

수학교육에서 ChatGPT-4o의 기능과 활용 가능성 탐색: 중학교 1학년 수학을 중심으로

박정민¹ · 손홍찬^{2†}

¹태장고등학교 교사, ²전북대학교 교수

Exploring the functionality and utilization potential of ChatGPT-4o in mathematics education: Focusing on 7th-grade mathematics

Park, Jeongmin¹ · Son, Hongchan^{2†}

¹Teacher, Taejang High School ²Professor, Jeonbuk National University

국문요약 본 연구에서는 ChatGPT-4o를 활용한 수학 문제해결 과정에서의 오류와 그 개선과정을 분석하여 이를 토대로 향후 ChatGPT와 같은 생성형 AI의 활용 가능성을 제시하고자 하였다. 이를 위해 중학교 1학년 '수학 1' 교과서의 5개 대단원에 대해 ChatGPT-4o의 단위별 정답률과 단원의 특성에 따른 오류 유형을 파악하고 프롬프트 엔지니어링의 효과를 확인하였다. 2015 개정 교육과정 중학교 1학년 '수학 1' 교과서 10종의 대단원 평가 문제에 대한 ChatGPT-4o의 정답률은 51.76%로 나타났다. 대단원 중 '문자와 식' 단원의 정답률이 가장 높았으며, '통계' 단원에서 가장 낮은 정답률을 보였다. 오류 유형으로는 '수와 연산', '문자와 식' 단위에서는 문제 인식 오류가 가장 많이 나타났으며 '좌표평면과 그래프'와 '통계' 단위에서는 그래프 인식 오류가 가장 많았고 '도형' 단위에서는 그림 인식 오류가 많이 나타났다. 정답률 향상을 위한 프롬프트 엔지니어링을 한 결과, 동일한 문제에 대해서는 정답률이 높아졌으나 유사 문제에 대해서는 정답률이 높아지지 않았다. 그리고 이전 모델들과 ChatGPT-4o의 정답률 변화를 비교 분석한 결과, 정답률이 증가함을 알 수 있었다. 현재 ChatGPT의 기능에는 한계가 있지만 발전 속도가 매우 빠르므로 가까운 미래에 ChatGPT가 학교 수업에 활용될 수 있을 것으로 기대되며 이에 대한 후속 연구가 필요하다.

주요용어 ChatGPT-4o, 정답률, 오류 유형, 프롬프트 엔지니어링

Abstract This study analyzed errors and improvements in mathematical problem-solving using ChatGPT-4o to explore the potential applications of generative AI like ChatGPT in education. To achieve this, the research examined ChatGPT-4o's accuracy across five major units of the 7th-grade 'Mathematics 1' textbook, identified error types based on unit characteristics, and evaluated the effectiveness of prompt engineering. ChatGPT-4o demonstrated an accuracy rate of 51.76% on major unit assessment problems from 10 versions of the 7th-grade 'Mathematics 1' textbook under the 2015 Revised Curriculum. Among the units, the 'Expressions and Equations' unit showed the highest accuracy rate, while the 'Statistics' unit had the lowest. In terms of error types, 'problem recognition errors' were most common in the 'Numbers and Operations' and 'Expressions and Equations' units. 'Graph recognition errors' were predominant in the 'Coordinate Plane and Graphs' and 'Statistics' units, while 'diagram recognition errors' were frequent in the 'Geometry' unit. Prompt engineering improved accuracy for identical problems but had little effect on similar problems. A comparative analysis with earlier models also revealed that ChatGPT-4o showed improved accuracy. Despite current limitations, the rapid development of ChatGPT suggests its potential for use in classroom settings in the near future, emphasizing the need for further research.

Key words ChatGPT-4o, accuracy rate, error types, prompt engineering

Received: August 31, 2024. **Revised:** November 24, 2024. **Accepted:** December 10, 2024.

†Corresponding to (교신저자): 손홍찬(Son, Hongchan), hcson@jbnu.ac.kr

2020 Mathematics Subject Classification : 97P80

I. 서론

오늘날 AI기술은 빠르게 발전하고 있고 이미 사회, 과학, 경제, 의료, 마케팅 등에서 많이 활용되고 있으며 교육 분야에도 많은 관심을 받고 있다(권미선, 2024; 박만구, 2020; 오세준, 2023; Davis, 2024). 교육계에 AI를 도입하여 혁신적인 변화가 생기는 것에 대한 우려도 있지만, 이미 교육 전문가들은 이를 학교 현장에 적용하기 위해 연구를 활발히 하고 있다. 현대 사회에서 정책 입안자, 연구자, 교육자 및 기술 전문가들은 ChatGPT를 활용하여 교육과 학습 효과를 극대화할 방법을 마련하기 위해 노력할 필요가 있다(Baidoo-Anu & Ansah, 2023).

ChatGPT-4o의 이전 모델인 ChatGPT-3.5의 경우, 수학적 연산을 안정적으로 수행하지 못하며 개념적 오류를 범하고 거짓된 내용을 그럴듯하게 인용문으로 생성할 수 있다는 한계를 가지고 있다(Tyson, 2023). 또한 ChatGPT-3.5를 활용하여 우리나라 3개년 국가수준 학업성취도 평가문제와 대학수학능력시험 문제의 정답률을 조사한 권오남 외(2023)의 연구에서도 정답률이 각각 37.1%, 15.97%로 나타나 수학 문제해결에 ChatGPT-3.5를 이용하는 것은 한계가 있음을 보여주고 있다.

그러나 ChatGPT의 기술은 빠르게 발전하고 있으며 교육 현장에서 교사와 학생들의 사용도 빠르게 증가하고 있다. 2023년 3월에 출시된 ChatGPT-4를 이용하여 현대 교육에 미치는 효과를 연구한 Gill et al. (2024)은 ChatGPT가 더욱 정확하고 안정적이며 반응이 빠른 교육 기술을 위한 발판이 될 수 있다고 보았으며, ChatGPT와 같은 AI기반 교육 도구를 수용하고 이해하는 것이 매우 중요해지고 있다고 하였다. OpenAI의 CEO Samuel H. Altman은 2024년 1월에 ChatGPT의 차기 모델 개발을 알리면서, 차기 ChatGPT 모델은 이전 모델보다 안정적으로 개선되어 전반적으로 모든 면에서 뛰어날 것이라고 예측하였다(Bill Gates. Podcast. Episode 6, 2024). 생성형 AI가 이와 같이 매우 빠르게 발전하고 있는 만큼, 우리는 학생들이 생성형 AI를 학습에 올바르게 활용할 수 있도록 미리 준비하고 대응할 필요가 있다(강동훈, 2023). 2024년 5월 13일에 새롭게 출시된 ChatGPT-4o는 이전 모델에 비해 새로운 기능이 추가되었고 정확도와 속도 등 모든 면에서 향상되었다. ChatGPT-4o의 기능이 크게 향상되면서 수학교육에서의 활용 가능성이 커졌으나, 비교적 최근에 출시되어 이에 관한 연구가 아직 부족하다. 따라서 수학교육에서 이를 활용할 수 있는지에 대한 연구가 필요하다.

ChatGPT-4o는 이미지 인식 기능을 갖추고 있어 이전 버전보다 학생들이 문제를 빠르고 정확하게 입력할 수 있다. 또한 이미지 인식뿐만 아니라 이미지 생성 기능도 있어 주어진 조건을 바탕으로 그림, 그래프, 표를 생성할 수 있다. 사용자가 원하는 함수를 그래프로 그리거나 그림, 도형도 생성할 수 있다. 또 ChatGPT-4o는 자연어 처리(NLP: Natural Language Processing) 기술을 기반으로 하여 학생들에게 개별적으로 맞춤형 학습 지원을 제공할 수 있다. 즉 학생들은 수학 문제를 이해하고 풀이하는 과정에서 학생 개개인의 수준과 필요에 맞게 피드백을 받을 수 있다. 즉각적인 피드백은 학생들의 오개념이나 오류에 대한 근본적인 원인을 해결하는 데 도움이 되는데(Gusukuma et al., 2018), ChatGPT-4o 환경에서 학습하는 학생들은 즉각적인 피드백을 받을 수 있다.

중학교 이차방정식과 이차함수에 대해 ChatGPT의 활용을 연구한 오세준(2023)은 ChatGPT가 수학 인공지능 교사의 역할을 하여 학생 개개인에게 맞춤형 교육이 가능하다고 보았다. 중학교 1학년 ‘수학 1’은 논리적·추론적 사고와 문제 해결 능력을 키우는 데 있어서 매우 중요하지만 많은 학생들이 중학교에 입학하면서 수학을 어려워하는 경우가 많다. 본고에서는 중학교 1학년 ‘수학 1’ 과목을 중심으로 ChatGPT-4o의 기능과 오류를 분석하고 프롬프트 엔지니어링을 통한 개선과정을 연구하였다. 이를 위해 중학교 1학년 ‘수학 1’ 교과서에 있는 대단원 평가문제를 대상으로 ChatGPT-4o의 단원별 정답률을 분석하고 오류 유형을 파악하였다. 그리고 낮은 정답률을 높이기 위해 프롬프트 엔지니어링을 적용한 후 정답률 변화를 분석하였다. 그리고 이전 모델인 ChatGPT-3.5, ChatGPT-4와 함께 ChatGPT-4o의 정답률의 변화를 비교 분석하였다.

II. 이론적 배경

1. 교육 분야에서 ChatGPT의 활용

산업혁명 이래로 기계 및 인공지능이 ‘지치지 않는 노동의 경제적 이익’으로 인간 노동자를 대체하여 자동화를 촉진했고 교육계에서도 많은 역할을 하기 시작했다(Diprose & Buist, 2016). ChatGPT는 OpenAI에서 딥 러닝 기반 모델로 개발한 인공지능 Chatbot으로 자연어 처리 및 생성이 가능하여 사용자가 원하는 목적에 따라 인간과 최대한 비슷하게 대화할 수 있다. 과학, 사회, 경제, 의료, 교육, 마케팅 등 다양한 여러 분야에서 주목받고 있고 현재 사용자 수가 매우 빠르게 증가하고 있다. 통계 포털인 Statista에 따르면 페이스북이 100만 명의 사용자를 돌파하는 데 걸린 시간은 10개월이었지만 이에 비해 ChatGPT의 경우는 불과 5일밖에 걸리지 않았다. 그리고 투자은행 UBS에 따르면 2023년 1월에 ChatGPT에 한 번이라도 접속한 사람 수(MAU)는 대략 1억 명이었다(김수민, 백선환, 2023).

앞에서 살펴본 것처럼 AI는 빠른 속도로 진화하고 있고 이를 사용하는 사용자도 급격히 증가하고 있으며 다양한 분야에서 활용되고 있다. 따라서 현대 사회에서 AI의 활용 방법을 익히는 것이 중요해지고 있고, 이는 교육 분야에서도 예외가 아니다.

Edwards & Cheok (2018)은 시간적 제한과 공간적 한계에서 겪는 어려움을 인공지능 교육(AIED)을 통해 극복할 수 있음을 주장하면서 교육 분야에서 인공지능 교사로서 역할을 할 수 있는 주요 기능을 다음과 같이 정리하였다.

가. 효율적인 시간 활용 가능

인공지능 교사는 교사들의 수업 자료 준비, 학생 평가와 같은 반복적인 업무를 줄이는 데에 도움을 줄 수 있다. 오랜 시간 걸리는 단순 업무를 인공지능 교사가 대신한다면 교사는 효율적으로 시간을 활용하여 학생과의 더 활발한 상호작용으로 학습 지도에 전념할 수 있다.

나. 개별화 학습 가능

한 명의 교사가 여러 명의 학생을 지도하여 개인별 맞춤 지도를 하기는 어렵다. 인공지능 교사는 개별 학생들의 학습 데이터를 분석하고 최적의 학습 방법을 제시하거나 개별 학생에게 맞춘 학습 자료를 제공할 수 있다. 따라서 인공지능 교사는 학생의 개별 학습 수준과 필요에 맞게 학습 경험을 만족시킬 수 있다.

다. 상호작용으로 피드백 제공

인공지능 교사는 시간적 제한과 공간적 한계를 극복하여 언제든지 어디서든지 학생과 상호작용하며 즉각적인 피드백을 제공할 수 있다. 따라서 학생들의 오류나 오개념을 빠르게 수정하도록 도와서 학생들의 학습 효과를 높일 수 있다. 인공지능 교사는 학생들의 응답을 분석하고 실시간으로 피드백하여 학생들의 이해도를 높이고 학습의 연속성을 유지하도록 도와준다.

신보미 외(2024)도 생성형 AI가 지닌 잠재성에 주목해야 하며 디지털 기기를 활용하는 자기주도적 수학 학습 능력을 배양해야 할 필요가 있음을 주장한 바 있다. 이를 위해 학생들이 인공지능을 사용하면서 상황에 맞는 인공지능 활용법을 선택하고 사용할 수 있도록 인공지능 리터러시 교육이 필요하다고 보았다.

