Print ISSN 1225-7044 / Online ISSN 2288-985X

JKNA

ORIGINAL ARTICLE

https://doi.org/10.17340/jkna.2024.0029
Journal of The Korean Neurological Association 42(3):233-240, 2024

ChatGPT가 한국 신경과 의사들에게 실제 임상에서 유용할까?

고판우[a,b]

경북대학교 의과대학 신경과학교실[a], 경북대학교병원 신경과[b]

Will ChatGPT be Useful for Korean Neurologists in Clinical Practice?

Pan-Woo Ko, MD[a,b]

Department of Neurology, School of Medicine, Kyungpook National University, Daegu, Korea[a]
Department of Neurology, Kyungpook National University Hospital, Daegu, Korea[b]

Address for correspondence
Pan-Woo Ko, MD
Department of Neurology, School of Medicine, Kyungpook National University, 680 Gukchaebosang-ro, Jung-gu, Daegu 41944, Korea
Tel: +82-53-200-5765
Fax: +82-53-422-4265
E-mail: panwoo.ko@gmail.com

Received April 26, 2024
Revised June 8, 2024
Accepted June 11, 2024

Background: Since the emergence of artificial intelligence (AI), such as ChatGPT, there has been extensive research on its utility in the medical field. This study evaluates the potential of utilizing the ChatGPT AI platform to provide specialized advice to Korean neurologists.
Methods: Through a multifaceted approach, it assesses both the knowledge and practical clinical utility using neurology training evaluation questions and clinical case reports.
Results: The accuracy of ChatGPT in answering medical exam questions (45.1%) is comparable to or slightly lower than previous studies. The study also examines performance of ChatGPT in providing advice based on clinical case reports, finding that while it can offer suitable diagnostic suggestions and propose relevant tests in some cases (8/12), its overall diagnostic accuracy is limited (2/12). Limitations include the platform's text-based nature, which hinders the interpretation of image-based questions, and the uneven distribution of case reports across subfields.
Conclusions: Despite these limitations, ongoing advancements in AI technology offer prospects for improved performance and expanded applications in the medical field, warranting continued research and monitoring of AI's role in medicine.
J Korean Neurol Assoc 42(3):233-240, 2024

Key Words: Artificial intelligence, Neurology, Case reports, Decision support technique, Examination questions

서 론

바야흐로 인공지능(artificial intelligence, AI)의 시대이다. 사물 인터넷을 활용한 가전제품에서부터 바둑, 체스, 엔터테인먼트 산업을 비롯한 다양한 분야에서 AI가 이용되고 있다. 특히 Open AI사에서 개발한 ChatGPT (Open AI, San Francisco, CA, USA)와 같은 텍스트 기반의 생성형 AI의 개발은 개인들이 다양한 분야에서 AI를 활용할 수 있도록 하였다.[1] ChatGPT는 대규모 텍스트 기반의 사전 학습된 데이터를 바탕으로 프롬프트라는 대화 형식의 알고리즘을 기반으로 사용자에게 적절한 답을 제공함으로써 사용이 간편하고 이해하기 쉬워 빠른 속도로 대중화되었다. 특히 많은 양의 데이터를 처리하고 해석해야 하는 과학 분야에서 ChatGPT의 유용성과 효과에 대해 연구와 논의가 진행되면서 연구자들은 미래 과학 분야의 유용한 도구로의 ChatGPT와 같은 AI의 잠재적 가치에 주목하고 있다.[2-4]

의학 분야에서도 ChatGPT의 유용성과 활용성에 대한 연구가 진행되고 있다. 최근 ChatGPT가 미국 의사자격시험 합격에 해당하는 정답률을 보인 연구는 AI가 의학 분야에서도 유용하게 활용될 수 있다는 가능성을 제시했다.[5] 해당 연구에

서는 ChatGPT가 의과대학 3학년 수준의 수행 능력을 보였다고 평가되었다. 보다 전문적인 분야에서도 우수한 성과를 보였는데 509개의 문제로 구성된 신경과 전문의 시험 문제 은행에서 65.8%의 정확도를 보였다는 연구도 있다.[6] 뿐만 아니라 실제 임상 현장에서 ChatGPT를 활용한 다양한 사례들이 소개되고 있으며 실제 연구에서도 그 유용성이 증명되고 있다.[7-11]

ChatGPT는 다양한 언어를 지원하지만 학습된 언어의 특성상 영어를 기반으로 한 프롬프트에서 답변 수행률이 우수한 것으로 알려져 있다. 본 연구에서는 의학 분야에서 ChatGPT 관련 선행 연구를 참조하여 신경학 분야에서 활용 가능성을 확인하고 특히 한국어를 사용하였을 때 AI의 유용성을 평가 문제를 통한 의학 지식 측면과 증례를 통한 문제 해결 측면으로 구분하여 평가하였다.

대상과 방법

AI 모델은 대중적으로 공개된 Open AI의 ChatGPT 3.5 모델을 사용하였다. 신경학 분야에서 의학 지식을 평가하는 도구로 신경학 전공의 수련평가 문항으로 사용된 2023년 인서비스 시험 120문항을 활용하였다. 또한 실제 임상 현장에서의 유용성 평가를 위해 가장 최근 발행된 대한신경과학회지의 증례보고 12건을 참고하였다. 저자의 선택 편견을 줄이고자 가장 최근 인서비스 시험 문항 전체와 가장 최근호인 2024년 2월호에 게재된 증례보고 논문을 모두 포함하였다.

