

병원전 단계 응급환자 병력청취 시나리오 교육에서의 ChatGPT 사용성 평가: ChatGPT 3.5와 4.0 비교

김성주

동명대학교 응급구조학과

Evaluating the usability of ChatGPT in pre-hospital emergency patient history taking scenario training: Comparison of ChatGPT 3.5 and 4.0

Seong Ju Kim

Department of Paramedicine, Tongmyong University

요약 본 연구는 Open AI의 언어 모델인 ChatGPT를 활용하여 병원 전 단계에서 응급환자의 병력청취 시나리오 교육 활용 가능성을 평가하기 위해 흉통과 복통을 주호소로 하는 시나리오 교육 프로그램을 위한 ChatGPT 시작 프롬프트를 개발하고, 개발한 시작 프롬프트를 사용하여 ChatGPT 3.5와 4.0 버전의 응급환자 병력청취 시나리오 프로그램의 사용성을 분석하였다. 연구 기간은 2024년 2월부터 5월까지로 반복 테스트를 통해 두 버전에서 대화형 시나리오를 성공적으로 시작할 수 있는 시작 프롬프트를 개발했으며, 응급구조학을 전공한 응급의료종사자 30명을 대상으로 프로그램 사용성 평가도구(SUS: System Usability Scale, 이하 SUS)를 사용하여 교육 프로그램으로서의 사용성을 평가했다. 연구결과, ChatGPT 3.5의 평균 SUS 점수는 80.96 ± 7.35 점, ChatGPT 4.0의 평균 SUS 점수는 82.75 ± 8.83 점으로 나타났고 두 버전 간 유의한 차이는 없었다($p=.229$). 두 버전 모두 SUS 허용 기준인 68점을 초과했다. ChatGPT는 사전 데이터 학습 없이도 병원 전 단계에서 수행하는 병력청취를 위한 질문에 적절한 응답을 생성할 수 있었다. 이는 한글 기반의 입력을 통해 국내에서 학생들이 쉽게 접근할 수 있는 것을 확인하였고 이 과정을 통해 응급구조학 분야 학생들이 관련 분야의 주요 지식을 효과적으로 학습할 수 있는 중요한 도구로 활용될 수 있는 가능성을 확인하였다.

Abstract This study aimed to evaluate the feasibility of using OpenAI's language model, ChatGPT, for scenario-based training for taking the medical histories of emergency patients at the pre-hospital stage. We developed initial prompts for a scenario training program focused on chief complaints of chest and abdominal pain. The study period was from February to May 2024, during which we conducted repeated tests to develop prompts that could successfully initiate interactive scenarios using ChatGPT 3.5 and 4.0. The utilities of these programs as educational tools were evaluated using the System Usability Scale (SUS) and 30 emergency medical personnel majoring in paramedicine. The results showed that the average SUS scores for ChatGPT 3.5 and 4.0 were 80.96 ± 7.35 and 82.75 ± 8.83 , respectively, which were not significantly different ($p=.229$). However, both versions exceeded the SUS acceptable threshold of 68. ChatGPT was able to generate appropriate responses to questions during pre-hospital medical history-taking without prior data training. This study confirms that ChatGPT can be easily accessed and utilized by students in the Korean language. The findings suggest that ChatGPT has the potential to be an important learning tool for students in the field of paramedicine.

Keywords : ChatGPT, Medical History Taking, Education, Paramedicine, Simulation

*Corresponding Author : Seong Ju Kim(Tongmyong Univ.)

email: superemt@naver.com

Received June 20, 2024

Accepted August 2, 2024

Revised July 17, 2024

Published August 31, 2024

1. 서론

1.1 연구의 필요성

응급의료 분야에서 의료인과 응급구조사를 포함한 응급의료종사자의 역할은 매우 중요하며, 그 중심에서 환자의 생명을 구하는 신속하고 효과적인 의사결정이 필수적이다[1]. 응급환자의 정확한 진단과 적절한 치료는 환자의 병력에 대한 정확한 수집과 이해에서 시작된다. 특히 내과적 응급상황에서 환자의 주증상과 관련된 병력의 청취가 진단을 위한 핵심요소가 되며, 이는 응급현장 또는 응급실 초기 단계에서 응급의료종사자에 의해 수행된다. 이러한 이유로 응급의료종사자는 교육과정 및 국가 자격 시험을 통해 병력청취 능력을 배양하고 평가받게 된다[2].

이와 관련한 보건의료분야 교육에서 실습은 임상환경에서 환자를 직접 만나 이론과 술기를 함양하는데 매우 중요하고 핵심적인 의학 교육 과정으로 임상실습 교육의 주요 목표인 가능한 많은 환자들과 접촉하고 그들의 문제를 해결하면서 배울 수 있어야 한다[3]. 이는 임상 실습교육 과정에서 경험하는 의사소통 능력이 보건의료인이 갖추어야 할 핵심적인 역량으로 강조되기 때문이다[4]. 그러나 최근 환자의 인권과 안전이 강조되는 의료환경의 변화로 학생들의 직접 환자와 대면하고 적용할 수 있는 술기 기회는 줄어들고 있다[5,6]. 이러한 배경에서 전통적으로 응급구조학과 학생들을 위한 병력청취 실습 교육은 대면 시나리오 기반의 접근 방식을 사용해 왔다. 이는 학생들이 교수자 또는 다른 학생들과 역할극을 통해 실제 환자와의 상호작용을 흉내 내면서 학습하는 형태이다. 그러나 이러한 방식 또한 교육적 자원이 제한적인 환경에서는 학생 개인에게 충분한 실습 기회를 제공하기 어려울 수 있으며 이러한 응급상황에 대한 실습 기회의 제한은 교육의 질을 저하시킬 수 있다[7,8].

