

중학교 지리 수업에서의 AI 챗봇 활용과 학습 데이터 분석: 학생들의 질문 유형 및 인식을 중심으로*

이소영**

An Analysis of AI Chatbot Utilization and Learning Data in Middle School Geography Education: Examining Student's Queries and Perceptions*

Soyoung Lee**

요약 : 본 연구는 중학교 지리 수업에서의 챗봇 활용과 관련하여 학생들의 질문 내용과 인식을 분석함으로써, 챗봇의 활용 가능성을 제고하는 것을 목적으로 한다. 이를 위해 경기도 소재 중학교 학생들을 대상으로 질문 데이터를 수집·분석하였고, 설문조사를 통해 챗봇에 대한 학생들의 인식을 파악하였다. 연구의 주요 결과는 다음과 같다. 첫째, 질문 분석 결과, 대부분의 질문이 '지리적 사실의 기억'이나 '개념의 이해' 수준에 편중되어 있었다. 즉, 고차원적 사고보다는 저차원적 사고를 요하는 질문이 주를 이루었다. 둘째, 인식 조사 결과, 학생들은 시간과 장소의 제약 없이 챗봇에 질문할 수 있어 지리 학습에 유용하다고 평가하였다. 그러나 부족한 데이터 세트로 인해 챗봇이 답변을 제대로 하지 못하는 점에 대해서는 불편함을 느끼는 것으로 나타났다. 본 연구는 향후 지리과 AI 챗봇의 초기 데이터 구축 및 설계를 위한 실질적 토대 마련에 기여할 것으로 기대된다.
주요어 : 챗봇, AI 튜터, 학습분석, 텍스트 마이닝, 블룸의 신교육목표분류체계

Abstract : This study aims to enhance the applicability of chatbots in middle school geography education, by analyzing students' use of and perceptions on chatbots. To this end, the researcher collected middle school student's chatbot queries in Gyeonggi Province, and also conducted a questionnaire survey to understand their perceptions on the chatbot as a geography education tool. The main findings of the study are as follows. First, the vast majority of student queries are focused on 'remembering geographical facts' or 'understanding conceptual knowledge', both of which indicate that most student chatbot queries are based on lower-order thinking. Second, most students regard the chatbot as a useful tool for geographical learning as they are able to ask questions to the online regardless of time and space. However, many students have found it inconvenient. For they are not satisfied with the chatbot's inability to provide little to no informative response, which is resulting from insufficient in-built datasets. These findings are expected to contribute to improve training data construction for the chatbots and also provide useful insights for the direction of educational chatbot designs.

Key Words : Chatbot, AI-tutor, Learning analytics, Text mining, Bloom's revised taxonomy

*이 논문은 2023년도 서울대학교 사범대학의 지원을 받은 「멀티모달 AI Agent 기반 학습자 맞춤형 교수학습 지원 시스템 개발을 위한 데이터 구축」의 연구결과로 수행되었음.

**가톨릭관동대학교 지리교육과 조교수(Assistant Professor, Department of Geography Education, Catholic Kwandong University, sylee03051@cku.ac.kr)

I. 연구배경 및 목적

오늘날 급속한 기술의 발전으로 인하여 사회는 빠르게 변화하고 있으며, 학교 현장도 이러한 변화의 흐름에서 예외는 아니다. 실제로 에듀테크 등 다양한 테크놀로지를 접목한 수업이 강조되고 있으며, 그러한 경향성은 코로나 19 이후 더욱 가속화되었다. 이에 따라 2022 개정 교육과정에서는 디지털 전환 및 미래 사회 변화에 대응할 수 있는 인재 양성을 목표로 삼고 있으며, 디지털 소양의 함양을 강조하고 있다(교육부, 2022a). 사회과 교육과정에서도 다양한 자료와 테크놀로지를 활용하여 정보를 수집, 해석, 활용, 창조할 수 있는 '정보활용능력'의 함양을 도모하고 있다(교육부, 2022b).

또한, 교육부는 2022 개정 교육과정에 따라 2025년부터 수학, 영어, 정보, 국어(특수교육) 교과에 AI 디지털교과서를 도입하고, 2028년까지 국어, 사회, 과학, 기술·가정 등 과목에 대해서도 단계적으로 확대 적용할 계획을 밝혔다(교육부, 2023b). 이에 따라 중학교 사회는 2027년, 고등학교 통합사회는 2028년부터 AI 디지털교과서가 도입될 예정이다.

AI 디지털교과서란, "인공지능을 포함한 지능정보기술을 활용한 다양한 학습자료 및 학습지원 기능을 탑재한 교과서"로, 학습자의 학습 진단 및 분석이 가능하다는 특징이 있다(교육부, 2023b). 여기서 학습분석(learning analytics)이란, 학습자와 학습 상황에 대한 데이터를 측정, 수집, 분석, 보고하는 것으로, 주로 교육의 효과성 증진 및 교수-학습 개선을 목적으로 한다. 이러한 학습분석은 컴퓨터 지원 학습 및 AI 기술의 발달 등에 따라 급속도로 성장하고 있는 학문 분야이다. 이때, 학습분석을 위한 학습 데이터는 학생의 인지 및 학습 과정을 파악할 수 있는 다양한 데이터를 의미한다. 학습 데이터를 수집·분석하는 플랫폼 역시 AI 디지털교과서(AI Digital Textbook, 이하 AIDT), 학습 관리 시스템(Learning Management System, 이하 LMS) 등 다양한 형태로 존재할 수 있다. 챗봇 역시 학습 데이터 수집을 위한 도구로 기능할 수 있으며(Hamzah *et al.*, 2021; Varnavsky and Popov, 2022), 이를 통해 수집한 학생들의 질문 데이터 역시 학습 데이터로 활용될 수 있다.

챗봇을 통해 수집한 학생들의 질문 데이터를 분석하는 것은 두 가지 측면에서 의의를 가진다. 첫째, 학생들의 질문에는 교과에 대한 흥미, 궁금증, 어려움, 오개념 등 학습

과 관련된 많은 정보가 내포되어 있다(Maskill and de Jesus, 1997, 강현태·노석구, 2018에서 재인용). 그러한 측면에서 학생들의 인지과정을 이해하고, 교수-학습 과정을 개선하는 데 기여할 수 있다.

둘째, AI 챗봇 개발을 위해서는 학습(training)을 위한 데이터 세트의 구축이 필수적이다(Caldarini *et al.*, 2022; 이소영 등, 2023). 과학기술정보통신부·한국지능정보사회진흥원(2023)은 AI 서비스 개발에 있어서 데이터 구축에만 자원의 80%가 투입되며, 데이터 세트의 부족은 AI 도입 및 개발의 진입장벽으로 작용한다고 밝혔다. 이때, AI 기능에 따라 필요로 하는 데이터 세트는 상이할 수 있는데, 교육용 챗봇 개발을 위해서는 각 교과 특성에 맞는 데이터 세트의 구축이 필요하다. 그러나 교사들 역시 학생들의 질문을 정확하게 예측하기는 어려운 실정이다(황홍섭, 2022). 그러한 측면에서 본 연구의 결과는 AI 챗봇의 학습 데이터 구축에 기여하는 바가 매우 클 것으로 기대된다. 또한, 「AI 디지털교과서 개발 가이드라인」에서도 학생들의 질의에 즉각적인 답변을 제공하는 AI 튜터의 중요성을 강조하고 있는 만큼(교육부·한국교육학술정보원, 2023), 본 연구는 사회과 AIDT 개발의 방향성 설정에도 도움이 될 것으로 생각된다.

따라서 본 연구의 목적은 중학교 지리 수업에서 학생들이 챗봇을 통해 질의하는 내용과 그 유형을 분석함으로써, 교수-학습 과정 개선을 위한 근거를 제공하고, AI 챗봇 개발을 위한 실제적인 토대를 마련하는 데에 있다. 또한, 챗봇 활용에 대한 학생들의 인식을 조사함으로써, 학교 현장에서의 활용 가능성에 대해 논의하고자 한다. 마지막으로, 본 연구에서 사용된 챗봇과 ChatGPT와의 비교를 통해 향후 AI 챗봇의 설계 및 고도화 방안에 대해서도 제안하고자 한다.

II. 이론적 배경: 학습분석

학습분석(Learning Analytics)이란, 학습자와 학습 상황에 대한 데이터를 측정(measurement), 수집(collection), 분석(analysis), 보고(reporting)하는 것을 의미한다(Long and Siemens, 2011; Nistor and Hernández-García, 2018; Varnavsky and Popov, 2022).

학습분석의 성장 배경은 컴퓨터 지원 학습(computer-supported learning)의 발전 및 LMS, 학습경험플랫폼

(Learning Experience platform, LXP) 등의 발달과 밀접한 관련이 있다. 즉, 온라인 학습 환경에서 다양한 플랫폼 등을 통해 이른바 다량의 학습 데이터를 수집하는 것이 학습 분석의 핵심이다(Long and Siemens, 2011; Nistor and Hernández-García, 2018). 학습분석의 개념적 모델은 그림 1과 같이 제시할 수 있는데(Nistor and Hernández-García, 2018), 모델에는 학습자의 인지과정이 데이터로 수집되고, 이를 분석하여 적절한 교육적 처치를 제공하는 주요 흐름이 잘 드러나 있다. Hamzah *et al.* (2021) 역시 학습분석은 도구, 기술, 적용 세 가지 주요 구성 요소로 이루어진다고 밝혔는데, 이때 도구는 학습분석에 사용되는 도구, 기술은 학습분석을 수행하는 방법, 적용은 그 결과로 교수-학습을 개선하는 것을 의미한다. 그들의 논의를 정리하면, 학습분석에서 가장 중요한 것은 ‘데이터’, ‘분석’, ‘적용’으로 정리할 수 있다.

우선, 데이터의 측면에서 살펴보았을 때, 데이터 수집 방법은 매우 다양할 것이다. 기존에는 LMS 등 데이터를 수집할 수 있는 플랫폼이 다소 제한적이었다면, 앞으로는 사물 인터넷, 웨어러블 디바이스, 가상현실, 챗봇, ADIT 등 데이터를 수집할 수 있는 도구(플랫폼)가 폭발적으로 증가할 것으로 예상된다(한국교육학술정보원, 2022).