1. 인서비스 시험 문항 입력 프로토콜

일반적으로 ChatGPT는 프롬프트 기반의 대화를 통해 대답을 도출하기 때문에 적절한 역할과 질문의 목적 기술이 보다 적절한 답을 도출하는 것으로 알려져 있다. 본 연구의 목적은 ChatGPT 자체의 의학 지식 평가 수행률을 보기 위한 것으로 별도의 역할은 제시하지 않고 "다음 문제의 정답은?"이라는 간략한 프롬프트를 제시하고 온라인에 공개된 시험 문항 원문과 답가지를 그대로 입력하였다. 단 의학 시험 문제의 특성상 사진과 그림이 포함된 문제가 다수 있어 텍스트 기반의 ChatGPT에 입력하는 데 제한이 있는 경우 출제자의 해설에 기술된 그림 설명을 Fig. 1과 같이 그대로 입력하여 질문하였다. 그림 설명을 별도 입력한 경우는 구분하여 결과를 도출하였고 전체 문항의 정답률과 세부 분과별 정답률을 구분하여 평가하였다. ChatGPT의 정답률과 실제 전공의들의 시험 결과를 비교하기 위하여 개인 정보를 포함하지 않은 성적 데이터를 활용하여 분석하였고 성적 분포 시각화를 위해 ggplot 패키지(R Studio, Boston, MA, USA)를 사용하였다.

2. 증례보고 분석

ChatGPT의 문제 해결 능력을 평가하고자 대한신경과학회지의 증례보고를 활용하였다. 일관된 형식으로 입력하고 저자의 선택적 편견을 배제하기 위해 논문을 작성하는 현 시점에

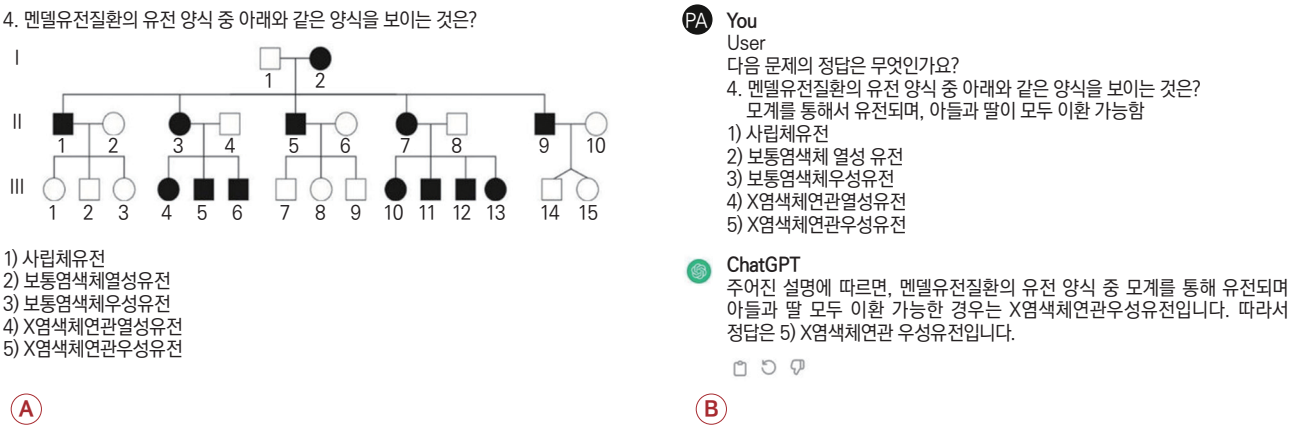

Figure 1. (A) Original neurology in-service examination question. (B) Example of ChatGPT prompt input. The figure description is entered exactly as in the in-service examination explanation section.

서 가장 최근에 발표된 대한신경과학회지 2024년 2월호의 증례보고 12건을 모두 포함하였다.[12-23] 증례보고의 경우 신경과 의사로써 문제 해결 능력 평가를 위한 목적으로 다음과 같은 입력 프롬프트를 적용하였다(Fig. 2-A).

<Input 2-1>의 프롬프트: 신경과 의사로서 다음과 같은 병력과 신경계 소견을 보이는 환자가 내원하였다. 감별해야 할 진단은 무엇인가?

<Input 2-2>의 프롬프트: 진단을 위해서 어떤 검사를 추가해야 하는가?

<Input 3>의 프롬프트: 시행한 검사 결과가 아래와 같다. 어떠한 진단을 고려할 수 있는가?

<Input 2-1>에서는 해당 프롬프트와 증례보고에 기술된 환자의 병력과 신경계 소견 원문을 그대로 입력하였다. ChatGPT가 제시한 감별 진단의 범위가 증례보고의 최종 진단을 포함한 경우에 적절한 답을 한 것으로 간주하였다. 이어서 <Input 2-2>의 질문을 연속으로 입력하였다. <Input 2-2>의 답변이 증례보고에 기술된 주요한 진단 검사 항목을 포함한 경우 적절한 답을 한 것으로 간주하였다. <Input 3>에서는 ChatGPT의 이전 답변에 상관없이 해당 프롬프트와 증례보고의 진단 검사 결과 원문을 연속된 질문으로 입력하였다. 이후 내린 진단이 증례의 진단과 일치하는 경우를 적절한 답을 한 것으로 간주하였다. 최초 제시한 감별 진단, 추가적인 검사 그리고 최종 진단 모두 적절한 답을 제시한 경우를 온전한 문제 해결 능력을 보인 것으로 평가하였고 최초 감별 진단과 검사법 제시가 미흡한 부분이 있어도 적절한 최종 진단을 제시한 경우를 부분적인 문제 해결 능력을 갖춘 것으로 평가하였다.

