM Neuropsychiatry course. Additionally, the question was raised about the average number of conversations required to learn when interacting with a generative AI chatbot that simulates a patient with a mental disorder. Finally, four open-ended questions assessed student satisfaction, dissatisfaction, suggestions for improvement, and the perceived

need for educational support related to chatbot-based role-playing.

4. Analysis of survey results

Survey responses regarding students' mock clinical experience and the usability of the AI chatbot were rated on a 5-point Likert scale. The results were analyzed using mean and standard deviation. Qualitative analysis was performed on the open-ended questions. Statistical analyses of the collected data were conducted using SPSS version 18.0 (SPSS Inc., Chicago, IL, USA).

III. RESULTS

1. Survey respondents

From October 24th to 30th in 2024, a total of 42 students participated in the survey, representing a response rate of 89.4% from the 47 eligible respondents.

2. Satisfaction assessment of simulated consultation experience

The evaluation of the simulated consultation experience yielded consistently high scores across multiple dimensions. Particularly notable were the educational outcome measures: clinical knowledge enhancement (4.24±0.79), knowledge self-assessment capability (4.17±0.76), and communication skills improvement (4.29±0.74). Students reported increased confidence in suicide risk assessment (3.98±0.84) and improved interviewing skills (4.24±0.73). Furthermore, they positively evaluated the platform's potential for real-world application (4.24±0.69) and future clinical practice (4.36±0.66), suggesting successful transfer of learning to practical clinical scenarios. These educational outcomes were complemented by high scores in other dimensions: appropriateness-related items received particularly high ratings (4.71±0.38), followed by interest (4.44±0.46). Notable individual items included: "I recommend applying the role-play using the chatbot (patient-physician) in other classes" (Q5, 4.71±0.46) and "I support using the chatbot for role-play in future training for juniors in Korean medicine neuropsychiatry theory and practice" (Q6, 4.76±0.43). The high scores for these items suggest strong student endorsement for expanding the use of this educational tool (Table 2).

3. Evaluation of AI chatbot usability

The usability evaluation revealed strong performance across multiple dimensions: usefulness (4.40±

Table 2. Satisfaction with Role Play

Characteristics	Q	Mean±standard deviation			
Interest	1	4.43±0.55	4.44±0.46	4.58±0.38	4.38±0.43
	2	4.64±0.48			
	3	4.24±0.76			
Appropriacy	4	4.67±0.53	4.71±0.38		
	5	4.71±0.46			
	6	4.76±0.43			
Knowledge	7	4.24±0.79	4.24±0.79	4.21±0.56	
Skill	8	4.17±0.76	4.14±0.62		
	9	4.29±0.74			
	10	3.98±0.84			
Attitude	11	4.24±0.73	4.24±0.73		
General ability	12	4.24±0.69	4.30±0.61		
	13	4.36±0.66			

These scores represent the mean scores on a Likert scale of 1 to 5 (strongly disagree to strongly agree).

0.54), attractiveness (4.42±0.56), accessibility (4.49±0.51), and convenience (4.42±0.42). The overall discoverability score was 4.43±0.63. However, reliability showed a relatively low score, with an average score of 4.13±0.55. The statement "I trust that the responses given by the chatbot will be similar to those of actual patients" (Q12) received a notably lower score (3.69±0.95) (Table 3).