학습 데이터의 유형 역시 아이 트래킹 데이터, 설문조사, 로그 데이터, 챗봇과의 대화 등 매우 다양하게 존재할 수 있다. 그리고 동일한 유형의 데이터라고 하더라도 목적에 따라 분석 및 적용 방법도 상이할 수 있다. 가령, 수집된 데이터를 실시간으로 분석하는 온라인 방식이 있는 반면(예: 대쉬보드 등으로 실시간 분석 제공), 모든 프로세스 종료 후 데이터 처리 및 분석이 이루어지는 오프라인 방식도 있다(Nistor and Hernández-García, 2018). 학습 분석의 구체적인 방법 역시 텍스트 마이닝, 네트워크 분석 등 다양하게 존재한다(Hoppe, 2017, 한국교육학술정

보원, 2022에서 재인용).

학습분석의 마지막 단계라고 할 수 있는 교육적 처치 및 적용 역시 다양한 목적으로 이뤄질 수 있는데, 주로 교수-학습 개선을 위해 사용되는 경우가 많다. 학습자에게는 학습에 대한 정보를 제공함으로써 자기주도적인 학습을 촉진할 수 있고, 교수자에게는 학습자의 학습 활동에 대한 데이터를 제공함으로써, 학생 맞춤형 처방을 제공하는 데에 도움을 줄 것이다. 이는 궁극적으로 다양한 학습 도구 및 콘텐츠 개발, 정책 결정 및 교육과정 개선까지 이어질 수 있을 것이다(안미리 등, 2016; Nistor and Hernández-García, 2018; Hamzah *et al.*, 2021; 한국교육학술정보원, 2022).

이러한 학습분석의 유용성에 따라 챗봇 역시 학습분석의 도구로서 활발히 논의되고 있다. Hamzah *et al.* (2021)은 챗봇을 활용한 학습분석 모델을 제시하고, 챗봇의 답변 성공률 및 학생들이 원하는 학습 콘텐츠를 파악하고자 하였으며, Varnavsky and Popov(2022) 역시 교육용 챗봇이 학습분석의 도구로 기능할 가능성에 대해 탐색하였다.

한편, 지리교육에서도 학습분석은 활발히 논의되고 있다. Papoušek *et al.* (2015)은 학생들에게 지리 지식(예: 국가나 도시의 이름 및 위치)을 질문하고, 응답 시간과 정답률을 분석하여 이를 반영한 온라인 적응형 시스템을 제안한 바 있다. Robinson *et al.* (2020)은 학습분석에서 학습의 ‘위치’도 매우 중요하다고 강조하며(Becker, 2013), 학습자가 학습하는 시공간 데이터를 수집하여 기존 ‘학습분석’을 ‘공간 학습분석(spatial learning analytics)’로 확장시키고자 하였다. Treves *et al.* (2015)은 아이트래킹 기술을 활용하여 가상 현장 학습(virtual field trip, VFT)에서 학생들의 움직임을 추적하고, 이를 시각화하여 학습에 대한 인사이트를 도출하고자 하였다.

이처럼 학습분석은 다양한 목적에 따라 다양한 도구를

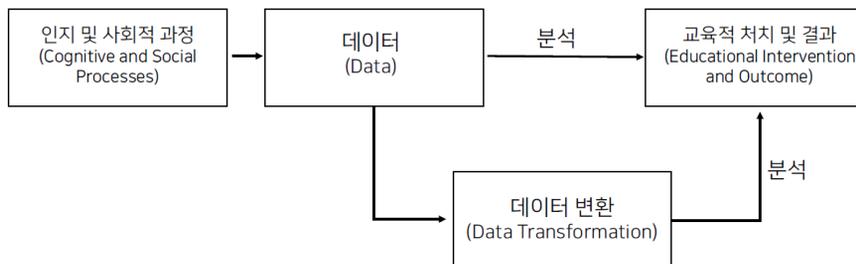


그림 1. 학습분석의 개념적 모델

출처 : Nistor and Hernández-García, 2018.

활용하여 데이터를 수집·분석하고 있다. 본 연구에서는 학습 데이터를 수집하는 도구로 ‘챗봇’을 활용하고, 학습 분석을 수행하는 방법으로 ‘텍스트 마이닝’기법을 활용하였다. 그리고 분석의 결과는 교수-학습 과정 개선 및 AI 챗봇 설계를 위한 데이터 세트 구축에 기여할 수 있을 것이다.

III. 연구 방법

1. 챗봇 개요 및 데이터 구축 현황

챗봇은 대화를 뜻하는 ‘챗(chat)’과 로봇을 의미하는 ‘봇(bot)’의 합성어로, 텍스트나 음성을 통해 인간과 상호작용하는 디지털 시스템을 뜻한다(Chocarro *et al.*, 2021; 이소영 등, 2023). 챗봇은 시간과 장소에 구애받지 않고 즉각적인 피드백을 제공할 수 있어 일종의 AI 튜터로서 기능할 수 있다. 최근에는 ChatGPT를 비롯한 다양한 챗봇들이 등장하고 있지만, 본 연구에서는 『멀티모달 AI Agent 기반 학습자 맞춤형 교수학습 지원 시스템 개발을 위한 데이터 구축』연구의 일환으로 개발된 챗봇을 활용하였다. 통상적으로 AI 기반 챗봇은 응답 생성 방식에 따라 검색 모델(retrieval-based)과 생성 모델(generative-based)으로 구분할 수 있는데(Li *et al.*, 2021), 본 연구에서 활용한 챗봇은 검색 모델에 해당된다. 검색 모델은 기존 학습된 데이터와의 유사도에 따라 적절한 응답을 산출하는 것이며, 생성 모델은 미리 정의된 응답에 의존하지 않

고, 학습된 데이터를 바탕으로 응답을 생성하는 방식으로(이현우·민덕기, 2018), OpenAI에서 출시한 ChatGPT(Generative Pre-trained Transformer)가 대표적인 예시이다.

따라서 본 연구에서 활용한 챗봇은 학생들의 질문에 대하여 미리 학습된 질의응답 데이터 세트 중 가장 유사한 답변을 산출하는 방식으로 작동한다. 초기에는 Doc2Vec 모델¹⁾을 기반으로 설계되었으며, 이후 2차년도 사업을 통해 엑소브레인 FAQ로 고도화되었다. 그리고 검색 모델의 경우, 특정 내용에 대한 학습 데이터가 부족한 경우, 적절한 대답의 산출이 어렵다는 한계가 있는데, 이를 보완하기 위하여 위키백과에서 답변을 불러오는 기능이 추가되었다.

해당 챗봇은 전술한 내용에서도 유추할 수 있듯이, 학습을 위한 다량의 데이터를 필요로 한다. 데이터 구축을 위한 방법으로는 크롤링, 직접 작성 등 다양한 방법이 있을 수 있는데, 본 연구에서는 중학교 『사회 1, 2』교과서 8종 및 사회와 교육과정의 내용체계를 참고하여 교과교육 전문가 및 교사가 직접 작성하는 방법을 선택하였다. 데이터 구축 작업은 약 2년에 걸쳐 이루어졌다. 1차년도(2022년) 사업을 통해 중학교 『사회』지리 영역에 대하여 5,700건, 2차년도(2023년) 사업을 통해 일반사회 영역에 대하여 4,000건을 구축하였다. 이렇게 총 9,700건의 질의응답 쌍(question-answer pairs)을 제작하였으며, 예시는 그림 2와 같다.²⁾ 그렇게 구축된 데이터 세트를 개발 챗봇에 학습시켰고, 2023년 8~12월까지 경기도 소재 중학교 학생들을 대상으로 수업 등에 적용하였다. 그 과정에서 챗봇

순번	중학교/고등학교	학년	교과	질문	답변
149	중학교	1	지리	오스트레일리아의 울룰루는 어떻게 형성되었나요?	오스트레일리아의 울룰루는 수억년 전 바다에서 퇴적된 사암중 이 지각 변동에 의해 융기한 후, 오랜 세월 자갈 침식을 받아 단단한 암석의 일부만 남게 되어 형성되었습니다. 울룰루는 오스트레일리아 원주민이 가장 신성하게 여기는 장소입니다.
150	중학교	1	지리	미국의 그랜드 캐니언은 어떤 과정으로 형성되었나요?	그랜드 캐니언은 남고 평평한 지역이 솟아올라 만들어진 고원에 하천의 침식이 더해져 형성된 현곡입니다.
152	중학교	1	지리	화산 지형에는 어떤 것이 있나요?	용암 대지, 화구호, 용암 동굴 등이 화산 지형에 속합니다. 화산 지형은 지구 내부의 마그마가 지각의 갈라진 틈을 뚫고 분출하면서 만들어지는 지형입니다.
153	중학교	1	지리	세계에서 가장 높은 폭포는 어디인가요??	베네수엘라 볼리바르에 위치한 앙헬 폭포입니다. 아우안테푸이산 정상에서 수직 절벽 아래로 떨어져 내리는, 세계에서 가장 높은 폭포입니다.
161	중학교	1	지리	고산 기후지역이 일교차가 큰 이유는 무엇인가요?	고산 지역은 공기 밀도가 낮아 낮에는 대기가 빠르게 가열되고 밤에는 빠르게 냉각되기 때문입니다.
162	중학교	1	지리	페루의 원주민들이 망토를 두르고 모자를 쓰는 이유는 무엇인가요?	고산 지역은 일교차가 커서 기온이 낮아지는 밤에 망토를 보온용으로 사용합니다. 모자는 강한 자외선을 막아주는 구실을 합니다.
163	중학교	1	지리	열대 고산기후지역에 난방 시설이 없는 이유는 무엇인가요?	일 년 내내 불과 같은 온화한 기후가 나타나기 때문입니다.
164	중학교	1	지리	스위스에 목초 재육이 많은 이유는 무엇인가요?	알프스 산지에 목재가 많이 때문입니다.
165	중학교	1	지리	앵그르브 숲이 생태계에 중요한 이유는 무엇인가요?	앵그르브 숲은 열대 기후지역의 해안에 분포하는데 나무들이 들쭉 깊숙이 뿌리를 내리는 것이 특징입니다. 유기물이 많아 영양분이 풍부하기 때문에 수많은 어패류, 파충류, 양서류, 조류의 부화장이자 보육장 역할을 합니다. 또한 나무뿌리가 복잡하게 얽혀 있어 바다생물들이 살기에 안전한 곳이기도 합니다.