2. 수학교육에서 활용 가능한 ChatGPT-4o의 기능

ChatGPT-4o의 GPT(Generative Pre-trained Transformer)는 사전 훈련된 생성 변환기라는 뜻이고 o는 ‘모든’, ‘전체의’라는 뜻을 전달하는 접두사인 ‘Omni’의 약자이다. Omni의 약자를 사용한 것은 ChatGPT-4o가 이미지, 음성, 파일 등의 여러 형태를 인식하고 이를 이해하고 처리할 수 있는 능력이 있음을 강조하기 위한 것이다 (OpenAI ChatGPT 공식 홈페이지1, 2024). 이러한 기능은 수학교육에 획기적인 변화를 가져올 수 있는 잠재력을 지닌다. 기존의 ChatGPT-3.5는 수학 문제를 입력할 때 수식을 입력하는 과정이 번거롭고, 표나 그래프, 그림을 인식하는 기능이 부족하여 문제를 제대로 입력하지 못하는 경우가 있었다. ChatGPT-3.5의 활용을 연구한 권오남 외(2023)와 강윤지(2024)의 연구와 ChatGPT-4의 활용을 연구한 오세준(2023)의 연구에서는 표, 그래프, 그림을 포함하는 문제는 제외하여 연구가 진행되었다. 이현수와 박영용(2024)연구에서도 중학교 2학년 단위 평가문제를 분석하였으나 그림을 포함하는 ‘기하’단원을 제외하고 표, 그래프, 그림 등이 포함되어 있는 문제도 제외하였다. 그러나 ChatGPT-4o에는 ChatGPT-3.5와 ChatGPT-4에서 입력하지 못했던 그림, 그래프, 표를 포함하는 문제를 쉽게 입력할 수 있다. 이기마와 김희정(2024)의 연구에서는 ChatGPT-4o를 활용해서 수학 문제를 이미지로 입력하여 정답률, 오류율, 오류의 특징 등을 확인하였다. OpenAI ChatGPT 공식 홈페이지3(2024)을 참고하여 수학교육에서 사용할 수 있는 ChatGPT-4o의 기능을 크게 4가지로 분류하면 다음과 같다.

가. 비전 기능 및 이미지 생성 모드

사용자가 이미지를 입력창에 입력하면 ChatGPT-4o는 이미지를 인식하고 이와 관련된 답변을 생성할 수 있다. 책이나 문서를 사진으로 찍었을 때 그 이미지 속의 한글을 텍스트 파일로 변환할 수 있다. 또 글자, 그림, 그래프, 표 등이 포함된 문제를 ChatGPT-4o에 이미지로 입력하면 이를 인식해서 분석할 수 있다. 또한 필기체 글씨를 인식하고 이를 디지털 텍스트로 변환할 수 있다. 이 기능을 수학교육에 적용하면, ChatGPT-4o가 학생들이 손으로 쓴 문제와 문제풀이 과정을 인식하고 분석할 수 있으며 자동으로 해석하여 오류를 찾아 피드백을 제공할 수 있다. 이를 통해 학생들은 자신의 풀이 과정을 디지털화해서 검토할 수 있고 풀이의 보충 설명도 받을 수 있다. 교사의 입장에서 ChatGPT-4o의 비전 기능 및 이미지 생성 모드 기능을 활용하면 학생들의 과제를 효율적으로 평가하고 개별적 맞춤 피드백을 제공할 수 있다.

나. 웹 검색을 통한 정보제공

ChatGPT-3.5는 모델 자체에서 사전 학습된 정보를 제공했으나 ChatGPT-4o는 웹검색으로 정보를 제공하는 것이 가능하다. 모델 자체에서 답변을 생성하는 대신 웹검색을 통해 더 정확하고 신뢰할 수 있는 답변을 제공하며, 이로 인해 hallucination(오류, 환각 문제) 현상을 줄일 수 있다. 또한 리서치 기능도 가능하여 ChatGPT-4o는 웹을 검색하고 가장 관련성 높은 자료를 조합하여 관련된 정보를 제공할 수 있다. 이 기능을 통해 사용자는 출처와 함께 신뢰할 수 있는 종합적인 정보와 요약본을 얻을 수 있다. 예를 들어 ChatGPT-4o에게 ‘아르키메데스의 업적과 그가 발견한 수학 이론’을 웹검색을 통해 알려달라고 요구하면 웹 검색 출처와 함께 종합적인 정보를 받을 수 있다.

다. 데이터 분석

사용자는 이미지 파일, CSV(Comma-Separated Values)파일, 엑셀 파일을 업로드하여 수학 문제를 해결할 수 있다. 예를 들어 사용자가 CSV파일을 입력하면 ChatGPT-4o는 파일 안에 담긴 내용을 분석하여 막대그래프로도

정리할 수 있으며 그 데이터를 바탕으로 다양한 질문에도 답변할 수 있다. 또한 유료 버전의 경우 PDF, 텍스트 파일, 스프레드시트, 프레젠테이션과 같은 문서를 업로드할 수 있으며 파일당 최대 512MB까지 업로드 가능하며 파일에 관련된 질문의 답과 요약본을 얻을 수 있다. 특히 이미지는 20MB까지 업로드할 수 있다. 텍스트를 분석할 때 텍스트를 단어, 구, 문장 등으로 나누는 단위를 토큰이라고 하는데 ChatGPT는 문서나 텍스트 파일을 최대 200만 토큰까지 처리할 수 있다.

라. 보이스 모드 플러스 기능

ChatGPT-3.5 모델에서도 음성 기능이 있었지만 대화 중에 자주 오류가 나고 반응 대기시간이 길어 불편했다. 그러나 ChatGPT-4o는 이러한 음성기능 반응속도를 개선하여 이제 사용자는 실시간에 가까운 속도로 대화할 수 있다. 이 기능은 실시간 음성 지원이 필요한 상황에서 매우 유용하며, 특히 수학교육에서는 음성으로 수학문제를 인식하거나 사용자의 문제를 풀이하는 과정을 음성으로 피드백하는 것이 가능해졌다.

3. 수학 수업에서 ChatGPT활용을 위한 프롬프트 유형

교육 분야에서 AI를 활용하기 위해서는 AI의 성능을 파악하는 분석과정이 필요하다. 벤치마크는 기업이 다른 기업의 우수사례를 배우고 참고하는 경영기법을 의미하는데, 컴퓨터 분야에서는 주로 전자기기의 성능을 분석하고 평가하기 위해 작업을 실행하는 것을 의미한다. 벤치마크는 경우에 따라 성능을 측정하기 위한 평가 문제로 언급하기도 한다(Davis, 2024). 수학교육과 관련된 벤치마크는 초등학교의 기초적인 문제부터 국제 수학 올림피아드나 대학교 고급 수학 문제까지 다양한 난이도의 문제들이 있다. 그러나 현재까지 벤치마크로 알려진 문제들이 인간을 대상으로 개발된 문제이므로 ChatGPT에 적용하고 얻은 결과를 인간에게서 얻은 결과와 같은 기준에서 평가하기는 어렵다(Davis, 2024), OpenAI 공식 홈페이지(2024)에는 다양한 벤치마크에서 평가한 결과가 공개되어 있고 이를 바탕으로 ChatGPT를 개선하기 위한 노력을 계속하고 있다. ChatGPT의 개선점이 매우 빠르게 수정되고 있어서 OpenAI홈페이지에 공개된 결과들이 현재 모델의 정확한 성능을 표현하지 못하지만, 앞으로의 발전 가능성을 제시하기 위한 자료로 사용될 수 있다(Open ChatGPT 공식 홈페이지2, 2024).

ChatGPT의 벤치 마크를 위해서는 명령어를 입력해야 하는데, 이 명령어는 아래의 4가지 프롬프트 유형으로 구분된다(Heston & Khun, 2023).

- 가. 제로샷 프롬프트(Zero-Shot Prompts): 특별히 훈련받지 않은 데이터에 대해 질문하는 것을 의미한다. ‘Zero’는 특정 작업이나 구체적인 훈련을 거의 받지 않았거나 전혀 받지 않았음을 의미한다. 아무 훈련 없이 주어진 문제에 대한 해답을 요구할 때 사용하는 명령이 제로샷 프롬프트에 해당된다.
- 나. 퓨샷 프롬프트(Few-Shot Prompts): 질문에 답하거나 요청된 작업을 수행하는데 도움이 되는 예시를 포함하는 프롬프트를 의미한다. 특별히 훈련되지 않았다는 점에서는 제로샷 프롬프트와 유사하지만 이 퓨샷 프롬프트에는 요청을 이해하는데 도움이 되는 예가 포함되어 있다. 예를 들어 ‘나에게 문제를 하나 알려줘’는 제로샷 프롬프트이지만 ‘중학교 1학년 함수에 대한 문제를 알려줘’는 퓨샷 프롬프트에 해당한다.
- 다. 반복적 프롬프트(Iterative Prompts): 프롬프트를 점진적으로 개선하기 위해 반복적으로 수정하도록 하는 프롬프트를 의미한다. 사용자는 초기 프롬프트를 생성하여 ChatGPT에게 요청한 후, 필요한 추가 정보를 제공하거나 피드백을 바탕으로 프롬프트를 개선할 수 있다. 이 과정을 반복하면서 프롬프트는 명확하고 구체적으로 개선되고, ChatGPT가 생성하는 응답은 점점 더 정확하고 유용해진다.
- 라. 구조화된 프롬프트(Structured Prompts): 일관되게 좋은 결과를 얻기 위해 프롬프트 핵심 구성요소를 구체적으로 제공하는 프롬프트를 의미한다. 질문하는 사람, 구체적인 질문 내용, 역할 부여, 작업 방식 등이

이 프롬프트에 해당한다. 예를 들어 ‘저는 중학교 1학년 학생으로 수학을 공부하고 있습니다.’라는 질문자에 대한 정보, ‘중학교 1학년 수학 문제를 이해하는 데 도움이 필요합니다.’와 같은 문제에 대한 구체적 내용, ‘중학교 1학년 수학 교사 역할을 해주세요.’와 같은 역할 부여, ‘그래프가 나오면 우선 그래프를 분석하고 관련된 수학적 개념을 설명해주세요.’와 같은 출력 형식 등이 구조화된 프롬프트이다. 이 프롬프트들은 ChatGPT가 보다 정확한 답을 하는 데 도움이 된다.

4. 프롬프트 작성의 중요성

ChatGPT와 같은 생성형 AI를 사용할 때, 프롬프트의 작성은 매우 중요하다. 프롬프트를 어떻게 작성하는지에 따라 사용자가 원하는 답변을 정확하고 구체적으로 얻을 수 있다. 프롬프트 작성의 중요성은 여러 교과에서 공통적으로 강조하고 있으며 수업의 보조도구로 ChatGPT를 효과적으로 사용하기 위해 프롬프트가 연구되었다(최숙기, 박종임, 2023, 이선, 2024, 오세준, 2023, 진준화, 2023, 정현우, 박건우, 2024). ChatGPT에 질문을 입력할 때의 핵심은 요구하는 바를 명확하게 프롬프트로 표현하는 것이다(정한민, 박정훈, 2023; 김동성, 2023). ChatGPT에 프롬프트를 어떻게 입력하는가에 따라서 답변하는 언어 모델의 성능에는 많은 차이가 있다(Reynolds & McDonell, 2021). 사용자가 별도의 요청 없이 수학 문제만 ChatGPT에 입력할 경우, ChatGPT는 그 문제와 관련된 개념을 설명하는 데에 초점을 맞추고 풀이 과정이나 정답을 생성하지 않을 수 있다. 또는 ChatGPT가 사용자의 연령과 국적을 고려하지 않고 영어로 답변하거나 교육과정에서 벗어나는 풀이 과정을 생성하기도 한다.

명확한 프롬프트 작성의 중요성은 다음 몇 가지로 정리할 수 있다. 첫째, 사용자의 의도를 정확하게 파악하고 이에 맞는 응답을 생성하도록 도와준다(Giray, 2023). 즉, 명확하게 작성된 프롬프트는 사용자의 의도에 맞게 정확한 결과를 생성하도록 도와준다. 예를 들어 ‘수학 문제 풀이를 한국어로 설명하세요’라고 입력하면 한국어로 설명한다. 문제를 입력하기 전에 따로 언급하지 않으면 보통 영어로 답변을 생성한다. 둘째, 효율적인 문제 해결을 하도록 도와준다(Nazari & Saadi, 2024). ChatGPT는 복잡한 문제를 체계적이고 논리적으로 정리하는 것을 도와준다. 예를 들어 ‘수학 문제를 입력할 때 Polya의 문제해결 4단계로 답변을 생성하세요’라고 입력하면 ‘문제 이해, 계획, 실행, 반성’의 단계로 풀이 과정과 답변을 생성할 수 있다. 셋째, 다양한 분야에서 활용할 수 있도록 도와준다. 다양한 형식의 프롬프트는 다양한 상황에 적용할 수 있게 도와준다. 예를 들어 수학 문제를 입력하고 서술형 문제에 대한 답변을 요구할 수도 있고 입력된 수학 문제와 관련된 수학 글쓰기를 요구할 수 있다.