결 과

1. 인서비스 시험 문항을 이용한 분석

2023년 대한신경과학회 인서비스 시험 평가 문항은 총 120문항으로 구성되어 있다. 그중 답가지가 그림으로 되어 있거나 ChatGPT에서 질문 프롬프트 입력이 불가능한 그림이나 동영상이 포함된 7문항은 제외하였다. 텍스트로만 구성된 문항은 61문항, 그림, 사진, 표를 포함한 문항은 52문항으로 구성되었다. 시험 문항 자체에는 세부 분과 분류가 되어 있지 않으나 일반적으로 세부 분과에 따라 문항이 구성되어 있었다. ChatGPT에 입력 가능한 문항은 일반신경학 15문항, 뇌졸중

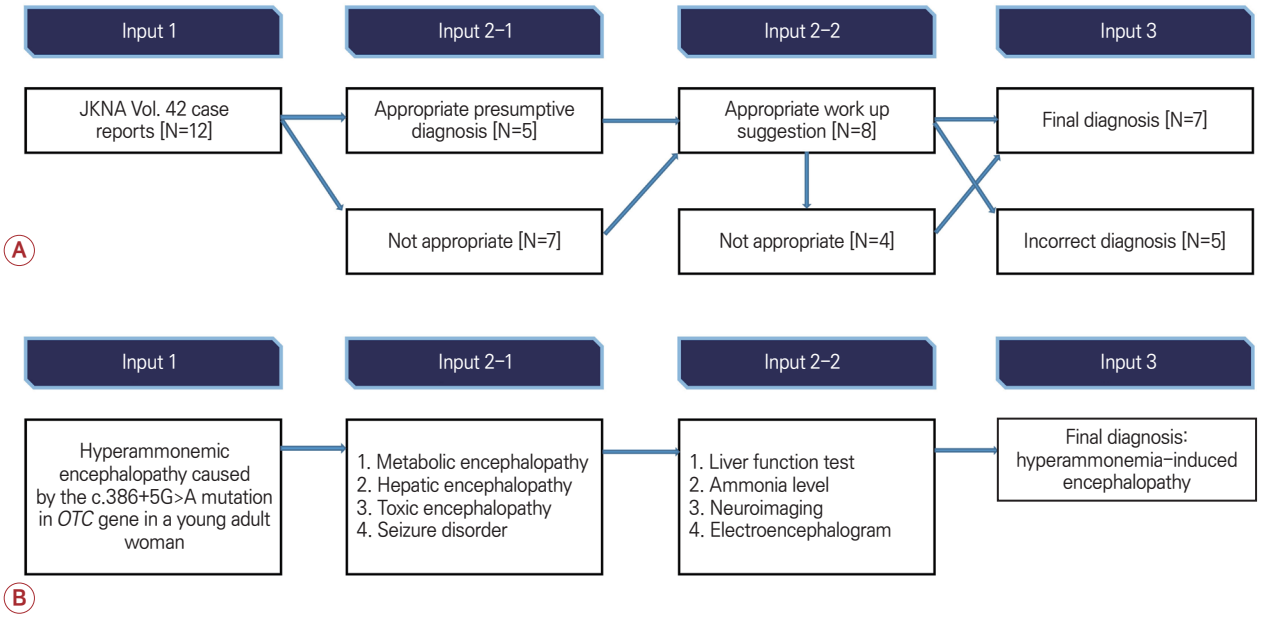

Figure 2. (A) Schematic representation of ChatGPT input prompts using case reports published in JKNA volume 42. (B) Summary of ChatGPT's response of an example case. JKNA; Journal of the Korean Neurological Association.

학 18문항, 말초신경질환 13문항, 뇌전증학 10문항, 중추신경감염학 6문항, 수면질환 5문항, 두통 8문항, 어지럼증 4문항, 이상운동질환 11문항, 치매 15문항, 중환자의학 8문항으로 구성되었다. 전체 문항 중 정답률은 45.1% (51/113)였으며 텍스트 원문 그대로 입력한 문항의 정답률은 42.6% (26/61)였다. Fig. 3의 예와 같이 정답 및 오답을 도출한 경우 모두에서 ChatGPT는 답을 제시하고 간략한 해설을 첨부한 결과를 나타내었다. 2023년 인서비스 시험의 응시 전공의 연차별 성적 분포는 Fig. 4와 같았다. 각 연차별 평균은 백분율(평균±표준편차)로 환산하였을 때 1년차 46.6±10.7, 2년차 55.3±10.6, 3년차 63.0±10.0, 4년차 71.2±8.8로 ChatGPT의 정답률은 전공의 1년차 평균 근사치로 측정되었다. 세부 분류에서는 Fig. 5와 같이 수면(80.0%), 중환자의학(62.5%), 치매(60.0%), 뇌전증(50.0%), 두통(50.0%), 중추신경감염학(50.0%), 이상운동질환(45.5%), 뇌졸중학(38.9%), 일반신경학(33.3%), 말초신경질환(30.8%), 어지럼증(0.0%) 순으로 높은 정답률을 보였다.