그림 2 구축 데이터 예시

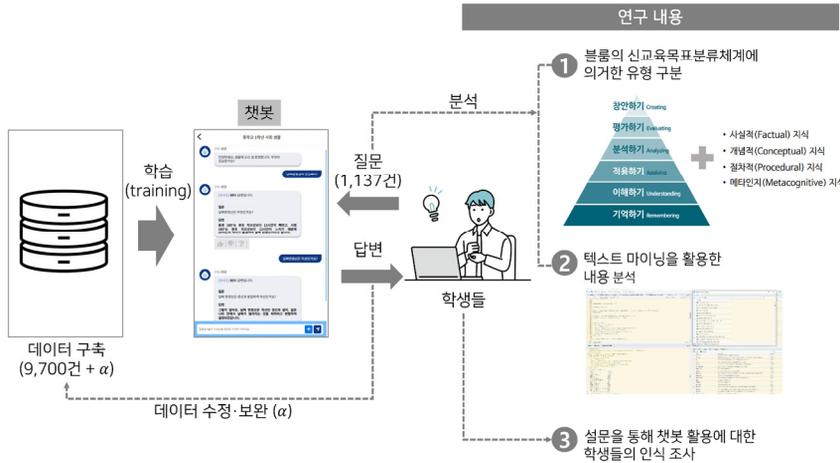


그림 3. 연구의 흐름

을 통해 학생들의 질문 데이터를 수집하였고, 본 연구에서는 그렇게 학생들로부터 수집한 1,137건의 데이터를 분석하고자 한다.³⁾ 이를 바탕으로 학생들이 어떤 질문을 하는지, 그리고 챗봇의 사용을 어떻게 인식하는지 등을 파악하고자 하였다. 본 연구의 흐름 및 주요 내용을 정리하면 그림 3과 같다.

2. 분석 방법

본 연구의 분석 대상은 전술한 바와 같이, 전체 데이터 세트가 아닌 학생들로부터 수집한 1,137건이다. 이에 따라 1,137건의 질문을 엑셀 파일로 정리 후, 사회과 교과 내용과 관련이 없는 질문(예: 안녕하세요, 선생님이나 친구의 이름을 넣은 장난 등)은 제외하였다. 최종적으로 1,060개의 데이터로 정리하였고, 이를 활용하여 질문의 유형과 내용을 분석하였다.

1) 질문 유형 분류: 블룸의 신교육목표분류체계 활용

먼저, 블룸(Bloom)의 신교육목표분류체계에 기초하여 학생들의 질문 유형을 분류하였다. 이는 블룸의 교육목표분류체계(Bloom *et al.*, 1956)를 앤더슨(Anderson)과 크레스호올(Krathwohl)이 수정·보완한 것으로, 6개의 인지과정과 함께 4개의 지식 차원으로 구성된다. 표 1과 같이 다양한 지식 유형(knowledge dimensions)과 인지과정(cognitive process)이 수준에 따라 어떻게 교차되는지 체계적으로 살펴볼 수 있다는 장점이 있다(Anderson *et*

al., 2005, 강현석 등 역, 2015; Wilson, 2016). 이러한 블룸의 신교육목표분류체계는 교육과정의 성취기준 혹은 교과서의 학습목표 등을 분석하는 데에 주로 활용된다(장연주, 2011; 최정인·백성혜, 2015; Şanlı, 2019). 최근에는 학생들이 ChatGPT에 질문한 내용을 인지과정 차원에 의거하여 6단계로 분류한 연구도 등장하였다(신화영·백성혜, 2024).

인지과정은 ‘기억하기’, ‘이해하기’, ‘적용하기’, ‘분석하기’, ‘평가하기’, ‘창안하기’와 같은 6개의 범주로 구분할 수 있다. ‘기억하기’의 경우, 이전에 학습하였던 사실이나 용어, 기본 개념 등을 기억에서 인출하고 회상하는 과정을 뜻하며, ‘이해하기’는 학생들이 새롭게 습득한 지식을 기존에 알고 있던 지식과 관련 짓는 것으로, 해석, 요약, 분류, 설명 등의 과정을 의미한다. ‘적용하기’는 배운 내용을 다른 상황에 적용해 보는 것으로, 시뮬레이션, 프레젠테이션 등과 관련이 있다. ‘분석하기’는 구성요소를 분해하고 요소 간 상호 관련성뿐만 아니라 각 요소와 전체 구조와의 관련성을 이해하고, 원인과 결과 등을 식별하는 과정을 의미한다. ‘평가하기’는 정보를 검토하여 판단하고, 평가하고, 비판하는 과정을 말하며, ‘창안하기’는 여러 정보를 종합하여 새로운 아이디어 등을 고안하거나 재구성하는 과정을 말한다(Anderson *et al.*, 2005, 강현석 등 역, 2015; Wilson, 2016; 신화영·백성혜, 2024).

지식의 수준 역시 네 개의 수준으로 구분할 수 있다(Anderson and Krathwohl, 2001; Kinder and Lambert, 2011; Wilson, 2016). 첫 번째, 사실적 지식(factual knowledge)은 학생들이 반드시 숙지해야 하는 기본적인

표 1. 블룸의 신교육목표분류체계(Bloom's Revised Taxonomy)

인지과정 차원		지식 차원	사실적 지식	개념적 지식	절차적 지식	메타인지 지식
저차	기억하기 (Remembering)	재인하기 (Recognizing) 회상하기 (Recalling)				
	↑ 이해하기 (Understanding)	해석하기 (Interpreting) 예증하기 (Exemplifying) 분류하기 (Classifying) 요약하기 (Summarizing) 추론하기 (Inferring) 비교하기 (Comparing) 설명하기 (Explaining)				
적용하기 (Applying)		집행하기 (Executing) 실행하기 (Implementing)				
분석하기 (Analyzing)		구별하기 (Differentiating) 조직하기 (Organizing) 귀속하기 (Attributing)				
		평가하기 (Evaluating)	점검하기 (Checking) 비판하기 (Critiquing)			
고차 창안하기 (Creating)		생성하기 (Generating) 계획하기 (Planning) 산출하기 (Producing)				

출처 : Anderson and Krathwohl, 2001; Wilson, 2016에서 수정.

요소들, 구체적인 사실, 전문적인 용어에 대한 지식의 미한다. 두 번째, 개념적 지식(conceptual knowledge)은 특정 학문 분야에서의 분류, 유목, 원리, 일반화에 대한 지식으로, 이론이나 모형, 구조 등에 대한 지식을 의미한다. 따라서 사실적 지식에 비하여 훨씬 일반적이고 추상적이다(강현석 등, 2015). 세 번째, 절차적 지식(procedural knowledge)은 방법, 일련의 절차나 단계에 대한 지식, 기법, 알고리즘 등에 대한 지식 혹은 정보를 의미한다. 마지막으로 메타인지 지식(Metacognitive knowledge)은 자신의 인지 전반에 대한 지식으로, 문제 해결 방법, 맥락과 조건을 포함한 인지과제에 대한 지식, 반성적 성찰 등이 포함된다.⁴⁾

본 연구에서는 학생들의 질문 내용(1,060건)을 표 1에서 제시한 블룸의 신교육목표분류체계에 의거하여 분류하였다. 분류작업에 있어서 질문의 예시는 장연주(2011), 강현석 등(2015), 최정인·백성혜(2015), 신화영·백성혜(2024) 등을 참고하였다. 질문 분류의 신뢰성을 높이기 위하여 연구자가 일정 기간을 두고 3회 반복을 통해 질문을 분석하였으며, 그 과정에서 판단하기 어려운 경우는 지리 교육전문가 1인에게 검토를 받아 교차 검증하였다.

2) 질문 내용 파악: 텍스트 마이닝 활용

다음으로, 학생들의 질문 내용 파악을 위해서는 텍스트 마이닝 기법을 활용하였다. 텍스트 마이닝은 데이터 마이닝(data mining)의 일종으로, 비정형화된 텍스트 자료로부터 유용하거나 가치 있는 정보를 추출하는 작업이라 할 수 있다(Nahm and Mooney, 2002; Feldman and Sanger, 2007; Gaikwad *et al.*, 2014). 이러한 장점에 따라 많은 연구에서 교과서, 교육과정, 학술지 논문 등 다양한 텍스트 자료를 수집하여 분석을 시도하였으며(이동민, 2017; 김재우, 2020; 하은혜 등, 2022), Cho *et al.* (2023)은 사회과 교육과정 공청회의 유튜브 댓글을 수집·분석하여 사회과 교육과정 개정과 관련된 주요 쟁점을 도출하기도 하였다. 그리고 이러한 데이터 마이닝 기법은 다양한 플랫폼으로부터 수집한 ‘학습 데이터 분석’의 방법으로도 주목받고 있다(Hoppe, 2017; 한국교육학술정보원, 2022).

데이터 정제 및 분석을 위해서는 R 프로그램의 KoNLP (Korean Natural Language Processing) 패키지를 활용하였다. 형태소 분석을 통해 명사와 형용사를 함께 추출하고, 그 과정에서 유사어나 띄어쓰기 통일 등 전처리 작업도 함께 거쳤다. 그 결과, 총 3,015개의 단어를 추출할 수

있었고, 이를 바탕으로 빈도분석을 실시하였다. 그리고 각 단어의 '동시출현(co-occurrence) 빈도'를 계산하여 단어 간 관계를 파악하였다. 동시출현이란, 하나의 질문에 두 단어가 함께 등장하는 것으로, 이를 통해 각 단어가 사용된 맥락을 파악할 수 있다(김영우, 2021). 그리고 그렇게 추출된 동시출현빈도 행렬을 바탕으로 단어 간 '연결 중심성(degree centrality)'을 계산하였다. 연결 중심성 산출은 tidygraph 패키지의 centrality_degree() 함수를 활용하였으며, 네트워크 그래프 시각화를 위해서는 ggraph 패키지를 활용하였다. 연결 중심성은 노드가 다른 노드들과 얼마나 밀접하게 연결되었는지를 나타낸 값으로, 네트워크 그래프로 표현하면 각 단어 간의 관계를 더욱 분명하게 파악할 수 있다(김영우, 2021).