4. Other responses

When talking to a generative AI chatbot that acts as a patient with a mental disorder, the average number of conversations needed for learning was 37.5, excluding two outliers (both answered 1 time) (Fig. 5). The qualitative responses revealed several key themes regarding students' experiences with the platform. Students expressed high satisfaction with the system's ability to provide varied patient responses based on question content and nuance, allowing them to experience potential clinical challenges in a low-pressure environment. One student noted, "This was the most interesting project I've been involved in over the past six years," while another appreciated that "it was possible to simulate interactions with types of patients that are hard to meet in real life." However, some limitations were identified, including occasional response uniformity and technical issues. As one student mentioned, "Some responses were somewhat uniform" and "In the latter part, the answers began to become increasingly similar." Technical challenges were also reported, with one student describing system errors that interrupted the learning process. Students suggested several improvements,

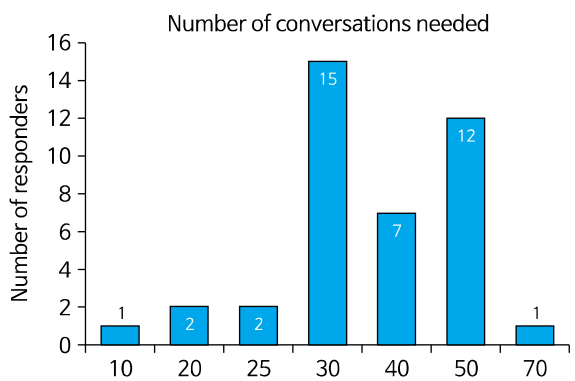


Fig. 5. The average number of conversations needed for learning (n=40). The data of two people who answered the number as 1 were considered outliers and were excluded.

Table 3. Experience in using Generative Artificial Intelligence Chatbots

Characteristics	Q	Mean±standard deviation		
Usefulness	1	4.48±0.59	4.40±0.54	4.38±0.41
	2	4.29±0.74		
	3	4.45±0.74		
Attractiveness	4	4.42±0.85	4.42±0.56	
	5	4.57±0.55		
	6	4.45±0.74		
Accessibility	7	4.31±0.68	4.49±0.51	
	8	4.62±0.49		
	9	4.55±0.77		
Reliability	10	4.38±0.58	4.13±0.55	
	11	4.31±0.30		
	12	3.69±0.95		
Convenience	13	4.33±0.82	4.42±0.42	
	14	4.10±0.73		
	15	4.48±0.59		
	16	4.76±0.43		
Discoverability	17	4.43±0.63	4.43±0.63	

These scores represent the mean scores on a Likert scale of 1 to 5 (strongly disagree to strongly agree).

including implementing time limits for consultations, developing more diverse patient modules, and adding visual feedback capabilities. As one student suggested, "Once familiar with the chatbot task, it might be beneficial to impose some time limits on the conversation, e.g., completing the conversation and making a diagnosis within 20 minutes." Regarding educational support, students expressed interest in expanding the platform's use across various medical subjects, particularly for rare cases or challenging diagnoses. One student recommended "exploring the use of chatbots in various subjects (such as pediatric patients, which students rarely encounter, or patients with cardiovascular concerns that require careful differential diagnosis) (Table 4)."

5. Instructor feedback

The instructor provided feedback on student-patient interactions across several domains: (1) interview structure and flow (e.g., "Good natural transition to suicide risk assessment"), (2) communication style (e.g., "Excellent demonstration of warm and considerate attitude toward the vulnerable patient"), and (3) clinical reasoning (e.g., "Need to strengthen the evidence for classifying this case as highest suicide risk").

IV. DISCUSSION

This study's development and implementation of a ChatGPT-based CPX platform represents a significant advancement in KM education, while also revealing several important considerations and limitations that warrant careful examination.

The high satisfaction scores across multiple dimensions (interest: 4.44±0.46, appropriateness: 4.71±0.38) demonstrate the platform's potential as a complementary tool in KM clinical education. Particularly noteworthy are the strong endorsements for expanding its use to other classes (4.71±0.46) and future training (4.76±0.43), indicating students' readiness to embrace AI-enhanced learning methodologies. This positive reception aligns with recent trends in medical education, where technological integration is increasingly recognized as crucial for preparing future healthcare professionals[8]. In particular, chatbots based on generative AI are expected to be able to assist KM students in their practical training in neuropsychiatry by imitating psychiatric patients[15].