2. 대한신경과학회지 증례보고를 이용한 분석

증례에서 기술된 현재 증상의 경과와 신경계진찰 소견을 기반으로 우선 감별해야 할 진단을 물었을 때 전체 12개의 증례 중 5개의 증례에 대해서 최종 진단에 해당하는 범주의 감별 진단들을 제시하였다. 한 가지 예로 고암모니아뇌병증의 증례[13]에서는 ChatGPT가 대사뇌병증을 감별 진단으로 제시하였는데 이와 같이 해당 범주에 포함된 진단인 경우 적절한 답을 한 것으로 평가하였다. 적절한 진단 검사를 제안한 경우는 전체 12예 중 8예에서 해당 논문의 본문에 제시된 주요 진단을 검사법으로 제안하였다. 최종 진단은 7예에서 증례와 일치하는 진단을 제시하였다(Fig. 2-A). 그러나 이는 각 단계의 프롬프트 입력에서 이전 감별 진단과 검사 제안의 결과에 관계없이 본문에 제시된 원문을 그대로 입력하여 다음 단계로 진행하였으므로 모든 단계에서 적절한 의학적 자문을 제공한 것으로는 볼 수 없었다. 전체 증례보고 중 모든 단계에서 적절한 답을 제시한 경우는 궤양대장염 환자에서 인플릭시맙 치료 후 발생한 시신경염 증례[18]와 젊은 성인 여성에서 *OTC*유전자의

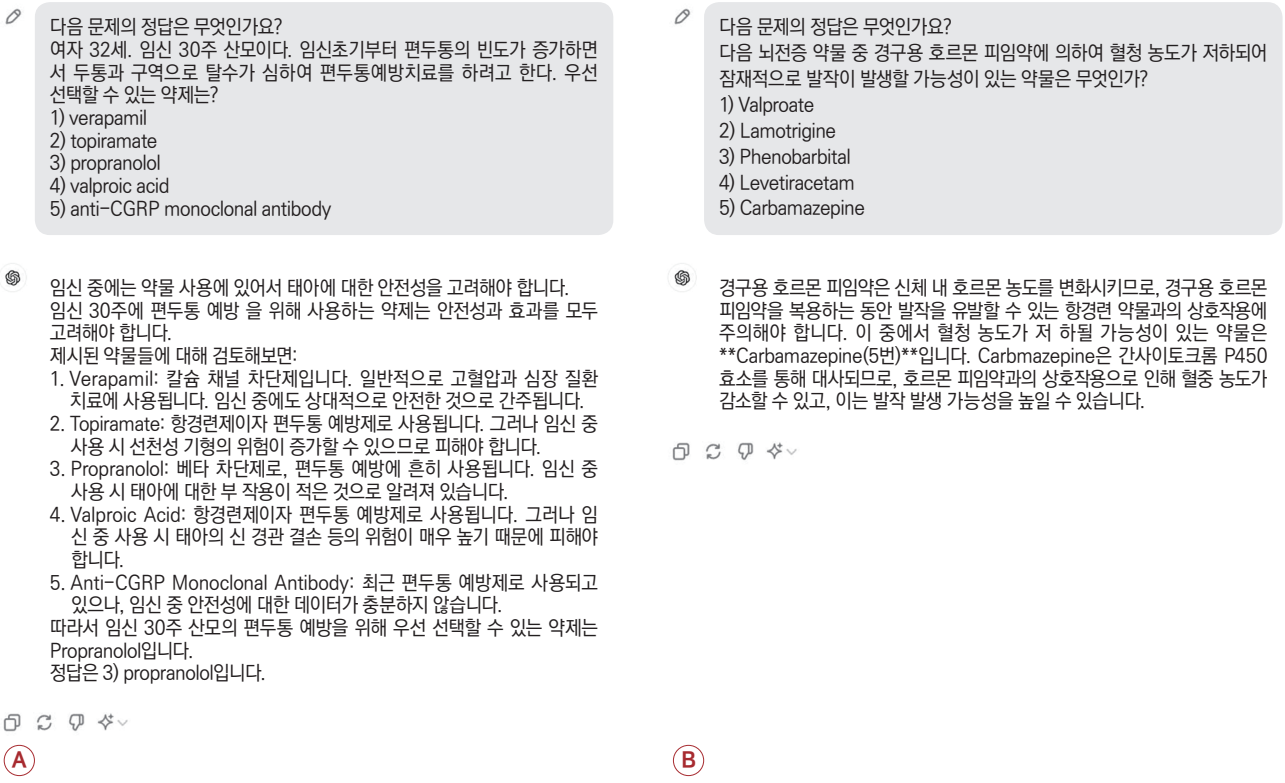

Figure 3. Examples of responses for correct (A) and incorrect (B) in-service exam questions. GCRP; calcitonin gene-related peptide.