3) 챗봇 활용에 대한 학생들의 인식: 설문 조사

마지막으로 챗봇 활용에 대한 학생들의 인식을 살펴보기 위하여 설문조사를 실시하였다. 문항 구성은 Topal *et al.* (2021)의 인터뷰 문항을 기본으로 하되, 기타 문항(Huang *et al.*, 2019; 황요한·이혜진, 2021)을 참고하여 수정·보완하였다. 설문조사는 접근의 용이성 및 편이성을 고려하여 구글폼으로 실시하였다.

IV. 연구 결과

1. 질문 유형 분석

학생들의 질문을 표 1에서 제시한 분류체계로 분류한 결과, 지식 차원에서는 사실적 지식에 대한 질문이 가장 많은 것으로 드러났다. 전체 1,060건의 질문 중 사실적 지식에 대한 질문이 583건으로 55%를 차지하였다. 사실적

지식은 기본적인 요소, 구체적인 사실, 전문적인 용어에 대한 지식 등을 의미하는데, 지리학에서는 지역의 위치, 인구, 면적, 수도 등과 같은 내용이 사실적 지식에 해당한다. 학생들은 특정 학교급 혹은 교육과정과 무관한 사실적 지식에 대해 질문하는 경우가 많았다. 예를 들어 '세계에서 가장 섬이 많은 국가는 어디인지', '세계에서 영토가 가장 넓은 국가는 어디인지', '서울특별시의 인구는 몇 명인지', '의정부시의 인구밀도는 얼마나 되는지' 등에 대한 질문을 하였다.

다음으로 개념적 지식에 해당하는 질문도 43.7%로 상당히 많았는데, 학생들은 사회 교과에 나오는 각종 개념이나 이론, 원리 등에 대하여 질문하는 경향이 있었다. 질문의 예시로는 '위도나 경도의 개념은 무엇인지', '경도에 따른 시간 차이', '편서풍은 무엇'이며, '난류와 강수량의 관계는 어떠한지' 등이 있었다.

세 번째로 절차적 지식은 일련의 절차나 방법 등에 대한 지식인데, 이와 관련된 질문은 1.3%로 많지 않았다. 예를 들어 '이산화탄소 배출량을 줄일 수 있는 방법은 무엇이 있나요?', '석유를 어떻게 시추하나요?', '일본과의 독도 갈등을 해결할 수 있는 방법은 무엇이 있을까요?' 등이 있었다.

다음으로 인지과정 차원에서 어떠한 유형의 질문이 많았는지 살펴해보도록 하겠다. 인지과정 차원에서는 '이해하기' 수준의 질문이 47.5%로 가장 많았으며, 다음으로 '기억하기'와 관련된 질문이 40.5%로 많았다. 그 외에 적용, 분석 수준에 해당하는 질문은 각각 4.1%, 4.7%를 차지하였으며, 평가 및 창안은 각각 2.0%, 1.6%에 불과하였다. '기억하기'는 단순히 개념이나 사실을 확인하는 질문으로, '사실적 지식'에 대한 질문이 대다수를 차지하였다(82.4%). 예컨대, '우리나라 최초의 지도는 무엇인지', '영국의 수도는 어디인지', '대한민국의 3대 대통령은 누구인지', '고조선은 언제 건국되었는지' 등의 질문이 존재하였

표 2. 블룸의 신교육목표분류체계에 의거한 질문 유형 분류

인지과정	지식		개념적 지식(B)		절차적 지식(C)		총합(D)	
	사실적 지식(A)		개념적 지식(B)		절차적 지식(C)		총합(D)	
	건수	(A/D)	건수	(B/D)	건수	(C/D)	건수	(D/E)
기억하기	351	82.4	74	17.4	1	0.2	426	40.2
이해하기	207	41.2	290	57.7	6	1.2	503	47.5
적용하기	4	9.3	35	81.4	4	9.3	43	4.1
분석하기	11	22.0	39	78.0	-	0.0	50	4.7
평가하기	6	28.6	15	71.4	-	0.0	21	2.0
창안하기	4	23.5	10	58.8	3	17.6	17	1.6
총합계(E)	583	55.0	463	43.7	14	1.3	1,060	100.0

다. 그 밖에도 ‘해리는 무엇인지’, ‘도심은 무엇인지’, ‘특레시는 무엇인지’ 등 전문용어에 대한 정의 및 개념에 대한 질문도 상당 부분 차지하였다.

인지과정의 6수준 중 가장 많은 부분을 차지하였던 ‘이해하기’와 관련된 질문으로는 ‘우리나라와 미국은 몇 시간차인지’, ‘위도에 따른 기온 차이’, ‘이동식 화전 농업을 하는 이유는 무엇인지’ 등 원리나 이론, 모형 등과 관련된 설명, 해석 등이 주를 이루었다. 즉, ‘이해하기’ 사고와 관련된 질문으로는 개념적 지식에 대한 질문이 많은 것을 확인할 수 있었다(57.7%).

학생들 질문의 대다수가 ‘기억하기’와 ‘이해하기’에 해당하는 저차적 사고와 관련된 질문이었으며(87.7%), 그보다 고차적 사고를 촉진하는 나머지 인지과정과 관련된 질문은 많지 않았다. 예컨대, ‘적용하기’는 ‘학습 내용을 새로운 상황에 적용’ 하는 것 혹은 ‘연습이나 문제 해결을 위해 여러 절차를 활용’하는 것이 될 수 있는데, 이와 관련된 질문은 4.1%에 불과하였다. 예시로는 ‘이산화탄소 배출량을 줄이기 위한 방법은 무엇이 있을까?’ 혹은 ‘쿠웨이트의 경우 물 수요량의 100%를 담수화 시설에 의존하는데, 시설이 부서지면 어떻게 될까요?’ 등의 질문이 있었다.

다음 ‘분석하기’는 ‘정보를 구성요소로 분해하고, 그들 간의 관계 혹은 원인과 결과를 식별하는 것’과 관련된 질문으로 분류하였다. 이는 전체 질문 중 4.7%에 해당하였고, 예시로는 ‘아프리카에는 우리나라보다 먼저 독립한

나라들도 많은데 왜 선진국이 없나요?’ 등이 있다.

‘평가하기’와 관련된 질문은 준거에 따른 판단, 비판 등과 관련된 것으로, 전체 질문의 2% 를 차지하였다. 관련 질문으로는 ‘저출산 현상이 나타나면 좋은 점은 무엇이 있는지’, ‘다문화적 변화가 일어나면 좋은 점은 무엇이고, 안 좋은 점은 무엇인지’, ‘국제하천은 상류에 있는 국가가 좋은지, 하류국가가 더 좋은지’ 등의 질문이 있었다.

마지막 ‘창안하기’는 가장 고차적인 사고와 관련이 있지만, 이는 전체 질문 중 1.6%로 가장 적은 비중을 차지하였다. ‘창안하기’는 답이 정해져 있는 수렴적 사고와 달리 다양한 답이 존재할 수 있는 확산적 사고와 관련된 질문으로, 창의적인 질문을 ‘창안하기’로 분류하였다. 예컨대, ‘페루는 알파카 털로 옷을 만드는데, 알파카가 멸종당하면 무엇으로 옷을 만들어요?’, ‘혹시 미래에는 새로운 종류의 자연재해가 생길 수 있을까요?’, ‘모든 자원이 고갈되면 어떻게 될까요?’ 등의 질문이 존재하였다.

2. 질문 내용 분석

다음으로는 텍스트 마이닝 기법을 통해 학생들이 어떤 내용의 질문을 많이 하였는지 살펴보고자 한다. 먼저, 학생들의 질문에서 단어를 추출하여 어떤 단어가 가장 많이 등장하였는지 출현 빈도를 계산하였다. 출현 빈도가 가장 높은 상위 30위까지의 단어를 정리하면 그림 4와 같다.

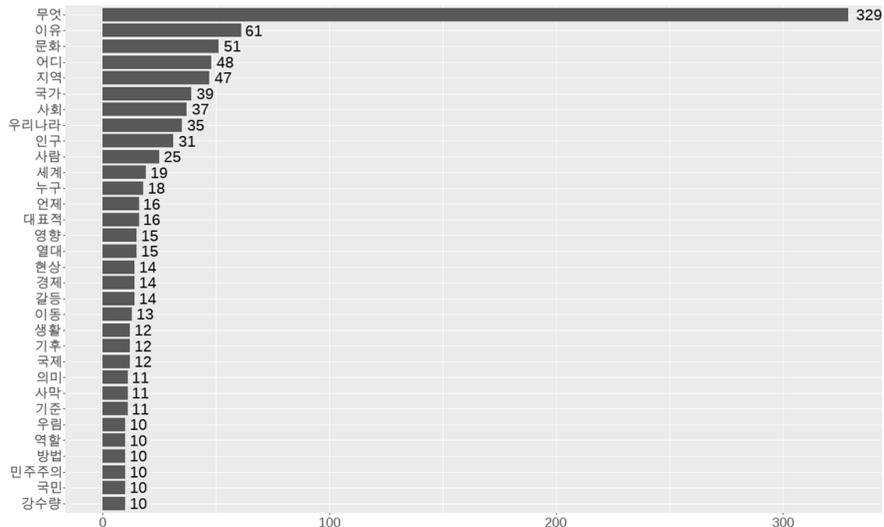


그림 4. 출현 빈도 상위 30개 키워드

* 상위 30위까지 제시하였으나, 출현 빈도가 동일한 경우가 있어 총 개수는 30개를 초과함.

앞선 분석에서도 사실이나 개념에 대한 질문이 87.7%로 압도적으로 많았는데, 텍스트 마이닝을 통해서도 그와 같은 경향성을 다시 한번 확인할 수 있었다. 학생들의 질문에서 가장 많이 출현한 단어는 ‘무엇’으로, 총 329회 등장하였다. 학생들은 사실이나 개념과 관련된 질문을 ‘~은 무엇입니까?’ 혹은 ‘~은 무엇인가요?’ 등으로 끝나는 질문을 많이 하는 경향이 있었다. 따라서 ‘무엇’이 포함된 질문을 중점적으로 살펴보기 위해 ‘무엇’과 동시에 한 문장에서 동시에 출현한 단어들을 표 3과 같이 추출해보았다. 그 결과, ‘무엇’과 가장 동시에 많이 등장한 단어는 총 50회 함께 등장한 ‘이유’였다. 즉, 학생들은 ‘무엇’에 해당하는 개념뿐만 아니라 어떤 현상이 발생하는 이유 혹은 원리에 대해서도 많은 관심을 가지고 있음을 확인할 수 있었다. 그리고 단어 추출 시, 두 글자 이상만 추출하는 것으로 설정하였기에 별도의 단어로 추출되지는 않았지만, 질문 분석 결과 ‘왜’에 대한 질문도 상당수 존재하였다. ‘왜’에 대한 질문의 예시로는 ‘왜 동해에는 갯벌이 안 만들어지나요?’, ‘이슬람교도들은 돼지고기를 왜 금기시 하나요?’ 등이 있었다(표 4).

