2025 | 12

26-B552111-000017-01

지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템 개발 연구보고서

(A Study on the Development of Generative AI-Based Science Inquiry Support System Integrated with the 'Intelligence Science Lab ON' Platform)

한국과학창의재단
Korea Foundation for Science and Creativity

CONTENTS

제1장 연구과제의 개요

1절 | 연구의 배경 및 필요성 ····· 2
2절 | 연구 방향 및 중점 추진 내용 ····· 5
3절 | 연구 방법 ····· 8

제2장 국내외 현황 조사

1절 | 데이터 기반 과학탐구의 개념과 의의 ····· 16
2절 | 생성형 AI의 교육적 활용과 효과 ····· 31
3절 | LLM 활용 전략 및 RAG 적용 사례 ····· 37

제3장 연구수행 내용 및 결과

1절 | 생성형 AI 기반 과학탐구 지원 시스템 개발 ····· 47
2절 | 교사·학생 설문 분석 ····· 53
3절 | 교사·학생 면담 분석 ····· 68
4절 | 시스템 로그데이터 분석 ····· 80
5절 | 시스템 성능 평가 ····· 86

제4장 시사점 및 제언

1절 | 결론 및 시사점 ····· 97
2절 | 후속 연구 제언 ····· 101

제5장 참고문헌 및 부록

1절 | 참고문헌 ····· 104
2절 | 부록 ····· 112

제 1 장 연구과제의 개요

1절 | 연구의 배경 및 필요성

2절 | 연구 방향 및 중점 추진 내용

3절 | 연구 방법

1절 연구의 배경 및 필요성

1 연구 배경

가 국내외 과학교육 정책의 변화

1) 국외 과학교육 정책

- 과학기술의 급격한 발달과 지능정보사회로의 변화에 대응할 수 있는 미래인재 양성을 위해 국제적으로 국가 차원의 과학적 소양 제고 정책을 강화하고 있음
  - 미국은 '23~'27년에 130억 달러 규모의 투자를 통해 유·초·중등 및 대학(원)의 STEM 교육 강화를 발표하였으며, 과학적 현상에 대해 조사·이해하는 과학교육-공학교육 연계 수업을 확대함. 또한 AI의 교육적 활용을 위해 교육자를 대상으로 인공지능 툴킷(AI Toolkit) 개발 및 민간 주도의 가상 실험 콘텐츠 개발·확산을 추진함(Government Technology, 2023).
  - 중국은 전 국민 과학적 자질 향상을 위해 전국 초·중·고 과학교육 실험학교를 추진하였으며, 이공계 석사학위 소지자를 과학 교사로 채용하는 등 과학교육 기초 강화와 교육의 질 제고 방안을 마련함
  - 영국 의회는 2030년대 변화된 미래를 위한 학교 교육 시스템 개정의 일환으로 과학 수업 시수를 확대하고, 중등 교육과정을 개편하는 한편, 교사의 디지털 리터러시 향상을 위한 프레임워크를 개발함으로써 기초과학 역량과 디지털 환경에 대한 수용성을 높이고자 함. 또한 '25년까지 학교 디지털 인프라 표준 수립을 목표로 4,500만 파운드를 투자하고 최대 92%의 정확도를 예상하는 AI 기반 자동 채점 및 맞춤형 피드백 도구 개발을 추진 중임(GOV.UK, 2025).
  - 싱가포르는 미래인재 양성을 위해 과학교육의 중요성을 강조하여 모든 학생의 과학적 소양 제고를 위해 영감, 탐구, 혁신을 위한 과학과 교육과정을 전면 개정하고, 학교의 온·오프라인 교육공간 개선과 지역 파트너 연계 과학실험·탐구를 강화함. 또한 SLS(Singapore Student Learning Space) 등 에듀테크를 활용한 교수학습 지원을 위해 교육부 차원의 계획 수립을 추진함(MOE, 2023; Govtech Singapore, 2025).

2) 국내 과학교육 정책

- 우리나라는 2022 개정 교육과정 및 제5차 과학교육 종합계획(2025~2029) 수립을 토대로 학생들의 과학탐구 역량 강화 및 온·오프라인 연계 과학탐구 환경 조성에 노력함
  - 2022 개정 교육과정은 전 교과에서의 디지털 소양 확대를 강조하며, 과학과에서는 디지털 과학탐구와 핸즈온 과학탐구의 조화 속에 학생 중심의 실험·탐구가 이루어질 수 있도록 '지능형 과학실' 구축 필요성을 제시함.
  - 제5차 과학교육 종합계획은 '탐구하는 즐거움, 미래에 도전하는 힘을 키우는 과학교육'이라는 목표하에 학생들이 기초과학과 과학탐구 역량을 제고할 수 있도록 '1학생 1탐구'와 같은 과학탐구 문화 조성, '지능형 과학실 ON' 연계 AI 기술을 활용한 맞춤형 과학탐구 지원, 모든 학교 대상 '지능형 과학실' 구축 완료 등을 골자로 함.
  - '지능형 과학실'은 오프라인 과학실 공간 구축과 온라인 플랫폼이 연계된 첨단 과학기술 기반 실험·탐구 공간으로서 데이터 기반 과학탐구가 이루어질 수 있도록 종합적인 사업 지원과 고도화가 추진 중임.

나 지능형 과학실 사업

- 첨단기술을 바탕으로 다양한 온·오프라인 참여형 과학탐구 실험 및 융·복합적 탐구활동이 가능한 과학 교수학습 환경 구축을 목적으로 하는 과학교육 대표 정책 사업
- 추진경과
  - 창의융합형 과학실 사업(2017년~2019년)
  - 지능형 과학실 모델학교 운영 및 확산(2021년~)
  - 지능형 과학실 ON(플랫폼) 개발·운영(2022년~)

1) 지능형 과학실

- 지능형 과학실은 학생 주도적 과학탐구를 강화한 온·오프라인 연계 혁신 탐구 공간으로서 전국 5,903교 구축 완료('24년 기준, 44.9%), 모델학교 78교 운영 중
  - 2027년까지 전국 학교의 70%에 지능형 과학실 구축 및 디지털 도구가 보급될 예정이며, '지능형 과학실 발전 협의체'를 통해 교수학습모델 및 수업자료 개발·보급, 교원 네트워크 활성화를 지원함.

구분	① 융합적 학습 경험 확대		② 협력적 탐구 환경 활성화		③ 맞춤형 피드백 강화
주요 내용	• 탐구 모듈별 유연한 공간 활용 • 핸즈온 및 디지털 탐구 도구 조화	+	• 온·오프라인 장점 결합 • 협력탐구/공동탐구/글로벌 탐구 활성화	+	• 실생활 맥락의 과학탐구 활성화 • 개별 맞춤형 피드백 제공 시스템 구축

2) 지능형 과학실 ON

- 지능형 과학실 ON은 학생 중심의 데이터 기반 과학탐구 지원 플랫폼으로서 지능정보기술을 도입하여 맞춤형 학습지원이 가능하도록 단계적 고도화 추진 중
  - 2025년 11월 기준 7만 여명의 가입자를 확보하였으며, 학교 현장에 디지털 도구를 활용한 과학탐구 운영이 이루어질 수 있도록 교원 연수 프로그램과 연계한 활성화 전략을 추진 중임.
  - 민간 에듀테크 업체와의 협업을 통해 17개 브랜드 432개 센서·기기 데이터를 연동하여 과학탐구가 가능하며, LRS 시스템 도입 등 맞춤형 서비스를 위한 인프라 도입을 추진하였음.

「지능형 과학실 ON」 개발방향

2022년	2023년	2024년	2025년	2026년~
과학탐구 플랫폼 개발		맞춤형 서비스 기반 마련		맞춤형 서비스 고도화
• 시범 운영 • 플랫폼 요구 분석	• 전면 운영 • 데이터 기반 과학탐구 기능 고도화 • 오픈 API 명세서 개발	• 협력적 과학탐구 환경 구축 • 맞춤형 수업 기반 마련 • 대시보드 개발 • IoT 센서 연동	• 맞춤형 수업 지원을 위한 AI 서비스 개발 • 종합역량 진단 도구 탑재 • IoT 센서 연동	• AI 기능 시범운영 • 맞춤형 수업지원 서비스 고도화 • 과학탐구 교육 데이터 센터 역할 확립

2 연구의 필요성 및 목적

- 2022 개정 교육과정 및 제5차 과학교육 종합계획('25~'29)은 지능정보기술 활용 학생주도 학습환경 구축 및 과학탐구 활성화를 강조하고 있지만, 초중등 학생의 과학 분야 정의적 성취는 국제 평균에 비해 낮고 격차 해소를 위한 맞춤형 교육 인프라는 부족한 실정
  - 수학·과학 성취도 추이 변화 국제비교 연구(TIMSS) 2023 결과 과학 기초 수준 이하 비율이 증가하였고('19, 5% → '23, 7%), 과학 흥미·자신감은 지속적으로 하위권에 머무르고 있음.
  - AX 시대의 미래인재 양성을 위해서는 다양한 분야와 연계하여 과학적으로 사고하며 현상을 탐구할 수 있는 역량과 함께 데이터 분석·해석 역량이 보다 중요하지만 과학에 대한 관심 부족은 학생들의 진로 탐색에도 영향을 미쳐 과학기술 분야 인재 부족으로 이어질 수 있음.
- 과학탐구의 즐거움과 과학의 가치를 이해하도록 하기 위해서는 과학탐구의 경험이 특별할 수 있도록 보다 개인화된 맞춤형 지원이 필요함.

진단	해결
☑ 지능형 과학실을 활용한 데이터 기반 탐구 활동을 통해 과학 긍정 경험이 상승하고 있으나 약 44.9% 학교에서만 구축 완료 ☑ 과학탐구 수행을 위해서는 다양한 지식과 기능이 필요하지만 학생들의 탐구역량을 제고할 수 있도록 교사가 모든 학생을 대상으로 맞춤형 피드백을 제공하기 어려움 ☑ '1학생 1탐구' 등 학생의 개별적인 탐구 호기심을 충족하여 자율탐구를 강화하도록 정책을 추진 중이지만 이를 실현하도록 할 지원 기반이 미흡	☑ 오프라인 공간 구축에 소요되는 시간을 감안하여 온라인 플랫폼과 연계하여 과학탐구 경험을 확대할 수 있는 다양한 전략 마련 필요 ☑ 첨단기술 활용 융복합 과학탐구가 가능한 지능형 과학실 ON 플랫폼에 AI 기술 기반 학생 맞춤형 서비스를 도입하여 학생별 과학탐구 역량 제고 지원 ☑ 학생별 탐구 주제에 맞게 데이터 수집, 분석, 보고서 작성 등의 단계마다 지원하여 모든 학생의 성장과 과학 자신감 제고

다 연구 목적

- 본 연구는 교육 환경이나 학생의 수준에 상관없이 누구나 과학탐구 과정을 통해 과학탐구 역량을 기르고 성장할 수 있도록 1) 지능형 과학실 ON 플랫폼 연계 AI 기반 맞춤형 서비스를 개발하고, 2) 이러한 접근이 학교 현장에 주는 효과를 분석하여 타당성을 확보하고자 함.
  - 본 연구는 생성형 AI 활용 맞춤형 지원을 위해 '대화형 서비스'와 '보고서 튜터링 서비스'를 개발하고자 하며, 전문가 협의 및 교육현장 사용성 평가 등의 의견수렴 절차를 통해 개발을 추진함.
  - 본 연구는 프로토타입 구현을 위한 개발 과정을 명시하고, 개발의 효과를 설문, 면담, 학생 보고서, 및 시스템 로그데이터를 통해 다각적으로 분석하여 후속 정책을 위한 시사점을 도출하였음.
- 본 연구는 공교육에서의 AI 도입의 가능성을 확인하고, 학생 개별화 교육을 실현하며, 과학탐구 과정평가 및 학생 중심 과학탐구 활성화에 기여할 수 있음.
  - '질문하는 과학 수업 조성', '학생 모두가 주인공인 과학교육'을 위해 공교육에서 어떻게 AI를 활용할 수 있는지에 대한 구체적인 사례가 될 수 있음.
  - 과학탐구 교육에 특화되어 안전하게 설계된 생성형 AI 도입을 통해 학생들이 과학적 사고와 지식의 확대를 경험할 수 있도록 방법을 구체화하고자 하며, 기존 과학수업에 비해 보다 면밀하게 실시간 과학탐구를 지원함으로써 과정중심, 학생중심의 과학탐구로의 개선을 이끌어내고자 함.