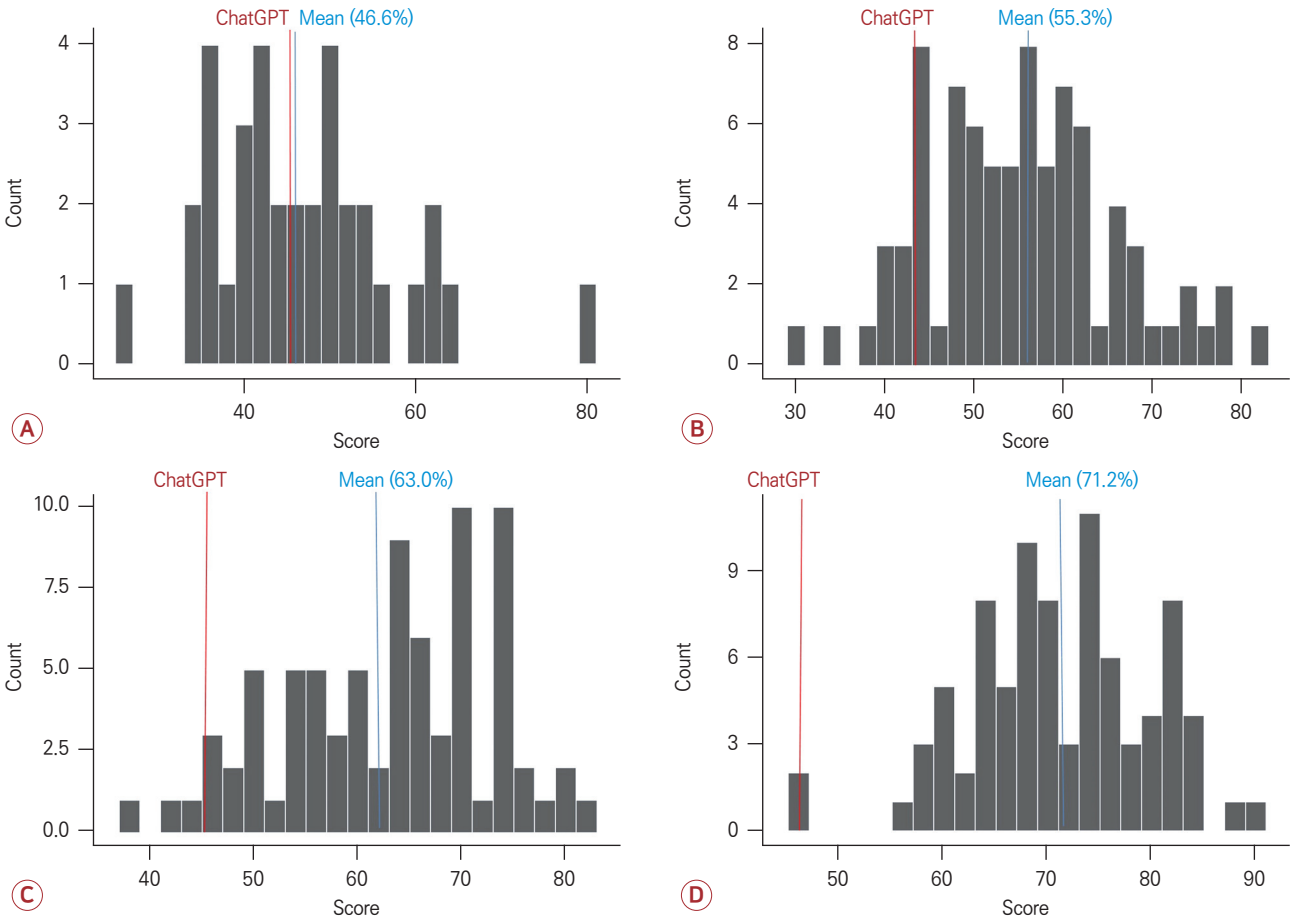

Figure 4. Histogram of 2023 in-service exam scores by year of residency. (A) First year, (B) second year, (C) third year, and (D) fourth year of residency. Blue line means average score of residents and red line means ChatGPT accuracy rate (%).

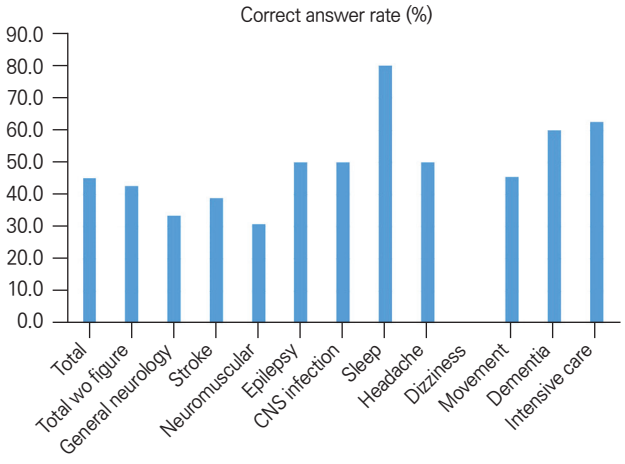

Figure 5. Subfield correct answer rates of neurology in-service examination questions. Total wo figure, total questions without figures (text only questions). wo; without, CNS; central nervous system.

C.386+5G>A 돌연변이로 발생한 고암모니아혈증 증례[13] 두 가지였다. ChatGPT의 답변 진행 과정의 예로 각 프롬프트 입력 단계마다 Fig. 2-B와 같이 최종 진단이 포함된 감별 진단, 논문에 기술된 진단 과정이 포함된 검사법을 추천하였고 증례와 일치하는 최종 진단을 도출하였다.