그리고 지리를 비롯한 사회과 전반에서 환경 혹은 사회의 변화, 역사적 사건 등이 우리 생활에 미친 영향 등을 파악하는 데에 주안점이 있다고 할 수 있는데, ‘영향’에 대한 질문도 ‘무엇’과 함께 11건 추출되었다. 예시로는 ‘종교는 개인의 신앙을 넘어 지역에 어떤 영향을 줄 수 있는지’, ‘경제성장이 우리 삶에 미치는 부정적인 영향은 무엇이 있는지’ 등이 있다.

아울러 사회과에서는 특정 현상 등에 대한 사례 역시 무척 중요한데, ‘대표적’ 역시 16건으로 높은 빈도로 출현하였음을 확인할 수 있었다. 예시로는 ‘대표적인 인구 유입 지역’, ‘우리나라로 인구 유출되는 대표적인 국가 5곳’, ‘지역 이기주의의 대표적인 예시는 무엇인가요?’ 등이 있었다.

다음으로는 각 단어 간 연결성을 살펴보기 위하여 연결 중심성을 산출하였다(표 5). 또한, 단어 간의 관계 파악을 위하여 동시출현이 4회 이상인 경우만 추출하여 그림 5와 같이 네트워크 그래프를 시각화하였다. 그림 5에서 노드(원)의 크기가 클수록 연결 중심성이 높고, 더 많은 단어들과 밀접하게 연결되어 있다고 해석할 수 있다(김재우, 2020; Cho *et al.*, 2023). 그리고 R의 tidygraph 패키지에서

표 3. ‘무엇’과 동시출현(co-occurrence) 단어 분석

	동시 출현 쌍		빈도
1	무엇	이유	50
2	무엇	사회	16
3	무엇	문화	15
4	무엇	경제	13
5	무엇	영향	11
6	무엇	대표적	10
7	무엇	국가	9
8	무엇	우리나라	9
9	무엇	지역	9
10	무엇	의미	9

표 4. ‘무엇’과 ‘이유’에 대한 질문 예시

영역	무엇	이유(왜)
지리	<ul style="list-style-type: none"> • 자원의 희소성이란 무엇인가요? • 다국적 기업의 특징은 무엇인가요? • 위도란 무엇인가요? • 행정구역은 무엇인가요? • GDP와 GNP의 차이점은 무엇인가요? 	<ul style="list-style-type: none"> • 지진이 일어나는 이유 • 지역마다 시간이 다른 이유는 무엇인가요? • 유목 생활을 하는 이유는 무엇인가요? • 왜 동해에는 갯벌이 안 만들어지나요? • 이슬람교도들은 돼지고기를 왜 금기시 하나요? • 오아시스는 왜 생기나요? • 왜 하천을 직선으로 만드나요?

기하는 것을 넘어 '중요성'을 이해하고, 국가의 영역이 가지는 의미 등을 보다 심층적으로 이해하고자 한다는 것을 알 수 있었다.

마지막으로 연결 중심성이 높은 단어들을 추려보았을 때, '지역', '우리나라', '세계(국제, 전세계)' 등의 단어가 등장하는데, 이는 학생들이 다양한 현상에 대해 특정 스케일에만 집중하는 것이 아니라, 다양한 스케일에서 현상과 원인을 함께 탐구하고자 한다는 것으로도 해석할 수 있다. 지리학에서 다중 스케일적 관점을 견지하는 것은 매우 중요한데, 이를 통해 학생들에게서도 그러한 태도를 엿볼 수 있었다.

그리고 학생들의 질문 중 논리적 오류가 포함된 질문도 일부 발견할 수 있었다. 학습분석학의 관점에서 학습 데이터는 그들의 인지과정 및 오개념을 파악하는 데에도 유용하게 사용될 수 있는데, 본 연구에서도 학생들의 질문 분석을 통해 오개념을 확인할 수 있었다.

“적도 부근은 왜 해발고도가 낮을까요?”
 “해발고도가 높으면 왜 평활한 땅이 없을까요?”

첫 번째 사례는 ‘적도 부근은 왜 해발고도가 낮을까요?’라는 질문으로, 이 질문을 한 학생은 적도 부근은 해발고도가 낮다는 잘못된 전제를 가지고 있었다. 두 번째 사례 역시, ‘해발고도가 높으면 왜 평활한 땅이 없을까요?’라는 질문으로, 해발고도가 높은 곳에는 고원과 같은 평탄한 지형이 나타나지 않는다는 가정하에 그 이유에 대해 질의하였다. 이처럼 교사는 챗봇의 질문 데이터 분석을 통해 학생들이 가진 오개념을 파악하고, 그에 걸맞은 피드백 제공하거나 교수-학습 과정을 개선할 수 있다.

3. 챗봇 활용에 대한 학생들의 인식

마지막으로 챗봇 활용에 대한 학생들의 인식을 살펴보기 위해서 설문조사를 실시하였다. 설문은 본 연구에서 활용된 챗봇을 사용한 학생들 중 동의한 학생들에 한하여 진행하였으며, 학생들이 쉽게 접근할 수 있도록 구글폼을 활용하였다. 총 21명의 학생이 설문에 응답하였고, 선택형 문항의 결과는 그림 6과 같다. 각 문항은 리커트 5점 척도로 이루어졌으며, 문항별 평균 비고를 위하여 ‘전혀 그렇지 않다’, ‘그렇지 않다’, ‘보통이다’, ‘그렇다’, ‘매우 그렇다’ 응답에 대해서 각각 1~5점으로 코딩하였다.

설문 분석 결과, 학생들은 챗봇이 사회 공부에 유용하다고 느끼는 것으로 나타났다. 이는 서술형 문항을 통해서도 확인할 수 있는데, 학생들은 ‘모르는 것을 물어볼 수 있어서 좋은 것 같다’, ‘교사에게 질문하기 어려울 때 챗봇으로 질문하니 마음이 편했다’ 등의 답변을 주었다. 특히, 교사에게 직접 질문하기 어려운 상황(주말, 학교가 아닌 곳)에서도 챗봇을 통해 궁금한 점을 해결할 수 있다는 점을 장점으로 꼽고 있었다. 챗봇의 가장 큰 장점 중 하나가 학생들이 처한 상황과 학습 스타일 유형과 무관하게 개인화된 학습을 지원할 수 있다는 점인데(Gupta and Chen, 2022), 그러한 점이 학생들의 설문 결과에도 잘 드러남을 확인할 수 있었다.

또한, 학생들은 챗봇을 사용하는 것 역시 흥미롭게 느끼는 것으로 나타났다. 이는 선행연구에서의 결과와도 일치하는 것으로(양혜진 등, 2019; 추성엽·민덕기, 2019), 학생들은 서술형 문항을 통해서도 ‘채팅’ 형태로 질의하는 것이 재미있다고 밝혔다.

그리고 본 연구에서 개발한 챗봇은 직관적인 인터페이스로 학생들이 어려움 없이 질문할 수 있도록 설계되어 있

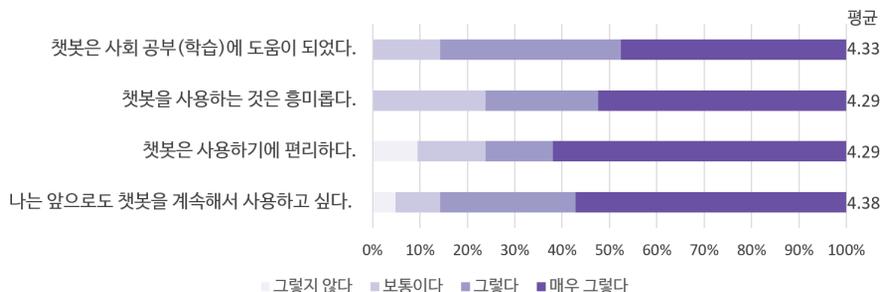


그림 6. 챗봇 활용에 대한 학생들의 인식(선택형 문항 결과)

* 모든 문항에 대해서 '매우 그렇지 않다'를 선택한 학생은 존재하지 않음.

는데, 이와 관련하여 많은 학생들이 ‘사용하기 편리하다’고 응답하였다. 서술형 문항을 통해서도 타 프로그램과 달리 한글로 구성되어 있어 사용하기 편리하다는 의견을 남겼다. 마지막으로 앞으로도 챗봇을 계속 사용하고 싶은지 의사를 확인하는 것이 중요한데, 이 부분에 대해서도 많은 학생들이 긍정적으로 답변하여 가장 높은 점수가 도출되었다.

반면, 한계점 및 개선이 필요한 지점도 분명히 존재하고 있었다. 챗봇 관련 연구에서 자주 언급되는 매우 중요한 문제 중 하나가 학습 데이터가 부족하다는 것인데 (Kuhail *et al.*, 2022), 본 연구에서 사용한 개발 챗봇 역시 예외는 아니었다. 특히, 본 연구에서 사용된 챗봇은 검색기반 챗봇이므로, 챗봇이 정상적으로 작동하기 위해서는 매우 많은 양의 데이터가 필요하고, 데이터가 부족하면 챗봇은 학습자의 질문에 적절한 응답을 할 수 없다(이소영 등, 2023). 1차년도에도 유사한 문제가 발생하였고, 2차년도에 데이터를 보완하였음에도 불구하고 동일한 문제를 겪고 있었다. 학생들은 챗봇이 적절한 답변을 해주지 못하는 상황에서 가장 큰 불편함을 느끼고 있는 것으로 드러났다. 그 밖에도 로그인 유지가 되지 않는 점 등 기술적인 측면과 관련하여 불편함을 느끼고 있었다. 학생들의 의견을 바탕으로 본 연구에서 활용된 챗봇의 장점 및 한계점을 정리하면 표 6과 같다.