The platform's high accessibility (4.49±0.51) and usefulness (4.40±0.54) scores address a critical gap in traditional standardized patient programs. As highlighted by Han et al. (2020)[4], conventional CPX implementation faces significant resource constraints, including financial burden and human resource limitations. Our platform offers a scalable, cost-effective solution that can be deployed across multiple institutions simultaneously, making clinical training opportunities more widely accessible. One of the advantages of this training platform is that the instructor can individually evaluate the learner's conversation and provide feedback. This is not only convenient for the instructor, but also something that has been difficult to do in traditional training to improve students' communication skills with patients with mental illness[19].

The survey results showed positive self-reported outcomes in knowledge acquisition (4.24±0.79) and skill development (4.14±0.62). Particularly noteworthy is the high rating for practical application (4.36±0.66), indicating that students felt more prepared to apply these skills in their future clinical practice. While these self-reported measures suggest the platform's potential as an educational tool, future studies incorporating objective knowledge assessments would be valuable to more rigorously evaluate its educational effectiveness. This finding is particularly relevant in the context of suicide risk assessment train-

Table 4. Responses to Open-Ended Questions (Optional)

Q	Responses
Satisfaction with chatbot role-play (n=16)	I appreciated the varied feedback from patients depending on the content, nuances, and tone of the questions. It allowed me to experience potential difficulties and confusion I might face when interacting with real patients.
	I found the responses to be more vivid than expected.
	This was the most interesting project I've been involved in over the past six years.
	I felt it was meaningful as I could engage as if I were conversing with a patient at risk of suicide.
	It seemed better for the learning process, as there was less pressure compared to interacting with actual patients, and I could conduct role-plays while researching information.
	I really liked that I could participate without constraints of time and space, which prompted me to think about how to respond if such a patient were to actually come in.
	The chatbot understood the questions better than I expected, and the responses were prompt.
	It was fun to receive patient feedback in a lively, conversational manner.
	I enjoyed the feeling of gaining clinical experience as a student.
	It is possible to simulate interactions with types of patients that are hard to meet in real life.
	It was fascinating and convenient to feel like I was talking to a real patient, accessible via computer or mobile phone.
	It was intriguing.
	I would like to encounter chatbot patients with various personalities beyond a cooperative/compliant version for treatment.
	It was very engaging.
	I believe that using chatbots will allow for efficient meetings with many patients. I hope we can have more simulated consultations using chatbots in the future.
	After only doing mechanical CPX with my peers, I found the chatbot virtual interviews to be engaging and focused. I wonder if I can send non-verbal expressions in parentheses, like what the chatbot does? The non-verbal cues in the chatbot's responses greatly helped me visualize the virtual patient.
Dissatisfaction (n=8)	Some responses were somewhat uniform.
	Occasionally, strange answers were generated, and there was repetition of previous statements.
	I felt that the chatbot's responses were not as specific and diverse as those of actual patients. Some answers seemed ambiguous. It felt like a person was zoning out and then giving a vague general response when suddenly asked a question.
	In the latter part, the answers began to become increasingly similar.
	Instead of written prompts in parentheses, it might be more realistic if the chatbot could visually express emotions and actions.
	When the patient expressed concern with every response, it felt somewhat AI-like.
	The answers seemed to vary slightly.
	At times, it seemed to freeze while generating a response, indicating an error.
Suggested improvements (n=8)	The conversation did not feel like a real-time exchange. I found myself looking up unknown information while proceeding.
	Once familiar with the chatbot task, it might be beneficial to impose some time limits on the conversation, e.g., completing the conversation and making a diagnosis within 20 minutes.
	The responses felt more monotonous and repetitive than I expected, so it would be good to improve them to be more specific and varied.
	It would be nice to encourage more concrete responses. Additionally, if possible, having the ability to provide image-based answers or a feature to indicate on images provided by the doctor could be interesting.
	It would be even better if there were various chatbot modules to experience different versions of patients.
	When the patient expressed concern with every response, it felt somewhat AI-like.
	In reality, there would be a back-and-forth exchange between the doctor and patient, but with a chatbot, I could take my time to think without worrying about what to ask next. Therefore, I think it's essential to practice a lot with chatbots for real application.
	During the project, an error occurred suddenly, displaying the message, "An error occurred while generating the response." Despite clicking regenerate several times, I kept getting the message, "An error occurred while regenerating the response," and even when I sent new questions, the same responses repeated. When I restarted the internet window, all conversations after the error were gone, and it seemed not to be reflected in the regeneration count, which was quite disconcerting at the time.
Desired educational support (n=8)	Since it was conducted online, I am still uncertain. I think that being aware it is AI and questioning in a command-like manner might contradict the essence of role-playing, so perhaps ignorance is bliss.
	I think it would be more meaningful to change the CPX exam format, which involves writing and memorizing scripts, to a chatbot exam format.
	It would be beneficial to explore the use of chatbots in various subjects (such as pediatric patients, which students rarely encounter, or patients with cardiovascular concerns that require careful differential diagnosis).
	I hope that role-playing education using chatbots will be conducted across various subjects.
	I positively view the idea of providing chatbots to practice interviews with patients across various diseases.
	Having exemplary responses, like recommended answers, could help develop conversation skills.
	Financial support for program development and utilization, as well as encouragement for student participation, would be beneficial.
	I would appreciate it if the school could support enabling various role-playing experiences or allowing students to create them.