고 찰

1. ChatGPT를 이용한 연구와 비교

본 연구에서는 대표적인 AI 플랫폼인 ChatGPT를 활용하여 AI가 한국어를 사용하는 신경과 의사에게 전문적인 조언을 할 수 있는지 잠재적 가능성을 평가하고자 하였다. 지식적인 측면에서의 평가뿐만 아니라 실제 임상적인 활용도를 평가

하기 위해 신경과 의사의 수련 척도를 평가하기 위한 인서비스 시험 문항과 임상 증례보고 문헌을 활용하여 다각도로 평가하였다. 이전 연구들에서는 의학 시험 문제를 활용하여 다양한 의학 분과 영역 및 언어로 ChatGPT의 성능을 확인하였다. 미국의사자격시험(United States medical licensing examination, USMLE) step 1과 step 2에 해당하는 문제은행을 이용한 연구에서 ChatGPT가 44.0-64.4%의 정확도로 수험생 평균을 8.15% 상회하는 결과를 보였고[5] 직접 USMLE 문항을 활용한 평가에서도 45.4-61.5% 정확도로 합격에 준하는 수행률을 보였다.[24] 의사자격시험보다 전문적인 세부 영역의 의학 분야에서도 ChatGPT의 시험 수행률을 측정한 평가 연구를 하였는데 성형외과학, 마취통증학, 당뇨학, 안과학, 신경외과학, 약리학 등 여러 의학 분야에서 연구에 따라 46-100% 범위의 정답 수행률을 보여주었다.[25-30] 본 연구에서는 이전의 연구 결과와 비교하였을 때 비슷하거나 낮은 정답률을 보였는데 이는 문제의 난이도나 사용한 플랫폼의 영향일 것으로 생각된다. 기존 연구에서도 USMLE와 같은 의과대학 학사 수준의 의사자격시험의 경우는 60% 이상의 수행률을 보였다. 하지만 당뇨학(100%)을 제외한 전문의 수준의 시험에서는 45.4-69.7% 정도로 상대적으로 낮은 정답률을 보였다. 마찬가지로 신경학 인서비스 시험은 신경과를 전공하는 수련의들의 평가를 위한 것으로 자격을 평가하는 전문의 시험보다 상대적으로 난이도가 높은 것으로 알려져 있어 이전의 전문 의학 시험 결과와 비교할 수 있겠다. 연차별 실제 시험 성적 분포와 비교하였을 때 ChatGPT가 신경과 전공의 1년차의 평균과 비슷한 정답 수행률을 보이는 것을 확인할 수 있었다.

본 연구에서는 공개된 모델의 ChatGPT를 사용하여 평가하였으나 사용한 AI 플랫폼에 따라 수행률에 차이가 있을 수 있다. USMLE 문제은행을 이용하여 ChatGPT와 GPT-4 (OpenAI)의 수행률을 비교한 연구에서는 GPT-4 (90.0%)가 ChatGPT (62.5%)에 비해 우수한 수행률을 보였다. 대만에서 정신건강의학과 전문의 자격시험에서는 다양한 AI 모델들의 시험 수행률을 비교 분석하였는데 GPT-4가 ChatGPT, Bard (Google, Mountain View, CA, USA) 그리고 Llama-2 (Meta, Menlo Park, CA, USA)에 비해 우수한 수행률을 보이는 것으로 나타났다.[31] 따라서 향후 사용한 AI 플랫폼에 따른 수행률에 대한 비교 연구도 필요할 것으로 생각된다.

단순한 의학 지식을 평가하는 수준보다 실제 임상에서 의학적인 조언을 받는 데 도움이 될 수 있을지 평가하기 위해 증례보고를 활용하여 평가하였다. 증례보고는 일반적인 표준 사례에 비해 드물거나 교육적 가치가 있는 증례를 보고한 것으로 실제 신경과 의사로서 임상적인 판단에 도움이 되는지를 평가하는 데 적합하다 볼 수 있다. 증례에 따라 진단법이나 치료, 영상적 특이성을 보고하는 등 목적의 차이가 있을 수 있어 일괄적인 척도 평가에 제한이 있을 수는 있으나 저자의 선택적 편견을 가능한 배제하고자 증례를 추출하지 않고 동일 학술지, 동일 권호수의 증례를 모두 사용하였다. 최초 감별 진단과 추가 검사 제안 그리고 최종 진단까지 모두 적절한 답을 한 증례는 12예의 증례 중 2예에 불과했다. 5예의 증례에서 앞선 감별 진단 혹은 추가 검사 제안은 부적절하였으나 최종 진단은 적절한 답을 하였는데 이는 보고된 원문을 그대로 입력하는 방법을 사용하여 진단에 필요한 결과를 입력한 것이므로 실제 임상에서 도움이 되는 사례로 보기는 어렵다. 다만 병력과 신경계진찰 결과를 입력한 상태에서 적절한 검사법을 제안한 사례는 12예 중 8예로 진단하는 과정에서 조언을 얻는 데는 도움이 될 수 있겠다. 기존 연구에서 실제 임상에서 ChatGPT를 활용한 사례는 임상 증례를 정리하거나 구조화하고 결과를 해석하는 데 주로 사용하였다.[32] 또한 해당 증례를 바탕으로 의학 보고서를 작성하는 데 유용성을 평가하기도 하였다.[33] 출판된 증례보고를 기반으로 한 ChatGPT의 수행률 분석 연구는 드물었는데 한 연구에서 63예의 증례보고를 기반으로 감별 진단의 정확도를 평가하였고 74.6%의 정확도를 보였다고 하였다.[34] 이 연구는 본 연구의 <Input 2-1>에 해당하는 병력을 입력하고 감별 진단의 정확성을 평가한 것으로 본 연구 결과인 최초 감별 진단 정확도인 41.7% (5/12)와 비교할 수 있는 결과라 할 수 있다.