4. 논의: ChatGPT와의 비교

전술한 바와 같이, 챗봇은 검색기반 모델과 생성기반 모델로 구분될 수 있는데, 본 연구에서 사용한 검색기반 모델 챗봇은 데이터에 의존적이므로 챗봇의 대화주제가

한정적이고, 데이터가 구축되어 있지 않으면 답변을 할 수 없다는 한계가 있다. 반면, ChatGPT와 같은 생성기반 모델은 어떤 질문에도 자연스럽게 답변을 제공한다는 장점이 있다. 이에 따라 본 연구에서 활용된 챗봇이 답변하지 못했던 일부 질문을 ChatGPT에게 질의하고, 비교·분석을 통해 향후 AI 챗봇의 고도화 및 설계 방안을 논의하고자 한다.

저자는 2024년 2월 14일, 10월 18일 두 차례 ‘한국에서 놀러 갈만한 석회동굴을 추천’해달라는 질문을 하였고, 그림 7은 해당 질문에 대한 ChatGPT의 답변이다. 2월 14일에 ChatGPT는 해당 질문에 대한 답변으로 ‘가사동굴’, ‘대진동굴’, ‘인제군굴’, ‘물아래동굴’을 석회동굴로 소개하였으나, 이 동굴들은 모두 실제 존재하지 않는 동굴이었다. 이는 많은 선행연구에서 다루고 있는 ChatGPT의 환각(hallucination) 현상을 확인할 수 있는 사례이다(Wilby and Esson, 2023; Zhou *et al.*, 2023). 이후 10월 18일에도 동일한 내용에 대해 질의하였는데, 여전히 잘못된 정보를 일부 포함하고 있었지만, 정확도가 향상된 것을 확인할 수 있었다(그림 7B).⁵⁾ 10월 18일 답변에서 ChatGPT가 제시한 ‘광명동굴’은 석회동굴이 아닌, 폐광을 관광지화한 것이며(한국관광 100선 광명동굴), ‘대이리 동굴지대’는 ‘강원도 평창군’이 아닌 ‘삼척시’에 위치하고 있다(한국민족문화대백과사전).

이를 통해 본 연구에서 다룬 챗봇과 ChatGPT의 장단점을 명확하게 확인할 수 있다. 본 연구에서 사용된 검색기반 챗봇은 학생들의 질문에 대해 정확한 답변을 제공하는 반면, 데이터 세트가 구축되지 않은 질문에 대해서는 답변을 하지 못한다. 반면, ChatGPT와 같은 생성기반 모델은 어떤 질문에도 유창한 답변이 가능하지만, 이른바 ‘환

표 6. 본 연구에서 활용된 챗봇에 대한 학생들의 의견

장점	한계점
<ul style="list-style-type: none"> • 선생님께 질문하기 어려울 때가 많았는데, 챗봇으로 질문하니 마음이 편했다. • 불편하게 인터넷 안에서 세세한 것들을 찾아낼 필요 없이 물어보면 세세히 답해주는 것이 좋았다. • 혼자 하는 것보다 편리해서 좋다 • 필요할 때 빠르고 쉽게 쓸 수 있어서 좋은 것 같다. • 답변의 정확도가 높고 채팅 형태인 게 재미있다. • 이거 공부할 때 되게 좋을 듯해요 • 사회 학습에 교과서만큼 큰 도움이 되었고, 타 과목에서도 사용할 수 있게 되면 좋을 것 같습니다. 	<ul style="list-style-type: none"> • 내 질문에 대해 정확한 답변이 안 나오는 경우도 생겨서 조금 아쉬운 부분이다. • 가끔씩 엉뚱한 대답을 한다. • 활화산이 이 세계에 몇 개 정도가 있냐 물어봤을 때 답을 안 해줘서 궁금한 게 풀리지 않았습니니다.
	<p>개선이 필요한 부분</p> <ul style="list-style-type: none"> • 모든 물음에 대해 다 대답해 줬으면 좋겠다. • 빨리 학습을 통해 발전하면 좋겠다. • 더 정확한 답을 해줬으면 좋겠다.

A	<p> ChatGPT</p> <p>한국에는 다양한 석회동굴이 있습니다. 명승지로 유명하며 관광객들에게 좋은 경험을 제공하는 몇 군데를 소개합니다:</p> <ol style="list-style-type: none"> 가사동굴 (Gasa Cave): 전남 보성에 위치한 가사동굴은 국내에서 가장 큰 석회동굴로 알려져 있습니다. 산속에 자리하고 있어 아름다운 경관과 함께 방문객들에게 다양한 동굴 속 이국적인 모습을 선사합니다. 대진동굴 (Daejin Cave): 강원도 정선에 위치한 대진동굴은 국내에서도 크기가 큰 편에 속합니다. 동굴 내부에는 다양한 동굴 생물과 석순이 그려져 있어 관광객들에게 신비로운 경험을 선사합니다. 인제군굴 (Inje Cave): 강원도 인제에 위치한 인제군굴은 국내에서 가장 규모가 큰 인공동굴 중 하나입니다. 동굴 내부는 조명과 설치된 보행로로 안전하게 관광할 수 있으며, 석순과 석채 등이 아름다운 모습을 이루고 있습니다. 몰아래동굴 (Moolahre Cave): 경기도 동부지역에 위치한 몰아래동굴은 동굴 내부에서 수중동굴을 경험할 수 있는 독특한 코스를 제공합니다. 동굴 내부의 수면에 반사된 석순과 석채는 아름다운 풍경을 이루고 있습니다.
B	<p> 한국에서 방문할 만한 석회동굴은 자연의 경이로움을 체험할 수 있는 장소로, 아름다운鍾유석과 석순을 감상할 수 있는 명소들입니다. 아래 몇 가지 추천 동굴을 소개합니다.</p> <ol style="list-style-type: none"> 광명동굴 (경기도 광명시) 광명동굴은 과거의 폐광을 관광지로 개발한 곳으로, 다양한 전시와 문화 행사를 통해 동굴의 신비로움을 더해줍니다. 와인 저장고, 미디어 아트 전시 등 다양한 볼거리도 함께 제공되며, 가족 여행이나 데이트 코스로도 인기가 많습니다. 백룡동굴 (강원도 평창군) 백룡동굴은 동굴 내부에 조명을 설치하지 않은 자연 그대로의 상태를 유지하는 동굴로, 자연 보호를 위해 안내원의 도움을 받아 체험하는 체험형 동굴입니다. 안전 장비를 착용하고 체험하는 과정이 독특하며, 자연 석회 동굴의 진수를 느낄 수 있습니다. 환선굴 (강원도 삼척시) 환선굴은 한국에서 가장 큰 규모의 석회 동굴로, 천연기념물로 지정된 동굴입니다. 웅장한鍾유석과 다양한 형태의 석회암 지형을 볼 수 있으며, 동굴 내부를 산책하며 동굴의 다양한 석순과 동굴 천장을 감상할 수 있습니다. 고수동굴 (충청북도 단양군) 고수동굴은 약 5억년의 역사를 가진 석회 동굴로, 다양한 형태의鍾유석과 석순을 볼 수 있는 관광지입니다. 비교적 쉽게 접근할 수 있어 가족 단위 관광객에게 적합하며, 동굴의 경관이 아름다워 사진 찍기 좋은 장소로도 유명합니다. 대이리 동굴지대 (강원도 평창군) 대이리 동굴지대는 백룡동굴을 포함한 여러 석회 동굴이 모여 있는 지역으로, 자연 상태가 잘 보존된 동굴들이 많습니다. 체험형 동굴 체험을 좋아하는 사람들에게 추천할 만한 장소입니다.

그림 7. 학생 질문에 대한 ChatGPT의 답변 비교

* A: 2024년 2월 14일 저자 검색, B: 2024년 10월 18일 저자 검색

각 현상에 대한 우려가 여전히 존재한다. 본 연구 결과, 학생들은 지리적 사실이나 개념에 대한 질문을 주로 하는 경향이 있었는데, ChatGPT는 이러한 사실적 지식에 대해서는 중간 수준의 신뢰도를 가진다고 평가되고 있다(Zhou et al., 2023). Lee and Lee(2023) 역시 ChatGPT가 지역지리 혹은 위상구조에 취약함을 밝힌 바 있다.⁶⁾

이처럼 검색-기반 모델과 생성-기반 모델의 장단점이 명확하기에 최근에는 두 모델을 연계하여 챗봇을 개발하는 방안이 대두되고 있다(정천수, 2023). 본 연구에서도 각 모델의 한계점을 보완하고, 보다 높은 확률로 정확한 답변을 제공하기 위해서 그와 같은 방안을 제안하는 바이다.

V. 결론 및 제언

본 연구는 챗봇을 통해 수집한 학생 질의 데이터를 분석함으로써, 교수-학습 개선 및 챗봇 설계의 방향성을 제시하고자 하였다. 학생들의 질문은 bloom의 신교육목표분류 체계에 의거하여 분류되었고, 보다 상세한 내용 파악을 위하여 텍스트 마이닝 분석도 병행하였다. 그리고 챗봇 활용성 제고를 위해 학생들의 인식을 조사하고, ChatGPT와의 비교를 통해 챗봇 고도화 방안을 논의하였다.

연구의 주요 결과는 다음과 같다. 첫째, 학생들의 질문은 특정 유형에 편중되는 경향이 있었다. bloom의 신교육목표분류체계는 '지식 유형'과 '인지과정' 두 차원으로 구성되는데, 지식 차원에서는 '사실적 지식'과 '개념적 지식'에 질문의 98.7%가 편중되어 있었다. 인지과정 차원에서는 87.7%의 질문이 '기억하기'와 '이해하기' 수준에 해당하는 것으로 나타났다. 반면, '평가 및 창안 수준'에서의 질문은 매우 적은 것으로 드러났다. 이는 학생들의 질문이 고차적 사고 과정과 관련되기보다는 단순 사실적 지식의 확인이나 이해 수준에 머물러있음을 시사한다.

둘째, 텍스트 마이닝을 통한 분석 결과에서도 유사한 경향을 확인할 수 있었다. 예컨대, 학생들의 질문에서 최다 빈도로 등장한 단어는 '무엇'이었으며, 그 외에도 '이유', '영향', '대표적', '사례' 등이 자주 등장하였다. 즉, 학생들은 주로 개념의 정의, 특정 현상 발생의 이유나 원리, 그것이 미치는 영향력 등에 대해 질의하는 경향이 있었다. 또한, '지역', '우리나라', '세계와 같은 단어들이 높은 연결 중심성을 가지는 것으로 나타났는데, 이를 통해 학생들이 특정 스케일에 국한하지 않고, 다양한 스케일에서 지리적

현상을 탐구하고자 하는 경향이 있음을 확인할 수 있었다. 그리고 질문 중 일부는 논리적 오류도 포함하고 있었는데, 이를 통해 학생들의 인지과정에 내재된 오개념도 파악할 수 있었다.