2절 연구 방향 및 중점 추진 내용

1 연구 방향

가 생성형 AI 기반 과학탐구 지원 시스템 개발

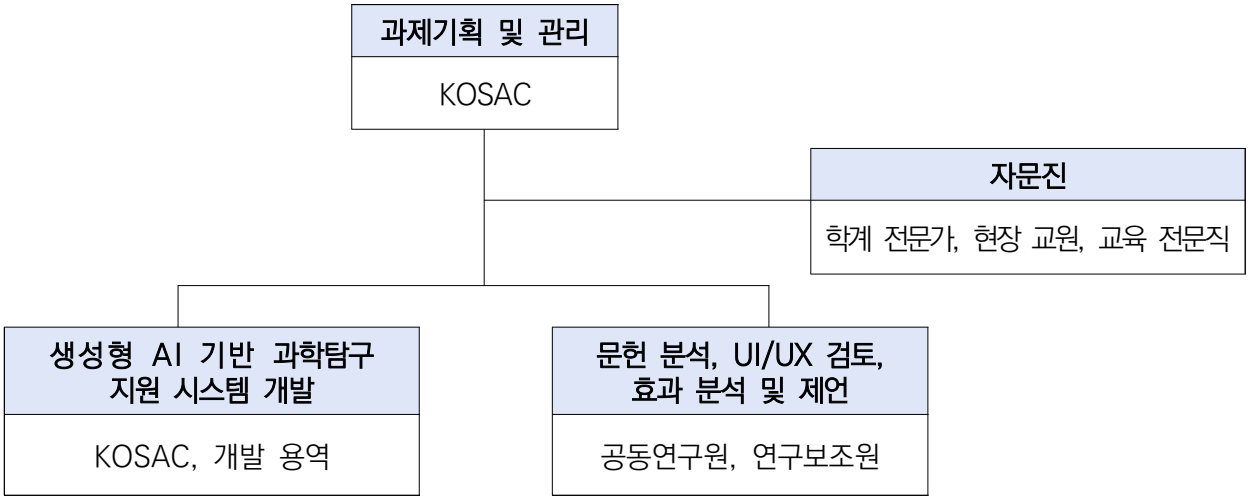
- 본 연구를 통해 개발하고자 하는 생성형 AI 기반 과학탐구 지원 시스템은 과학탐구의 본질에 따라 교사와 학생을 종합적으로 지원하고 과학탐구 과정과 결과를 두루 지원할 수 있도록 두 가지 서비스를 개발함.
  - 과학탐구는 정답만이 중요하지 않고, 과학적으로 사고하여 결론에 이르는 전 과정에 참여하는 것에서 학습이 이루어질 수 있음. 현실적인 제약에 의해 실험 결과가 이론과 다르게 도출되었더라도 이러한 현상을 해석하고 과학 개념을 이해하려는 노력을 기울이는 과정을 진작하고 독려할 필요가 있음.
  - 학생을 대상으로 과학탐구 과정을 개별적으로 지원하기 위해서는 '대화형 서비스' 개발을 통해 데이터 수집, 분석, 보고서 제출의 전 과정에서 질문하고 탐구를 심화할 수 있도록 하고자 함.
  - 교사의 수행평가를 지원하기 위해 학생이 제출한 보고서의 서술내용, 그래프, 표, 이미지 등에 대해 피드백을 제공하는 '보고서 튜터링 서비스' 개발을 추진함. 탐구의 목적과 보고서의 취지에 따라 피드백을 제공하며, 교사들의 평가 업무 경감을 지원할 수 있음. 또한 과학탐구 산출물에 대한 세밀한 피드백은 학생들의 과학탐구 역량 제고에 도움을 줄 수 있음.
- 제5차 과학교육 종합계획('25~'29)의 비전과 목표를 계승하고, 세부 전략 5개 중 유관한 3개를 채택하여 추진과제가 강조하는 취지에 맞게 시스템의 개발 방향을 설정함.
  - 본 연구는 학생의 과학 정의적 성취와 탐구역량을 제고하고, 교실 수업의 혁신을 촉진하고자 함.

비 전	첨단과학기술 사회를 누리고 더 나은 미래를 개척하는 인재 양성
목 표	탐구하는 즐거움, 미래에 도전하는 힘을 키우는 과학교육

전 략	AI기반 학생 맞춤형 과학탐구 지원 시스템 개발 원리
모든 학생의 과학 자신감 제고	① 지능정보기술을 활용하여 과학학습에서의 학생 주도성을 강화한다. ② 학생의 질문을 이끌어내어 깊이 있는 과학적 사고를 경험하도록 한다. ③ 인지적·정의적 성취의 간극을 좁히고 맞춤형 성장을 지원한다.
기초과학 역량 강화를 위한 체계 고도화	① 학습이력에 기반한 학습 콘텐츠 제공[1]으로 사전·사후 학습경험을 확장한다. ② 과학탐구의 과정평가를 강화하여 과학 역량 함양을 지원한다.
과학 교원이 주도하는 과학교육 발전 체계 운영	① 교사의 수업계획에 따라 AI 시스템을 활용·관리할 수 있도록 한다. ② 지식전달에서 탈피하여 과학교사의 학습 설계 및 코칭 역량 강화를 지원한다.

1) 2025년 지능형 과학실 ON 고도화 계획에 따라 ①보고서 튜터링, ②대화형 서비스, ③콘텐츠 추천 기능을 개발하나, 본 연구에서는 생성

나 추진체계



2 중점 추진 내용

가 보고서 튜터링 서비스

- 과학탐구 활성화를 위해 학생의 탐구 산출물에 대한 개별화된 피드백 제공 및 맞춤형 성장 지원
- 지능형 과학실 ON에 제출된 학생 보고서에 제시된 데이터 수집·분석·해석의 방법과 결과 등 과학탐구 산출물을 학교급, 학습목표에 맞게 맞춤형 피드백 제공
  - 보고서 작성 과제에 대해 텍스트, 이미지(그래프), 표 등의 학생 제출 내용을 기준으로 적절성을 분석하여 맞춤형 피드백 제공
  - 교사의 답변 채택·재구성이 가능하고, 피드백을 개선할 수 있어 인간-AI 상호작용을 통한 교사의 과학탐구 수업 업무 경감 및 학생의 세밀한 과정평가 지원

◆ 교사가 개별 학생의 보고서에 피드백을 충분히 제공하기 위해서는 많은 시간과 노력이 요구되며, AI 도입 시 공평하고 효율적으로 학생별 보고서 튜터링 가능

나 대화형 서비스

- 깊이있는 과학탐구에 필수적인 고차적 사고를 촉진하고 미래인재의 핵심역량으로 요구되는 AI와의 협력 역량을 초중등 학생들이 격차 없이 함양하기 위해서는 공교육에서의 AI 상호작용 학습환경 필요
- 학생 주도적 과학탐구 경험 확장을 위해 개인화된 질의응답을 통해 자율탐구의 성공 가능성을 높이고, 과학개념 이해와 탐구수행의 과정 기능이 함께 향상될 수 있도록 지원
- 학생의 질문에 대한 답변뿐만 아니라 연관 개념학습 지원, 후속 질문의 유도, 학생에게 질문 제시 등의 다양한 질의응답 활동을 통해 깊이 있는 과학탐구 촉진
  - (예시 1) 수업 전 : 선수 학습 점검 및 후속 학습을 위한 연관 개념 이해

형 AI 기반 서비스에 한정하여 두 가지 기능을 개발하고자 함.

- (예시 2) 수업 후 : 보고서 내용 검토, 성찰을 통해 반성적 사고 촉진 및 과학탐구 보고서 개선

- 대화형 서비스의 안전성을 확보하기 위해 교사의 모니터링 기능 개발
  - 대화형 서비스 활용 여부 선택(미사용/시작/종료), 학생 질의내용 실시간 확인, 부적절 사용 시 경고 메시지 전송 등

◆ 과학탐구 과정에서 개인화된 학습 문제를 즉각적으로 해결함으로써 과학탐구의 자신감을 제고하고, AI와의 상호작용을 통해 탐구학습이 확대, 심화되도록 흥미 유발 가능

다 LLM의 교육적 활용을 위한 시스템 안정성 강화

- 생성형 AI의 도입을 위해 LLM의 활용방식과 모델의 선택이 필요함. 본 연구에서는 상용화된 생성형 AI를 API 연동 방식으로 활용하며, 안정성, 보안성, 멀티모달 기능 지원 여부와 정부 정책을 검토하여 HyperClovaX를 개발에 활용하는 것으로 선정함
  - 보고서 튜터링의 경우 학생이 탑재한 이미지(그래프)에 대한 이해가 필요하므로 멀티모달 상호작용이 가능한 HyperClovaX 005 모델을 활용하며, 대화형 서비스의 경우 과학·수학 데이터 학습이 고도화된 추론 모델인 HyperClovaX 007 모델을 활용하였음
- 또한 생성형 AI의 환각효과 등의 부작용을 해소하고 신뢰성 높은 상호작용, 우리나라의 교육과정에 맞는 과학 개념 안내 등을 위해 RAG(검색증강생성, Retrieval Augmented Generation) 기술을 활용하여 교육적 활용도를 높이고자 하였음.
- LLM 보완을 위한 RAG 기술의 활용에 더해 학생들의 대화형 시스템 질의응답 내용은 전부 암호화하여 시스템의 안정성을 강화하고, 과학교육 범위를 넘어선 질의에는 답변하지 않도록 개발함으로써 교육적 목적으로 시스템 활용을 제한함.
  - 학생들에게 과학탐구 시 생성형 AI 활용에 대한 윤리적 기준을 자연스럽게 지능형 과학실 ON 활용 과정에서 습득할 수 있도록 하고자 하였으며, '과학탐구역량'과 '디지털 리터러시'를 함께 향상시킬 수 있는 사용자 경험을 제공하도록 개발하는 것이 필요함.
- 개인정보보호 강화, 학생 모니터링 도구 지원 등 개발 전략 체계화와 함께 학생이 지능형 과학실 ON에서 일반적인 생성형 AI 이상의 학습 경험을 얻도록 교육적 전략 정교화가 중요함.