본 연구에서는 한국어를 사용한 신경학 분야에서 ChatGPT의 잠재적 활용성을 평가하기 위하여 서로 다른 접근법을 사용하였으나 몇 가지 한계점이 있다. 첫째, 신경학 분야의 특성상 평가 문항에서 많은 사진과 영상 자료가 제공되는데 ChatGPT의 경우 텍스트 기반의 프롬프트만 입력이 가능하여 이러한 자료 해석에 대한 평가가 제한적이었다. 그리고 텍스트 기반의

문항만 평가하더라도 정답률이 42.6%에 지나지 않아 실제 전문의 수준의 의학 분야에서 조언을 얻기에는 어렵다고 생각된다. 둘째, 연구의 목적이 한국어를 사용하였을 때 활용성을 평가하기 위한 것이므로 입력 자료로 한국어를 사용한 자료를 활용하여 실제 ChatGPT의 의학 분야 활용성에 비해 과소평가되었을 가능성이 있다. 단 연구 결과가 기존 연구들과 비교하였을 때 큰 차이가 없고 사용한 언어에 따른 의학 시험 수행률 차이를 연구한 다른 연구에서 사용 언어에 따른 큰 차이가 없었다는 보고가 있어 결과 해석에 참고할 수 있겠다.[35] 셋째, 선택 편견을 제거하고자 동일 학술지, 동일 권호수의 모든 증례를 포함하였으나 분과별 배분이 일정하지 않아 신경학 분야 전체에 대한 일반적인 결과로 해석하기에는 무리가 있다. 해당 학술지의 해당 권호수의 증례보고에서는 뇌졸중, 뇌전증 그리고 치매 세부 증례는 없었고 대사뇌병증, 뇌염 증례가 상대적으로 많았다. 따라서 향후 더 많은 증례와 세부 분과별로 배분된 분석이 필요할 것이다.

ChatGPT는 2022년 말 처음 공개된 이후 폭발적인 대중의 관심 속에 성장해 왔다.[1] 현재까지 발표된 연구를 종합하면 의학 분야에서는 의사자격시험의 합격선은 통과할 정도의 지식 수준은 갖추었으나 아직까지 보다 전문적인 의학적 판단과 분석에는 미흡함이 있는 것으로 보인다. 본 연구에서는 전문적인 의학 분야 중 신경학에서 한국어를 사용하였을 때 신경학 지식과 신경학 분야의 임상적 적용에서 유용성을 다각도로 분석하였다. 결과적으로 기존 연구 결과와 유사하거나 다소 부족한 결과를 보였다. 하지만 현재 수준에서도 전문적인 경험과 판단이 필요한 분야가 아닌 의학 교육이나 환자 교육에서의 활용성은 뛰어난 것으로 보고되고 있다.[36,37] AI 학습의 특성상 학습하는 데이터 양이 많아지고 고도화될수록 더욱 발전된 답을 보인다는 점을 고려할 때 향후 의학 분야에서도 잠재적인 활용성과 발전 가능성을 기대할 수 있다. 따라서 향후 다양한 AI 도구의 발전에 따라 의학 분야에서의 활용성과 유용성에 대해서는 지속적인 관심과 추적 연구가 필요할 것이다.

REFERENCES

1. OpenAI. Introducing ChatGPT. [online] [cited 2023 Nov 30]. Available from: https://openai.com/blog/chatgpt.
2. Nature. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 2023;613:612.
3. Else H. Abstracts written by ChatGPT fool scientists. *Nature* 2023;613:423.
4. Gordijn B, Have HT. ChatGPT: evolution or revolution? *Med Health Care Philos* 2023;26:1-2.
5. 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* 2023;9:e45312.
6. Chen TC, Multala E, Kearns P, Delashaw J, Dumont A, Maraganore D, et al. Assessment of ChatGPT's performance on neurology written board examination questions. *BMJ Neurol Open* 2023;5:e000530.
7. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, et al. Assessing the value of ChatGPT for clinical decision support optimization. *medRxiv* 2023 Feb 23. [Epub ahead of print]
8. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 Chatbot for clinical vignettes with common chief complaints: a pilot study. *Int J Environ Res Public Health* 2023;20:3378.
9. Potapenko I, Boberg-Ans LC, Hansen MS, Klefter ON, van Dijk EHC, Subhi Y. Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. *Acta Ophthalmol* 2023;101:829-831.
10. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr* 2023;7:pkad015.
11. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol* 2023;29:721-732.
12. Choi JH, Seo BS, An YE, Kim JM, Park MS, Lee SH. Neuromyelitis optica spectrum disorder after COVID-19 infection. *J Korean Neurol Assoc* 2024;42:35-39.
13. Choo YS, Koo GE, Kang YJ, Kang D, Ko YJ, Park JY, et al. Hyperammonemic encephalopathy caused by the c.386+5G>A mutation in OTC gene in a young adult woman. *J Korean Neurol Assoc* 2024;42:62-65.
14. Eun MY, Hah H, Hwang J. Coexistence of anti-Hu and anti-SOX1 autoantibodies in atezolizumab-related encephalitis. *J Korean Neurol Assoc* 2024;42:48-52.
15. Je Y, No JW, Park YE. Atypical presentation of Elsberg syndrome caused by herpes simplex virus type 2 infection. *J Korean Neurol Assoc* 2024;42:27-30.
16. Jeong BK, Lee JY, Son W, Na SJ. Reversible homonymous hemianopia associated with focal hyperperfusion in hyperglycemic state. *J Korean Neurol Assoc* 2024;42:57-61.
17. Kim HJ, Kim DR, Kim JG, Song SK, Kang CH. Non-cirrhotic hyperammonemic encephalopathy with portosystemic shunt. *J Korean Neurol Assoc* 2024;42:53-56.