AI: artificial intelligence, CPX: clinical performance examination.

ing, where students need safe environments to practice sensitive conversations without risking patient well-being[15]. The platform's ability to provide standardized, repeatable scenarios while allowing for varied student responses represents a significant advantage over traditional training methods. A systematic review in 2023 concluded that simulation training is promising for practical training in suicide crisis intervention[20]. Simulation-based education assisted by AI is expected to be more promising in practical training for high-risk groups such as suicide crisis patients due to its potential to meet the three basic points of simulation-based education: 'basic knowledge acquirement', 'briefing', and 'psychological safety assurance'[21].

The qualitative feedback from students provided valuable insights into both the strengths and limitations of the platform. Students particularly valued the ability to practice clinical interactions without time pressure and the opportunity to encounter rare or challenging cases, addressing common limitations in traditional clinical education. As one student noted, "It was meaningful as I could engage as if I were conversing with a patient at risk of suicide" while another appreciated that they "could participate without constraints of time and space." However, the feedback also revealed important technical and pedagogical challenges, including response uniformity and occasional system instability. Students' suggestions for implementing time limits and developing more diverse patient modules indicate the need for a balance between educational flexibility and clinical realism. The desire for expanded applications across various medical subjects, particularly for rare conditions and pediatric cases, suggests potential broader applications in KM education. This aligns with current trends in medical education that emphasize using technology to supplement areas where traditional clinical exposure may be limited[21].

Despite these promising findings, several challenges emerged from our study. Importantly, the relatively low authenticity rating (3.69±0.95) for virtual patient responses highlights a significant limitation in current AI technology. This aligns with previous study about the challenges of achieving naturalistic patient-doctor interactions with LLMs[7]. In addition, this low authenticity rating is a result that was found identically in suicide risk assessment training using another generative AI model, the Claude 3 Sonnet model[15]. Likewise, the gap between AI-generated responses and real patient behavior remains a crucial area for improvement. However, it is not possible to distinguish whether these results are due to students' pre-existing attitudes toward generative AI or their experiences in this study, and qualitative interviews with students could help understand their perceived low authenticity. Also, as described in our previous study[15], if the students watch educational videos of actual patients with mental illness in parallel with the educational period using a generative AI chatbot, it may be possible that students' evaluations of the similarity between the chatbot's responses and the actual patient's responses will change.