다양한 상황에서 사용자의 의도에 맞게 효율적으로 문제 해결을 하기 위해서는 명확하고 구체적인 프롬프트의 작성이 중요하다. ChatGPT는 수정 기능을 통해 자신의 실수를 적극적으로 인정하고 사용자의 피드백에 따라 답변을 최적화하여 사용자의 의도에 맞게 답변을 수정할 수 있으며, 답변의 정확도를 높이거나 낮출 수 있다(Zhou et al., 2023)

위에서 언급한 다양한 프롬프트들을 사용하여 출력 또는 결과를 원하는 방향으로 생성하도록 프롬프트를 설계하고 최적화하는 과정이 ‘프롬프트 엔지니어링’이다. 즉, 언어 모델이 입력된 텍스트인 프롬프트를 기반으로 최대한 정확한 최종 결과물을 얻기 위해 반복적인 테스트와 피드백으로 프롬프트를 지속적으로 개선하는 작업을 말한다.

III. 연구 방법

1. 연구 대상

본 연구에서는 2015 개정 교육과정 중학교 1학년 ‘수학 1’ 과목의 교과서 10종의 대단원 평가 문제를 대상으로 하였다. 중학교에 들어서면서 초등학교에서 배운 수학적 개념을 바탕으로 확장한 심화 과정을 배우게 된다. 예를 들어 초등학교에서는 구체적인 수를 다뤘다면 중학교에서는 방정식, 함수, 기하, 등 더 복잡하고 추상적이고 논리적인 문제까지 해결해야 한다. 수학적 사고력과 문제 해결 능력을 기르는 데 매우 중요한 부분이다. 중학교 1학년 ‘수학 1’의 대단원 평가 문제를 대상으로 하되, 교과서에 따라 대단원 평가 문제가 없는 경우 중단원 평가 문제로 대체하였고, 총 917문제를 선정하였다. 10종의 교과서는 익명으로 A, B, C, D, E, F, G, H, I, J로 표시하였다. 10종의 교과서가 같은 대단원명을 가지는 것이 아니어서 내용 요소를 기준으로 하여 대부분의 교과서가 채택한 대단원명을 기준으로 수와 연산, 문자와 식, 좌표평면과 그래프, 도형, 통계 단원으로 구분하였다. 교과서 대단원 평가의 모든 문제를 다루면서 ChatGPT-4o가 그림, 그래프, 표 등을 포함한 문제를 잘 인식하고 이해하는지, 그리고 풀이 과정에서 그래프, 그림, 표로 잘 표현하는지를 확인하였다. 연구 대상으로 선정한 문제는 <표 1>과 같다.

<표 1> 중학교 1학년 ‘수학 1’ 교과서 10종 단위 종합평가의 분석 문제 수(단위: 개)

	A	B	C	D	E	F	G	H	I	J	계
수와 연산	16	27	11	18	18	22	13	14	16	37	192
문자와 식	16	13	10	18	16	19	14	15	15	18	154
좌표평면과 그래프	16	12	8	13	14	19	11	6	13	14	126
도형	47	40	16	32	31	41	25	28	30	46	336
통계	10	10	9	12	14	13	13	5	10	13	109
합계	105	102	54	93	93	114	76	68	84	128	917

2. 연구 방법

가. ChatGPT-4o의 단위별 정답률 확인

2015 개정 교육과정 중학교 1학년 ‘수학 1’ 과목의 교과서 10종에서 5개의 대단원 평가 문제를 대상으로 ChatGPT-4o에게 풀이 과정과 답을 요구하고 단위별로 그 정답률을 분석하였다. ChatGPT-4o의 기능이 이전 버전 보다 개선되어 대단원 평가에 있는 모든 문제를 대상으로 정답률 분석이 가능해졌다. ChatGPT-4o에 수학 문제를 입력하고 풀이 과정과 답을 요구하여 5개의 대단원 중 정답률이 가장 높은 단위와 낮은 단위를 확인하고 그 특징을 분석하였다. ChatGPT-4o의 단위별 정답률 확인을 위해서 문제를 입력하기 전에 구조화된 프롬프트를 적용하여 “중학교 1학년 수준의 학생에게 설명하는 것처럼 문제 풀이 과정과 정답을 한국어로 설명하기”를 요구하였다. ChatGPT-4o는 같은 문제에 대하여 다른 답변을 할 수 있기 때문에 정확한 정답률을 확인하기 위해서 다른 연구자들처럼(권오남 외, 2023; 강윤지, 2024) 같은 문제를 여러 차례 입력하여 그 답변을 분석하는 방법을 사용하였다. 오세준(2023)의 연구에서는 GPT-4모델에서 올프람 플러그인(PLUG-IN) 기능을 사용하였는데 ChatGPT-4o에 일부 문제를 입력한 결과 플러그인의 여부로 인한 답변의 차이가 없어 올프람 플러그인 기능을 사용하지 않고 문제를 입력하였다. 이 연구에서는 917문제에 대한 답변을 3회씩 요구하여 총 2751회의 답변을 받아 교과서에 제시된 풀이 과정과 답을 비교하였다. 정답률의 계산은 2751회 중에 맞춘 경우를 계산하였고, 전체 정답률과 단위별 정답률을 비교 분석했다.

나. ChatGPT-4o의 단원별 오류 분석

ChatGPT-4o에 수학 문제를 이미지로 입력한 후 생성된 풀이 과정과 답에서 나타난 오류 유형을 단원별로 분류하였다. 오류 유형 중 정답을 맞힌 경우는 두 가지로 분류하였다. ChatGPT-4o의 답변이 중학교 1학년 교육과정을 벗어난 유형과 틀린 풀이 과정이지만 답이 맞는 경우이다. 정답이 틀린 경우의 오류 유형은 총 11가지로 분류하였다. 문제, 그림, 표, 좌표평면, 수직선, 그래프를 잘못 인식하는 오류, 계산 오류, 잘못된 수학적 개념 사용, 잘못된 수학적 기호 사용, 잘못된 식 세우기, 맞는 풀이 과정이지만 답이 틀린 경우이다.

다. ChatGPT-4o의 정답률 변화 확인

1) 프롬프트 엔지니어링을 통한 ChatGPT-4o의 정답률 변화 확인

중학교 1학년 ‘수학 1’ 대단원 평가 917문제를 3회씩 입력하여 풀이 과정과 답을 얻을 때 오류가 나타났지만 별도의 피드백을 제시하지 않았다. 이 과정에서 ChatGPT-4o가 두 가지 풀이를 생성하며 선호하는 답변을 선택하고 한 경우에도 피드백 없이 다음 문제 풀이를 진행하였다. 4회 차에 문제를 입력할 때는 정답률 변화를 확인하기 위해 퓨샷 프롬프트와 반복적 프롬프트를 활용한 프롬프트 엔지니어링 과정을 진행하였다. 정답률을 높이기 위한 프롬프트 엔지니어링 과정은 동일한 문제에 대한 효과와 유사한 문제에 대한 효과로 두 가지 측면에서 조사되었다. 첫 번째의 경우에는 5종의 교과서(B, D, F, H, J)의 각각의 문제에 대해서 입력하고 오류가 나타날 경우 ChatGPT-4o에게 수정을 즉각적으로 요구하였다. 수정 이후 동일한 문제를 새롭게 입력했을 때를 4회차라고 하였으며 이때의 정답률 변화를 확인하였다. 두 번째의 경우에는 5종의 교과서(B, D, F, H, J)에서 수정을 요구한 이후 초기화하지 않고 앞의 문제와는 다른 나머지 5종의 교과서(A, C, E, G, I)의 문제, 즉 유사한 문제를 각각 4회 차에 입력하였다. 이 방법으로 특정한 유형의 문제들을 훈련한 후 이와 유사한 문제의 정답률 변화를 확인할 수 있다.

2) ChatGPT-3.5와 ChatGPT-4 그리고 ChatGPT-4o의 정답률 비교

Plevris et al. (2023)의 연구에서는 수학 문제 30문제를 대상으로 ChatGPT-3.5, ChatGPT-4, Google Bard에 입력하여 정답률을 비교하였다. 세트 A는 온라인에서 찾을 수 없는 독창적인 수학 문제 15문제이고, 세트 B는 일반적으로 풀이와 정답을 온라인에서 찾을 수 있는 수학 문제 15문제로 구성되어 있다. Plevris et al. (2023)의 연구에서 사용된 문제를 ChatGPT-4o에 입력하여 ChatGPT-3.5, ChatGPT-4, ChatGPT-4o의 정답률을 비교했다. 이 방법으로 다양한 버전의 ChatGPT의 정답률 변화를 확인하고, 향후 수학교육에서 ChatGPT-4o의 활용 가능성을 평가할 수 있다.

IV. 연구 결과

1. ChatGPT-4o의 정답률(단원별 비교)

중학교 1학년 ‘수학 1’ 과목의 대단원 평가 문제에 대한 정답률은 <표 2>와 같다. 총 917문제를 3회씩 입력하여 총 2751회 입력한 결과, 전체 정답률은 51.76%로 나타났다. ‘문자와 식’, ‘수와 연산’ 단원에서 높은 정답률을 보였고 ‘통계’ 단원에서 가장 낮은 정답률을 보였다. ‘수와 연산’ 단원과 ‘문자와 식’ 단원에서 다른 교과서보다 낮은 정답률을 보인 교과서가 있었는데, 그 이유는 ChatGPT가 수직선이나 그림 등이 포함된 문제를 풀이하는 과정에서 잘못 인식했기 때문이다. 또한 ChatGPT-4o는 ‘보기에서 옳은 것을 고르면?’인 문구가 들어있는 문제를 잘 해결하지 못하는데 이와 같은 문제가 있는 교과서에서 오류가 많이 나타났다. ‘통계’ 단원에서 낮은 정답률을 보인 이유는 주로 줄기 잎 그림, 도수분포다각형, 히스토그램, 표 등을 제대로 인식하지 못한 데에서 그 원인을 찾을 수 있었다.

<표 2> 중학교 1학년 수학 1 교과서 10종 대단원 평가 문제 정답률(단위: %)

	A	B	C	D	E	F	G	H	I	J	계
수와 연산	87.50	86.42	51.52	66.67	70.37	57.58	58.97	69.05	79.17	63.96	69.79
문자와 식	68.75	69.23	80.00	75.93	68.75	36.84	66.67	86.67	93.33	85.19	72.29
좌표평면과 그래프	56.25	80.56	58.33	58.97	61.90	71.93	60.61	38.89	82.05	50.00	63.49
도형	39.72	40.83	29.17	31.25	39.78	31.71	36.00	33.33	31.11	45.65	36.81
통계	20.00	13.33	22.22	27.78	35.71	7.69	33.33	20.00	13.33	33.33	23.55
합계	52.06	58.50	46.30	50.18	53.41	41.52	48.68	51.96	57.14	55.73	51.76

2. 오류 유형

가. 오류 유형 분류

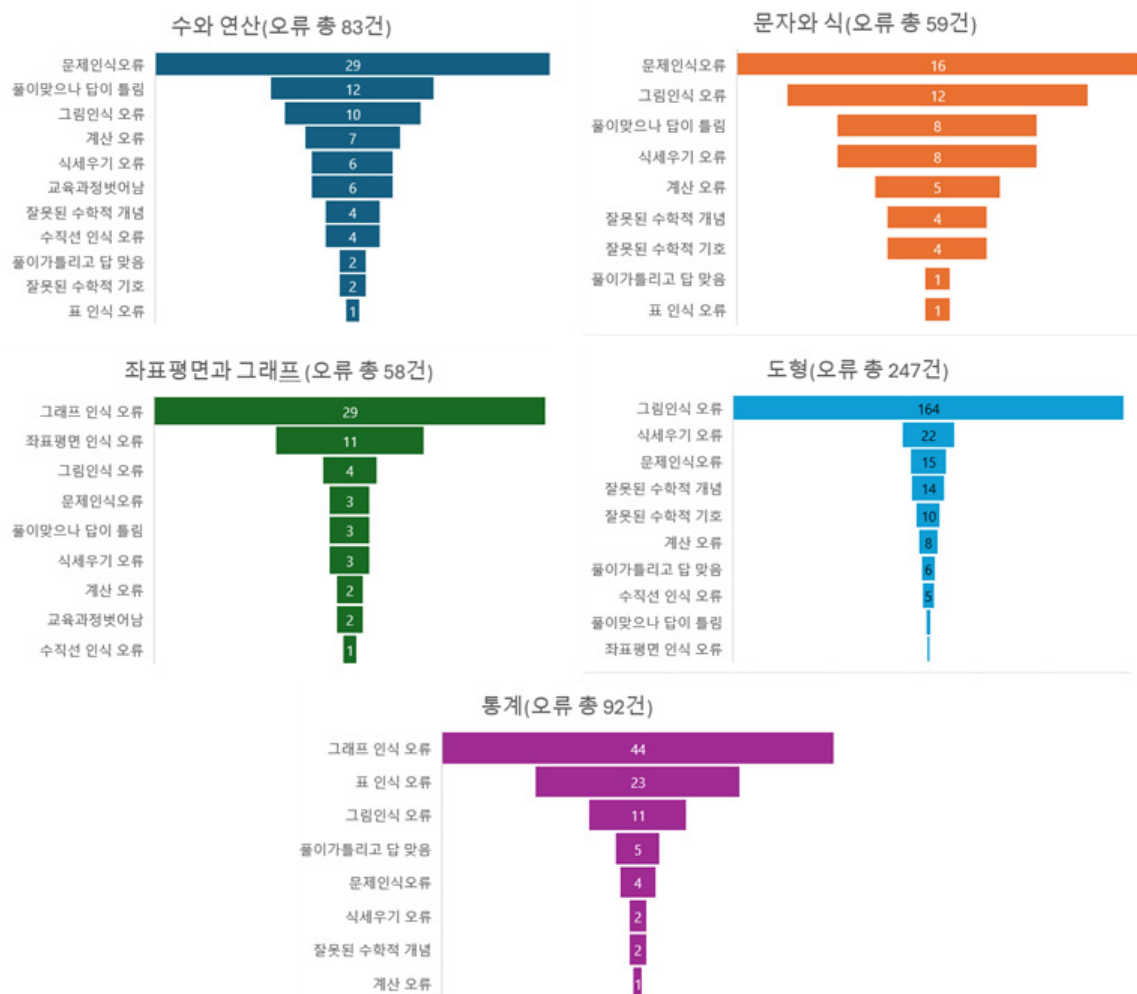
오류 유형은 답이 틀린 경우 11가지, 답이 맞은 경우 2가지로 분류하였다. 오류 유형을 범주화하는 데는 선행연구(권오남 외 2023; 강윤지, 2024)를 참고하였다. 또한 이제까지의 선행연구에서는 그림이 포함된 문제를 입력할 수 없었는데, 본 연구에서는 예전에 입력하지 못했던 문제를 입력했을 때 나타나는 오류 유형을 그림, 좌표평면, 그래프, 수직선, 표로 구분하였다. 오류 유형에 관련된 예시를 간략하게 <표 3>에 정리하였다. 답이 틀린 경우에는 문제 인식 오류, 그림 인식 오류, 좌표평면 인식 오류, 그래프 인식 오류, 수직선 인식 오류, 표 인식 오류와 같이 ChatGPT 모델 자체의 문제로 나타내는 오류 유형 6가지와 풀이는 맞으나 답이 틀린 경우, 계산 오류, 식 세우기 오류, 잘못된 수학적 개념 사용, 잘못된 수학적 기호 사용처럼 학생들에게 나타날 수 있는 오류 유형 5가지 총 11가지로 분류하였다.