18. Kim N, Han J, Park SJ, Chun YS, Ju Y, Kim HJ. Optic neuritis after infliximab treatment in a patient with ulcerative colitis. *J Korean Neurol Assoc* 2024;42:31-34.
19. Ko PW, Lee JM. A case of spinal muscular atrophy type 3a showing improvement in neurofilament light chain of cerebrospinal fluid: real-world evidence. *J Korean Neurol Assoc* 2024;42:40-43.
20. Lee JS. Maternally inherited diabetes and deafness presenting with memory impairment and gait disturbance. *J Korean Neurol Assoc* 2024;42:71-75.
21. Lee S, Kim JR, Kim YK, Bae H, Park SR, Kim K, et al. Atypical paroxysmal kinesigenic dyskinesia with paroxysmal exercise-induced dyskinesia. *J Korean Neurol Assoc* 2024;42:66-70.
22. Shin J, Jeon T, Oh K, Han JH, Kim CK, Lee KJ. Primary central nervous system lymphoma mimicking lacunar infarction. *J Korean Neurol Assoc* 2024;42:23-26.
23. Song Y, Lee S, Kwon Y, Kim B, Kim HK. Anti-NMDA receptor encephalitis presenting with language impairment and bradycardia. *J Korean Neurol Assoc* 2024;42:44-47.
24. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;2:e0000198.
25. Humar P, Asaad M, Bengur FB, Nguyen V. ChatGPT is equivalent to first-year plastic surgery residents: evaluation of ChatGPT on the plastic surgery in-service examination. *Aesthet Surg J* 2023;43:NP1085-NP1089.
26. Birkett L, Fowler T, Pullen S. Performance of ChatGPT on a primary FRCA multiple choice question bank. *Br J Anaesth* 2023;131:e34-e35.
27. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence Chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol* 2023;141:589-597.
28. Nakhleh A, Spitzer S, Shehadeh N. ChatGPT's response to the diabetes knowledge questionnaire: implications for diabetes education. *Diabetes Technol Ther* 2023;25:571-573.
29. Hopkins BS, Nguyen VN, Dallas J, Texakalidis P, Yang M, Renn A, et al. ChatGPT versus the neurosurgical written boards: a comparative analysis of artificial intelligence/machine learning performance on neurosurgical board-style questions. *J Neurosurg* 2023;139:904-911.
30. Taesotikul S, Singhan W, Taesotikul T. ChatGPT vs pharmacy students in the pharmacotherapy time-limit test: a comparative study in Thailand. *Curr Pharm Teach Learn* 2024;16:404-410.
31. Li DJ, Kao YC, Tsai SJ, Bai YM, Yeh TC, Chu CS, et al. Comparing the performance of ChatGPT GPT-4, Bard, and Llama-2 in the Taiwan psychiatric licensing examination and in differential diagnosis with multi-center psychiatrists. *Psychiatry Clin Neurosci* 2024;78:347-352.
32. Lin KC, Chen TA, Lin MH, Chen YC, Chen TJ. Integration and assessment of ChatGPT in medical case reporting: a multifaceted approach. *Eur J Investig Health Psychol Educ* 2024;14:888-901.
33. Lee PY, Salim H, Abdullah A, Teo CH. Use of ChatGPT in medical research and scientific writing. *Malays Fam Physician* 2023;18:58.
34. Shieh A, Tran B, He G, Kumar M, Freed JA, Majety P. Assessing ChatGPT 4.0's test performance and clinical diagnostic accuracy on USMLE STEP 2 CK and clinical case reports. *Sci Rep* 2024;14:9330.
35. Fang C, Wu Y, Fu W, Ling J, Wang Y, Liu X. et al. How does ChatGPT-4 preform on non-English national medical licensing examination? An evaluation in Chinese language. *PLOS Digit Health* 2023;2:e0000397.
36. Lee H. The rise of ChatGPT: exploring its potential in medical education. *Anat Sci Educ* 2024;17:926-931.
37. Kianian R, Sun D, Giaconi J. Can ChatGPT aid clinicians in educating patients on the surgical management of glaucoma? *J Glaucoma* 2024;33:94-100.