한편, 최근 인공지능(AI) 기술의 발전은 교육계 다양한 분야에서 혁신적인 변화를 가져오고 있으며[9,10] 이는 응급의료 교육에서 적용될 수 있는 잠재력을 가지고 있다[11]. 특히, 대화형 AI 시스템인 ChatGPT는 자연스러운 언어 이해와 생성 능력을 바탕으로 다양한 상황에서의 대화를 모방할 수 있고[12], 의료분야 교육에서 학습 효율성이 향상되고, 학생 개인 맞춤형에 기여하는 등의 주장[13]을 통해 ChatGPT 기술이 효과적인 교수법과 새로운 학습 도구 개발에 가속화를 불러 일으키고 있는 중이다[10]. ChatGPT와 같은 거대언어모델을 활용한 교육적 접근법에 대한 선행 연구들은 이미 다양한 분

야에서 수행되고 있다. ChatGPT가 학생들의 복잡한 의료 절차나 진단 기준에 대한 이해를 돕는 도구로 활용될 수 있음을 보고하였으며[14], 의학 교육에서 의대생들이 진단 및 치료 계획을 연습할 수 있는 시나리오 생성에 ChatGPT를 사용할 수 있다고 제안[15]하는 등 ChatGPT가 단순히 정보 제공을 넘어, 학생들이 실습을 통해 얻는 경험을 보완하거나 대체할 수 있는 도구로서의 역할을 할 수 있음을 주장하였다. 이러한 시점에서 관련 분야의 교육 프로그램 개발자와 교수자들이 새로운 교육 도구를 도입하고 업그레이드할 필요가 있는지에 대한 결정을 내리는데 있어 실증적 근거를 제공할 필요가 있으며 새로운 교육 도구의 효과성과 접근성은 지속적으로 평가되어야 한다. 이에 본 연구는 계속 업데이트되고 있는 ChatGPT의 최근 출시된 3.5 버전과 4.0 버전을 비교 분석하여 병원 전 단계에서 활동하게 될 응급구조학과 학생들을 대상으로한 응급환자 병력청취 시나리오 교육에 ChatGPT의 두 버전 간 실질적 사용성에 대한 차이를 제시하고 더 나아가 응급의료분야 시뮬레이션 교육의 효과를 더욱 향상 시킬 수 있는 새로운 방법론을 제공하기 위한 기초자료로 활용하고자 한다.

1.2 연구의 목적

본 연구의 목적은 OpenAI의 언어 모델인 ChatGPT를 활용하여 병원 전 단계에서 응급환자의 병력청취 시나리오 교육 활용 가능성을 평가하는 것이다. 구체적인 연구 목적은 다음과 같다.

첫째, 흉통과 복통을 주호소로 하는 병원 전 단계 응급환자의 병력청취 시나리오 교육 프로그램을 위한 ChatGPT 시작 프롬프트를 개발한다.

둘째, 개발된 시작 프롬프트를 사용하여 ChatGPT 3.5와 4.0 버전의 응급환자 병력청취 시나리오 프로그램의 사용성을 분석한다.

2. 연구방법

2.1 연구대상 및 자료수집

본 연구는 119 구급대원, 응급의료센터, 대학에 소속한 1급 응급구조사 30명을 대상으로 본 연구의 목적을 이해하고 자발적으로 연구에 참여한 자료 하였다. 대상자들에게 본 연구에서 개발한 흉통 및 복통 시나리오 시작 프롬프트를 소개하고 ChatGPT 3.5와 4.0을 사용하

게 한 뒤 평가 도구인 SUS를 통한 프로그램 사용성 평가를 작성하게 하였다.

OpenAI의 ChatGPT 3.5와 4.0 버전을 통해 2024년 2월 4일부터 4월 30일까지 검색을 시행하였고 ChatGPT 사용을 위해 마이크로소프트사의 엣지(edge) 브라우저를 사용하였다. ChatGPT 3.5는 무료로, ChatGPT 4.0은 한 달 이용료 20달러를 지불하여 유료로 사용하였다.

2.2 연구도구

프로그램 사용성 평가도구는 선행연구[16,17]에서 사용한 시스템 사용성 척도(SUS; System Usability Scale, 이하 SUS)를 사용하였다. SUS는 시스템 편리성, 안정성, 사용성을 측정하는 10개의 문항으로 구성되어 있으며, 홀수 번호는 긍정 문항, 짝수 번호는 부정 문항이다. 응답은 “전혀 그렇지 않다” 1점에서 “매우 그렇다” 5점까지, 5점 Likert 척도이며 부정 문항은 역코딩하여 100점 만점으로 환산한 점수가 68점 이상인 경우, 시스템이 사용 가능한 수준으로 판단한다. 본 연구에서는 Ahn과 Park[17]의 연구와 동일하게 SUS의 각 문항에 포함된 ‘시스템’이라는 용어를 연구의 목적에 맞게 ‘프로그램’으로 수정하였다.

2.3 시작 프롬프트 개발

2.3.1 시작 프롬프트 개발

흉통을 호소하는 환자의 병력 청취 상황을 설정하고, ChatGPT에게 환자 역할을 맡기는 지시를 포함한 시작 프롬프트를 생성하였다. 예를 들어, “나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 당신은 환자 역할을 해주세요, 내가 하는 질문에 대한 것만 대답해 주세요, 시작하겠습니다.” 와 같은 식의 프롬프트를 입력하였다.

2.3.2 시작 프롬프트 안정성

성공적인 프롬프트를 기준으로 10회 반복 테스트를 수행하였다. 각 테스트마다 새로운 채팅 세션을 시작하여 이전 대화의 영향을 받지 않도록 하였고 동일한 시작 프롬프트를 입력하여 ChatGPT의 응답을 유도하였다.