<표 3> 중학교1 ‘수학 1’ 단원 평가 문제에서 나타난 오류 유형 분류

	관련 오류 유형	관련 오류 예시
답이 틀린 경우	계산 오류	단순 산술 오류
	식 세우기 오류	문제의 상황을 식으로 세우는 과정에서 오류
	잘못된 수학적 개념 사용	소인수를 소수로 인식한 경우 할선, 현의 의미 잘못 사용한 경우 넓음과 합동의 수학적 개념을 잘못 사용한 경우
	잘못된 수학적 기호 사용	반직선, 직선의 기호를 잘못 인식한 경우 호의 길이를 기호로 인식하지 못한 경우 $0.1 \times a = 0.a$ 과 같은 표현을 사용하는 경우
	문제 인식 오류	옳은 것을 고르는 문제에서 옳지 않은 보기를 선택하는 경우 문제의 숫자를 잘못 인식한 경우
	그림 인식 오류	그림 속 단서를 이용하지 못하는 경우 복잡한 도형을 인식하지 못하는 경우
	좌표평면 인식 오류	좌표평면의 점의 좌표를 읽지 못하는 경우 잘못 구한 좌표를 이용해서 다음 문제를 풀이하는 경우
	그래프 인식 오류	여러 그래프 중 특정 그래프를 인식하지 못하는 경우 그래프에 나타난 점의 좌표를 읽지 못하는 경우 잘못된 그래프 분석
	수직선 인식 오류	수직선의 점의 위치를 인식하지 못하는 경우 잘못 구한 점으로 풀이를 시도하는 경우
	표 인식 오류	표를 잘못 인식하여 잘못된 수를 이용하여 풀이하는 경우
	풀이는 맞으나 답이 틀린 경우	보기의 내용이 옳다고 하고 답에서 옳지 않다고 하는 경우
답이 맞는 경우	교육과정에서 벗어난 경우	교육과정에서 벗어난 삼각형의 넓이 공식을 이용한 경우
	풀이는 틀리고 답이 맞는 경우	줄기임 그림에서 내용을 잘못 파악하고 답만 맞는 경우

나. 단위별 오류 유형 비교

중학교 1학년 ‘수학 1’ 대단원 평가 문제에서 나타난 오류 유형은 [그림 1]과 같다. 917문제 중 한 문제를 3회 입력할 때, 한 문제에서 같은 오류 유형이 중복되면 1건으로 계산하고 한 문제에서 2가지 오류 유형이 나타나면 2건으로 산출하였다. ‘수와 연산’, ‘문자와 식’ 단위에서는 문제 인식 오류가 가장 많이 나타났다. ‘보기 중 옳은 것을 고르는 문제’에서 틀린 보기를 답으로 선택하기도 하고 바르게 설명해 놓고 답은 옳지 않다고 하는 경우가 있었다. “다음 중 대소관계가 옳은 것은?”의 문제를 ChatGPT-4o를 통해 이미지로 입력해서 번역을 요구하니 “Which of the following inequalities is incorrect?”로 번역하여 문제 인식의 오류를 나타냈다. 또한 ChatGPT-4o에 문제를 이미지로 입력했을 때 제공과 같이 작게 적힌 숫자를 잘못 인식하는 경우가 있었다. 텍스트로 입력하면 오류가 나타나지 않고 이미지로 입력할 때 잘못된 숫자로 인식하는 오류를 보였다. ‘문자와 식’, ‘도형’ 단위에서 그림 인식 오류가 많이 나타났다. ChatGPT-4o가 그림을 인식하는 기능이 추가되었다고 하지만 아주 간단한 그림만 인식할 수 있었다. ‘좌표평면과 그래프’ 단위에서는 그래프 인식의 오류가 가장 많았는데 그래프를 잘못 분석하거나 그래프 위에 나타나 있는 점의 좌표를 읽어내지 못한 경우가 있었다. ‘통계’ 단위에서는 그래프 인식 오류와 표 인식 오류가 가장 많았다. 일부 수학 문제에서는 문제해결 과정에서 그래프나 표를 정확하게 생성하고 문제의 조건을 잘 분석하지만 많은 경우에 표, 그래프를 제대로 분석하지 못한 경우가 많았다.



[그림 1] 단위별 오류 유형 빈도수

다. 단원별 오류 유형 예시

1) ‘수와 연산’ 단원에서의 오류 예시

‘수와 연산’에서 가장 많이 나타난 세 가지 오류 유형은 문제 인식 오류(29건), 풀이가 맞으나 답이 틀린 경우(12건), 그림 인식 오류(10건)이었다. [그림 2]는 ‘수와 연산’ 단원에서 문제 인식 오류를 보여주는 예시로, ChatGPT에 문제를 입력했을 때 문제는 오른쪽 위에 제시되고 풀이 과정과 답은 왼쪽 아래에 생성된다. 보기에서 ‘옳은 것을 고르면?’의 문구가 포함된 문제에서 ①~⑤ 보기의 설명은 바르게 하였으나 옳지 않은 것을 정답으로 선택했다. 즉, 옳지 않은 것을 고르는 문제로 인식한 것이다. 이미지 안에 있는 한글을 인식하는 기능이 있지만 문제를 인식하는 과정에서 오류가 나타났다. 이와 같이 옳은 것을 고르는 문제에서 옳지 않은 것을 골라 틀린 경우가 많았다.

1
 다음 중 옳은 것은? 99

① 가장 작은 소수는 1이다.
 ② 모든 짝수는 합성수이다.
 ③ 서로 다른 두 소수는 서로소이다.
 ④ 소수의 제곱인 수는 소수이다.
 ⑤ 10보다 작은 소수는 모두 5개이다.

주어진 문제에서 옳지 않은 것을 찾는 문제입니다. 각 선택지를 하나씩 살펴보겠습니다.

1. 가장 작은 소수는 1이다.

- 풀이: 소수는 1과 자기 자신만을 약수로 가지는 1보다 큰 자연수입니다. 따라서 가장 작은 소수는 2입니다. 선택지 1은 옳지 않습니다.

2. 모든 짝수는 합성수이다.

- 풀이: 짝수는 2로 나누어 떨어지는 수입니다. 모든 짝수가 합성수라는 것은 틀렸습니다. 예를 들어, 2는 짝수이지만 소수입니다. 따라서 이 문장은 틀렸습니다.

3. 서로 다른 두 소수는 서로소이다.

- 풀이: 서로 다른 두 소수는 1 외에는 공통 약수가 없으므로 항상 서로소입니다. 따라서 이 문장은 맞습니다.

4. 소수의 제곱인 수는 소수이다.

- 풀이: 소수의 제곱은 1과 자기 자신뿐만 아니라 다른 약수(예: $2^2=4$ 일 경우 2와 2)가 생기므로 합성수가 됩니다. 따라서 이 문장은 틀렸습니다.

5. 10보다 작은 소수는 모두 5개이다.

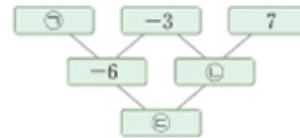
- 풀이: 10보다 작은 소수는 2, 3, 5, 7로 4개입니다. 따라서 이 문장은 틀렸습니다.

옳지 않은 것은 선택지 1, 2, 4, 5입니다. 이 중에서 선택지 1이 가장 먼저 나오므로, 답은 **1번**입니다.

[그림 2] G교과서 ‘수와 연산’ I-1 문제 인식 오류 예시

[그림 3]은 ‘수와 연산’ 단원에서 그림 인식 오류의 예시를 보여준다. 그림에서 아래쪽의 수가 위쪽에 연결된 두 수의 합일 때 빈칸에 들어가는 수를 구하는 문제이다. 그러나 그림 모양을 반대 모양의 트리로 인식하고 풀이를 시도하여 잘못된 답을 생성하였다. 비슷한 예로 가로 3칸, 세로 3칸으로 총 9개의 칸이 있는 그림에서 가로, 세로, 대각선의 합이 같은 경우 빈칸의 수를 구하는 문제에서도 마찬가지로 그림을 제대로 인식하지 못하였다. ChatGPT-4o는 그림 모양이 복잡하지 않은 문제에서도 이를 제대로 인식하지 못하는 경우가 많았다.

13 다음 그림에서 아래쪽의 수가 위쪽에 연결된 두 수의 합일 때, ㉠, ㉡, ㉢에 들어갈 수들의 곱을 구하십시오.



주어진 트리에서 두 수의 합을 구하고 곱하는 문제를 해결하겠습니다.

주어진 트리는 다음과 같습니다:



- 아래쪽의 수가 위쪽에 연결된 두 수의 합일 때, ㉢에 들어갈 수들의 곱을 구합니다.

트리 구조를 분석해 보겠습니다:

[그림 3] B교과서 ‘수와 연산’ II-13 그림 인식 오류

2) ‘문자와 식’ 단원에서의 오류 예시

‘문자와 식’ 단원에서 가장 많이 나타난 네 가지 오류 유형은 문제 인식 오류(16건), 그림 인식 오류(12건), 잘못된 식 세우기 오류(8건), 풀이는 맞으나 답이 틀린 경우(8건)이었다. [그림 4]는 ‘문자와 식’ 단원에서 문제 인식 오류의 예시를 보여준다. 어떤 다항식에서 더해야 할 것을 잘못 빼서 계산한 값을 보고 바르게 계산한 식을 구하는 문제이다. 이 문제에서 ChatGPT-4o는 어떤 다항식까지 구하고 답이라고 생성하면서 바르게 계산한 식까지 구하지 못했다. 이와 같은 유형의 문제가 다른 교과서에서도 있었는데 ‘어떤 수’, ‘어떤 다항식’까지 구하고 ‘바르게 계산한 값’, ‘바르게 계산한 식’까지 구하지 않았다. ChatGPT-4o는 문제가 요구하는 질문을 잘못 이해하여 중간 과정까지만 계산하고 다음 단계까지는 시도하지 않은 경우가 많았다.

[그림 5]는 ‘문자와 식’ 단원에서 잘못된 수학적 기호 사용 오류와 문제 인식 오류의 예시를 보여준다. 바르게 나타낸 것을 고르는 문제에서 옳지 않은 것을 답으로 생성하여 문제 인식의 오류를 나타냈다. 또한 보기 ①번에서 $0.1 \times a = 0.a$ 로 잘못된 수학적 기호 사용을 바르게 나타냈다고 인식하는 오류를 보였고, 보기 ②번과 ③번에 대한 풀이 과정에서 옳지 않다고 설명하면서 옳은 풀이를 제시하였지만 바르게 나타낸 보기를 찾아야 하는 문제에서 옳지 않은 보기를 답으로 제시하였다. 보기 ④번 보기의 내용 $a - b \div 5 = \frac{a-b}{5}$ 를 $a - b \div 5 = a - \frac{b}{5}$ 이라고 옳게 수정하였다. 바르게 제시된 보기⑤은 옳은 답으로 제시해야하는데 언급을 하지 않았다.