라 사용성 검토 및 효과 분석

- 지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템을 현장의 요구에 부합하도록 개발하기 위해 현장 교사와 학생의 의견 수렴 절차가 중요하며, 시스템 개발 단계에 따라 사용성 검토를 추진함.
- 1차 사용성 검토는 교사를 대상으로 보고서 튜터링 서비스의 초기 프로토타입을 검토하도록 하였으며, 지능형 과학실 ON에 탑재된 학생 보고서를 학교급별로 무작위 추출하여 튜터링을 적용하였고, 적용 결과에 대해 교사의 평가 의견을 수렴하였음
  - 교사의 의견수렴을 토대로 튜터링 제공 방식을 결정하였으며, 이를 반영하여 2차 사용성 검토를 추진함
- 2차 사용성 검토는 실제 교실 수업에서 AI기능이 도입된 지능형 과학실 ON 활용 수업을 진행하면서 보고서 튜터링과 대화형 서비스를 활용하도록 하였고, 사용성 검토에 참여한 교사, 학생을 대상으로 설문과 면담을 추진하고 시스템 로그 데이터를 추출함
- 본 연구에서는 1차 사용성 검토는 개발 과정에 담고, 학교 현장에 실제적으로 시스템을 적용한 2번째 사용성 검토 결과를 토대로 효과를 분석하였음

3절 연구 방법

1 연구 절차

가 시스템 개발 절차 및 일정

- 본 연구는 〈2025년 지능형 과학실 ON 운영 및 콘텐츠 개발 용역〉 과제의 수행과 연계하여 연구진의 검토 및 교육현장의 의견수렴을 토대로 시스템을 개발하고 효과를 분석하고자 아래의 절차를 수행함
  - (‘25년 4월) 추진계획 수립 및 연구심의 요청
  - (’25년 4~5월) 서비스 개발에 대한 개인정보보호 차원의 전문가 자문 수렴
  - (‘25년 5월) 연구심의 완료 및 프로토타입 개발 추진
  - (’25년 6월) 프로토타입 시연을 통한 수정·보완 계획 수립, LLM의 선택
  - (‘25년 7월) 보고서 튜터링 서비스 사용성 검토단 운영 및 의견 수렴 반영
  - (’25년 8월) 개인정보보호위원회 사전적정성 검토 신청 및 정부차원의 보안성 검토 추진
  - (‘25년 9월) 교육현장 의견수렴 간담회 및 현장적용 사용성 검토 추진
  - (’25년 10월) 현장적용 사용성 검토를 통한 교사, 학생 의견수렴
  - (‘25년 11월) 현장적용 사용성 검토 데이터 분석
  - (’25년 12월) 현장적용 사용성 검토 의견 반영 및 보고서 작성
- 서비스를 개발하는 것과 더불어 안정적인 AI 도입을 위한 개인정보보호 및 보안성 측면의 검토를 추진함
  - 개인정보보호 측면에서의 특이사항이 없음을 확인하였으며, 보안성 강화와 관련한 검토 의견을 준용하여 개발을 추진하였음

〈추진 일정〉

<table>
<tr><th>구분</th><th>5월</th><th>6월</th><th>7월</th><th>8월</th><th>9월</th><th>10월</th><th>11월</th><th>12월</th></tr>
<tr><td>AI 기반 과학탐구 지원 시스템 프로토타입 개발</td><td colspan="2">– 초기 프로토타입 개발 및 시연
• LLM 선택·테스트</td><td colspan="3">– 1차 프로토타입 개선
– 2차 사용성 검토를 위한 시스템 환경 마련</td><td colspan="3">– 2차 프로토타입 개선
– ‘25년도 최종 개발방향 확정</td></tr>
<tr><td>시스템 보안성 및 안정성 강화</td><td colspan="3">– 정보보호 자문 의견 수렴</td><td colspan="2">– 사전 적정성 검토
– 보안성 검토</td><td colspan="2">– 검토 의견 반영</td></tr>
<tr><td>시스템 사용성 강화</td><td colspan="2">– 교육현장 요구분석</td><td colspan="2">– 1차 사용성 검토
• 보고서 튜터링
• 교사 대상</td><td>– 검토 의견 반영</td><td colspan="2">– 2차 사용성 검토
• 보고서 튜터링
• 대화형 서비스
• 교사·학생 대상</td><td>– 검토 의견 반영</td></tr>
<tr><td>연구진 협의 및 보고서 작성</td><td colspan="5">– 서비스 개발 테스트 및 방향성 협의
– 효과분석 방향 협의</td><td colspan="2">– 효과분석</td><td>– 최종 평가</td></tr>
</table>

나 효과분석 절차 및 일정

- 현장적용 사용성 검토 참여자를 대상으로 지능형 과학실 ON AI 기능의 사용성과 효과를 분석하기 위해 시스템 활용 전·후의 사용자 인식 변화, 온라인 활동 분석 및 사후 사용성 분석을 실시하기 위해 교사와 학생에게 각각 일련의 절차에 참여하도록 하였음
- '25년 10월부터 11월 사이에 사용성 검토 및 효과분석이 이루어졌으며, 연구 참여자가 지능형 과학실 ON 계정으로 접속하였을 시 AI 기능을 활용할 수 있도록 계정에 권한을 부여하여 연구를 추진하였음

1) 교사

- 교사가 2차시 이상의 과학탐구 수업계획을 수립하고 수업계획서에 따라 지능형 과학실 ON의 탐구수업 메뉴에서 탐구 개설 및 시간표 생성
- 사전 설문조사 참여
- 탐구수업을 운영하고 실제 수업 운영 결과를 반영하여 결과보고서 제출
- 사후 설문조사 참여
- FGI 참여

2) 학생

- 사전 설문조사 참여
- 탐구수업 참여(보고서 총 2회 제출, 보고서 튜터링 적용 후 수정 · 보완하여 재제출)
- 사후 설문조사 참여
- 희망자에 한 해 FGI 참여

〈사용성 검토 절차〉

2 연구 방법

가 연구 참여자

- 연구 참여자는 지능형 과학실 ON을 활용한 과학수업 운영 경험이 있는 현직 교사 중 연구 참여를 희망하는 교사와 해당 교사의 수업에 참여하는 학생들로 선정하였음. 본 연구에 참여하기 위해서는 연구 기간 중 다음의 조건을 충족하여야 함
  - 교사와 학생 모두 지능형 과학실 ON 활용 경험이 있어 탐구수업 메뉴 활용에 어려움이 없을 것
  - 2차시 이상의 과학탐구 수업을 계획하고 탐구수업 메뉴를 통해 수업을 운영할 것
  - 수업 중 AI가 적용된 대화형 서비스와 보고서 튜터링 서비스를 활용할 것
  - 학생들은 보고서 제출 후 AI 튜터링 결과를 반영하여 수정된 보고서를 재제출할 것
  - 사전·사후 설문에 참여할 것
- 본 생성형 AI 기반 과학탐구 지원 기능을 교실 수업에 적용하여 사용성 검토 및 효과분석을 위한 연구에 참여하는 교사는 13명으로 선정됨. 초등학교 교사 7명, 중학교 교사 3명, 고등학교 교사 3명 등 총 13개교의 교사로 구성되었음
- 학생의 경우 총 313명이 연구에 참여할 예정이었으나 연구 참여 교사의 수업에서 과학탐구를 수행하면서 298명의 학생이 보고서 튜터링 전·후로 총 2회 보고서 제출을 완료하였으며, 사전·사후 설문은 그중 288명의 학생이 완료하였음. 다만, 설문 1회의 응답이 충분치 못했던 도움반 학생 1인을 제외하고 학생 287명을 최종 연구 참여자로 확정함. 학생 287명의 구성은 초등학교 91명, 중학교 79명, 고등학교 117명과 같음.
  - 학생 대상 면담은 초·중·고 각각 면담 여건이 가능한 1개교를 선정하여 참여 희망자를 대상으로 이루어졌으며 C 초등학교의 학생 5명, H 중학교 학생 3명, K 고등학교 학생 4명이 면담에 참여하였음.

〈연구 참여자〉

연번	구분	지역	학교명	교사	학생	
					신청자	연구참여자
1	초등학교	강원	A 초등학교	교사 1	16명	16명
2		전남	B 초등학교	교사 2	3명	3명
3		충남	C 초등학교	교사 3	26명	26명
4		강원	D 초등학교	교사 4	5명	5명
5		전북	E 초등학교	교사 5	5명	5명
6		경기	F 초등학교	교사 6	22명	22명
7		제주	G 초등학교	교사 7	14명	14명
8	중학교	경남	H 중학교	교사 8	26명	24명
9		경기	I 중학교	교사 9	34명	23명
10		경기	J 중학교	교사 10	32명	32명
11	고등학교	전남	K 고등학교	교사 11	68명	65명
12		전북	L 고등학교	교사 12	24명	24명
13		경남	M 고등학교	교사 13	38명	28명
합계				13명	313명	287명

나 연구 방법

- **설문 분석**
  - 교사를 대상으로 수업 전 서비스에 대한 기대에 대해 조사하였고, 수업 운영 후 시스템 사용성 평가 설문을 실시함. 기초통계분석과 주관식 응답에 대한 정성적 분석을 통해 교사의 인식에 대해 분석함
  - 학생을 대상으로 수업 전·후 과학긍정경험을 조사하였으며, 수업 후 사용성 평가 설문을 실시함. 학생들의 과학긍정경험은 대응표본 t-검정을 통해 사전·사후 변화를 검증하였으며, 사용성 평가 결과는 기초통계를 통해 확인함. 학생들의 주관식 응답에 대한 정성적 분석을 통해 시스템의 장점과 개선점에 대한 인식을 분석함
- **면담 분석**
  - 연구에 참여한 교사 전원을 대상으로 시스템에 대한 효과와 개선점에 대해 면담을 실시함. 초등교사와 중등교사로 구분하여 FGI를 실시함. 면담은 각 1시간 남짓 소요되었음.
  - 연구에 참여한 학생 중 희망 학생을 대상으로 시스템에 대한 효과와 개선점에 대해 면담을 실시함. 초등학생, 중학생, 고등학생 별로 면담을 실시하였으며, 각 FGI는 40분~60분 사이로 추진되었음
  - 면담 내용은 녹음하여 전사하였고, 주제분석(Braun & Clark, 2006) 방법을 통해 의미를 파악하고자 하였음.
- **로그데이터 분석**
  - 대화형 서비스의 로그데이터를 수집하여 총 2,845건의 이벤트 로그를 분석함. 내용 상의 하위 주제, 응답시간, 질문 유형에 대해 분석을 실시함.
  - 보고서 튜터링 서비스를 활용하여 연구 참여 학생들이 제출한 보고서를 분석하고자 하였음. 287명의 학생들이 각 2개의 보고서를 제출하였으므로 총 574개의 보고서 내용 중 결측값을 제외하고 분석에 활용하였음. 대응표본 t-검정과 일원분산분석을 통해 학생 집단의 학업 성취도의 변화를 분석하였으며, 성취도 변화에 있어 학교별 차이를 확인하였음.
- **시스템 성능 평가**
  - 교사가 수업 계획 시 구성한 평가 루브릭과 보고서 튜터링 결과 값 간의 유사도를 분석함. 전통적 유사도 평가 지표(KF1, sacreBLEU, BERTScore)를 통해 표면적 어휘·구문 유사도를 G-Score를 통해 내용적·개념적 유사도를 분석하여 지표 간 상호 보완적 성격을 고려하여 분석을 실시함.
  - 대화형 서비스에 적용된 RAG 시스템의 실질적 성능을 검증하기 위해 평가 데이터 셋을 구성하고 LLM-as-a-judge(RAGAS) 평가 메커니즘을 활용하여 성능을 분석함.

다 연구 도구

1) 설문조사

- **과학긍정경험**
  - 과학긍정경험 지표(TIPES)(신영준 외, 2017)는 학습자의 과학 관련 긍정 경험 측정을 위해 개발된 표준화 도구로서 5개 영역(학습정서, 자아개념, 동기, 진로포부, 태도)에 대해 각각 응답하도록 구성되었으며, 4점 척도의 총 35문항으로 구성됨.
- **사용성 평가지**
  - 사용성 평가지는 기술수용이론(Technology acceptance model)(Fred Davis, 1989)을 바탕으로 본 연구에 맞게 문항을 구성하여 개발하였음. 응답자 정보 6~7문항, 유용성 4문항, 사용용이성 4문항, 사용의도 3문항, 사용태도 3문항, 신뢰·정확도 2문항, 학습효과 및 교육적 가치 4문항, 장점과 개선점

에 대한 주관식 응답 2문항으로 구성하였음.