셋째, 챗봇에 대한 인식 조사 결과, 대부분의 학생들이 챗봇 활용이 지리 학습에 유용하다고 응답하였다. 특히, 교사에게 질문하기 어려울 때 챗봇을 통해 궁금한 점을 해결할 수 있는 것을 장점으로 꼽고 있었는데, 이는 학생이 처한 상황과 학습 스타일과 무관하게 개별화된 학습을 지원할 수 있다는 챗봇의 장점(Gupta and Chen, 2022)이 잘 드러나는 부분이다. 그리고 학생들은 챗봇을 활용하는 것이 흥미롭다고 응답하였는데, 이는 많은 선행연구에서의 결과와 일맥상통한다(양혜진 등, 2019; 추성엽·민덕기, 2019; Okonkwo and Ade-Ibijola, 2021). 하지만, 본 연구에서 사용한 챗봇은 데이터 구축 및 기술적 한계로 인해 적절한 답변을 하지 못하는 상황이 종종 발생하였는데, 이를 해결하기 위한 방안으로 검색 기반 모델과 ChatGPT와 같은 생성형 AI의 연계를 제안하는 바이다. 두 모델이 연계될 때, 각 모델이 가지는 한계를 보완하여 정확성이 높으면서도 자연스러운 대답이 가능한 챗봇의 개발이 가능할 것이다. 더 나아가 사진, 그래프, 지도 등 시각 자료도 활용하는 방식으로 고도화된다면(이소영 등, 2023), AI 튜터로서의 기능을 충실히 이행할 수 있을 것으로 기대된다.

본 연구는 세 가지 측면에서 의의를 가진다. 첫째, 학생들이 지리 수업 시간에 궁금해하는 내용, 잘못 이해하고 있는 오개념 등을 파악함으로써, 교수-학습 과정을 효과적으로 개선할 수 있다는 점이다. 본 연구는 챗봇의 대화 내용이 학습 데이터로서, 기능할 수 있음을 분석한 시론적 연구로, 연구 결과는 교수-학습 과정 개선의 근거자료로 쓰일 수 있다.

둘째, AI 챗봇 구축을 위해서는 데이터 세트의 구축이 필수적인데, 학생들의 질의 내용을 분석함으로써 향후 개발될 지리과 AI 챗봇 설계 및 개발을 위한 초석을 마련하였다는 점이다. 데이터 세트 부족 시, 적절한 답변 제공이 어렵고, 교사들 역시 학생들의 질문을 정확히 예측하기 어렵다는 점은 선행연구에서도 잘 알려진 결과이다(황홍섭, 2022). 따라서 본 연구는 실제 학생들의 질의 데이터를 분석함으로써 데이터를 선제적으로 보완할 수 있다는 측면에서 실용적으로도 의의를 가진다.

셋째, 챗봇과 같은 AI 기술의 도입을 위해서는 구성원

들의 인식이 중요한데, 그중에서도 직접적인 교육 수요자인 학생들의 인식이 매우 중요하다. 그러한 측면에서 챗봇 사용에 대한 학생들의 인식을 직접 조사하고, 이를 바탕으로 고도화 방안을 제안한 본 연구는 학교 현장에서의 AI 챗봇 활용을 위한 기초 자료로 사용될 수 있을 것이다.

다만, 본 연구에서 학생들의 질문은 전반적으로 사실적 지식이나 개념 등 저차원적 질문에 집중된 경향이 있었는데, 이는 세 가지 가능성을 내포하고 있다. 첫 번째는 챗봇 구동을 위해 미리 구축된 데이터 세트의 상당수가 사실적 지식에 기반하고 있다는 점이다. 두 번째는 챗봇의 기술적 한계와 관련된 것으로, 본 연구에 사용된 챗봇은 앞서 언급한 바와 같이, 데이터 세트가 구축되지 않은 질문에 대해서는 답변을 하지 못한다. 이에 따라 학생들은 몇 번의 질문 경험을 통해 챗봇이 답변하지 못할 것이라 예상되는 질문은 자체적으로 하지 않았을 가능성도 존재한다. 세 번째는 학교 교육, 특히 한국 교육의 특성상 학생들이 스스로 고차원적 질문을 던져보는 연습이 부족했을 가능성이 있다. 신화영·백성혜(2024)는 과학 수업 시간에 학생들로 하여금 궁금한 내용을 ChatGPT에 질의하도록 하였는데, 이 연구에서도 학생들의 질문 대부분이 저차원적 사고에 해당하는 것으로 나타났다. 다만, 주목할 점은 ChatGPT를 긍정적으로 인식한 학생일수록 차시가 거듭될수록 고차원적 질의를 하는 경향을 보였다는 점이다. 이러한 선행연구의 결과는 우리나라 학생들이 교과를 불문하고 고차원적 질문을 능동적으로 구성하는 경험이 부족했음을 시사한다. 그러나 동시에 지속적으로 그러한 환경에 노출될 경우, 학생들 스스로 고차원적 질문을 하는 방법을 스스로 습득할 수 있다는 것을 의미하기도 한다. 실제로 본 연구에서도 한 학생은 챗봇에 질문하는 방법에 대한 교육이 필요함을 언급하였다. 또한, 한국교육학술정보원(2024) 역시 생성형 AI를 활용한 수업에서 학생들에게 요구되는 가장 중요한 역량 중 하나로 ‘질문하는 능력’을 강조하였다. 이러한 점을 종합해 볼 때, 앞으로의 학교 교육은 학습자가 자기주도적으로 학습 과정을 성찰하고, 깊이 있는 탐구를 통해 고차원적 질문을 할 수 있도록 설계되는 것이 필요할 것이다.

다만, AI 튜터로서, 챗봇의 효과적인 활용 혹은 학교 현장에서의 정착을 위해서는 다음 세 가지 부분에 대한 고려 역시 필요할 것이다. 첫째, 단순히 수업 시간에 챗봇을 사용하는 것을 넘어 수업에 적용 가능한 교수-학습 모듈 혹은 프레임워크의 개발이 필요하다.

둘째, 본 논문에서도 다룬 바와 같이 AI 챗봇에는 여전히 환각 현상과 같은 불확실성이 존재한다. 따라서 학습자가 AI 챗봇에 지나치게 의존하지 않고, 자기주도적으로 올바르게 활용할 수 있도록 AI 리터러시 교육이 필요할 것이다(Scheider *et al.*, 2023).

셋째, 학습분석학과 관련된 것으로, 교수자는 학습 데이터 해석에 있어 각별한 주의를 요한다. 챗봇 등을 활용한 학습분석의 결과는 다방면에서 유용하게 사용될 수 있지만, 동시에 학습자의 인지과정을 완전히 이해하는 데에는 한계가 있음을 인지해야 한다. 다시 말하면, 수집된 학습 데이터는 관찰 혹은 측정 가능한 인지과정만을 반영하였을 수 있다는 것이다. 즉, 관찰이 불가능한 부분에 대해서는 마치 학습이 발생하지 않은 것처럼 판단할 수 있다(Nistor and Hernández-García, 2018). 그러한 점은 교수자로 하여금 학습자에 대한 왜곡된 인식을 심어줄 수 있고, 학습자에게도 부정적인 영향을 미칠 수 있으므로 각별한 주의가 필요하다. 따라서 학습분석의 결과는 교육적 처치를 위한 근거로 사용될 수 있지만, 불확실성과 한계 역시 동시에 존재한다는 사실을 염두에 두어야 할 것이다.

註

- 1) 초기 챗봇은 민경모·유준희(2022)에 명시된 Doc2Vec 모델의 코드를 활용하여 개발되었다.
- 2) 챗봇 개요와 데이터 구축에 대한 자세한 사항은 이소영 등(2023) 참고
- 3) 개발된 챗봇의 특성상 학생들의 질문에 대하여 챗봇이 답변을 아예 하지 못하거나 잘못된 답변을 하는 경우도 발생하였다. 그러한 경우는 연구자 및 담당 교사가 시스템에 접속하여 수시로 질문에 대한 답변을 수정·보완하였다(그림 3의 α 에 해당). 다만, 챗봇의 답변 내용 및 정확성 등은 본 연구의 주제와는 다소 벗어나므로, 본 연구에서는 학생들의 질문만을 분석의 대상으로 삼는다.
- 4) 통상적으로 챗봇은 응답 방식에 따라 싱글턴(single-turn)과 멀티턴(multi-turn)방식으로 구분할 수 있는데(이소영 등, 2023), 본 연구에서 사용한 사회과 챗봇은 싱글턴 방식으로 개별 질문에 대해 짧은 답변만을 제공할 뿐, 전체적인 대화 맥락을 고려하기는 어렵다. 이에 따라 학생들이 메타인

지와 관련된 질문은 구조적으로 하기 어렵다고 판단되어 연구에서 제외하였다.

- 5) ChatGPT의 답변 정확도의 변화에는 질의 시기의 차이뿐만 아니라 모델 버전의 차이도 영향을 미쳤을 것이다. 10월 18일에는 GPT-4모델을 사용하였는데, 이는 GPT-3.5모델에 비해 환각 현상이 크게 감소하였으며, 사실성 평가에서도 40% 높은 점수를 받은 것으로 알려져 있다(정천수, 2023).
- 6) 가령, A라는 도시가 포항의 북쪽에 위치하는지, 남쪽에 위치하는지에 대한 질문에 정확히 응답하지 못하였으며, 각 도시 간 거리 비교에 대한 질문에도 어떤 쌍이 더 멀리 떨어져 있는지에 대하여 적절한 답변을 제공하지 못하였다(Lee and Lee, 2023).

참고문헌

강헌태·노석구, 2018, “초등학교 과학수업에서 학생의 사전질문 분석틀 개발 및 적용,” 한국과학교육학회지, 38(2), 235-247.

과학기술정보통신부·한국지능정보사회진흥원, 2023, 2023년도 인공지능 학습용 데이터 구축 지원사업 공고문(2023.4).