8

어떤 다항식에 $3x+2$ 를 더해야 할 것을 잘못 하여 빼었더니 $-2x+10$ 이 되었다. 바르게 계산한 식을 구하시오. [5점]



문제에서 제시된 상황은 다항식에 $3x+2$ 를 더해야 할 것을 잘못하여 빼었고, 그 결과로 $-2x+10$ 이 되었다는 것입니다. 이를 통해 원래의 다항식을 구하겠습니다.

문제 풀이

- 주어진 상황: 잘못된 계산은 다음과 같습니다.

$$\text{원래 다항식} - (3x + 2) = -2x + 10$$

- 원래 다항식 구하기: 이 식을 정리하면,

$$\text{원래 다항식} = (-2x + 10) + (3x + 2)$$

- 계산:

$$\begin{aligned} \text{원래 다항식} &= -2x + 10 + 3x + 2 \\ &= (3x - 2x) + (10 + 2) \\ &= x + 12 \end{aligned}$$



따라서, **바르게 계산한 식은 $x+12$ **입니다.

[그림 4] G교과서 ‘문자와 식’ II-8 문제 인식 오류

- 다음 중 기호 \times , \div 를 생략하여 바르게 나타낸 것은? [5점]

- ① $0.1 \times a = 0.a$
- ② $a \times a \times a = 3a$
- ③ $a \div 2 \times b = \frac{a}{2b}$
- ④ $a - b \div 5 = \frac{a-b}{5}$
- ⑤ $x \times x \times x \times y = x^3y$

문제 2

다음 중 기호 \times , \div 를 생략하여 바르게 나타낸 것은?

1. $0.1 \times a = 0.a$
2. $a \times a \times a = 3a$
3. $a \div 2 \times b = \frac{a}{2b}$
4. $a - b \div 5 = a - \frac{b}{5}$
5. $x \times x \times x \times y = x^3y$

옳지 않은 것은 2번과 3번입니다. $a \times a \times a = a^3$ 이고 $a \div 2 \times b = \frac{a}{2} \times b = \frac{ab}{2}$ 입니다.

[그림 5] E교과서 ‘문자와 식’ II-2 잘못된 수학적 기호 사용, 문제 인식 오류 예시

3) ‘좌표평면과 그래프’ 단원에서의 오류 예시

‘좌표평면과 그래프’ 단원에서 가장 많이 나타난 두 가지 오류 유형은 그래프 인식 오류(29건)와 좌표평면 인식 오류(11건)였다. [그림 6]은 ‘좌표평면과 그래프’ 단원에서 그래프 인식 오류의 예시를 보여준다. 그래프에서 5시간 동안 거리가 0이 되는 경우가 두 번 나타나므로 노선을 2회 운행하는데 걸리는 시간이 5시간이라고 해석해야 한다. 그러나 1회 운행하는데 걸리는 시간을 5시간이라고 해석하여 오류를 나타냈다.

[4-5] 다음 그래프는 어느 버스가 종점을 출발하여 다시 종점으로 돌아올 때까지 시간에 따른 버스와 종점 사이의 직선거리를 나타낸 것이다. 물음에 답하시오.



4 이 버스가 노선을 1회 운행하는 데 걸리는 시간은? [6점]

- ① 2시간 30분 ② 3시간
- ③ 3시간 30분 ④ 4시간
- ⑤ 5시간



문제 4

주어진 그래프는 버스가 출발해서 다시 종점으로 돌아오는 시간을 나타냅니다. 이를 통해 한 번 운행하는 데 걸리는 시간을 알아내야 합니다.

그래프를 보면, 버스가 다시 종점으로 돌아오는 데 5시간이 걸립니다.

따라서, 정답은 5번 (5시간)입니다.

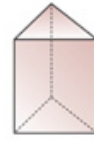
[그림 6] E교과서 ‘좌표평면과 그래프’ III-4 그래프 인식 오류 예시

4) ‘도형’ 단원에서의 오류 예시

‘도형’ 단원에서 가장 많이 나타난 세 가지 오류 유형은 그림 인식 오류(164건), 잘못된 식 세우기 오류(22건), 문제 인식 오류(15건)였다. [그림 7]은 ‘도형’ 단원에서 그림 인식 오류의 예시를 보여준다. 입체도형 그림을 해석하고 주어진 문제의 답을 구하는 문제이다. 교점의 개수는 꼭짓점의 수와 동일하므로 6개, 교선의 개수는 모서리의 수와 동일하므로 9개이다. 문제의 답은 $6 + 9 = 15$ 이다. 그러나 ChatGPT는 그림을 잘못 이해한 상태로 풀었다. 삼각기둥을 ‘지붕이 삼각형 모양인 집 모양’으로 인식하여 밑면의 꼭짓점을 4개, 지붕 위의 꼭짓점을 1개로 잘못 판단하였다.

1 오른쪽 그림과 같은 입체도형에서 교점의 개수를 a , 교선의 개수를 b 라고 할 때, $a+b$ 의 값은?

- ① 6 ② 9
③ 12 ④ 15
⑤ 18



문제 1

오른쪽 그림과 같은 입체도형에서 교점의 개수를 a , 교선의 개수를 b 라고 할 때, $a+b$ 의 값을 구하시오.

도형 분석

주어진 도형은 지붕이 삼각형 모양인 집 모양입니다. 이 도형을 분석해보겠습니다.

1. 교점 (꼭짓점, a)의 개수:

- 밑면의 네 꼭짓점: 4개
- 지붕의 꼭짓점: 1개 (맨 위의 점)
- 지붕의 삼각형 모양을 구성하는 각 모서리의 끝점: 2개

$$\text{총 교점 수: } 4 + 1 + 2 = 7$$

2. 교선 (모서리, b)의 개수:

- 밑면의 모서리: 4개
- 옆면에서 지붕까지 연결된 모서리: 4개
- 지붕의 모서리: 3개

$$\text{총 교선 수: } 4 + 4 + 3 = 11$$

$a+b$ 계산

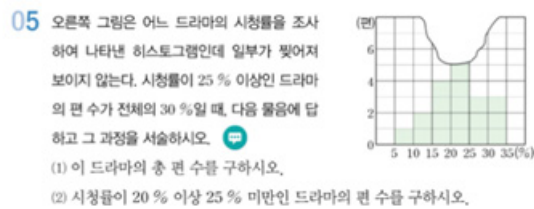
$$a+b = 7 + 11 = 18$$


답: 18

[그림 7] D교과서 ‘도형’ IV-1 그림 인식 오류 예시

5) ‘통계’ 단원에서의 오류 예시

‘통계’ 단원에서 가장 많이 나타난 세 가지 오류 유형은 그래프 인식 오류(44건), 표 인식 오류(23건), 그림 인식 오류(11건)였다. [그림 8]은 ‘통계’ 단원에서 그래프 인식의 오류에 대한 예시를 보여준다. 히스토그램에서 세로 축인 드라마의 수를 구하는 과정에서 오류가 나타났다. ChatGPT-4o는 앞에서 설명했듯이 수직선, 좌표평면 위에 있는 점의 좌표를 잘 구하지 못했는데 히스토그램의 막대 크기도 잘 구하지 못했다. 아래의 예시에서 25% 이상인 드라마는 25% 이상 30% 미만에 3명, 30% 이상 35% 미만에 3명이므로 총 6명이어야 하는데 3명이라고 잘못 파악해서 풀이했다. 또한 찢어져 보이지 않는 부분의 세로 축의 값을 계산해서 구해야 하는데 이를 구하지 않고 15% 이상 20% 미만의 4편으로 선택하여 잘못 풀이하였다.



 이 문제는 어떤 드라마의 시청률을 조사한 히스토그램을 바탕으로 총 편 수와 특정 시청률 구간에 속하는 편 수를 구하는 것입니다.

1. 이 드라마의 총 편 수를 구하는 방법:

- 문제에서 주어진 조건에 따르면, 시청률이 25% 이상인 드라마의 편 수가 전체의 30%라고 합니다.
- 히스토그램에서 25% 이상인 구간을 보면, 이 구간에서 드라마의 편 수는 3편입니다.
- 이 3편이 전체의 30%에 해당한다는 말이므로, 전체 편 수는 다음과 같이 계산할 수 있습니다:

$$\text{전체편수} = \frac{3\text{편}}{0.3} = 10\text{편}$$

2. 시청률이 20% 이상 25% 미만인 드라마의 편 수를 구하는 방법:

- 히스토그램에서 시청률이 20% 이상 25% 미만인 구간을 보면, 이 구간에 해당하는 편 수는 4편입니다.

따라서 답은 다음과 같습니다:

1. 이 드라마의 총 편 수는 10편입니다.
2. 시청률이 20% 이상 25% 미만인 드라마의 편 수는 4편입니다.

[그림 8] H교과서 ‘통계’ IV-5 히스토그램 그래프 인식 오류 예시

3. 프롬프트 엔지니어링에 의한 정답률 변화

가. 프롬프트 엔지니어링으로 변화된 정답률

ChatGPT 입력창 아래에는 ‘ChatGPT는 실수를 할 수 있으니 중요한 정보를 확인하세요’라는 문구가 표시되어 있다. ChatGPT는 자신의 실수를 인정하고 답변을 수정하여 다시 제공할 수 있다. 따라서 퓨샷 프롬프트와 반복적 프롬프트의 방법을 활용한 프롬프트 엔지니어링을 통해 정답률이 향상되는지 확인하였다. 교과서 5종(교과서B, 교과서D, 교과서F, 교과서H, 교과서J)의 대단원 평가 문제를 입력했을 때, 풀이나 답이 틀린 경우에는 올바른 풀이로 수정하기를 요구하거나 올바른 풀이 과정과 답을 피드백으로 제시하였다. 피드백 직후 동일한 문제를 다시 입력했을 때 연속적으로 올바른 풀이와 정답을 생성하면 다음 문제를 입력하는 과정을 반복했다. 틀린 풀이에 대해 수정을 요구했을 때 수정이 되지 않거나 연속으로 옳은 풀이 과정과 정답을 생성하지 못하면 오답으로 결정하고 정답률에 반영하였다. 그 결과를 <표 4>에 나타냈다. 모든 단원에서 피드백 이후의 정답률이 향상됨을 알 수 있다. 수와 연산 단원에서 93.22%의 정답률까지 향상되었고 가장 낮은 정답률이었던 통계 단원에서도 73.58%까지 정답률을 올릴 수 있었다. 즉 ChatGPT는 자신의 실수를 인정하고 자신의 답변을 옳게 수정하려고 시도한다. 그러나 표, 그림, 그래프를 포함한 문제는 수정 및 피드백을 하더라도 동일한 오류가 반복되는 경우가 있었다.

교과서 5종(교과서B, 교과서D, 교과서F, 교과서H, 교과서J)에 대해 옳은 풀이과정과 정답에 대한 피드백을 시도한 후 이를 초기화하지 않은 상태에서 나머지 교과서 5종(교과서A, 교과서C, 교과서E, 교과서G, 교과서I)의 문제를 다시 입력하였다. 일반적으로 앞의 5종의 교과서로 수정, 피드백하여 정답률을 올렸으므로 표현이나 숫자만 변형된 유사한 유형의 문제들을 포함한 나머지 교과서 5종에서도 정답률이 올라갈 것으로 기대했다. 오세준(2023)의 연구에서 비슷한 예제로 풀이 방법을 알려주고 바로 비슷한 유형의 문제에 대한 풀이 과정과 답을 요구했을 때, 정답률이 90% 정도로 높았다. 이를 참고하여 나머지 교과서 5종의 문제를 4회째 입력하여 프롬프트 엔지니어링의 작업으로 인한 정답률의 변화는 <표 4>에 나타나 있다. 교과서 B, D, F, H, J의 같은 문제에 대한 프롬프트 엔지니어링의 효과로 정답률이 향상되었지만, 교과서 A, C, E, G, I의 유사한 문제에 대한 프롬프트 엔지니어링의 효과는 정답률 향상으로 이어지지 않았다.


<표 4> B, D, F, H, J교과서 문제로 수정, 피드백 후 4회 정답률 변화(단위: %)

	교과서 B, D, F, H, J				교과서 A, C, E, G, I			
	1회	2회	3회	4회	1회	2회	3회	4회
수와 연산	67.80	69.49	69.49	93.22	67.57	72.97	72.97	55.41
문자와 식	63.86	73.49	72.29	92.77	71.83	77.46	76.06	76.65
좌표평면과 그래프	57.81	65.63	65.63	81.25	61.3	62.9	67.7	69.4
도형	36.90	37.97	36.90	72.19	36.2	34.2	38.3	30.9
통계	28.30	12.09	18.87	73.58	25.00	26.79	26.79	14.29

ChatGPT-4o의 프롬프트 엔지니어링의 효과를 확인해 보았을 때 같은 문제에 대한 효과는 뚜렷하였지만 유사한 문제에 대한 효과는 크지 않았다. ChatGPT-4o에 같은 문제를 4회째 입력했을 때 프롬프트 엔지니어링을 통해 정답률이 향상되는 것을 확인하였다. 오류에 대한 즉각적인 수정을 요구하면 반영이 되며 옳은 풀이 과정과 답을 생성할 수 있었다. 동일한 문제에 대한 프롬프트 엔지니어링의 효과로 ‘수와 연산’, ‘문자와 식’ 단원의 정답률은 90% 이상, ‘좌표평면과 그래프’ 단원은 80% 이상, ‘도형’, ‘통계’ 단원에서는 70% 이상으로 향상되었다. 하지만 동일한 문제가 아닌 유사한 문제를 다시 입력했을 때 이전에 했던 오류를 나타내어 프롬프트 엔지니어링의 효과는 나타나지 않았다.