Some specific solutions can be implemented to enhance response authenticity. For example, to improve clinical authenticity, actual de-identified patient interview transcripts from KM practice could be used to train the response patterns. This would help ensure that the virtual patient's responses more closely mirror real clinical encounters while maintaining patient privacy. Another potential dataset for fine-tuning could be the publicly available counseling dataset from AI Hub (https://aihub.or.kr/), operated by the Ministry of Science and ICT and the National Information Society Agency. Regular updates of the prompt engineering based on student and instructor feedback could help address specific authenticity issues as they are identified. For exam-

ple, when students reported that 'some responses were somewhat uniform' and 'the answers began to become increasingly similar', the prompts could be modified to generate more varied and contextually appropriate responses. Additionally, providing students with audiovisual educational materials showing real patient interviews before their interaction with the virtual patient could enhance their understanding of authentic patient responses and behaviors. This could include video recordings of standardized patient interviews or clinical case demonstrations, which would help students better recognize the nuances of patient communication and improve the authenticity of their virtual patient interactions. This multi-modal educational approach could serve as a valuable reference point for students to compare and calibrate their expectations during virtual consultations.

While our platform successfully demonstrated general clinical education capabilities, its application in KM education requires careful consideration of the unique context of mental health care in KM. Given South Korea's high suicide rate[13] and the frequent encounters between KM doctors and patients with psychological conditions[14], there is a pressing need for standardized mental health assessment training that integrates both psychiatric approaches and KM perspectives. The current scenario, while focused on suicide risk assessment, could be enhanced by incorporating KM-specific elements such as tongue and pulse examination that could provide additional insights into the patient's mental state, and KM-specific therapeutic approaches that could be discussed with patients at risk. This integration of KM-specific elements while maintaining standardized suicide risk assessment protocols is crucial not only for educational authenticity but also for preparing KM doctors to effectively contribute to mental health care and suicide prevention within Korea's healthcare system.

This study has several important limitations that should be acknowledged. First, our evaluation was limited to a single clinical scenario focusing on suicide risk assessment. While this provided valuable insights into the platform's capabilities, it may not fully represent its effectiveness across the broader spectrum of KM diagnoses and treatments. In KM education, the development of CPX scenarios is being developed to reflect unique KM diagnoses and treatments[22]. A study in 2010 implemented an interactive KM pattern diagnosis capability assessment software, which presents 36 categories of symptoms/signs including tongue and pulse examination that are key to pattern diagnosis, which is worth referencing[23]. In addition, 'the Clinical Practice Guideline of Korean Medicine: The Pattern Identification of the Eight Principles' developed in 2023 present a standard KM pattern diagnosis, which can be referenced on this educational platform in the future[24]. Second, methodological limitations include the absence of a control group using traditional standardized patient methods, making it difficult to objectively compare the educational effectiveness of our platform. Additionally, our evaluation relied primarily on subjective satisfaction measures rather than objective assessments of clinical competency improvement. Therefore, future studies could evaluate the educational effectiveness of the current educational platform by assessing changes in knowledge, attitudes, and communication skills regarding suicide prevention as assessed in previous studies[11,12]. Third, technical limitations inherent to the ChatGPT system were observed. The AI sometimes provided inconsistent responses or deviated from the intended clinical context, requiring continuous monitoring and feedback from instructors. Additionally, the system's understanding of specialized KM terminology and diagnostic methods may be limited, potentially affecting the authenticity of simulated consultations. This

could potentially be improved by using a terminology standard, such as 'International Standard Terminologies on Traditional Chinese Medicine' by the World Health Organization, to fine-tune the LLM model[25]. Fourth, as this was a short-term pilot study, the long-term educational impact and transferability of skills to real clinical settings remain unverified. Future longitudinal studies would be valuable in assessing the sustained effectiveness of this educational approach. In this context, recent research has identified improvements in communication-related competencies and anxiety over time in medical students[26], and the synergistic effects of these longitudinal changes and current suicide risk assessment training could also be investigated. Lastly, the study's sample was limited to fourth-year students at a single KM college, which may restrict the generalizability of our findings to other academic years or institutions. A more diverse sample across multiple institutions would provide more robust evidence of the platform's effectiveness. These limitations suggest several directions for future research, including the need for controlled comparative studies, longitudinal assessments of clinical competency development, and broader implementation across various KM educational contexts.