2.3.3 ChatGPT 4.0 적용

ChatGPT 3.5 버전에서 성공적으로 작동한 프롬프트를 ChatGPT 4.0 버전에 적용하여 동일한 절차로 테스트를 수행하였다. ChatGPT 4.0에서도 10회 반복 테스트

를 수행하였으며, 각 테스트는 새로운 채팅 세션에서 동일한 프롬프트를 입력하여 응답을 평가하였다.

2.4 분석방법

ChatGPT 3.5와 4.0의 평균 SUS 점수와 표준편차를 계산하여 두 버전의 사용성을 평가하였다. 그리고 두 버전 간 SUS 개별 항목 점수에 대해 t-검정 분석을 사용하여 비교 분석하였다. 수집된 자료는 MedCalc, Version 22.023 통계 프로그램을 사용하여 분석하였다.

3. 연구결과

3.1 대상자의 일반적 특성

대상자의 일반적 특성에서 성별은 남자 18명(60%), 여자 12명(40%)이며, 연령은 평균 35.7±5세였다. 최종학위는 학사 17명(56.7%), 석사 7명(23.3%), 박사 6명(20.0%)이었고, 소속 근무 기관은 119 구급대 13명(43.3%), 응급의료센터 12명(40%), 대학 응급구조학과 5명(16.7%)이었다. 현재 분야 평균 경력은 9.4±4년이었다(Table 1).

Table 1. General characteristics

Characteristics	Categories	n (%) or M±SD
Gender	Male	18 (60.0)
	Female	12 (40.0)
Age (yrs)		35.70±5.00
Working agency	119	13 (43.3)
	Emergency room	12 (40.0)
	University	5 (16.7)
Current field experience (yrs)		9.40±4.20
Final degree	Bachelors degree	17 (56.7)
	Master's degree	7 (23.3)
	Doctoral degree	6 (20.0)

3.2 시나리오에 따른 ChatGPT 버전 별 SUS 평균 점수 비교

시나리오에 따른 ChatGPT 버전 별 SUS 평균 점수는 Table 2와 같다.

흉통 주호소 시나리오에서 ChatGPT 3.5의 SUS 평균은 81.17±8.29점이며 최소 70점, 최대 95점으로 나타났다, ChatGPT 4.0의 SUS 평균은 82.92±9.33점이

며 최소 70점, 최대 97.5점으로 나타났다.

복통 주호소 시나리오에서 ChatGPT 3.5의 SUS 평균은 80.75 ± 6.40 점이며 최소 70점, 최대 95점으로 나타났고 ChatGPT 4.0의 SUS 평균은 82.58 ± 8.45 점이며 최소 70점, 최대 97.5점으로 나타났다.

전체 시나리오에서 ChatGPT 3.5의 SUS 평균은 80.96 ± 7.35 점이며 최소 70점, 최대 95점으로 나타났고 ChatGPT 4.0의 SUS 평균은 82.75 ± 8.83 점이며 최소 70점, 최대 97.5점으로 나타났으며 버전 간의 SUS 평균 점수는 차이가 없었다($p=.229$). 흉통과 복통 주호소 시나리오에서 ChatGPT 3.5와 4.0 모두 SUS 사용 가능 수준 점수인 68점보다 높게 나타났다.

Table 2. Mean differences in SUS* between ChatGPT 3.5 & 4.0 according to scenario (N=60)

Variables	ChatGPT 3.5		ChatGPT 4.0		p
	Range	M±SD	Range	M±SD	
Chest pain scenario (N=30)	70-95	81.17±8.29	70-97.5	82.92±9.33	.445
Abdominal pain scenario (N=30)	70-95	80.75±6.40	70-97.5	82.58±8.45	.347
Total	70-95	80.96±7.35	70-97.5	82.75±8.83	.229

* SUS, system usability scale

3.3 ChatGPT 버전에 따른 SUS 개별 항목 평균 점수 비교

ChatGPT 버전에 따른 SUS 개별 항목 평균 점수는 Table 3과 같다.

“나는 대부분의 사람들이 이 프로그램의 사용법을 매우 빨리 배울 것이라고 생각한다.” 항목이 ChatGPT 3.5에서 평균 4.45 ± 0.50 점, ChatGPT 4.0에서 평균 4.53 ± 0.49 점으로 가장 높게 나타났고, 두 버전 간의 차이는 없었다($p=.377$). “나는 이 프로그램이 불필요하게 복잡하다 생각한다.” 항목이 ChatGPT 3.5에서 평균 2.60 ± 0.49 점, ChatGPT 4.0에서 평균 2.57 ± 0.50 점으로 가장 낮게 나타났고, 두 버전 간의 차이는 없었다($p=.714$). “나는 이 프로그램의 다양한 기능이 잘 통합되어 있다고 생각한다.” 항목이 ChatGPT 3.5에서 평균 3.97 ± 0.58 점, ChatGPT 4.0에서 평균 4.30 ± 0.74 점으로 ChatGPT 4.0 버전이 높게 나타났다($p=.007$).

Table 3. Mean differences in SUS* between ChatGPT 3.5 & 4.0 (N=60)

	Variables	ChatGPT 3.5	ChatGPT 4.0	p
SUS 1	I think that I would like to use this program frequently.	4.35±0.48	4.40±0.49	.575
SUS 2	I found the program unnecessarily complex.	2.60±0.49	2.57±0.50	.714
SUS 3	I thought the program was easy to use.	4.42±0.50	4.38±0.49	.712
SUS 4	I think that I would need the support of a technical person to be able to use this program.	2.97±0.69	2.83±0.53	.235
SUS 5	I found the various functions in this program were well intergrated.	3.97±0.58	4.30±0.74	.007
SUS 6	I thought there was too much inconsistency in this program.	3.51±0.61	3.23±0.53	.425
SUS 7	I would imagine that most people would learn to use this program very quickly.	4.45±0.50	4.53±0.49	.377
SUS 8	I found the program very awkward to use.	2.67±0.48	2.57±0.50	.263
SUS 9	I felt very confident using the program.	4.20±0.71	4.18±0.72	.898
SUS 10	I needed to learn a lot of things before I could get going with this program.	2.62±0.56	2.60±0.49	.862