1) 프롬프트 엔지니어링을 통한 오류 수정

아래의 [그림 9]는 ‘수와 연산’ 단원에서 어떤 수에서 잘못 계산한 결과값을 이용하여 바르게 구한 값을 계산하는 문제이다. ‘어떤 수’까지만 구하는 오류를 수정하여, ‘어떤 수’를 구하고 ‘바르게 계산한 값’까지 구해야 한다고 요구하였더니 ‘바르게 계산한 값’까지 구하게 되었다. 어떤 수를 네모 칸으로 표현한 뒤, 곱해야 할 것을 잘못 더한 정보를 이용하여 네모 칸의 값(어떤 수)을 구한 다음, 그 이후에 곱셈 수행으로 올바르게 계산함으로써 오류가 수정되었다.

07 어떤 유리수에 $-\frac{3}{5}$ 을 곱해야 할 것을 잘못하여 더하였더니 $\frac{1}{3}$ 이 되었다. 이때 옳게 계산한 값을 구하고 그 과정을 서술하시오. 



이미지를 참고하여 문제를 다시 풀어보겠습니다.

문제 풀이 과정:

1. 문제 식 설정: 주어진 식은 $\square + \left(-\frac{3}{5}\right) = \frac{1}{3}$ 입니다.
2. \square 값 구하기: \square 를 구하기 위해 $-\frac{3}{5}$ 를 오른쪽으로 이항합니다.

$$\square = \frac{1}{3} + \frac{3}{5}$$

3. 공통 분모를 구하여 더하기:

$$\square = \frac{5}{15} + \frac{9}{15} = \frac{14}{15}$$

4. 곱셈 수행: 이제 올바르게 계산한 식은 $\square \times \left(-\frac{3}{5}\right)$ 입니다.

$$\frac{14}{15} \times \left(-\frac{3}{5}\right) = -\frac{42}{75}$$

하지만, 문제에 나온 참고 자료에서는 최종 답이 $-\frac{14}{25}$ 로 되어 있으므로 올바른 과정은 아래와 같습니다:

$$\frac{14}{15} \times \left(-\frac{3}{5}\right) = -\frac{14 \times 3}{15 \times 5} = -\frac{42}{75}$$

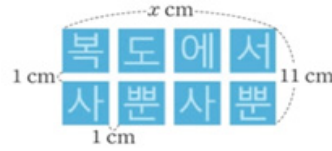
따라서, 올바른 계산 결과는 $-\frac{14}{25}$ 입니다.

[그림 9] H교과서 ‘수와 연산’ I-2-7 수정이 반영된 오류 예시

2) 프롬프트 엔지니어링 후에도 계속된 오류

‘수와 연산’, ‘문자와 식’ 단원에서 가장 많이 나타났던 오류는 문제 인식 오류였다. 옳은 것을 고르는 문제에서 옳지 않은 것을 고르는 오류를 반복적으로 수정, 피드백하였으나 수정이 반영되지 않은 경우가 많았다. ‘좌표평면과 그래프’와 ‘통계’ 단원에서는 그래프, 수직선, 좌표평면 위에 있는 점의 좌표를 구하는 문제에서 오류가 자주 발생하였다. ChatGPT-4o는 좌표평면, 그래프, 수직선 위의 점의 좌표를 알려주면 옳은 문제 풀이가 가능하지만 새롭게 입력한 문제에서 점의 좌표를 찾는 데는 어려움이 있었다. ‘도형’ 단원에서도 그림을 잘못 인식하는 경우는 수정이 되지 않아 반복적인 오류를 보였다. [그림 10]은 ‘문자와 식’ 단원에서 그림 인식 오류를 보여주는 예시로 반복적인 수정과 피드백을 해도 오류 수정이 되지 않는 결과를 보였다. 문제해결을 위해서는 세로에서 1cm, 가로에서 3cm를 빼야 하는데 가로, 세로에서 각각 2cm씩 빼는 오류를 나타냈다. 이 예시는 사용자가 문제 풀이에서 오류 수정을 요구했을 때 ChatGPT-4o가 자신의 오류를 인정하였지만, 동일한 문제를 다시 입력하면 같은 오류가 반복되어 오류가 수정되지 않았음을 보여준다.

- 4 학교 안전 표어 공모전에 제출하기 위해 다음 그림과 같이 가로 길이가 x cm, 세로 길이가 11 cm인 직사각형 모양의 종이를 같은 간격으로 나누어 표어를 작성하려고 한다. 색칠한 부분의 넓이를 식으로 나타내시오.



주어진 문제에 따라 색칠된 부분의 넓이를 구하는 방법을 다시 설명드리겠습니다.

색칠된 부분의 넓이 계산:

1. 가로 길이 계산:

- 전체 가로 길이는 x cm입니다.
- 양쪽에 1 cm씩의 여백이 있으므로, 색칠된 부분의 가로 길이는 $x - 2$ cm가 됩니다.

2. 세로 길이 계산:

- 전체 세로 길이는 11 cm입니다.
- 위아래에 1 cm씩의 여백이 있으므로, 색칠된 부분의 세로 길이는 $11 - 2 = 9$ cm가 됩니다.

3. 색칠된 부분의 넓이:

- 색칠된 부분의 넓이는 가로 길이와 세로 길이를 곱하여 구할 수 있습니다.

$$\text{색칠된 부분의 넓이} = (x - 2) \times 9 = 9(x - 2) \text{ cm}^2$$

따라서, 색칠된 부분의 넓이는 $9(x - 2) \text{ cm}^2$ 로 나타낼 수 있습니다.

[그림 10] D교과서 ‘문자와 식’ II-4 수정이 안 되는 오류

3) 프롬프트 엔지니어링을 통해 얻은 우수한 풀이와 다양한 풀이

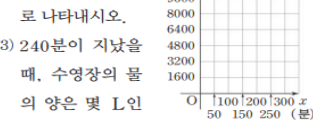
ChatGPT-4o의 수학 문제 풀이 과정에서 다양한 오류가 종종 발견되지는 하지만 조건에 맞게 우수한 풀이 과정과 답을 생성한 예시도 있다. [그림 11]은 ‘좌표평면과 그래프’ 단원에서 우수한 풀이를 생성한 예시를 보여준다. 문장으로 표현된 문제를 분석하여 (1) 표의 빈칸을 채우고 (2) 표를 참고하여 그래프로 나타내고 (3) 그래프를 참고하여 방정식을 구해서 특정한 값을 구해야 한다. 문제를 정확히 이해한 내용을 바탕으로 표의 빈칸을 완성하였으며 완성한 표를 바탕으로 좌표의 점을 찾았다. ChatGPT-4o의 풀이 과정에서 그래프를 그리겠다고 하였지만 그리지는 않았고, 그래프를 그려달라고 다시 요구했을 때는 그래프까지 정확하게 완성하였다. 마지막으로 그래프를 방정식으로 표현하고 문제에 대한 답을 구하였다. 이러한 풀이 과정은 아무리 늦어도 1분이 걸리지 않았다. ChatGPT는 사용자가 요구하는 내용을 반영하며 원하는 답을 생성할 수 있어 상호작용이 가능한 피드백을 즉각적으로 제공해 줄 수 있다.

8. 수영장에 물을 1분에 32 L씩 채우고 있다. x 분 동안 채운 수영장의 물의 양을 y L라고 할 때, 다음 물음에 답하시오. (단, 수영장의 물을 모두 채우는 데에는 5시간이 소요된다.) [9점]

(1) 표를 완성하시오.

x (분)	50	100	150	200	250	300
y (L)						

(2) (1)의 표를 그래프



로 나타내시오.

(3) 240분이 지났을

때, 수영장의 물

의 양은 몇 L인

지 구하시오.



주어진 문제에 따라 수영장에 물을 채우는 속도와 시간을 바탕으로 표를 완성하고, 그래프를 그리며, 특정 시간 이후의 물의 양을 계산하겠습니다.

(1) 표 완성하기

표는 다음과 같이 완성됩니다.

x (분)	50	100	150	200	250	300
y (L)	1600	3200	4800	6400	8000	9600

(2) 그래프 그리기

위 표의 데이터를 기반으로 그래프를 그려겠습니다.

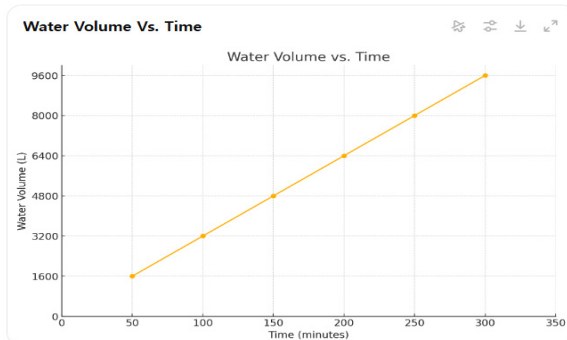
(3) 240분이 지났을 때 물의 양 계산

240분 동안 물을 채웠다면, 물의 양은 다음과 같습니다.

$$y = 32 \times 240 = 7680 \text{ L}$$

따라서 240분 동안 채운 물의 양은 7680L입니다.

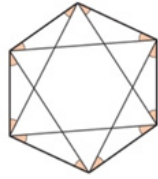
그래프는?



[그림 11] C교과서 ‘좌표평면과 그래프’, III-8

[그림 12]는 ‘도형’ 단원 문제에 대해 ChatGPT-4o가 생성한 다양한 풀이의 예시이다. 다각형에서 색칠된 각의 크기를 구하는 문제로 첫 번째 풀이에서는 외각을 이용하여 풀이 과정을 생성했다. 풀이 과정에서 육각형을 정육각형이라고 잘못 언급하였으나 정육각형의 특징을 사용하지는 않았다. 색칠된 각을 포함하는 작은 삼각형에서 두 내각의 합은 외각이 되고 이것이 작은 육각형의 외각이 된다는 사실과 육각형의 외각의 합이 360° 가 된다는 사실을 이용하였다. 그리고 다른 풀이 방법을 요구했을 때 색칠된 각의 합은 큰 육각형 내각의 합에서 두 큰 삼각형의 내각의 합을 뺀 것과 같다는 원리를 이용하며 옳은 풀이 과정과 정답을 생성하였다. 다른 방법을 요구할 때마다 다른 풀이 방법을 시도하였다. 따라서 학생들이 하나의 풀이 방법이 이해되지 않을 때 ChatGPT-4o는 더 쉬운 다른 풀이 방법을 생성할 수 있어 학생들의 이해를 도울 수 있다.

11 다음 그림에서 색칠한 각의 크기의 합을 구하시오.



이 문제를 풀기 위해 색칠된 각의 합을 구하는 방법을 간단히 설명드리겠습니다.

이 그림에서 색칠된 각들은 모두 정육각형 내부의 삼각형들의 외각에 해당합니다. 삼각형의 외각의 합은 삼각형 내부의 두 내각의 합과 같으며, 다각형의 모든 외각의 합은 항상 360° 입니다.

따라서, 주어진 정육각형에서 색칠된 각들의 합도 정육각형의 외각의 합과 동일하게 360° 가 됩니다. 이는 육각형의 대칭성과 외각의 성질을 이용한 간단한 풀이입니다.

결론적으로, 색칠된 각들의 합은 **360°** 입니다.

다른 풀이 방법 알려줘

육각형의 내각의 합은 $(6 - 2) \times 180^\circ = 720^\circ$ 입니다. 여기서 6은 육각형의 변의 개수입니다. 문제에서 색칠된 각들의 합은, 이 720° 에서 두 개의 큰 삼각형의 내각의 합($2 \times 180^\circ = 360^\circ$)을 뺀 값입니다. 따라서 색칠된 각들의 합은 $720^\circ - 360^\circ = 360^\circ$ 가 됩니다. 결과적으로, 색칠된 각들의 합은 **360°** 입니다.