V. CONCLUSION

The ChatGPT-based CPX platform developed in this study demonstrates promising potential as an accessible and standardized tool for clinical training in KM education. While the high satisfaction rates and usability scores validate its educational utility, the identified limitations in response authenticity suggest its optimal role as a complementary tool rather than a replacement for traditional clinical education methods. The findings provide valuable insights for the future development of AI-enhanced medical education platforms, particularly in the context of unique educational requirements in KM and neuropsychiatric education. Future iterations of the platform should focus on enhancing response authenticity, expanding scenario diversity, and implementing objective assessment measures while maintaining the high levels of accessibility and user satisfaction demonstrated in this pilot study.

CONFLICT OF INTEREST

None.

REFERENCES

1. Gwak D, Kim M, Kwon J, Yin C. Trends in Korean Medicine education research from 2003 to 2022. J Korean Med Hist. 2023;36(1):89-98. https://doi.org/10.15521/jkmh.2023.36.1.089
2. Jeong SH, Kim JP, Kang YJ, Jeong HI, Kim KH. A Survey of Recognitions and Satisfaction with Education in Traditional Korean Medicine. J Soc Prev Korean Med. 2020;24(3):49-56. https://doi.org/10.25153/spkom.2020.24.3.005
3. Cho E, Kim Y, Ha J, Park J, Jung H. Developing a manual for clinical practice on real patients in Korean Medicine. J Kor Med Edu. 2023;1(1):15-22. https://doi.org/10.1016/j.imr.2023.100924
4. Han CY, Kang DW, Park JG, Kim BH, Kim KS, Kim YB, et al. An Analysis of Clerkship Satisfaction in College of Korean Medicine: Focusing on Doctor-patient Role-play and mock CPX. J Korean Med Ophthalmol Otolaryngol Dermatol. 2020;33(1):12-24. https://doi.org/10.6114/jkood.2020.33.1.012
5. Han SY, Lee SH, Chae H. Developing a best practice framework for clinical competency education in the traditional East-Asian medicine curriculum. BMC Med Educ. 2022;22(1):352. https://doi.org/10.1186/s12909-022-03398-4
6. Jo HJ, Min SH. The current status and future operations of Clinical Performance Evaluation (CPX) in the nationwide colleges (graduate schools) of Traditional Korean Medicine. J Korean Med Hist. 2020;33(2):9-21. https://doi.org/10.15521/jkmh.2020.33.2.009
7. Park SY, Kim CE. Enhancing Korean Medicine Education with Large Language Models: Focusing on the Development of Educational Artificial Intelligence. J Physiol & Pathol Korean Med. 2023;37(5):134-8. http://dx.doi.org/10.15188/kjopp.2023.10.37.5.134
8. Shoja MM, Van de Ridder JMM, Rajput V. The Emerging