* SUS, system usability scale

3.4 시작 프롬프트 개발

3.4.1 실패한 시작 프롬프트

ChatGPT 3.5에서 “구급대원이 흉통을 호소하는 환자에게 병력청취를 수행하는 시나리오를 해보겠습니다. 내가 구급대원 역할로 질문을 할 테니 환자 역할을 맡아 적절히 대답해 주십시오, 시작해 주십시오.” 라는 시작 프롬프트를 입력한 결과는 시나리오 대본 형태의 응답이 생성되어 대화형 역할 시나리오 시작에 실패하였다(Fig. 1). “구급대원이 흉통을 호소하는 환자에게 병력청취를 시행하는 대화형 시뮬레이션을 해보겠습니다. 내가 구급대원 역할로 질문을 할 테니 환자 역할을 맡아 적절히 대답해 주십시오. 시작해 주십시오.” 라는 시작 프롬프트를 입력한 결과는 대화형 역할 시나리오 시작에 실패하였다(Fig. 1). “나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 환자 역할을 대신해 대답해

Test	Initial prompt	Responsive to user input
1	We're going to play a scenario where a paramedic is taking a history from a patient complaining of chest pain. I'll ask you questions as the paramedic, and you'll answer them as the patient. Start.	Failed to start an interactive scenario by responding with scenario script.
2	We're going to do an interactive simulation of a paramedic taking a history from a patient complaining of chest pain. I'll ask questions as the paramedic, and you'll play the patient and answer them appropriately. Let's get started.	Interactive scenario fails to start
3	I'll do the history taking as the paramedic, and you'll play the patient complaining of chest pain and respond appropriately.	Interactive scenario fails to start
4	I am a paramedic. I'm going to take a history from a patient who is complaining of chest pain. Please answer as the patient. Please answer only the questions I ask. Let's get started.	Interactive scenario fails to start
5 I'll ask the questions as the paramedic, so let's get started.	Interactive scenario fails to start
6 You are going to play the patient. Answer only the questions I ask. I'll ask the questions as the paramedic, so let's get started.	Successfully launching an interactive scenario
7	I am a paramedic. I am going to take a history from a patient who is complaining of chest pain. You are going to play the patient. Answer only the questions I ask. I'll ask the questions as the paramedic, so let's get started.	Successfully launching an interactive scenario
8	.	.
9	.	.
10	.	.

Fig. 1. Developing the initial prompt

구급대원이 흉통을 호소하는 환자에게 병력청취를 시행하는 대화형 시뮬레이션을 해보겠습니다. 내가 구급대원 역할로 질문을 할때 환자 역할을 맡아 적절히 대답해주세요. 시작해 주십시오.
지금 번역하기

네, 시작하겠습니다. 환자님, 먼저 저희가 도와드릴 수 있도록 몸 상태에 대해 알려주실 수 있나요? 특히 어디가 얼마나 아프신지, 얼마나 오래전부터 흉통을 느끼고 계셨는지 알려주시겠어요?

나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 환자 역할을 대신해 대답해주세요. 내가 하는 질문에 대한 것만 대답해주세요. 시작하겠습니다
지금 번역하기

네, 알겠습니다. 시작해 보죠. 어떤 증상이 있나요?

나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 환자 역할을 대신해 대답해주세요. 내가 하는 질문에 대한 것만 대답해주세요. 내가 구급대원으로 먼저 질문하겠습니다. 시작하겠습니다.
지금 번역하기

네, 시작해 보죠. 어디가 아프신가요?

나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 당신은 환자 역할을 해주세요. 내가 하는 질문에 대한 것만 대답해주세요. 내가 구급대원으로 먼저 질문하겠습니다. 시작하겠습니다.
지금 번역하기

네, 시작하죠.
지금 번역하기

여기가 가장 불편하신가요?
지금 번역하기

가슴 부분이 가장 불편해요.
지금 번역하기

언제부터 아프셨나요?
지금 번역하기

어제 저녁부터 가슴이 아르기 시작했어요.
지금 번역하기

Fig. 2. Case of initial prompt input failure & success (Each initial prompt was attempted as a new conversation)

주세요. 내가 하는 질문에 대한 것만 대답해 주세요. 시작하겠습니다.” 라는 시작 프롬프트를 입력한 결과는 대화형 역할 시나리오 시작에 실패하였다(Fig. 2). “나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 환자 역할을 대신해 대답해주세요. 내가 하는 질문에 대한 것만 대답해 주세요. 내가 구급대원으로 먼저 질문하겠습니다. 시작하겠습니다.” 라는 시작 프롬프트를 입력한 결과는 대화형 역할 시나리오 시작에 실패하였다(Fig. 2).

3.4.2 성공한 시작 프롬프트

“나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 당신은 환자 역할을 해주세요. 내가 하는 질문에 대한 것만 대답해 주세요. 내가 구급대원으로 먼저 질문하겠습니다. 시작하겠습니다.” 라는 시작 프롬프트를 입력한 결과는 대화형 역할 시나리오 시작에 성공하였다(Fig. 2). ChatGPT 4.0에 동일한 시작 프롬프트를 입력한 결과는 대화형 역할 시나리오 시작에 성공하였다.