교육부, 2022a, 2022 개정 교육과정 총론 해설(중학교), 교육부 고시 제2022-33호(2022.12.22).

교육부, 2022b, 사회과 교육과정, 교육부 고시 제 2022-33호 [별책 7].

교육부, 2023a, 디지털 기반 교육혁신 방안, 교육부 보도자료(2023.2.23.).

교육부, 2023b, AI 디지털교과서 추진방안(안), 교육부 정책자료(2023.6.)

교육부·한국교육학술정보원, 2023, 「AI 디지털교과서 개발 가이드라인」, 한국교육학술정보원 교육자료 GM 2023-11.

김영우, 2021, 「Do it! 쉽게 배우는 R 텍스트 마이닝」, 이지스퍼블리싱(주).

김재우, 2020, “텍스트 마이닝을 활용한 사회과 교육과정 시기별 학술 담론의 특징 분석,” 서울대학교 대학원 박사학위논문.

민경모·유준희, 2022, “Doc2Vec을 이용한 중학교 과학 질문-답변 챗봇 개발 및 학생 질의 분석,” The SNU

Journal of Education Research, 31(3), 115-145.

신화영·백성혜, 2024, “ChatGPT 활용한 초등 과학 수업에서 질문 단계의 변화 및 수업에 대한 인식 분석,” 초등과학교육, 43(2), 322-336.

안미리·최윤영·배윤희·고윤미·김민하, 2016, “학습분석학 국내 문헌 고찰: 로그 데이터를 이용한 실증연구를 중심으로,” 교육공학연구, 32(2), 253-291.

양혜진·김혜영·신동광·이장호, 2019, “인공지능 음성챗봇기반 초등학교 영어 말하기 수업 연구,” 멀티미디어언어교육, 22(4), 184-205.

이경한, 2009, 「사회과 지리 수업과 평가」, 경기: 교육과학사.

이동민, 2017, “한국 초등지리교육 학위논문의 연구 동향에 대한 언어 네트워크 분석,” 사회과교육, 56(4), 75-90.

이소영·이상일·박의현·최학모, 2023, “사회과 챗봇의 설계와 교육적 활용 방안에 관한 연구 - 중학교 지리 영역을 중심으로 -,” 한국지리환경교육학회지, 31(1), 147-163.

이현우·민덕기, 2018, “건국봇: 검색모델과 생성모델을 결합한 챗봇,” 한국정보처리학회 학술대회논문집, 25(1), 449-452.

장연주, 2011, “초등 사회과 수업목표 분석: Bloom의 신교육목표분류학에 근거하여,” 한국교원대학교 석사학위논문.

정천수, 2023, “전통적인 챗봇과 ChatGPT 연계 서비스 방안 연구,” Journal of Information Technology Applications and Management, 30(4), 11-28.

최정인·백성혜, 2015, “Bloom의 신교육목표분류체계에 기초한 2007 및 2009 개정 초등학교 과학과 교육과정과 미국의 차세대 과학 표준(Next Generation Science Standards)의 성취기준 비교 분석,” 한국과학교육학회지, 35(2), 277-288.

추성엽·민덕기, 2019, “영어 상호작용 촉진을 위한 과업 기반 AI 챗봇 활용 및 학생 발화 분석,” 초등영어교육, 25(2), 27-52.

하은혜·이채현·조철기, 2022, “언어 네트워크 분석 기법을 활용한 중등학교 지리 교과서의 지리정보체계(GIS) 내용 분석,” 한국지역지리학회지, 28(3), 367-381.

한국교육학술정보원, 2022, 「학습분석을 위한 데이터 수집 및 저장 가이드라인 연구」, 한국교육학술정보원 연구보고 RR 2022-07.

- 황요한·이혜진, 2021, “AI 기술을 활용한 영어교육의 가능성: 영어 예비교사들의 인공지능 챗봇 사용과 개발을 중심으로,” *Multimedia-Assisted Language Learning*, 24(1), 104-133.
- 황홍섭, 2022, “초등 사회과 마이크로러닝을 위한 챗봇의 현장 수업 적용 및 검증,” *사회과교육*, 61(2), 69-94.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., and Wittrock, M.C., 2005, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Boston: Allyn & Bacon (Pearson)(강현석·강이철·권대훈·박영무·이원희·조영남·주동범·최호성 역, 2015, 『교육과정 수업 평가를 위한 새로운 분류학: Bloom 교육목표분류학의 개정』, 서울: 아카데미프레스).
- Becker, B., 2013, Learning analytics: Insights into the natural learning behavior of our students, *Behavioral & Social Sciences Librarian*, 32(1), 63-67.
- Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R., 1956, *The Taxonomy of educational objectives, handbook I: The Cognitive domain*, New York: David McKay.
- Caldarini, G., Jaf, S., and McGarry, K., 2022, A literature survey of recent advances in chatbots, *Information*, 13(1), 41.
- Cho, C.-K., Kim, H., and Lee, S., 2023, Analysis of the issues that emerged in the revision of the national social studies curriculum in South Korea: Text mining and semantic network analysis of the comments at the public hearing on YouTube, *Journal of Education and E-Learning Research*, 10(3), 463-473.
- Chocarro, R., Cortiñas, M., and Marcos-Matás, G., 2021, Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics, *Educational Studies*, 1-19.
- Feldman, R. and Sanger, J., 2007, *The Text Mining Handbook*, UK: Cambridge University Press.
- Gaikwad, S.V., Chaugule, A., and Patil, P., 2014, Text mining methods and techniques, *International Journal of Computer Applications*, 85(17), 42-45.
- Gupta, S. and Chen, Y., 2022, Supporting inclusive learning using chatbots? A chatbot-led interview study, *Journal of Information Systems Education*, 33(1), 98-108.
- Hamzah, W.W., Ismail, I., Yusof, M.K., Saany, S.M., and Yacob, A., 2021, Using learning analytics to explore responses from student conversations with chatbot for education, *International Journal of Engineering Pedagogy*, 11(6), 70-84.
- Hoppe, H.U., 2017, Computational methods for the analysis of learning and knowledge building communities, *Handbook of learning analytics*, 23-33.
- Huang, W., Hew, K.F., and Fryer, L.K., 2022, Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning, *Journal of Computer Assisted Learning*, 38(1), 237-257.
- Kinder, A. and Lambert, D., 2011, The National Curriculum Review: what geography should we teach?, *Teaching Geography*, 36(3), 93-95.
- Kuhail, M.A., Alturki, N., Alramlawi, S., and Alhejori, K., 2022, Interacting with educational chatbots: A systematic review, *Education and Information Technologies*, 1-46.
- Lee, S. and Lee, S., 2023, Designing an Educational Chatbot for Social Studies Classes in the Middle School, *Proceedings of the 23th International Conference on Education Research*, October 13, Seoul, South Korea.
- Long, P.D. and Siemens, G., 2011, Penetrating the fog: Analytics in learning and education, *Educause Review*, 46(5), 31-40.
- Maskill, R. and de Jesus, H.P., 1997, Pupils' questions, alternative frameworks and the design of science teaching, *International Journal of Science Education*, 19(7), 781-799.
- Nahm, U.Y. and Mooney, R.J., 2002, Text Mining with Information Extraction, *Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases*, Stanford, CA, 60-67.
- Nistor, N. and Hernández-García, Á., 2018, What types of data are used in learning analytics? An overview of six cases, *Computers in Human Behavior*, 89, 335-338.
- Okonkwo, C.W. and Ade-Ibijola, A., 2021, Chatbots applications in education: A systematic review, *Computers*

- and Education: Artificial Intelligence, 2, 100033.
- Papoušek, J., Pelánek, R., Řihák, J., and Stanislav, V., 2015, An analysis of response times in adaptive practice of geography facts, *Proceedings of the 8th international conference on educational data mining*, Madrid, Spain, 562-563.
- Robinson, A.C., Anderson, C.L., and Quinn, S.D., 2020, Evaluating geovisualization for spatial learning analytics, *International Journal of Cartography*, 6(3), 331-349.
- Şanlı, C., 2019, Investigation of question types in high school geography coursebooks and their analysis in accordance with the Revised Bloom's Taxonomy, *Ege Coğrafya Dergisi*, 28(2), 111-127.
- Scheider, S., Bartholomeus, H., and Verstegen, J., 2023, ChatGPT is not a pocket calculator - Problems of AI-chatbots for teaching Geography. arXiv:2307.03196.
- Topal, A.D., Eren, C.D., and Kolburan, A.K., 2021, Chatbot application in a 5th grade science course, *Education and Information Technologies*, 26(5), 6241-6265.
- Treves, R., Viterbo, P., and Haklay, M., 2015, Footprints in the sky: using student track logs from a "bird's eye view" virtual field trip to enhance learning, *Journal of Geography in Higher Education*, 39(1), 97-110.
- Varnavsky, A.N. and Popov, D.A., 2022, Research of the possibility of using educational chatbots as a learning analytics tool, In 2022 *Dynamics of Systems, Mechanisms and Machines (Dynamics)*, IEEE, 1-5.
- Wilby, R.L. and Esson, J., 2024, AI literacy in geographic education and research: Capabilities, caveats, and criticality, *The Geographical Journal*, 190(1), 1-8.
- Wilson, L.O., 2016, Anderson and Krathwohl-Bloom's taxonomy revised, *Understanding the new version of Bloom's taxonomy*.
- Zhou, T.Y., Huang, Y., Chen, C., and Xing, Z., 2023, Red teaming ChatGPT via jailbreaking: Bias, robustness, reliability and toxicity, arXiv preprint arXiv:2301.12867.
- 한국관광 100선 광명동굴, https://www.gm.go.kr/cv/gc/11/CVGC_11100.jsp
- 한국민족문화대백과사전, <https://encykorea.aks.ac.kr/Article/E0026846>,
- 교신 : 이소영, 25601, 강원도 강릉시 범일로 579길 24 가톨릭관동대학교 사범대학 지리교육과(이메일: sylee03051@cku.ac.kr)
- Correspondence: Soyoung Lee, Department of Geography Education, College of Education, Catholic Kwandong University, 24 Beomil-ro 579beon-gil, Gangneung-si, Gangwon-do, 25601, Republic of Korea (Email: sylee03051@cku.ac.kr)

투고접수일: 2024년 11월 28일
심사완료일: 2024년 12월 10일
게재확정일: 2024년 12월 20일