Role of Generative Artificial Intelligence in Medical Education, Research, and Practice. Cureus. 2023;15(6):e40883. https://doi.org/10.7759/cureus.40883
9. Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. JMIR Med Educ. 2023;9:e46885. https://doi.org/10.2196/46885
10. Kim SJ. Evaluating the usability of ChatGPT in pre-hospital emergency patient history taking scenario training: Comparison of ChatGPT 3.5 and 4.0. J Korea Academia-Industrial Cooperation Soc. 2024;25(8):230-8. https://doi.org/10.5762/KAIS.2024.25.8.230
11. Desai ND, Chavda P, Shah SH, Shah N, Shah SN, Sharma E. A novel approach to suicide prevention - Educating when it matters. Ind Psychiatry J. 2018;27(1):115-23. https://doi.org/10.4103/ipj.ipj_10_18
12. Nebhinani N, Kuppili PP, Paul K. Effectiveness of Brief Educational Training on Medical Students' Attitude toward Suicide Prevention. J Neurosci Rural Pract. 2020;11(4):609-15. https://doi.org/10.1055/s-0040-1716769
13. Suicide rates: OECD; Available from: https://data.oecd.org/healthstat/suicide-rates.htm.
14. Kwon CY, Lee B. The potential role of Korean medicine doctors in Korea's suicide prevention policy: The first web-based survey. Asian J Psychiatr. 2024;96:104040. https://doi.org/10.1016/j.ajp.2024.104040
15. Kwon C-Y. Utilization of Generative Artificial Intelligence Chatbot for Training in Suicide Risk Assessment of Depressed Patients: Focusing on Students at a College of Korean Medicine. J of Oriental Neuropsychiatry. 2024;35(2):153-62. https://doi.org/10.7231/jon.2024.35.2.153
16. Horowitz LM, Bridge JA, Teach SJ, Ballard E, Klima J, Rosenstein DL, et al. Ask Suicide-Screening Questions (ASQ): a brief instrument for the pediatric emergency department. Arch Pediatr Adolesc Med. 2012;166(12):1170-6. https://doi.org/10.1001/archpediatrics.2012.1276
17. Hong J, Kim J, Kim Y. A Survey of Students' Satisfaction on Participation Learning Using Role-play in Clerkship. J Korean Acad Pediatr Dent. 2015;42(1):1-9. https://doi.org/10.5933/JKAPD.2015.42.1.1
18. Min Y, Kwak S. A Comparative Study on the Usability by the Platfrom of Artificial Intelligence Chatbot Service in Library. J Korean Biblia Soc Library Inf Sci. 2023;34(2):183-203. http://dx.doi.org/10.14699/kbiblia.2023.34.2.183
19. Iezzoni LI, Ramanan RA, Lee S. Teaching medical students about communicating with patients with major mental illness. J Gen Intern Med. 2006;21(10):1112-5. https://doi.org/10.1111/j.1525-1497.2006.00521.x
20. Richard O, Jollant F, Billon G, Attoe C, Vodovar D, Piot MA. Simulation training in suicide risk assessment and intervention: a systematic review and meta-analysis. Med Educ Online. 2023;28(1):2199469. https://doi.org/10.1080/10872981.2023.2199469
21. Komasawa N, Yokohira M. Simulation-Based Education in the Artificial Intelligence Era. Cureus. 2023;15(6):e40940. https://doi.org/10.7759/cureus.40940
22. Jo HJ, Jo NY, Park JS. Consistency of Student Self-Assessment of Role Play Using the Syndrome Differentiation-Based Clinical Performance Examination Scenario and Improvement Measures. J Soc Prev Korean Med. 2022;26(2):37-53.
23. Kim K. Development of the Software to test Pattern Diagnosis Ability in Oriental Medicine. J Korea Instit Orient Med Diagn. 2010;14(1):70-8. https://doi.org/ 10.25153/spkom.2022.26.2.004
24. The Ministry of Health and Welfare, National Institute for Korean Medicine Development. The Clinical Practice Guideline of Korean Medicine: The Pattern Identification of the Eight Principles. 2023.
25. Integrated Health Services (IHS), Traditional, Complementary and Integrative Medicine (TCI). WHO international standard terminologies on traditional Chinese medicine. 2022.
26. Kerr AM, Thompson CM. A longitudinal study of third-year medical students' communication competence, communication anxiety, and attitudes toward patient-centered care. Patient Educ Couns. 2022;105(5):1298-304. https://doi.org/10.1016/j.pec.2021.09.004