- 응답자 정보를 제외한 20개의 설문 문항과 주관식 응답 2문항은 보고서 튜터링과 대화형 서비스에 대해 각각 사용성 평가 의견을 제시할 수 있도록 별도의 질문으로 구성하여 사용성 검사지를 제작하였으므로 44개의 설문 문항으로 구성됨.

〈설문 문항 구성〉

대상	구분	주요내용
학생	사전	1) 과학긍정경험 + 기타 사전설문(총 47~49문항) · 참여자 정보 : 6문항 · 과학 긍정경험 : 35문항 (5항목) · AI 경험 : 3문항 · 참여 혜택 증정을 위한 정보 입력 – 14세 이상 : 3문항 – 14세 미만 : 5문항
	사후	2) 과학긍정경험 + 사용성 평가지(총 88~90문항) · 참여자 정보 : 6문항 · 과학 긍정경험 : 35문항 (5항목) · 사용성 평가 : 44문항 (6항목) · 참여혜택 증정을 위한 정보 입력 – 14세 이상 : 3문항 – 14세 미만 : 5문항
교사	사전	1) 참여자 정보 및 기대(총 21문항) · 참여자 정보 : 7문항 · AI 경험 : 3문항 · 사전설문 문항 : 11문항 (2항목)
	사후	2) 사용성 평가지 · 참여자 정보 : 7문항 · 사용성 평가 : 44문항 (6항목)

2) 면담조사

- **면담 프로토콜**
  - 교사와 학생의 면담을 위해 면담 프로토콜을 개발하였으며, 지능형 과학실 ON의 AI 기능의 효과와 개선점에 대해 질문하고자 하였음. 효과와 개선점을 다양한 방식으로 확인하고자 AI 기능 활용 경험, 장점, 개선점 및 교수학습 활동 참여에 대한 인식에 대한 질문으로 세분화하여 면담 프로토콜을 구성하였음.

〈면담 프로토콜 구성〉

교사 면담 프로토콜 질문 예시
1. AI를 과학탐구 수업에 어떻게 활용하셨나요? 2. AI가 과학탐구 수업에 도입되었을 때 어떠한 변화가 나타났나요? 어떠한 변화가 앞으로 발생할 것으로 예측하나요? - 교사와 학생의 측면에서 역할, 활동상의 변화가 있었나요? - 어떠한 학생의 성장이 있었나요, 혹시 부작용이 있었나요? 3. AI가 과학탐구 수업에 도움이 되기 위해서는 어떠한 시스템의 개선이 필요한가요? 4. AI를 과학탐구에 활용할 때 교사효능감은 어떠했나요? 5. AI를 과학탐구에 활용하기 위해 앞으로 어떠한 교사 전문성 향상 노력이 필요하다고 생각하시나요? 6. AI를 과학탐구에 활용한 학생들의 과학 학습에 대한 태도는 어떠했나요?
학생 면담 프로토콜 질문 예시
1. 지능형 과학실 ON 활용한 수업은 어떤 점이 좋은가요? 2. 지능형 과학실 ON 활용한 수업은 어떤 점이 어려운가요? 3. AI 기능을 활용하니 그 전의 수업과는 어떤 점이 다른가요? - 학생들의 과학탐구 활동에 도움이 되었나요? - 나의 과학탐구에 대한 생각, 흥미, 즐거움, 자신감 등은 어떠했나요? 4. 어떠한 과학탐구 수업을 할 때 즐거운가요? 5. 지능형 과학실 ON에 대한 개선 의견이 있다면 말씀해주세요.

제2장 국내외 현황 조사

1절 | 데이터 기반 과학탐구의 개념과 의의

2절 | 생성형 AI의 교육적 활용과 효과

3절 | LLM 활용 전략 및 RAG 적용 사례

1절 데이터 기반 과학탐구의 개념과 의의

1 데이터 기반 과학탐구의 중요성

가 과학 연구 패러다임의 전환

- 21세기 과학 연구는 전통적인 이론 중심, 실험 중심 방법론에서 데이터 집약적 패러다임(data-intensive paradigm)으로 근본적인 전환을 경험하고 있음(Tansley & Tolle, 2009)
- 이러한 변화는 과학의 '제4 패러다임(fourth paradigm)'으로 명명되며, 이론(제1 패러다임), 실험(제2 패러다임), 시뮬레이션(제3 패러다임)에 이어 데이터 중심 과학 발견의 새로운 시대를 열고 있음(Hey et al., 2009)
- 전통적으로 과학적 진보는 가설이나 이론을 먼저 수립한 후 이를 검증하거나 반증하기 위한 데이터를 수집하는 방식으로 이루어졌으나, 빅데이터 시대에는 사전 이론 없이 지속적으로 축적된 방대한 데이터로부터 새로운 지식을 발견할 수 있는 기회가 제공됨(Karpatne et al., 2017)
- 연구자들은 이제 대규모 데이터셋을 활용하여 시뮬레이션을 수행하고, 모델링에 참여하며, 이전에 알려지지 않은 인과관계를 발견하고, 새로운 이론을 구축할 수 있음(Kelling et al., 2009)

나 과학 데이터의 폭발적 증가

- 현대의 과학 기기와 센서 기술의 발전으로 전 세계적으로 방대한 양의 과학 데이터가 일상적으로 생성되고 있음(Fataliyev & Mehdiyev, 2019)
- 이러한 데이터의 양과 다양성은 수동 분석의 능력을 훨씬 초과하며, 실제로 전통적인 데이터베이스 관리 시스템의 처리 능력도 넘어섬(Provost & Fawcett, 2013)
- 컴퓨팅 파워의 증가, 네트워크의 편재성, 대규모 데이터셋을 처리하고 분석하기 위한 계산 및 통계 방법론의 정교화에 힘입어 데이터 과학(data science) 분야가 크게 확장됨(Karpatne et al., 2017)
- 전 세계의 연구 기관과 실험실에서 대형 과학 장비를 통해 지속적으로 상당한 양의 과학 데이터를 생성하고 있으며, 이러한 데이터는 종종 공개적으로 보급됨(Michener & Jones, 2012)

다 과학교육에의 시사점

- 과학 연구 패러다임의 변화는 과학교육에도 중대한 영향을 미치고 있으며, 특히 STEM(Science, Technology, Engineering, Mathematics) 전공 학생들의 데이터 관련 역량 강화가 시급한 과제로 부상함(Qiao et al., 2024)
- 전통적인 과학 리터러시(science literacy)의 데이터 관련 요구사항은 주로 실험을 통해 얻은 소규모 데이터를 다루는 기술 습득에 중점을 두었으나, 이는 더 이상 대규모 과학 데이터가 주도하는 진화하는 연구 패러다임의 요구를 충족시키기에 충분하지 않음(Qin & D'ignazio, 2010)
- 미래의 연구자들은 실험실 내에서 프로젝트별 데이터를 설계하고 수집하는 것뿐만 아니라, 전 세계적으로 공개 가능한 방대한 과학 데이터 저장소에서 잠재적 가치를 추출하여 관심 있는 과학 문제를 해결하는

능력을 갖추어야 함(Mustafee et al., 2020)

라 국내외 과학교육 표준의 변화

- 미국의 차세대 과학교육표준(Next Generation Science Standards, NGSS)은 8가지 과학 및 공학 실천(Science and Engineering Practices) 중 하나로 '데이터 분석과 해석(Analyzing and Interpreting Data)'을 명시적으로 포함함(NGSS Lead States, 2013)
- NGSS는 과학 탐구가 단순히 사실적 지식의 습득이 아니라 실제 과학자들이 수행하는 실천(practices)에 학생들을 참여시키는 것을 강조함
- 과학적 조사는 분석을 위해 데이터를 생성하며, 데이터의 패턴과 경향이 항상 명확하지 않기 때문에 과학자들은 표작성(tabulation), 그래픽 해석, 시각화, 통계 분석을 포함한 다양한 도구를 사용하여 데이터의 유의미한 특징과 패턴을 식별함(NGSS Lead States, 2013)
- 국내에서도 2022 개정 교육과정에서 디지털 소양과 데이터 리터러시의 중요성이 강조되고 있으며, 과학교육에서도 데이터 기반 탐구 역량 함양이 중요한 목표로 설정됨

마 연구의 필요성

- 지능형 과학실 ON과 같은 첨단 디지털 학습 환경의 구축으로 다양한 센서와 측정 도구를 통한 실시간 데이터 수집과 분석이 가능해짐
- IoT(Internet of Things) 기반 과학실험 장비의 도입으로 학생들이 다루는 데이터의 양과 복잡성이 증가하고 있음
- 최근 생성형 AI 기술의 급속한 발전은 데이터 분석, 시각화, 해석 등 탐구의 각 단계에서 학생들을 지원할 수 있는 새로운 가능성을 제공함
- 생성형 AI는 복잡한 데이터 분석 과정에서 학생의 인지적 파트너 역할을 수행하며, 개별화된 학습 지원을 가능하게 함
- 이러한 맥락에서 '데이터 기반 과학탐구'의 개념을 명확히 정의하고, 그 교육적 의의를 체계적으로 탐색할 필요가 있음
- 특히 지능형 과학실 환경에서 생성형 AI를 활용한 데이터 기반 탐구 지원 시스템 개발을 위해서는 이론적 기반이 되는 개념 정립이 선행되어야 함
- 데이터 리터러시, 과학 데이터 리터러시, 계산적 사고, 과학적 실천 등 관련 개념들 간의 관계를 명확히 하고, 데이터 기반 과학탐구의 고유한 특징과 구성요소를 밝힐 필요가 있음
- 이를 통해 한국과학창의재단의 지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템 개발에 견고한 이론적 토대를 제공할 수 있을 것으로 기대됨

2 핵심 개념 정의

가 과학 데이터

- 과학 데이터는 과학 연구에서 실험, 관측, 계산을 통해 수집되고 분석되는 정보로 정의됨(Demchenko et al., 2012; Fox & Hendler, 2011)

- 과학 데이터의 구체적 사례로는 물리학, 화학, 생물학의 실험 데이터, 행성 운동에 대한 관측 데이터, 대기 데이터 등이 포함됨(Qiao et al., 2024)
- 과학 데이터와 일반 데이터의 주요 차이점은 그 목적에 있음: 과학 데이터는 과학적 가설을 검증하거나 반증하고, 이론을 지지하거나 도전하며, 새로운 지식을 발견하는 데 사용되는 반면, 일반 데이터는 상업적, 정치적, 사회적 또는 기타 비과학적 응용에 사용될 수 있음(Fox & Hendler, 2011; Tenopir et al., 2011)
- 과학 데이터는 신뢰성과 재현성을 보장하기 위한 엄격한 기준을 준수함(Wilkinson et al., 2016)
- Wilkinson 등(2016)은 과학 데이터 관리와 관리를 위한 FAIR 원칙을 제시함: Findable(발견 가능), Accessible(접근 가능), Interoperable(상호운용 가능), Reusable(재사용 가능)
- 빅데이터 시대에 과학 데이터의 활용은 일반 데이터의 활용과 유사해짐: 사람들은 일반적으로 문제 해결에 있어서 데이터의 유용성을 기반으로 데이터의 가치를 평가하며, 이러한 목적을 위해 방대한 데이터 풀에서 필요한 데이터를 선택함(Pangrazio & Selwyn, 2019)