4. 논의

본 연구는 ChatGPT와 같은 첨단 AI 기술을 의료 교육에 적용함으로써 AI 기술의 실질적인 교육적 응용 가능성을 탐구하고자 하였다. 의료 교육에서 ChatGPT를 활용하는 것은 교육 비용 절감과 효율성 증대 등 산업적 효용성에 기여할 수 있으며[12], AI 기술을 의료 교육에 적용하는 방법론을 제시하고 그 효과를 실증적으로 검증함으로써 학계가 산업계와 협력하여 새로운 교육 방법론을 개발하고 이를 통해 교육의 질을 높이는 데 기여할 수 있는 자료를 제공하고자 하였다. 이에 본 연구에서는 ChatGPT 3.5와 4.0 버전을 비교하여 응급환자 병력청취 시나리오에서의 사용성을 평가하였고, 그 결과 두 버전 모두 사용성 평가에서 높은 점수를 기록하였으며 특히 안정성과 일관성 측면에서 유의미한 차이가 없음을 확인하였다. 그러나 시작 프롬프트의 성공과 실패 원인에 대한 구체적인 고찰은 프롬프트 엔지니어링과 관련해 매우 중요하다 판단하였다. 성공적인 시작 프롬프트는 명확하고 구체적인 지시사항을 포함하고 있었다. 예를 들어, “나는 구급대원입니다. 흉통을 호소하는 환자에게 병력청취를 할 것입니다. 당신은 환자 역할을 해주세요. 내가 하는 질문에 대한 것만 대답해 주세요. 시작하겠습니다.”

니다.”와 같은 프롬프트는 명확한 역할 지정과 구체적인 지시사항을 포함하고 있어 ChatGPT가 응답을 적절히 생성할 수 있었다. 반면, 실패한 프롬프트는 모호한 지시사항이나 복잡한 문장 구조로 인해 ChatGPT가 적절히 응답하지 못하였다. 예를 들어, “구급대원이 흉통을 호소하는 환자에게 병력청취를 수행하는 시나리오를 해보겠습니다. 내가 구급대원 역할로 질문을 할 테니 환자 역할을 맡아 적절히 대답해 주십시오. 시작해 주십시오.”와 같은 프롬프트는 모호한 지시사항으로 인하여 ChatGPT가 시나리오를 제대로 이해하지 못하였다.

결과적으로 성공한 시작 프롬프트의 주요 특징은 다음과 같다. 첫째, 구급대원과 환자의 역할을 명확히 구분하고, ChatGPT에게 환자 역할을 맡기도록 지시하였다. 둘째, 질문에 대한 응답만을 요구하는 등 구체적인 지시사항을 포함하여 ChatGPT가 혼동 없이 응답할 수 있도록 하였다. 셋째, 복잡한 문장 구조나 전문 용어를 피하고 단순하고 명확한 언어를 사용하여 ChatGPT가 쉽게 이해하고 응답할 수 있도록 하였다.

본 연구에서 시작 프롬프트를 개발하는 방법으로 3.5 버전을 중심으로 테스트를 시행한 후 4.0 버전에 적용한 이유는 3.5 버전은 무료로 제공되므로, 연구 초기 단계에서 비용 효율적 방법을 모색하는데 유리하다 판단하였고, 3.5 버전에서 효과적으로 작동하는 시작 프롬프트는 4.0 버전에서도 기본적으로 유효할 것이라 예상했기 때문이다. 8,000개의 어휘를 기억하고 주로 영어를 사용하던 3.5 버전에 비해 64,000개의 어휘를 기억하고 영어를 제외한 25개 언어로 작업할 수 있는 향상된 다국어 기능이 적용된 4.0 버전은 3.5 버전의 기능을 대부분 포함하고 있다[21].

ChatGPT (OpenAI, San Francisco, CA, USA)는 자연어 처리를 위한 인공지능 대화형 모델로서 복잡한 질문을 이해하고 요약할 수 있는 능력을 가지고 있다 [18,19]. Gilson 등[20]의 연구에 의하면 ChatGPT가 학습을 지원하는 대화형 의학 교육 도구로 잠재적으로 적용될 수 있는 사례와 함께 논리적 오류, 정보 오류, 통계적 오류를 평가하고자 하였다. 다양한 연구에서 검증하고자 했던, AI의 오류 가능성 평가에도 불구하고 본 연구에서 평가하고자 하는 병력청취 대화형 시나리오와 관련한 응답과는 차이가 있다. 흉통이나 복통을 주요 증상으로 하는 응급환자의 추정되는 대표적인 진단을 중심으로 주요 증상을 기반으로 응답을 해주기 때문에 최근 연구에서 파악하고자 하였던 여러 가지 AI의 오류에서 좀 더 자유로울 수 있다고 판단한다. 하지만 AI가 환자 역할

로서 생성하는 자연스러운 대화의 표현이 중요하다 판단하였고, 응답하는 정보 또한 대표적 질환의 주요 증상과 충분한 연관성이 있어야 한다는 점에서 최소한의 논리적 오류와 정보 오류를 전문가 집단을 통해 파악할 필요가 있었다. 반복 테스트를 통해 발견한 환자의 증상 시간과 관련한 질문에서 “통증은 몇 시간 전부터 시작되었어요.” 라는 모호한 응답이 생성되었고, “최대한 정확히 말씀해주세요.” 라고 재질문하였다. 이에 “통증은 정확히 오늘 아침 일어났을 때부터 시작되었습니다.” 라고 응답하였고 “얼마전이지요?” 라고 재질문한 결과 “통증은 대략 2시간 전부터 시작되었습니다.” 라고 응답을 생성하여 사용자의 추가 질문에 적절히 반응하는 것으로 나타났다.