[그림 12] B교과서 ‘도형’ 단원 VI -11 에서 우수한 풀이 예시

나. ChatGPT-3.5, Chat-GPT4 그리고 ChatGPT-4o의 정답률 비교

ChatGPT-3.5의 문제점을 보완하여 다음 모델인 ChatGPT-4와 ChatGPT-4o가 출시되었다. 이들 모델은 매개변수의 개수가 늘어났으며 맥락 이해 능력이 향상되었고, 관련성 높은 응답 생성이 가능해졌다. 또한 언어지원이 강화되어 다국어 사용이 가능해졌고, 웹 검색이 가능하여 오류가 감소하였다. 또 더 정확한 정보를 제공할 수 있어서 전문화된 작업을 지원할 수 있다(Plevris et al. 2023). 이러한 개선 사항은 수학교육에 긍정적인 영향을 미칠 수 있으며, 이를 확인하기 위하여 수학 문제 풀이에 대한 정답률 변화를 확인하였다. 현재는 ChatGPT-4, ChatGPT-4o mini, ChatGPT-4o만 사용 가능하기 때문에 이들을 비교 연구하기 위해서 Plevris et al.(2023)이 ChatGPT3.5, ChatGPT4, Google Bard를 비교하여 연구한 문제를 ChatGPT-4o에 똑같이 입력하여 비교 분석하였다. 선행연구에서 선택한 수학 문제 30문제는 두 세트로 나뉜다. 세트 A는 온라인에서 찾을 수 없는 독창적인 수학 문제 15문제이고 세트 B는 일반적으로 풀이와 정답을 온라인에서 찾을 수 있는 수학 문제 15문제로 구성되어 있다. <표 5>와 <표 6>은 ChatGPT-3.5, ChatGPT-4, ChatGPT-4o의 정답률의 변화를 보여준다. <표 5>는 세트 A 15문제를 세 모델에 각각 3회씩 입력하고 구한 정답률이다. ChatGPT-3.5의 정답률은 53.3%, ChatGPT-4는 80.00%, ChatGPT-4o는 93.33%로 ChatGPT-4o의 정답률이 가장 높았다. <표 6>은 세트 B 15문제를 세 모델에 3회씩 입력하고 구한 정답률이다. ChatGPT-3.5의 정답률은 42.22%, ChatGPT-4는 62.22%, ChatGPT-4o는 84.44%를 기록하여, 역시 ChatGPT-4o의 정답률이 가장 높았다. ChatGPT3.5에서 ChatGPT4.0, ChatGPT-4o로 발전하면서 정답률이 급속히 상승함을 알 수 있다. 이러한 발전 추세로 보면 chatGPT는 앞으로 수학교육에서 활발히 활용될 가능성이 높아 보인다.

<표 5> ChatGPT-3.5, ChatGPT-4, ChatGPT-4o의 setA 정답률 비교(맞으면 1 틀리면 0)

문제 번호		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	합	정답률
3.5	1회	1	0	0	1	1	1	1	0	1	1	0	1	0	0	0	8	53.33
	2회	1	0	1	0	1	0	0	0	1	1	1	1	0	0	1	8	53.33
	3회	1	0	0	0	1	0	1	1	1	1	1	1	0	0	0	8	53.33
	합계	3	0	1	1	3	1	2	1	3	3	2	3	0	0	1	24	53.33
4	1회	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	11	73.33
	2회	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	13	86.67
	3회	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	12	86.67
	합계	1	0	1	3	3	3	3	3	3	3	3	2	2	3	3	36	80.00
4o	1회	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	14	93.33
	2회	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	14	93.33
	3회	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	14	93.33
	합계	3	0	3	3	3	3	3	3	3	3	3	3	3	3	3	42	93.33

<표 6> ChatGPT-3.5, ChatGPT-4, ChatGPT-4o의 setB 정답률 비교(맞으면 1 틀리면 0)

문제 번호		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	합	정답률
3.5	1회	1	1	0	0	1	0	0	0	0	0	1	0	1	1	0	6	40.00
	2회	1	0	0	0	1	0	0	1	0	0	1	1	1	1	0	7	46.67
	3회	0	1	0	1	1	0	0	0	0	0	1	1	0	1	0	6	40.00
	합계	2	2	0	1	3	0	0	1	0	0	3	2	2	3	0	19	42.22
4	1회	1	1	0	1	1	1	1	0	0	0	1	1	1	1	0	10	66.67
	2회	1	1	0	0	1	0	1	0	0	0	1	1	1	1	0	8	53.33
	3회	1	1	0	0	1	1	1	1	0	0	1	1	1	1	0	10	66.67
	합계	3	3	0	1	3	2	3	1	0	0	3	3	3	3	0	28	62.22
4o	1회	1	1	0	1	1	1	1	0	1	0	1	1	1	1	1	12	80.00
	2회	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	13	86.67
	3회	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	13	86.67
	합계	3	3	0	3	3	3	3	2	3	0	3	3	3	3	3	38	84.44

V. 결론

현재의 ChatGPT-4o는 이미지를 입력하면 바로 문제를 인식할 수 있어서 다른 수학 관련 프로그램들에서 요구하는 수학적 기호와 수식을 작성하는 법을 따로 익히지 않아도 된다. 또 상호작용이 가능하여 즉각적인 피드백이 가능하고, 학생이 다른 문제 풀이 방법을 요구하면 다른 방법을 바로 시도하여 알려줄 수 있다. 또 개별 사용자의 요구에 맞게 답할 수 있고, 답변 생성시간이 빠르고, 표, 그래프, 그림 등을 단시간에 작성할 수 있다. 또한 ChatGPT-4o는 학생이 어떠한 문제를 입력해도 수학적 지식을 바탕으로 그 문제에 대한 풀이를 시도하는데, 이는 다른 AI 프로그램이 문제해결을 할 수 없을 때 풀이를 회피하거나 유사한 문제의 풀이를 제시하는 것과 비교해 보면 ChatGPT의 장점을 잘 이해할 수 있다. 이와 같은 장점으로 ChatGPT-4o는 수학교육에서도 효과적으로 활용될 수 있는 잠재력을 가지고 있다.

본 연구에서는 수학교육에서 ChatGPT-4o의 기능과 활용 가능성을 탐구하였다. 활용 가능성을 탐구하기 위해 ChatGPT-4o의 기능과 정답률, 오류 유형, 프롬프트 엔지니어링의 효과를 분석하여 다음 같은 결론과 시사점을 얻었다.

첫째, ChatGPT-4o의 정답률을 분석한 결과 ChatGPT-4o의 전체 정답률은 51.76%로 나타났다. 단위별로 비교한 결과 “문자와 식”(72.29%)과 수와 연산’ 단원에서 다른 단위보다 정답률이 더 높게 나왔다. 반면에 표, 그래프, 그림이 많이 나오는 ‘통계’(23.55%) 단원의 정답률이 가장 낮게 나왔다. 이는 ChatGPT-4o가 특정 유형의 문제에 강점을 보이지만 다른 유형에서는 여전히 어려움을 겪고 있음을 시사한다. 교육 현장에서 ChatGPT를 적용할 때 교사가 AI 프로그램의 문제점과 장점을 정확하게 파악하고 ChatGPT와 같은 생성형AI에 대한 올바른 활용 방안을 마련할 필요가 있음을 시사한다.

둘째, ChatGPT-4o가 생성하는 풀이 과정과 정답에서 발생한 오류를 유형별로 분류하고 분석한 결과, ‘수와 연산’과 ‘문자와 식’ 단원에서는 문제 인식 오류가 가장 많았으며 ‘좌표평면과 그래프’와 ‘통계’ 단원에서는 그래프 인식 오류가 가장 많았고 ‘도형’ 단원에서는 그림 인식 오류가 가장 많이 나타났다. 그림, 그래프, 표 등 시각적 자료를 해석하는 일부 문제에서 우수한 풀이 과정을 생성하기도 했지만 잘못된 해석으로 오류가 나타나기도 했다. 특히 ChatGPT-4o가 새롭게 비전 기능을 갖추어 문제를 이미지로 입력하거나 다양한 시각적 자료를 인식할 수 있지만 아직은 정확히 인식하는 능력이 부족하여 학교 현장에 적용할 때 주의가 필요하다.

셋째, ChatGPT-4o의 프롬프트 엔지니어링의 효과를 확인해 보았을 때 동일한 문제에 대해 수정한 효과는 나타났지만 유사한 문제에 대한 수정의 효과는 크지 않았다. 따라서 현재로서는 수학교육 전문가가 유사한 많은 수학 문제를 프롬프트 엔지니어링 한 후 학생들에게 사용하게 하더라도 효과를 기대하기 어렵다. 그러나 프롬프트 엔지니어링의 효과를 검증하기 위해 수학 문제를 입력하고 생성된 풀이 과정에서 오류인 부분을 찾아내고 수정하는 과정이 학생들에게도 유용할 수 있다. 즉, 학생들에게 올바른 풀이 과정과 답을 도출하기 위해 프롬프트 엔지니어링의 수정 과정을 거치는 활동을 제시함으로써 학생과 활발히 의사소통하면서 학생의 수학적 개념과 이해를 더 확고하게 할 수 있다.

넷째, 세 모델 ChatGPT-3.5와 ChatGPT-4, 그리고 ChatGPT-4o의 정답률을 비교한 결과, ChatGPT-4o는 다른 모델보다 월등히 높은 정답률을 보였다. 세트 A, 세트 B 문제의 정답률이 각각 93.33%, 84.44%까지 향상된 결과로 미루어보았을 때 ChatGPT는 앞으로 수학교육에서 사용이 활발해질 것으로 추측된다. 그러나 교육 현장에서 ChatGPT-4o를 사용할 때는 단원의 성격에 따른 정답률의 차이, 시각적 자료 인식에서의 오류, 그리고 문제에 따라 프롬프트 엔지니어링의 효과가 다르다는 점 등에 대해 유의할 필요가 있다.

ChatGPT-4o와 같은 생성형 AI는 향후 수학교육에서 활용할 수 있는 잠재력을 가지고 있음을 확인할 수 있었다. 이제까지의 분석을 바탕으로 다음과 같이 제언하고자 한다.

첫째, ChatGPT-4o의 수학 문제 풀이에 대한 정답률이 많이 개선되었지만, 학교 현장에서 학생들에게 안정적으로 활용할 수 있기 위해서는 정답률 개선을 위한 후속 연구가 필요하다. 한글로 작성된 수학 문제를 인식할 때 오류가 나타났으며, 같은 문제를 영어로 풀이하는 것과 한국어로 풀이하는 것에 정답률의 차이가 나타나 이에 관련된 연구도 필요하다.

둘째, ChatGPT-4o의 비전 기능에 관련된 후속 연구 또한 필요하다. ChatGPT에 표, 그래프, 그림 등의 입력이 가능해져서 그림을 포함하는 문제를 입력하는 것이 쉬워졌지만 연구의 결과에서 볼 수 있는 것처럼 여러 오류가 나타났다. 또한 수학 문제에서 주어진 조건을 바탕으로 표, 그래프, 그림 등으로 정확하게 나타내는 것과 관련된 연구를 통해 이와 같은 기능이 개선되어 제대로 기능한다면, 그림, 표, 그래프 등을 보다 정확히 표현함으로써 학생들이 수학 문제를 더 잘 이해하는 데 도움이 될 것이다.

셋째, 연구 수행 중에 ChatGPT는 다양한 문제를 유창하게 풀어내는 능력을 갖고 있음을 볼 수 있었다. 이러한 장점을 활용하여 서술형 평가, 수학 에세이 쓰기, 수학 독서에도 활용될 가능성이 있다고 생각되어 이에 대한 연구가 필요하며, 생성형 AI와 같은 ChatGPT가 교육에 도입될 경우 학생들에게 미치는 영향에 대한 연구가 부족하므로 이에 대한 연구도 아울러 필요하다.