나 데이터 리터러시(Data Literacy)

- Shields(2004)는 데이터 리터러시의 초기 옹호자 중 한 명으로, 이를 "데이터에 접근하고, 조작하며, 제시하는 능력"으로 정의함
- 이후 데이터 리터러시는 연구자들로부터 점점 더 많은 관심을 받으며 개념적 프레임워크가 더욱 발전함(Calzada & Marzal, 2013; Kippers et al., 2018)
- Wolff 등(2016)은 데이터 리터러시를 "데이터 기반 사회를 위한 데이터의 이해 창출"로 확장하여 정의함
- Gummer와 Mandinach(2015)는 교육자를 위한 데이터 리터러시 개념적 프레임워크를 구축하면서, 데이터 식별, 수집, 분석, 해석, 그리고 의사결정으로의 전환을 포함하는 포괄적 개념으로 제시함
- 최근 연구들은 데이터 리터러시가 단순한 기술적 능력을 넘어 데이터 인식(data awareness), 데이터 윤리(data ethics), 데이터 보안(data security)에 대한 이해를 포함해야 함을 강조함(Sander, 2020; Pangrazio et al., 2019)
- Sander(2020)는 비판적 빅데이터 리터러시(critical big data literacy)의 개념을 제안하며, 데이터의 권력 관계, 편향, 사회적 영향에 대한 비판적 이해를 포함시킴
- 현재 데이터 리터러시는 다차원적 개념으로 발전하여, 데이터 관련 기술(skills), 인식(awareness), 윤리(ethics)를 모두 포괄함

다 과학 데이터 리터러시(Science Data Literacy, SDL)

- 과학 데이터 리터러시(SDL)는 과학 리터러시와 데이터 리터러시의 교차점에서 등장한 개념임
- Qin과 D'ignazio(2010)는 SDL을 "과학 데이터를 이해하고, 사용하며, 관리하는 능력"으로 정의함
- 그러나 보편적으로 인정되는 SDL의 개념적 프레임워크는 여전히 부재한 상황이었으며, 이는 SDL의 전체 범위에 대한 명확한 이해를 저해함(Qiao et al., 2024)
- Carlson 등(2011)은 지리정보학 커리큘럼을 기반으로 SDL 프레임워크를 구축하면서, 그래프와 차트 해석, 데이터로부터 결론 도출, 데이터 오용 인식과 같은 기술을 강조함
- 과학 데이터 생명주기(science data lifecycle) 이론은 SDL에 대한 또 다른 중요한 기여를 제공함: 이 이론은 데이터 생성부터 폐기까지 과학 데이터의 전체 생명주기에 대한 포괄적인 문서화와 관리를 옹호함(Ball, 2012; Michener & Jones, 2012)
- 최근 Qiao 등(2024)은 STEM 전공 대학생을 위한 체계적이고 포괄적인 SDL 개념적 프레임워크를 제시함

- 이 프레임워크는 SDL을 세 가지 주요 차원으로 구성함:
- 과학 데이터 인식(Science Data Awareness): 과학 데이터에 대한 인간의 인식과 이해로 정의되며, 세 가지 요소를 포함함
  - 데이터 민감성 인식(Data Acuity Awareness, DAA): 특정 문제 해결에 필요한 데이터를 식별하고 인식하는 능력
  - 데이터 가치 인식(Data Value Awareness, DVA): 데이터의 가치와 중요성을 인식하는 능력
  - 데이터 보안 인식(Data Security Awareness, DSA): 데이터 보호와 보안의 중요성을 인식하는 능력
- 과학 데이터 기술(Science Data Skills): 과학 탐구와 실천에서 과학 데이터를 활용하는 능력으로 정의되며, 여섯 가지 요소를 포함함
  - 데이터 발견 및 수집 기술(Data Finding and Collection Skills, DFACS)
  - 데이터 조직 및 관리 기술(Data Organization and Management Skills, DOAMS)
  - 데이터 처리 및 분석 기술(Data Processing and Analysis Skills, DPAAS)
  - 데이터 해석 기술(Data Interpretation Skills, DIS)
  - 데이터 추론 기술(Data Deduction Skills, DDS)
  - 데이터 소통 및 공유 기술(Data Communication and Sharing Skills, DCASS)
- 과학 데이터 규제와 윤리(Science Data Regulations and Ethics): 과학 데이터의 수집, 사용, 공유를 규율하는 법적, 규제적, 윤리적 규범에 대한 준수로 정의되며, 두 가지 요소를 포함함
  - 데이터 법률 및 규정(Data Laws and Regulations, DLAR)
  - 데이터 윤리(Data Ethics, DE)
- 이러한 SDL 프레임워크는 전문가 델파이 조사를 통해 검증되었으며, STEM 전공 학생들이 과학 데이터를 다루는 데 필요한 필수 리터러시를 체계적으로 제시함(Qiao et al., 2024)
- SDL은 일반 데이터 리터러시와 구별되는 특징을 가짐: SDL은 과학 데이터의 특수성(엄격성, 재현성, 과학적 목적)을 고려하며, 과학 연구자, 엔지니어, STEM 전공 학생 등 과학 분야 종사자를 주요 대상으로 함

라 데이터 기반 과학탐구(Data-Driven Science Inquiry)

- 데이터 기반 과학탐구는 앞서 논의한 개념들을 통합하여 과학교육의 맥락에서 구체화한 탐구 방식으로 이해할 수 있음
- 본 연구에서는 데이터 기반 과학탐구를 "과학적 현상에 대한 데이터의 수집, 처리, 분석, 시각화, 해석을 통해 패턴을 발견하고 과학적 지식을 구성하는 탐구 과정"으로 정의함
- 데이터 기반 과학탐구는 전통적 과학탐구의 논리적 구조(질문 제기, 가설 설정, 실험 설계, 데이터 수집, 분석, 결론 도출)를 유지하면서도, 데이터 과학의 방법론과 도구를 통합하는 특징을 가짐
- 특히 다음과 같은 요소들이 데이터 기반 과학탐구의 핵심을 구성함:
  - 실세계의 대규모 과학 데이터 활용
  - 계산적 도구와 통계적 방법의 사용
  - 데이터 시각화를 통한 패턴 발견
  - 증거 기반 논증(evidence-based argumentation)
  - 데이터 생명주기 전반에 걸친 체계적 접근
- 데이터 기반 과학탐구는 귀납적 접근과 연역적 접근을 모두 포함함: 학생들은 데이터로부터 패턴을

발견하여 가설을 생성할 수도 있고(귀납), 기존 이론에 기반하여 가설을 설정한 후 데이터로 검증할 수도 있음(연역)

- 이러한 탐구 방식은 NGSS의 과학적 실천, 특히 '데이터 분석과 해석', '수학적 및 계산적 사고 활용', '증거로부터 논증에 참여하기' 등과 밀접하게 연결됨(NGSS Lead States, 2013)
- 데이터 기반 과학탐구는 단순히 기술적 능력만을 강조하는 것이 아니라, SDL의 세 가지 차원(인식, 기술, 윤리)을 모두 포괄하는 종합적 탐구 활동임

3 데이터 기반 과학탐구 연구 동향

가 과학 연구의 패러다임 변화: 제4 패러다임

- 과학적 발견의 방법론은 역사적으로 세 가지 주요 패러다임을 거쳐 발전해 옴
- 제1 패러다임(경험적 과학): 수천 년 전부터 과학자들은 자연 현상을 관찰하고 기술하는 경험적 방법을 사용함
- 제2 패러다임(이론적 과학): 지난 수백 년 동안 과학자들은 일반화와 모델을 사용하여 자연 현상을 설명하는 이론적 접근을 발전시킴
- 제3 패러다임(계산 과학): 지난 수십 년 동안 컴퓨터 시뮬레이션을 통해 복잡한 현상을 모델링하는 계산 과학이 등장함
- 제4 패러다임(데이터 집약적 과학): 21세기에는 데이터 집약적 과학 발견이 새로운 패러다임으로 부상함(Tansley & Tolle, 2009; Hey et al., 2009)
- 제4 패러다임의 핵심 특징은 이론이나 가설 없이 방대한 데이터로부터 직접 패턴과 지식을 추출할 수 있다는 점임
- Kelling 등(2009)은 생물다양성 연구에서 데이터 집약적 과학이 어떻게 새로운 패러다임으로 작동하는지 구체적으로 제시함: 전통적 방식에서는 연구자가 가설을 세우고 이를 검증하기 위한 데이터를 수집했으나, 데이터 집약적 접근에서는 먼저 대규모 데이터를 수집하고 이로부터 패턴을 발견하여 새로운 가설을 생성함
- Karpatne 등(2017)은 이를 '이론 기반 데이터 과학(theory-guided data science)'으로 발전시켜, 데이터 주도적 발견과 과학 이론을 통합하는 새로운 패러다임을 제안함
- 이러한 패러다임 전환은 과학교육에 중요한 시사점을 제공함: 학생들은 단순히 확립된 과학 지식을 학습하는 것을 넘어, 데이터로부터 새로운 지식을 발견하는 과정에 참여해야 함

나 NGSS와 과학적 실천(Science Practices)

- 미국의 차세대 과학교육표준(NGSS)은 과학교육의 새로운 방향을 제시하며, 과학적 실천(science practices)을 핵심 요소로 강조함(NGSS Lead States, 2013)
- NGSS는 '과학적 과정(science processes)' 또는 '탐구 기능(inquiry skills)'이 아닌 '실천(practices)'이라는 용어를 의도적으로 사용함: 이는 과학적 탐구가 단순한 기능뿐만 아니라 특정 맥락에서 적용되는 지식을 요구한다는 점을 강조함(NRC, 2012)
- NGSS는 8가지 과학 및 공학 실천을 제시함:
  - 질문하기(과학용) 및 문제 정의하기(공학용)
  - 모델 개발 및 사용하기

- 탐구 계획 및 수행하기
- 데이터 분석 및 해석하기
- 수학적 및 계산적 사고 활용하기
- 설명 구성하기(과학용) 및 해결책 설계하기(공학용)
- 증거로부터 논증에 참여하기
- 정보 획득, 평가, 소통하기

- 이 중 '데이터 분석 및 해석하기(Analyzing and Interpreting Data)'는 데이터 기반 과학탐구의 핵심 실천임
- NGSS는 데이터 분석과 해석에 대해 다음과 같이 설명함: "과학적 조사는 분석을 위해 데이터를 생성함. 데이터의 패턴과 경향이 항상 명확하지 않기 때문에, 과학자들은 표작성, 그래픽 해석, 시각화, 통계 분석을 포함한 다양한 도구를 사용하여 데이터의 유의미한 특징과 패턴을 식별함"(NGSS Lead States, 2013)
- 또한 '수학적 및 계산적 사고 활용하기(Using Mathematics and Computational Thinking)' 실천은 데이터 기반 탐구와 직접적으로 연결됨: "현대 기술은 대규모 데이터셋의 수집을 훨씬 용이하게 하여 분석을 위한 2차 자료를 제공함"(NGSS Lead States, 2013)
- NGSS는 이러한 실천들이 분리된 것이 아니라 의도적으로 중첩되고 상호 연결된다는 점을 강조함: 데이터 기반 과학탐구에서 학생들은 질문하기, 데이터 수집 계획하기, 데이터 분석하기, 증거로부터 논증하기 등 여러 실천을 통합적으로 수행함
- NGSS의 3차원 학습 프레임워크는 과학적 실천(Science and Engineering Practices), 교차 개념(Crosscutting Concepts), 핵심 개념(Disciplinary Core Ideas)을 통합함: 이는 학생들이 실천에 참여하지 않고는 과학적 개념을 완전히 이해할 수 없으며, 특정 내용의 맥락 없이는 실천에서 역량을 학습하거나 보여줄 수 없다는 인식에 기반함(NRC, 2012)