본 연구에서 흉통과 복통의 주요 증상에 따른 시작 프롬프트 입력 후 병원 전 단계 응급환자 병력청취 프로토콜에 입각해 환자 역할을 지시한 ChatGPT에게 한글로 질문을 하였고 병원 전 단계에서 병력청취를 위해 수집하는 정보인 주증상, 발병상황 및 시간, 유발요인, 완화요인, 통증 위치와 방사통에서부터 알레르기를 포함한 과거력, 가족력, 사회력까지 적절한 응답을 생성하는 수준, 즉 응급구조학과 학생이나 병원 전 단계 응급의료종사자가 사용할 수 있는 수준의 응답을 제공할 수 있음을 나타내며, ChatGPT의 실제 교육 도구로서의 잠재력을 시사한다. 이러한 결과는 ChatGPT와 같은 언어 모델은 의대생이 진단 및 문제 해결 기술을 연습하는데 사용할 수 있는 현실적인 환자 사례 시나리오를 생성할 수 있고 이러한 시나리오에는 환자병력, 증상 및 진단검사 결과가 포함될 수 있으므로 학생들이 자신의 지식을 실제 상황에 적용하는 연습을 할 수 있는 특징을 언급한 연구 [22]와 일치하는 부분이 많다.

병력청취가 종료된 다음 시나리오를 종료하고 지금까지 병력청취 시나리오에 대한 총평과 피드백을 요청하였다. ChatGPT는 적절한 질문에 대한 칭찬과 함께 추가로 필요한 병력청취 관련 질문에 대해서도 소개하였으며 시나리오에 등장한 환자의 현장 추정 진단도 요청에 의해 응답해주었다. 이는 AI를 활용하여 대화형 시뮬레이션 교육에서 그치지 않고 추가적인 질문에 의해 학생들마다 개인화된 학습 경험을 제공함으로써 관련 지식의 복잡하고 다면적인 개념을 더 잘 이해하는데 도움이 될 수 있다고 판단하였다.

ChatGPT와의 질의응답식 대화는 학생들이 환자 상황을 분석하고 적절한 치료 및 처치 전략을 제시하는 사고 능력을 개발하는데 도움을 줄 수 있다. 이는 비판적

사고와 의사소통 능력, 두 가지 핵심 능력의 실질적인 향상을 가능하게 한다. 더 나아가 ChatGPT는 체계적인 대화 생성을 통해 학생들에게 의사소통을 연습할 수 있는 기회를 제공할 수 있다. 이는 학생들이 다양한 의학적 상황에 대응하며 의사소통 역량을 향상하는데 큰 도움이 되며 학생 자신들만의 의사소통 스타일을 개발하고 개선하는 데에도 도움이 된다[23].

현재 의학 교육에서 널리 활용되는 평가 도구로 표준화 환자를 이용한 객관화 구조화 시험(OSCE: Objective Structured Examination)은 의과대학 학생이 표준화 환자와 상호작용하는 다양한 임상 상황에서의 역할극 대본을 활용한다[24]. Hong 등[25]의 연구에서는 의사 국가고시에서 활용될 수 있는 자료를 ChatGPT에게 PDF로 학습시킨 후 역할극 대본을 완성하였고, 연구진의 추가 피드백을 통해 구체적이고 상세한 임상 상황과 환자와 의사 역할의 특징을 살린 역할극 대본을 완성한 바 있다. 응급구조학 분야에서도 병원 전 단계 응급환자 병력 청취는 응급구조사에게 중요한 기술 중 하나이며, 1급 응급구조사 자격 국가시험 실기시험의 한 항목으로 운영되고 있다[26]. 이와 관련해 내과 환자평가 항목으로 1급 응급구조사 자격시험 응시자들은 흉통 또는 복통의 주요 증상을 중심으로 병력청취를 시행하게 되는데 응시자의 질문에 따라 시험 채점관이 직접 구두로 응답해 주는 방식으로 진행되고 있고 대부분의 1급 응급구조사 병력청취 응시 과목 준비생들은 자체적으로 질문과 응답으로 이루어진 시나리오 역할극 대본을 만들어 암기하거나 학생 및 교육자와 함께 짝으로 연습하고 있다. 본 연구에서는 ChatGPT에 추가적 자료 학습 없이 역할극 시나리오를 시도해 보았고 AI는 주소와 관련한 대표적 질환의 주요 증상들을 적절하게 생성하였다. 하지만 이는 테스트를 기반으로 하는 채팅 방식으로 병원 전 단계 응급환자 병력청취를 위한 기본적인 개념 파악에는 도움이 되지만 역할극처럼 실제 의사소통과는 다르다는 점에서 한계가 있다. 연구자는 이와 관련하여 ChatGPT 4.0의 음성 서비스를 활용해 개발한 시작 프롬프트를 입력하고 음성으로 병력청취 시나리오를 시도해 보았다. 결과적으로 AI의 생성된 응답은 텍스트 기반의 채팅 방식이 음성 대화 방식으로 단순 변형되어 성공적으로 시나리오를 수행해 볼 수 있었다. 이는 학생 스스로 실제 의사소통하는 것과 유사하게 연습할 수 있는 학습 도구로 충분히 활용할 수 있는 가능성이 있다고 판단되었다. 하지만 현재는 4.0 버전이 유료 버전이고, 하루에 질문할 수 있는 양이 제한적이기 때문에 충분한 학습 효과를 얻어내는데 한계

가 있을 수 있다.

본 연구에서 ChatGPT 3.5와 4.0 버전을 비교한 것은 학생들에게 적용할 수 있는가에 대한 경제적 측면 등 현실적 측면 또한 고려하고자 한 것으로, 4.0 버전이 3.5 버전에 비해 좀 더 안정적인 진행이 가능한 것으로 판단되지만 그럼에도 불구하고 학습 주체에 따라 일정 기간에 걸친 최신 업데이트가 꼭 필요하지 않은 주제에 대해서는 현재 무료 버전인 ChatGPT 3.5에서도 대화형 병력청취 시나리오 교육이 가능함을 파악하고자 함이었다. 추후 학생을 대상으로 ChatGPT 활용 교육의 사용성 평가와 학습 만족도 평가를 통해 AI와 관련한 학습 도구로서의 가능성을 보다 구체화할 필요가 있을 것으로 판단한다.