참고문헌

- 강동훈. (2023). 챗지피티 (ChatGPT) 의 등장과 국어교육의 대응. **국어문학**, 82, 469-496. [Kang, D. (2023). The advent of ChatGPT and the response of Korean language education. *Korean Language and Literature*, 82, 469-496.]
- 강옥기, 권언근, 황혜정, 전대열, 노지화, 우희정, 윤상혁, 이형주, 유승연, 윤혜미, 홍창섭, 정경호(2020a). **중학교 수학 1**. 서울: 동아출판(주). [Kang, O., Kwon, E., Hwang, H., Jeon, D., Noh, J., Woo, H., Yoon, S., Lee, H., Yoo, S., Yoon, H., Hong, C., & Jeong, G. (2020a). *Middle school math 1*. Seoul: Donga Publisher.]
- 강윤지(2024). 수학 문장제 해결과 관련한 ChatGPT 의 교수학적 활용 방안 모색. **수학교육 논문집**, 38(1), 49-67. [Kang, Y. J. (2024). A study on the didactical application of ChatGPT for mathematical word problem solving. *Communications of Mathematical Education*, 38(1), 49-67.] <https://doi.org/10.7468/jksmee.2024.38.1.49>
- 고호경, 김응환, 김인수, 이봉주, 한준철, 최수영, 김정현, 김화영, 정시훈, 조준모, 최화식, 최화정(2019). **중학교 수학 1**. 서울: (주)교학사. [Ko, H., Kim, E., Kim, I., Lee, B., Han, J., Choi, S., Kim, J., Kim, H., Jeong, S., Jo, J., Choi, H., & Choi, H. (2019). *Middle school math 1*. Seoul: Kyohaksa.]
- 권미선(2024). GPTs 기반 문제해결 맞춤형 챗봇 제작 및 수학적 성능 분석. **수학교육**, 63(3), 467-484. [Kwon, M. (2023). Development of a customized GPTs-based chatbot for pre-service teacher education and analysis of its educational performance in mathematics. *The Mathematical Education*, 63(3), 467-484.] <https://doi.org/10.7468/mathedu.2024.63.3.467>
- 권오남, 오세준, 윤정은, 이경원, 신병철, 정원(2023). ChatGPT 의 수학적 성능 분석: 국가수준 학업성취도 평가 및 대학수학능력시험 수학 문제 풀이를 중심으로. **수학교육 논문집**, 37(2), 233-256. [Kwon, O., Oh, S., Yoon, J., Lee, K., Shin, B., & Jung, W. (2023). Analyzing mathematical performances of ChatGPT: Focusing on the solution of national assessment of educational achievement and the college scholastic ability test. *Communications of Mathematical Education*, 37(2), 233-256.]
- 김동성(2023). 챗GPT 프롬프트 엔지니어링을 활용한 상식 추론결과에서 나타난 언어 자질들 분석. **언어**, 48(2), 687-711. [Kim, D. S. (2023). The analysis on the linguistic features of common sense inference using Chat-GPT prompt engineering. *Korean Journal of Linguistics*, 48(2), 687-711.] <http://dx.doi.org/10.18855/lisoko.2023.48.3.009>
- 김수민, 백선환(2023). **챗GPT 거대한 전환**. 서울:알에이치코리아
- 김원경, 조민식, 방금성, 임석훈, 김동화, 강순자, 배수경, 지은정, 김윤희(2018). **중학교 수학 1**. 서울: (주)비상교육. [Kim, W., Cho, M., Bang, S., Rim, S., Kim, D., Kang, S., Bae, S., Ji, E., Kim, Y. (2018). *Middle school math 1*. Seoul: Visang Education.]
- 김육태, 김영식(2024). 수학과 ‘확률과 통계’영역에서 ChatGPT 를 활용한 서답형 평가 피드백이 학생들의 피드백 리터러시에 미치는 영향. **컴퓨터교육학회 논문지**, 27(3), 19-30. [Kim, W. T., & Kim, Y. S. (2024). The influence of ChatGPT-enhanced evaluation feedback on students' feedback literacy in descriptive assessments within the 'Probability and statistics' domain of mathematics education. *The Journal of Korean Association of Computer Education*, 27(3), 19-30.] <https://doi.org/10.32431/kace.2024.27.3.002>
- 김화경, 나귀수, 이미라, 이해경, 권영기(2020). **중학교 수학 1**. 서울: (주)좋은책신사고. [Kim, H., Na, G., Lee, M., Lee, A., & Kwon, Y. (2020). *Middle school math 1*. Seoul: Shinsago.]
- 류희찬, 선우하식, 신보미, 정동승, 장영훈, 설정수, 박슬히(2018). **중학교 수학 1**. 서울: (주)천재교육. [Ryu, H., Seonwoo, H., Shin, B., Jeong, D., Jang, Y., Seol, J., & Park, S. (2018). *Middle school math 1*. Seoul: Chunjae Education.]
- 박교식, 이종희, 김진환, 남진영, 김남희, 권석일, 임재훈, 유연주, 김선희, 장미라, 장미선, 최수연, 황지연, 김재원, 박소현, 양수영, 이은영, 정주연, 주미(2018). **중학교 수학 1**. 서울: 동아출판(주). [Park, K., Lee, J., Kim, J., Nam, J., Kim, N., Kwon, S., Lim, J., Yoo, Y., Kim, S., Jang, M., Jang, M., Choi, S., Hwang, J., Kim, J., Park, S., Yang, S., Lee, E., Jeong, J., & Joo, M. (2018). *Middle school math 1*. Seoul: Donga Publisher.]
- 박만구(2020). 수학교육에서 인공지능의 활용 동향. **한국초등교육**, 31(특집), 91-102. [Park, M. G. (2020). The trends of using artificial intelligence in mathematics education. *Korean Journal of Elementary Education*, 31(S), 91-102.]
- 손태권(2023). ChatGPT의 수학교육 활용 가능성 탐색: 분수 문제에 관한 학생의 산출물과 예비교사의 담화 사례를 중심으로. **초등수학교육**, 26(2), 99-113. [Son, T. K. (2023). Exploring the possibility of using ChatGPT in mathematics education: Focusing on student product and pre-service teachers' discourse related to fraction problems. *Education of Primary School Mathematics*, 26(2), 99-113.] <https://doi.org/10.7468/jksmec.2023.26.2.99>

- 신보미, 류경민, 김경용, 강성모, 주재은(2024). 생성형 AI를 활용한 디지털 리터러시 개발 방안 모색. **한국학교수학회논문집**, 27(2), 151-176. [Shin, B., Ryu, K., Kim, K., Kang, S., & Joo, J. (2024). A study on exploring strategies for developing digital literacy using generative AI: Focusing on the development of mathematics teaching and learning materials based on ChatGPT. *The Korean School Mathematics Society*, 27(2), 151-176.] <https://doi.org/10.30807/ksms.2024.27.2.003>
- 오세준(2023). 수학 문제 해결에서 효과적인 ChatGPT 의 프롬프트 고찰: 이차방정식과 이차함수를 중심으로. **수학교육 논문집**, 37(3), 945-967. [Oh, S. J. (2023). Effective ChatGPT prompts in mathematical problem solving: Focusing on quadratic equations and quadratic functions. *Communications of Mathematical Education*, 37(3), 545-567.] <https://doi.org/10.7468/jksmee.2023.37.3.545>
- 이기마, 김희정(2024). ChatGPT 의 수학적 성과와 오류 분석: 다중양식적 접근을 통한 수학교육 활용 가능성 탐색. **한국학교수학회논문집**, 27(3), 363-392. [Lee, G. M., & Kim, H. J. (2024). Exploring the potential of ChatGPT in mathematics education: Analyzing mathematical performance and errors through a multimodal approach. *The Korean School Mathematics Society*, 27(3), 363-392.]
- 이준열, 최부림, 김동재, 이정례, 김상미, 원유미, 강해기, 김성철, 강순구(2020). **중학교 수학 1**. 서울: (주)천재교육. [Lee, J., Choi, B., Kim, D., Lee, J., Kim, S., Won, Y., Kang, H., Kim, S., & Kang, S. (2020). *Middle school math 1*. Seoul: Chunjae Education.]
- 이선(2024). 영어 쓰기 수업에서 피드백 보조도구로서의 ChatGPT 활용 가능성. **외국어교육**, 31(2), 27-54. [Lee, S. (2024). Exploring ChatGPT's potential as a supplementary feedback tool in English writing classes. *Foreign Languages Education*, 31(2), 27-54.] <https://doi.org/10.15334/FLE.2024.31.2.27>
- 이현수, 박영용(2024). ChatGPT 3.5 와 4.0 의 수학적 능력 비교 분석-중학교 2학년 수학 교과서를 중심으로. **한국콘텐츠학회논문지**, 24(7), 436-449. [Lee, H.-S., & Park, Y.-Y. (2024). A comparative analysis of the mathematical capabilities of ChatGPT 3.5 and 4.0 - Focused on the 8th grade mathematics textbooks. *Journal of the Korea Contents Association*, 24(7), 436-449.] <https://doi.org/10.5392/JKCA.2024.24.07.436>
- 장경윤, 강현영, 김동원, 안재만, 이동환, 박진형, 정경희, 홍은지, 김민정, 박정선, 지영명, 구나영(2017). **중학교 수학 1**. 서울: (주)지학사. [Jang, K., Kang, H., Kim, D., Ahn, J., Lee, D., Park, J., Jeong, K., Hong, E., Kim, M., Park, J., Ji, Y., & Gu, N. (2017). *Middle school math 1*. Seoul: Jihaksa.]
- 정한민, 박정훈(2023). ChatGPT를 이용한 문헌 작성 설계 및 이슈. **한국지식정보기술학회 논문지**, 18(1), 31-40. [Jung, H., & Park, J. (2023). Design and Issues of Writing Literatures Using ChatGPT. *Journal of Knowledge Information Technology and Systems (JKITS)*, 18(1), 31-40.] <https://doi.org/10.34163/jkits.2023.18.1.004>
- 정현우, 박건우(2024). 대규모 언어 모델의 소수 검증 능력 심층 평가: ChatGPT와 PaLM 2를 중심으로. **정보과학회논문지**, 51(8), 699-705. [Jung, H., & Park, K. (2024). In-depth evaluations of the primality testing capabilities of large language models: With a focus on ChatGPT and PaLM 2. *Journal of KIISE*, 51(8), 699-705.] <https://doi.org/10.5626/JOK.2024.51.8.699>
- 주미경, 강은주, 강소영, 이현구, 강석주, 오화평, 권상순(2017). **중학교 수학 1**. 서울: (주)금성출판사. [Joo, M., Kang, E., Kang, S., Lee, H., Kang, S., Oh, H., & Kwon, S. (2017). *Middle school math 1*. Seoul: Geumseong Publisher.]
- 진준화(2023). 중국어 문법 연구에 있어 '생성형 AI 언어 모델 서비스: ChatGPT'를 어떻게 활용할 것인가? **중국과 중국학**, (50), 165-196. [Jin, J. (2023). How to Use 'Generative AI Language Model Service: ChatGPT' in Chinese Grammar Research? *China and Sinology*, (50), 165-196.] <http://dx.doi.org/10.17935/Chinan.2023.50.165>
- 최숙기, 박종임(2023). 인공지능 시대의 작문 평가를 위한 ChatGPT 활용 방안 연구. **청람어문교육**, (95), 65-109. [Choi S. K., & Park, J. I. (2023). Exploring the utilization of ChatGPT for composition assessment in the era of artificial intelligence. *Cheongram Korean Language Education Association*, (95), 65-109.] <http://dx.doi.org/10.26589/jockle.95.202309.65>
- 황선옥, 강병개, 윤갑진, 인광연, 장홍월, 정종식, 조성률(2020a). **중학교 수학 1**. 서울: (주)미래엔. [Hwang, S., Kang, B., Yoon, G., In, G., Jang, H., Jeong, J., & Cho, S. (2020a). *Middle school math 1*. Seoul: MiraeN.]
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62. <https://doi.org/10.2139/ssrn.4337484>
- Bill Gates Podcast. (2024). Episode 6: Sam altman <https://youtu.be/PkXELH6Y2IM?si=01dycrmPZtMI56UG> (최종 검색 일: 2024.8.26.)

- Cain, W. (2024). Prompting change: exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, 68(1), 47-57. <https://doi.org/10.1007/s11528-023-00896-0>
- Davis, E. (2024). Mathematics, word problems, common sense, and artificial intelligence. *Bulletin of the American Mathematical Society*, 61(2), 287-303. <https://doi.org/10.1090/bull/1828>
- Diprose, W., & Buist, N. (2016). Artificial intelligence in medicine: humans need not apply? *The New Zealand Medical Journal (Online)*, 129(1434), 73-76.
- Edwards, B. I., & Cheok, A. D. (2018). Why not robot teachers: artificial intelligence for addressing teacher shortage. *Applied Artificial Intelligence*, 32(4), 345-360. <https://doi.org/10.20944/preprints201712.0022.v1>
- Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Stephanie Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., Uhlig, S., & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI Chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19-23. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of Biomedical Engineering*, 51(12), 2629-2633. <https://doi.org/10.1007/s10439-023-03272-4>
- Gusukuma, L., Bart, A. C., Kafura, D., & Ernst, J. (2018, August). Misconception-driven feedback Results from an experimental study. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*, 160-168. ICER. <https://doi.org/10.1145/3230977.3231002>
- Heston, T. F., & Khun, C. (2023). Prompt engineering in medical education. *International Medical Education*, 2(3), 198-205. <https://doi.org/10.3390/ime2030019>
- Nazari, M., & Saadi, G. (2024). Developing effective prompts to improve communication with ChatGPT: a formula for higher education stakeholders. *Discover Education*, 3(1), 45. <https://doi.org/10.1007/s44217-024-00122-w>
- OpneAI ChatGPT (2024a). Hello GPT-4o. <https://openai.com/index/hello-gpt-4o/> (최종검색일: 2024.8.12.)
- OpneAI ChatGPT (2024b). GPT-4. <https://openai.com/index/gpt-4-research/> (최종검색일: 2024.8.12.)
- OpneAI ChatGPT (2024c). How can I access GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o mini? Retrieved August 16, 2024, from <https://help.openai.com/en/articles/7102672-how-can-i-access-gpt-4-gpt-4-turbo-gpt-4o-and-gpt-4o-mini>
- Plevris, V., Papazafeiropoulos, G., & Jiménez Rios, A. (2023). Chatbots put to the test in math and logic problems: A comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google bard. *AI*, 4(4), 949-969. <https://doi.org/10.3390/ai4040048>
- Reynolds, L., & McDonell, K. (2021, May). Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1-7). <https://doi.org/10.1145/3411763.3451760>
- Tyson, J. (2023). Shortcomings of ChatGPT. *Journal of Chemical Education*, 100(8), 3098-3101. <https://doi.org/10.1021/acs.jchemed.3c00361>
- Zhou, J., Ke, P., Qiu, X., Huang, M., & Zhang, J. (2023). ChatGPT: potential, prospects, and limitations. *Frontiers of Information Technology & Electronic Engineering*, 25, 1-6. <https://doi.org/10.1631/fitee.2300089>