다 데이터 리터러시 연구의 발전

- 데이터 리터러시에 대한 연구는 초기에는 주로 데이터를 다루는 기본적 기술에 초점을 맞추었으나, 점차 발전하면서 데이터 윤리와 데이터 인식에 대한 관심이 증가함(Qiao et al., 2024)
- Shields(2004)의 초기 연구 이후, 데이터 리터러시는 정보 리터러시(information literacy), 통계 리터러시(statistical literacy)와 구별되는 독립적 개념으로 발전함
- Calzada와 Marzal(2013)은 데이터 리터러시를 정보 리터러시 프로그램에 통합하기 위한 핵심 역량과 내용을 제시함: 데이터 인식, 데이터 보존, 데이터 전환, 데이터 처리, 데이터 분석, 데이터 시각화 및 프레젠테이션 등
- Kippers 등(2018)은 교육자들이 데이터 활용 개입 동안 무엇을 배우고 어떤 어려움을 겪는지 연구하며, 데이터 리터러시가 실제 현장에서 어떻게 적용되는지 탐구함
- Wolff 등(2016)은 데이터 리터러시를 단순한 기술적 능력을 넘어 데이터 기반 사회에 대한 비판적 이해를 포함하는 개념으로 확장함
- Sander(2020)는 비판적 빅데이터 리터러시(critical big data literacy) 개념을 제안하며, 데이터의 수집, 분석, 활용 과정에서 발생하는 권력 관계, 편향, 프라이버시 침해 등에 대한 비판적 인식의 중요성을 강조함
- 최근 연구들은 데이터 리터러시를 다차원적 구조로 개념화함: 데이터 관련 인식(awareness), 기술(skills), 윤리(ethics)를 통합하는 포괄적 프레임워크가 제시됨(Pangrazio et al., 2019; Gebre, 2022)

라 과학 데이터 리터러시(SDL)의 등장과 발전

- 과학 데이터 리터러시(SDL)는 과학 리터러시(science literacy)와 일반 데이터 리터러시의 교차점에서 등장한 특수화된 개념임
- 과학 리터러시 연구는 오랫동안 데이터의 분석, 이해, 해석을 과학 리터러시의 중요한 측면으로 인식해 옴(Shaffer et al., 2019; Sholikah & Pertiwi, 2021)
- OECD(2019)의 PISA 2018 프레임워크는 과학 리터러시를 "현상을 과학적으로 설명하고, 과학 탐구를 평가 및 설계하며, 데이터와 증거를 과학적으로 해석하는 능력"으로 정의하며 데이터의 중요성을 명시함
- 그러나 기존 과학 리터러시 프레임워크의 데이터 관련 요구사항은 주로 탐구 실험에서 파생된 소규모 데이터를 다루는 기술 습득에 중점을 두어, 대규모 과학 데이터 시대의 요구를 충족시키기에 부족함(Qin & D'ignazio, 2010)
- Qin과 D'ignazio(2010)는 SDL 개념을 처음 제안하며 메타데이터의 중심적 역할을 강조함: 과학 데이터 리터러시 교육에서 학생들이 메타데이터를 이해하고 활용하는 능력이 핵심적임
- Carlson 등(2011)은 대학생과 연구 교수진을 대상으로 데이터 정보 리터러시 요구를 조사하고, 학문 분야별로 차이가 있음을 발견함
- 과학 데이터 생명주기 이론은 SDL에 대한 중요한 이론적 기여를 제공함(Ball, 2012; Michener & Jones, 2012): 이 이론은 데이터 계획, 수집, 관리, 분석 및 시각화, 공유 및 보존, 발견, 재사용을 포함하는 전체 생명주기를 다룸
- Faundeen 등(2014)은 미국 지질조사국(USGS)의 과학 데이터 생명주기 모델을 제시하며, 각 단계에서 필요한 역량을 구체화함
- 최근 Qiao 등(2024)의 연구는 SDL에 대한 가장 포괄적이고 체계적인 프레임워크를 제시함: 이들은 문헌 검토와 전문가 델파이 조사를 통해 SDL의 세 차원(인식, 기술, 윤리)과 11개 요소를 도출하고, 이를 측정할 수 있는 평가 도구를 개발함
- 이러한 SDL 프레임워크는 데이터 기반 과학탐구의 이론적 기반을 제공하며, 학생들이 발달시켜야 할 구체적 역량을 명확히 함

마 계산적 사고와 데이터 기반 탐구의 연결

- Wing(2006)은 계산적 사고(computational thinking)를 "컴퓨터 과학의 기본 개념을 활용하여 문제를 해결하고, 시스템을 설계하며, 인간 행동을 이해하는 것"으로 정의함
- 계산적 사고는 데이터 기반 과학탐구와 밀접하게 연결됨: 대규모 데이터를 처리하고 분석하기 위해서는 알고리즘적 사고, 패턴 인식, 추상화, 자동화 등 계산적 사고의 핵심 요소가 필요함
- Karpatne 등(2017)은 이론 기반 데이터 과학에서 계산적 방법이 과학적 발견을 가능하게 하는 방식을 설명함: 기계 학습, 데이터 마이닝, 통계적 모델링 등의 계산적 도구가 데이터로부터 패턴을 추출하고 예측 모델을 구축하는 데 핵심적 역할을 함
- NGSS의 '수학적 및 계산적 사고 활용하기' 실천은 이러한 연결을 명시적으로 반영함: 학생들은 데이터 분석에서 수학적 개념뿐만 아니라 코딩, 알고리즘, 시뮬레이션 등 계산적 도구를 활용해야 함
- 과학교육 연구자들은 계산적 사고를 과학 탐구에 통합하는 방법을 탐구해 옴: 예를 들어, 학생들이 Python이나 R과 같은 프로그래밍 언어를 사용하여 데이터를 처리하고 분석하는 활동은 계산적 사고와 과학적 실천을 동시에 발달시킴
- 데이터 기반 과학탐구는 계산적 사고를 authentic한 과학적 맥락에서 적용할 수 있는 기회를 제공함

바 실세계 데이터(Authentic Data) 활용의 교육적 가치

- Kjelvik과 Schultheis(2019)는 과학 연구로부터 얻은 실제 데이터(authentic data)를 사용하는 것이 학생들의 SDL을 지원하는 잠재력을 탐구함
- 실세계 데이터는 교과서나 실험 매뉴얼에 제시된 '깨끗한(clean)' 데이터와 달리 복잡성, 불완전성, 불확실성을 포함함
- Schultheis와 Kjelvik(2020)은 'messy data'를 사용하는 것이 데이터 리터러시를 촉진하고 과학의 본성(nature of science)을 드러낸다고 주장함: 학생들은 결측치, 이상치, 측정 오차 등 실제 연구에서 마주치는 데이터의 특성을 경험함
- 실세계 데이터 활용은 다음과 같은 교육적 이점을 제공함:
  - 과학 연구의 authentic한 경험 제공
  - 데이터 청소(data cleaning)와 전처리(preprocessing) 기술 발달
  - 비판적 사고와 문제 해결 능력 향상
  - 과학적 지식이 불확실성 속에서 구성됨을 이해
  - 학생들의 동기와 참여 증대
- 공개 과학 데이터(open science data)의 증가로 교육 현장에서 실세계 데이터 접근이 용이해짐: 예를 들어, NASA, NOAA, GenBank 등의 공개 데이터베이스는 학생들이 실제 과학 데이터로 탐구할 수 있는 기회를 제공함
- 데이터 기반 과학탐구는 이러한 실세계 데이터를 교육적으로 활용하는 체계적 접근을 제공함

4 데이터 기반 과학탐구의 특징

가 대규모 과학 데이터의 활용

- 데이터 기반 과학탐구의 가장 두드러진 특징은 대규모 과학 데이터를 탐구의 중심에 두는 것임
- 전통적 과학탐구에서는 학생들이 직접 수행한 소규모 실험에서 생성된 데이터(일반적으로 수십 개 이하의 데이터 포인트)를 주로 다루었으나, 데이터 기반 탐구에서는 수백, 수천, 또는 그 이상의 데이터 포인트를 포함하는 대규모 데이터셋을 다룸(Kelling et al., 2009)
- 현대 과학 기기는 엄청난 양의 데이터를 생성함: 예를 들어, 차세대 염기서열 분석(Next-Generation Sequencing)은 단일 실험에서 수십억 개의 DNA 서열을 생성하며, 천문학의 대형 망원경은 밤마다 테라바이트 단위의 이미지 데이터를 생성함
- 지능형 과학실 환경에서는 다양한 센서(온도, 습도, 압력, pH, 빛의 세기 등)가 실시간으로 데이터를 수집하여 대규모 데이터셋을 생성할 수 있으며, 이는 학생들에게 authentic한 데이터 기반 탐구 경험을 제공함

나 데이터 생명주기의 전 과정 경험

- 데이터 기반 과학탐구는 데이터 생명주기(data lifecycle)의 전 과정을 포괄함
- Michener와 Jones(2012)가 제시한 데이터 생명주기는 다음 단계를 포함함:
  - 계획(Plan): 연구 질문 설정, 데이터 수집 전략 수립, 필요한 자원 파악

- 수집(Collect): 실험, 관측, 시뮬레이션 등을 통한 데이터 획득
- 보증(Assure): 데이터 품질 확인, 품질 관리(QA/QC) 절차 적용
- 기술(Describe): 메타데이터 생성, 데이터 문서화
- 보존(Preserve): 데이터 백업, 장기 저장 전략 수립
- 발견(Discover): 기존 데이터셋 검색 및 접근
- 통합(Integrate): 여러 출처의 데이터 결합
- 분석(Analyze): 통계적 분석, 시각화, 모델링 수행

- 전통적 과학교육에서는 주로 데이터 수집과 분석에만 초점을 맞추었으나, 데이터 기반 탐구는 전체 생명주기를 다룸으로써 더 완전한 과학 연구 경험을 제공함
- 특히 데이터 관리와 보존, 메타데이터 생성, 데이터 발견과 재사용 등은 빅데이터 시대의 과학 연구에서 필수적이지만 전통적 과학교육에서는 소홀히 다루어졌던 영역임
- 학생들은 데이터 생명주기 전반을 경험함으로써 데이터의 지속가능한 관리와 재사용의 중요성을 이해하게 됨

다 계산적 도구와 프로그래밍의 활용

- 데이터 기반 과학탐구는 계산적 도구와 프로그래밍 언어의 활용을 필수적으로 요구함
- 대규모 데이터를 효과적으로 처리하고 분석하기 위해서는 Excel과 같은 기본적인 스프레드시트 도구를 넘어 Python, R, MATLAB 등의 프로그래밍 언어나 Jupyter Notebook과 같은 계산 환경이 필요함
- 생성형 AI는 학생들이 프로그래밍을 학습하고 활용하는 데 강력한 지원을 제공할 수 있음: 코드 생성, 오류 디버깅, 개념 설명 등을 통해 프로그래밍 진입 장벽을 낮출 수 있음