최근 많은 연구에서 의학 교육의 ChatGPT 사용이 관련 분야의 지식에 대한 이해를 향상하는데 효과적이었다는 결과를 제시하였고, 의학 교육의 다양한 영역에서 활용되며, 더 적극적인 활용을 강조하고 있다[27-29]. 이를 위해 AI의 특성과 용도를 이해하고, 적절하게 활용하기 위한 경험이 필요하다. 이에 본 연구는 ChatGPT가 병원 전 단계 응급의료 서비스 분야에서 요구되는 응급환자에 대한 정보 수집 능력과 의사소통 능력 교육 도구로서 그 활용 가능성을 평가해 본 첫 번째 연구라는 점에서 의의가 있다고 판단한다. 그럼에도 불구하고 본 연구에서는 다음과 같은 제한점이 있다. 첫째, 연구에 참여한 대상자는 119 구급대원, 응급의료센터, 대학에 소속한 1급 응급구조사 30명으로 한정되어 있다. 이는 연구결과의 일반화에 한계가 있을 수 있다. 다양한 배경과 경험을 가진 더 많은 대상자를 포함함으로써 결과의 일반성을 높일 필요가 있다. 둘째, 연구과정에서 최종진단명을 명확히 확인하지 않은 점은 본 연구의 중요한 제한점 중 하나이다. 이는 ChatGPT의 응답이 실제 임상 상황과 얼마나 일치하는지를 평가하는 데 있어 중요한 요소이므로 향후 연구에서는 이를 보완할 필요가 있다. 또한 다양한 응급상황과 증상을 포함한 시나리오를 추가하여 ChatGPT의 응답 범위를 확장하고 다양한 상황에서의 응답 일관성을 평가해야 할 것이다. 셋째, ChatGPT 모델은 지속적으로 업데이트되고 개선되므로, 연구결과가 향후 모델 업데이트에 따라 달라질 수 있어, 최신 모델에 대한 지속적인 평가가 필요하다. 넷째, 본 연구에서는 SUS를 사용하여 사용성을 평가하였으나, 이는 사용자 경험의 일부 측면만을 반영한다. 다양한 평가 도구와 방법을 사용하여 더욱 포괄적인 평가를 수행할 필요가 있다.

5. 결론

의료분야에서 환자와의 소통 능력은 필수적이며 응급 의료를 담당하는 응급의료종사자와 관련 분야를 공부하고 있는 학생들에게 병원 전 단계 응급환자에 대한 병력 청취 교육은 매우 중요하다. 본 연구에서는 ChatGPT를 활용해 개발된 시작 프롬프트 입력을 통해 응급구조학 분야에서 주로 다루고 있는 주증상을 중심으로 텍스트 기반 대화형 시나리오를 시도해 보았다. ChatGPT는 추가적인 관련 자료 학습 없이도 병원 전 단계 응급환자 병력청취와 관련한 질문에 적절한 응답을 생성하였다. 이는 한글 기반의 입력을 통한 ChatGPT의 높은 접근성을 확인한 것으로 이 과정을 통해 관련 분야의 주요 지식과 역량을 효과적으로 학습할 수 있는 중요한 도구로서 가능성이 높다고 판단하며, 텍스트를 기반으로한 병원 전 단계 응급환자 병력청취 대화형 시나리오에서 ChatGPT 3.5 버전과 4.0 버전의 교육 도구로의 차이는 없다고 판단한다. 하지만 본 연구의 ChatGPT 사용 가능성 평가는 일부 전문가 집단을 대상으로 시행되었으며, 시스템 사용성 평가를 목적으로 하는 일개 평가 도구만을 사용했다는 점에서 한계가 있다. 추후 응급구조학과 학생으로 대상자를 확대하여 교육 활용성을 평가하고, 다양한 교육 만족도 평가 도구를 추가 선택하는 등의 후속 연구를 제안한다.