라 데이터 시각화의 중심적 역할

- 데이터 기반 과학탐구에서 데이터 시각화(data visualization)는 단순히 결과를 제시하는 도구를 넘어 탐구 과정 자체의 핵심 요소임
- Fox와 Hendler(2011)는 과학 데이터 시각화가 과학적 발견의 방정식을 어떻게 변화시키는지 논의하며, 시각화가 패턴 발견과 가설 생성을 촉진한다고 주장함
- 데이터 시각화는 다음과 같은 기능을 수행함:
  - 탐색적 데이터 분석(Exploratory Data Analysis, EDA): 데이터의 전반적 특성, 분포, 이상치, 패턴을 빠르게 파악
  - 가설 생성: 시각적 패턴으로부터 새로운 질문과 가설을 도출
  - 모델 검증: 통계 모델이나 시뮬레이션 결과를 데이터와 비교
  - 소통과 설득: 발견한 패턴과 결론을 효과적으로 전달
- 학생들은 다양한 시각화 유형(산점도, 히스토그램, 박스플롯, 히트맵, 네트워크 그래프 등)을 경험하고, 연구 질문과 데이터 특성에 맞는 적절한 시각화를 선택하는 능력을 발달시켜야 함

마 통계적 사고와 불확실성의 이해

- 데이터 기반 과학탐구는 통계적 사고(statistical thinking)와 불확실성(uncertainty)에 대한 이해를 요구함
- 대규모 데이터에서는 측정 오차, 표본 변동성, 무작위성 등 다양한 불확실성의 원천이 존재하며, 이를 정량화하고 해석하는 것이 중요함
- 학생들이 발달시켜야 할 통계적 개념은 다음을 포함함:
  - 기술 통계: 평균, 중앙값, 표준편차, 사분위수 등을 통한 데이터 요약
  - 분포의 이해: 정규분포, 왜도, 첨도 등 데이터 분포의 특성
  - 상관과 인과: 변수 간 관계의 본질 이해
  - 표본과 모집단: 표본으로부터 모집단에 대한 추론
  - 가설 검정: p-값, 신뢰구간, 유의수준 등의 개념
  - 효과 크기: 통계적 유의성과 실질적 중요성의 구별
- 불확실성의 정량화는 과학의 본성을 이해하는 데 핵심적임: 과학적 지식은 절대적 확실성이 아니라 증거에 기반한 확률적 진술임을 학생들이 인식해야 함
- NGSS는 학생들이 "조사에서 오류의 원인을 식별하고 결과의 확실성 정도를 계산"해야 함을 명시함(NGSS Lead States, 2013)

바 증거 기반 논증과 과학적 추론

- 데이터 기반 과학탐구는 데이터를 증거로 활용하여 논증(argumentation)을 구성하는 과정을 포함함
- NGSS의 '증거로부터 논증에 참여하기(Engaging in Argument from Evidence)' 실천은 이러한 측면을 강조함: "논증은 증거 기반 결론과 해결책에 도달하는 과정임. 과학과 공학에서 증거에 기반한 추론과 논증은 자연 현상에 대한 최선의 설명이나 설계 문제에 대한 최선의 해결책을 식별하는 데 필수적임"(NGSS Lead States, 2013)
- 과학적 논증의 구조는 Toulmin(1958)의 논증 패턴을 따름:
  - 주장(Claim): 데이터로부터 도출된 결론이나 답변
  - 증거(Evidence): 주장을 지지하는 데이터
  - 추론(Reasoning): 증거가 주장을 어떻게 지지하는지 설명하는 논리
  - 반증(Rebuttal): 대안적 설명이나 제한점 고려
- 데이터 기반 과학탐구에서 학생들은 다음을 수행해야 함:
  - 데이터로부터 패턴을 식별하고 이를 바탕으로 주장을 구성함
  - 주장을 지지하는 적절한 증거를 선택하고 제시함
  - 과학적 원리나 개념을 활용하여 증거와 주장을 연결하는 추론을 제공함
  - 경쟁하는 주장들을 비교하고 평가함
  - 데이터의 한계와 대안적 해석을 고려함
- 이러한 논증 과정은 비판적 사고와 메타인지를 발달시키며, 과학 지식이 증거와 논리적 추론을 통해 구성됨을 이해하게 함

사 협력적 탐구와 데이터 공유

- 대규모 데이터를 다루는 프로젝트는 종종 여러 사람의 전문성과 노력을 필요로 함: 데이터 수집, 전처리, 분석, 해석, 시각화 등 각 단계에서 협력이 요구됨
- 데이터 공유(data sharing)는 현대 과학의 핵심 원칙임: Tenopir 등(2011)의 연구에 따르면, 과학자들은 데이터 공유가 과학 발전을 촉진한다는 데 동의하지만, 실제로는 다양한 장벽(시간, 자원, 지적 재산권 우려 등)이 존재함
- Wilkinson 등(2016)의 FAIR 원칙은 과학 데이터가 Findable(발견 가능), Accessible(접근 가능), Interoperable(상호운용 가능), Reusable(재사용 가능)해야 함을 강조하며, 이는 데이터 공유와 재사용의 중요성을 반영함
- 데이터 기반 과학탐구 교육에서 협력과 공유는 다음과 같이 구현될 수 있음:
  - 학생 팀이 공동으로 대규모 데이터 프로젝트를 수행함
  - 데이터, 코드, 분석 결과를 팀 내에서 공유하고 검토함
  - GitHub와 같은 버전 관리 시스템을 사용하여 협력적 작업을 관리함
  - 다른 팀이나 클래스가 수집한 데이터를 활용하거나 자신의 데이터를 공유함
  - 데이터 공유의 윤리적, 법적 측면(개인정보 보호, 저작권 등)을 학습함
- 공개 과학 데이터베이스의 활용은 학생들이 실제 과학 연구에서 다루는 authentic한 문제(기후 변화, 전염병 확산, 생물다양성 감소, 대기 오염 등)를 탐구할 수 있게 함. 실세계 맥락은 다음과 같은 교육적 이점을 제공함:
  - 학생들의 동기와 참여를 높임
  - 과학 학습의 관련성(relevance)을 증대시킴
  - 과학-기술-사회(Science-Technology-Society, STS)의 연결을 이해하게 함
  - 시민 과학(citizen science)과 사회적 실천으로서의 과학을 경험함

5 교육적 의의

가 과학 데이터 리터러시(SDL) 함양

- 데이터 기반 과학탐구의 가장 직접적인 교육적 의의는 학생들의 과학 데이터 리터러시를 체계적으로 발달시킨다는 점임
- Qiao 등(2024)이 제시한 SDL의 세 가지 차원이 데이터 기반 탐구 과정에서 자연스럽게 통합됨:
- 과학 데이터 인식(Science Data Awareness) 측면:
  - 학생들은 탐구 질문에 답하기 위해 어떤 데이터가 필요한지 식별하는 능력(데이터 민감성 인식)을 발달시킴
  - 방대한 데이터 중에서 유용하고 신뢰할 수 있는 데이터를 선별하는 능력(데이터 가치 인식)을 기름
  - 데이터 보호, 개인정보, 연구 윤리의 중요성(데이터 보안 인식)을 이해함
- 과학 데이터 기술(Science Data Skills) 측면:
  - 데이터 발견 및 수집: 공개 데이터베이스 탐색, 효과적인 검색 전략 사용
  - 데이터 조직 및 관리: 체계적인 파일 명명, 메타데이터 생성, 데이터베이스 구조화
  - 데이터 처리 및 분석: 데이터 청소, 통계 분석, 기계 학습 알고리즘 적용

- 데이터 해석: 분석 결과의 의미를 과학적 맥락에서 이해
- 데이터 추론: 데이터에 기반한 결론 도출 및 의사결정
- 데이터 소통 및 공유: 효과적인 시각화, 명확한 설명, 데이터와 코드의 공유

- 과학 데이터 규제와 윤리(Science Data Regulations and Ethics) 측면:
  - 데이터 사용에 관한 법적 규제(저작권, 라이선스 등)를 이해함
  - 연구 윤리(데이터 조작 금지, 출처 명시, 이해 충돌 관리 등)를 내면화함
  - FAIR 원칙과 개방 과학의 가치를 인식함
- 이러한 SDL의 함양은 학생들이 미래의 과학자, 엔지니어, 또는 데이터 중심 직업에서 성공하기 위한 필수 역량을 제공함
- 더 나아가 SDL은 과학 분야를 넘어 일반 시민으로서 데이터 기반 사회에서 정보에 기반한 의사결정을 내리는 데 필요한 기초 소양을 제공함

나 21세기 핵심역량 개발

1) 비판적 사고(Critical Thinking)

- 학생들은 데이터의 신뢰성과 타당성을 평가하는 비판적 안목을 발달시킴: 데이터의 출처, 수집 방법, 표본 크기, 측정 오차 등을 고려하여 데이터의 품질을 판단함
- 상관관계와 인과관계를 구별하고, 데이터로부터 정당화되지 않은 결론을 도출하지 않도록 주의하는 능력을 기름
- 데이터 시각화가 어떻게 오해를 유발할 수 있는지 이해하고, 데이터의 조작이나 왜곡을 식별하는 능력을 발달시킴
- 대안적 설명과 해석을 고려하며, 증거의 강도를 평가하는 능력을 향상시킴

2) 문제 해결(Problem Solving)

- 복잡하고 구조화되지 않은(ill-structured) 실세계 문제를 다루는 경험을 제공함
- 큰 문제를 관리 가능한 하위 문제로 분해하고, 각 단계를 체계적으로 해결하는 전략을 학습함
- 예상치 못한 결과나 오류에 직면했을 때 문제를 진단하고 해결책을 찾는 회복탄력성(resilience)을 발달시킴
- 여러 접근 방법을 시도하고 최적의 해결책을 찾는 창의적 문제 해결 능력을 기름

3) 협력(Collaboration)

- 데이터 기반 프로젝트는 종종 팀 작업을 요구하며, 이를 통해 협력 능력을 자연스럽게 발달시킴
- 역할 분담, 의사소통, 갈등 해결 등 효과적인 팀워크의 요소를 경험함
- 동료의 작업을 검토하고(peer review), 건설적인 피드백을 주고받는 능력을 발달시킴
- 다양한 관점과 전문성을 통합하여 더 나은 결과를 도출하는 협력적 지식 구성을 경험함

4) 의사소통(Communication)

- 복잡한 데이터 분석 결과를 다양한 청중(동료 학생, 교사, 지역사회 등)에게 효과적으로 전달하는 능력을 발달시킴
- 데이터 시각화를 통해 정보를 명확하고 설득력 있게 제시하는 기술을 학습함

- 과학적 글쓰기(scientific writing)를 통해 탐구 과정과 결과를 체계적으로 문서화하는 능력을 기름
- 구두 발표(oral presentation)를 통해 자신의 발견을 청중에게 명확하게 전달하는 능력을 향상시킨

5) 디지털 리터러시(Digital Literacy)

- 다양한 디지털 도구와 플랫폼을 효과적으로 활용하는 능력을 발달시킴
- 프로그래밍, 데이터베이스 관리, 클라우드 컴퓨팅 등 현대 과학 연구에 필수적인 디지털 기술을 습득함
- 디지털 정보의 신뢰성을 평가하고, 온라인 자원을 책임감 있게 사용하는 능력을 기름
- 디지털 도구를 창의적으로 활용하여 새로운 문제를 해결하는 역량을 발달시킴

6) 자기주도적 학습(Self-Directed Learning)

- 데이터 기반 탐구의 개방적이고 비선형적인 특성은 학생들이 스스로 학습 경로를 설정하고 관리하는 능력을 요구함
- 필요한 지식이나 기술을 스스로 식별하고, 적절한 자원을 찾아 학습하는 능력을 발달시킴
- 자신의 학습 과정을 모니터링하고, 필요에 따라 전략을 조정하는 메타인지 능력을 향상시킴
- 실패와 오류를 학습 기회로 활용하며, 지속적인 개선을 추구하는 성장 마인드셋(growth mindset)을 발달시킴