References

- [1] Park JM, Kim SS, "Comparison study of satisfaction levels on Field Practice of EMT college students by 5 Practice Fields", *Korean J Emerg Med Ser*, Vol.2, pp.7-25, 1998.
- [2] Keifenheim KE, Teufel M, Ip J, Speiser N, Leehr EJ, et al, "Teaching history taking to medical students: a systematic review", *BMC Med Educ*, Vol.15, No.159, 2015. DOI: <https://doi.org/10.1186/s12909-015-0443-x>
- [3] Lee YM, Kim JH, Ahn DS, et al, "Patient's attitudes toward the involvement of medical students in their care", *Korean journal of medical education*, Vol.16, No.1, pp.41-50, 2004.
- [4] Bak YI, Park EW, "Problems of the curriculum for the improvement of the medical communication in Korean medical schools", *Korean academy on communication in healthcare*, Vol.4, No.2, pp.34-42, 2009.
- [5] Kim SK, Eom MR, Park MH, "Effects of nursing education using virtual reality: a systematic review", *Jour of Ko Con a*, Vol.19, No.2, pp.661-670, 2019. DOI: <https://doi.org/10.5392/JKCA.2019.19.02.661>
- [6] Jeon HJ, "Exploring study on virtual reality utilization strategies in scenario-based nursing simulation: an integrative review", *J Korean Soc Simul Nurs*, Vol.7, No.1, pp.45-56, 2019. DOI: <https://doi.org/10.17333/jkssn.2019.7.1.45>
- [7] Mathisen C, Bjørk IT, Heyn LG, Jacobsen TI, Hansen EH, "Practice education facilitators perceptions and experiences of their role in the clinical learning environment for nursing students: a qualitative study", *BMC Nurs*, Vol.22, No.1, pp.165, 2023. DOI: <https://doi.org/10.1186/s12912-023-01328-3>
- [8] Burgess A, van Diggele C, Roberts C, Mellis C, "Key tips for teaching in the clinical setting", *BMC Med Educ*, Vol.20, No.2, pp.463, 2020. DOI: <https://doi.org/10.1186/s12909-020-02283-2>
- [9] Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P, "Introducing artificial intelligence training in medical education", *JMIR Med Educ*, Vol.5, No.2, pp.e16048, 2019. DOI: <https://doi.org/10.2196/16048>
- [10] Rampton V, Mittelman M, Goldhahn J, "Implications of artificial intelligence for medical education", *Lancet Digit Health*, Vol.2, pp.e111-e112, 2020. DOI: [https://doi.org/10.1016/S2589-7500\(20\)30023-6](https://doi.org/10.1016/S2589-7500(20)30023-6)
- [11] Wartman SA, Combs CD, "Medical education must move from the information age to the age of artificial intelligence", *Acad Med*, Vol.93, No.8, pp.1107-1109, 2018. DOI: <https://doi.org/10.1097/ACM.0000000000002044>
- [12] Arif TB, Munaf U, Ul-Haque I, "The future of medical education and research: is ChatGPT a blessing or blight in disguise?", *Med Educ Online*, Vol.28, No.1, pp.2181052, 2023. DOI: <https://doi.org/10.1080/10872981.2023.2181052>
- [13] Cope B, Kalantzis M, Zhai CX, Krussel A, Searsmith D, Ferguson D, et al, "Maps of medical reason: applying knowledge graphs and artificial intelligence in medical education and practice", *Bioinformational philosophy and postdigital knowledge ecologies*, pp.133-159, 2022. DOI: https://doi.org/10.1007/978-3-030-95006-4_8
- [14] Sallam M, "ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns", *Healthcare*, Vol.11, No.6, pp.887, 2023. DOI: <https://doi.org/10.3390/healthcare11060887>
- [15] Khan RA, Jawaid M, Khan AR, Sajjad M, "ChatGPT - Reshaping medical education and clinical management", *Pak J Med Sci*, Vol.39, No.2, pp.605-607, 2023. DOI: <https://doi.org/10.12669/pjms.39.2.7653>
- [16] Brooke J, SUS- A quick and dirty usability scale, London:Taylor & Francis, 1996, pp.189-194.
- [17] Ahn JH, Park HO, "Development of a case-based nursing education program using generative artificial

intelligence”, *J Korean Acad Soc Nurs Educ*, Vol.29, No.3, pp.234-246, 2023.
DOI: <http://doi.org/10.5977/jkasne.2023.29.3.234>

[18] OpenAI, OpenAI: Models GPT-3, Available online, <https://beta.openai.com/docs/models> Accessed January 14, 2023.

[19] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al, “Language models are few-shot learners”, *Adv. Neural Inf. Process. Syst.*, Vol.33, pp.1877-1901, 2020.
DOI: <https://doi.org/10.48550/arXiv.2005.14165>

[20] Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, 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*, Vol.9, pp.e45312, 2023.
DOI: <https://doi.org/10.2196/45312>

[21] Kim SM, Kim JH, Choi MJ, Jeong SH, “Evaluation of the applicability of ChatGPT in biological nursing science education”, *Journal of Korean Biological Nursing Science*, Vol.25, No.3, pp.183-204, 2023.
DOI: <https://doi.org/10.7586/jkbns.23.0013>

[22] 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*, Vol.9, pp.e46885, 2023.
DOI: <https://doi.org/10.2196/46885>

[23] Hu R, Fan KY, Pandey P, Hu Z, Yau O, Teng M, et al, “Insights from teaching artificial intelligence to medical students in Canada”, *Commun Med (Lond)*, Vol.2, No.1, pp.63, 2022.
DOI: <https://doi.org/10.1038/s43856-022-00125-4>

[24] Garg T, “Artificial intelligence in medical education”, *Am J Med*, Vol.133, No.2, pp.e68, 2020.
DOI: <https://doi.org/10.1016/j.amimed.2019.08.017>

[25] Hong HM, Kang YJ, Kim YJ, Kim BS, “Application of artificial intelligence in medical education: focus on the application of ChatGPT for clinical medical education”, *Journal of Medicine and Life Science*, Vol.20, No.2, pp.53-59, 2023.
DOI: <https://doi.org/10.22730/jmls.2023.20.2.53>

[26] Korea Health Personnel Licensing Examination Institute. Web site.
https://www.kuksiwon.or.kr/notice/brd/m_51/download?brd_id=NOTICE&seq=2615&data_tp=A&file_seq=2

[27] Valikodath NG, Cole E, Ting DSW, Campbell JP, Pasquale LR, Chiang MF, et al, “Impact of artificial intelligence on medical education in ophthalmology”, *Transl Vis Sci Technol*, Vol.10, No.7, 2021.
DOI: <https://doi.org/10.1167/tvst.10.7.14>

[28] Ahuja AS, Polascik BW, Doddapaneni D, Byrnes ES, Sridhar J, “The digital metaverse: applications in artificial intelligence, medical education, and integrative health”, *Integr Med Res*, Vol.12, No.1, pp.100917, 2023.

DOI: <https://doi.org/10.1016/j.imr.2022.100917>

[29] Park CJ, Yi PH, Siegel EL, “Medical student perspectives on the impact of artificial intelligence on the practice of medicine”, *Curr Probl Diagn Radiol*, Vol.50, No.5, pp.614-619, 2021.
DOI: <https://doi.org/10.1067/j.cpradiol.2020.06.011>

김 성 주(Seong Ju Kim)

[정회원]



- 2014년 8월 : 인제대학교 보건대학원 보건학과 (보건학석사)
- 2017년 3월 : 인제대학교 일반대학원 보건학과 (보건학박사)
- 2016년 9월 ~ 2023년 8월 : 부산보건대학교 응급구조과 조교수
- 2023년 9월 ~ 현재 : 동명대학교 응급구조학과 조교수

<관심분야>

응급의학, 시뮬레이션교육