다 과학의 본성(Nature of Science) 이해

1) 과학 지식의 증거 기반

- 학생들은 과학적 주장이 데이터와 증거에 기반해야 함을 직접 경험함
- 충분한 증거 없이 결론을 내리는 것의 위험성을 이해하며, 증거의 질과 양이 결론의 신뢰도에 영향을 미침을 인식함
- 과학 지식이 개인적 의견이나 권위가 아니라 객관적 증거에 의해 정당화됨을 이해함

2) 과학 지식의 잠정성

- 새로운 데이터나 더 나은 분석 방법이 등장하면 기존 결론이 수정될 수 있음을 경험함
- 과학 지식이 절대적 진리가 아니라 현재까지의 최선의 설명임을 이해함
- 과학적 불확실성의 존재가 과학의 약점이 아니라 정직성과 엄격성의 표현임을 인식함

3) 과학적 방법의 다양성

- 과학 연구가 단일한 '과학적 방법'을 따르지 않으며, 연구 질문과 맥락에 따라 다양한 접근이 사용됨을 이해함
- 귀납적 접근(데이터로부터 패턴 발견)과 연역적 접근(가설 검증)이 모두 과학 연구에서 중요함을 인식함
- 탐색적 연구와 확인적 연구의 차이를 이해함

4) 과학의 창의성

- 데이터 분석과 해석 과정에서 창의성이 중요한 역할을 함을 경험함: 어떤 시각화를 선택할지, 어떤 패턴에 주목할지, 어떻게 결과를 해석할지 등에서 창의적 사고가 필요함

- 과학이 단순히 규칙을 따르는 기계적 과정이 아니라 상상력과 독창성을 요구하는 창조적 활동임을 이해함

5) 과학의 사회문화적 특성

- 데이터 해석이 연구자의 배경, 가정, 관점에 영향을 받을 수 있음을 인식함
- 과학 연구가 사회적 맥락 안에서 이루어지며, 연구 주제의 선택, 자원 배분, 결과의 적용 등이 사회적 가치와 우선순위에 영향을 받음을 이해함
- 과학자들 간의 협력, 경쟁, 동료 검토 등 과학 공동체의 사회적 과정을 이해함

라 실제 과학 연구 경험 제공

1) 참된(Authentic) 과학 실천

- 학생들은 교과서의 검증된 실험을 재현하는 것을 넘어, 실제 과학자들이 사용하는 데이터, 도구, 방법을 경험함
- 공개 과학 데이터베이스를 활용함으로써 현재 진행 중인 과학 연구에 기여하거나 그 결과를 활용할 수 있음
- 실제 연구에서 마주치는 도전(불완전한 데이터, 예상치 못한 결과, 모호한 패턴 등)을 경험함

2) 과학 연구 과정의 전체성(holism)

- 전통적 과학 수업에서는 종종 실험과 데이터 수집에만 초점을 맞추지만, 데이터 기반 탐구는 연구의 전 과정(질문 설정, 문헌 검토, 데이터 수집/발견, 분석, 해석, 소통)을 포괄함
- 각 단계가 어떻게 연결되며, 전체 연구의 질에 어떻게 기여하는지 이해함
- 연구 과정의 반복적이고 순환적인 특성을 경험함

2) 과학 연구 커뮤니티로의 진입

- 데이터와 결과를 공유하고, 동료의 작업을 검토하며, 건설적인 비판을 주고받는 과정을 통해 과학 공동체의 실천에 참여함
- 과학 커뮤니케이션의 형식(논문, 포스터, 발표 등)을 학습하고 활용함
- 시민 과학(citizen science) 프로젝트에 참여하여 실제 과학 연구에 기여할 수 있음

마 STEM 진로에 대한 관심과 준비도 향상

1) STEM 진로의 현실적 이해

- 학생들은 데이터 기반 탐구를 통해 현대 과학 및 공학 분야에서 데이터와 계산이 중심적 역할을 함을 이해함
- 데이터 과학자, 생물정보학자, 기후 과학자, 공중보건 연구자 등 데이터 중심 STEM 진로의 다양성을 인식함
- 실제 과학 연구의 도전과 보람을 경험하며, 자신의 관심과 적성을 탐색할 수 있음

2) 대학 및 직업을 위한 기술 습득

- 데이터 분석, 프로그래밍, 통계 등 현대 STEM 분야에서 필수적인 기술을 미리 습득함
- 대학 수준의 연구나 인턴십에서 요구되는 역량을 개발함
- 이력서, 대학 지원서, 장학금 신청서 등에서 강조할 수 있는 구체적 프로젝트 경험을 쌓음

3) 자기효능감(Self-Efficacy) 향상

- 복잡한 데이터 프로젝트를 성공적으로 완수하는 경험은 학생들의 STEM 자기효능감을 향상시킴
- "나는 과학 연구를 할 수 있다"는 자신감을 발달시키며, 이는 STEM 진로 선택에 긍정적 영향을 미침
- 특히 전통적으로 STEM 분야에서 소외되었던 집단(여성, 소수 인종 등)의 학생들에게 STEM 정체성을 형성할 기회를 제공함

바 개별화된 학습 기회 제공

1) 수준별 차별화

- 같은 데이터셋이라도 학생들의 수준에 따라 다른 깊이와 복잡성으로 탐구할 수 있음
  - 초보 학습자: 기본적 기술 통계와 간단한 시각화
  - 중급 학습자: 다변량 분석과 상관관계 탐색
  - 고급 학습자: 회귀 분석, 기계 학습, 복잡한 모델링
- 학생들은 자신의 속도로 학습하며, 준비가 되었을 때 더 도전적인 과제로 나아갈 수 있음

2) 관심사 기반 학습

- 학생들은 자신의 관심사와 연결된 데이터를 선택하여 탐구할 수 있음: 스포츠에 관심 있는 학생은 스포츠 통계 데이터를, 환경에 관심 있는 학생은 환경 데이터를 다룰 수 있음
- 개인적으로 의미 있는 질문을 설정하고 탐구하는 기회는 동기와 참여를 크게 높임

3) 다양한 학습 스타일 지원

- 시각적 학습자는 데이터 시각화에 강점을 발휘할 수 있음
- 논리적/수학적 학습자는 통계 분석과 모델링에 흥미를 느낄 수 있음
- 언어적 학습자는 결과 해석과 과학적 글쓰기에서 능력을 발휘할 수 있음
- 대인관계 학습자는 협력적 데이터 프로젝트에서 강점을 보일 수 있음
- 이러한 다양성은 모든 학생이 자신의 강점을 발견하고 발휘할 기회를 제공함

2절 생성형 AI의 교육적 활용과 효과

1 생성형 AI의 교육적 활용 유형

가 생성형 AI의 교육적 활용 개요

- 생성형 AI(Generative AI)는 2022년 말 ChatGPT의 출시 이후 교육 분야에서 가장 빠르게 확산된 기술 중 하나로, 교수학습 방식에 근본적인 변화를 가져오고 있음(Kasneci et al., 2023).
  - 생성형 AI는 자연어 처리 기술을 기반으로 인간과 유사한 텍스트를 생성하고, 맥락에 적합한 응답을 제공함으로써 교육적 상호작용을 지원하는 새로운 가능성을 제시함(Baidoo-Anu & Owusu, 2023).
  - UNESCO는 교육에서의 기술 활용에 대한 책임 있는 접근을 강조하며, 기술 변화의 급격한 속도에 맞추어 거버넌스와 규제의 필요성을 제기하였음(Azoulay, 2023).
  - 생성형 AI 기술은 전통적인 "one-size-fits-all" 교수 모델의 한계를 극복하고, 학습자의 다양한 요구에 맞춤형으로 대응할 수 있는 잠재력을 보유함(Samala et al., 2024).
  - 최근 체계적 문헌 연구에 따르면, 2023년부터 2024년 사이 고등교육에서의 생성형 AI 활용에 관한 실증 연구가 급격히 증가하였으며, 교수학습 및 기관 운영에 미치는 영향에 대한 학술적 관심이 확대되고 있음(Weng et al., 2024).

- **학습 성과에 대한 효과**
  - Deng et al. (2024)의 메타분석 연구에서는 ChatGPT를 활용한 69개의 실험 연구를 분석한 결과, 생성형 AI 활용이 학업 성취도, 정의적-동기적 상태, 고차적 사고 성향을 전반적으로 향상시키는 것으로 나타남.
  - 중국 대학생 323명을 대상으로 한 종단 연구에서 생성형 AI의 상호작용 품질과 출력 품질이 학습 동기와 창의적 자기효능감에 긍정적인 영향을 미치는 것으로 확인됨. 또한 실험 연구에서 AI 튜터를 활용한 학생들이 교실 내 능동적 학습에 비해 더 짧은 시간에 유의미하게 더 많은 학습을 달성한 것으로 보고됨.

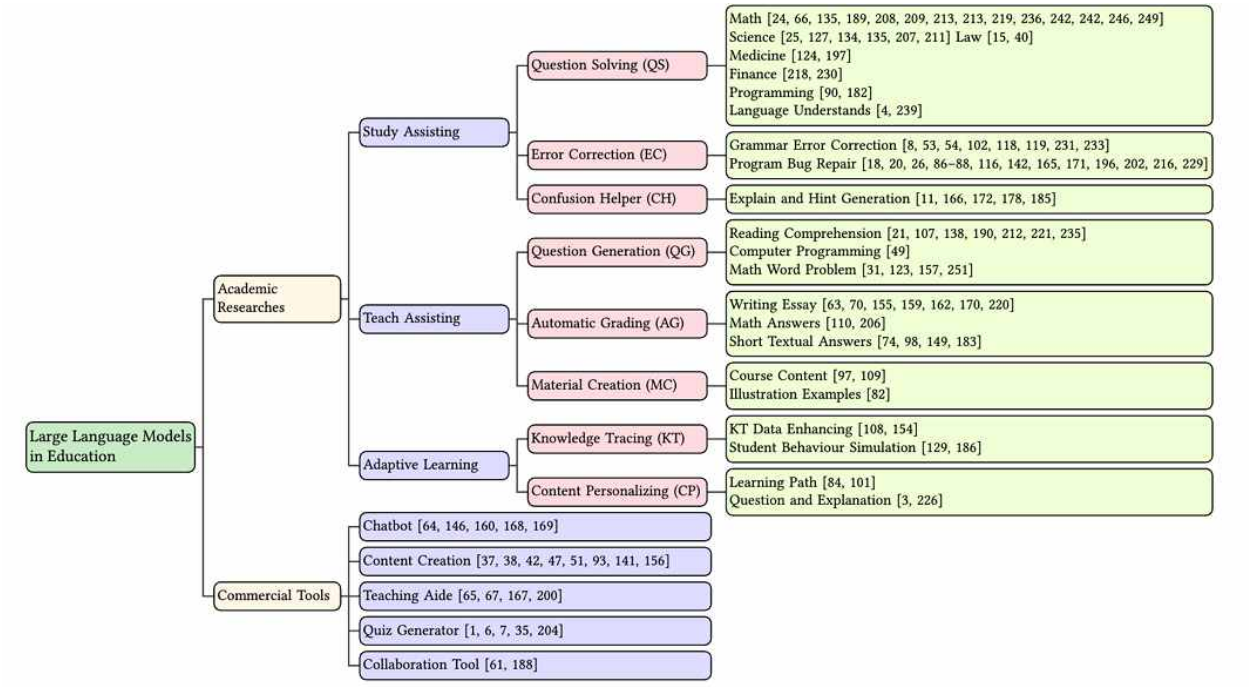
- **학습 참여도 및 동기 향상**
  - 중국 25개 대학 72,615명의 학부생을 대상으로 한 대규모 조사 연구에서 생성형 AI의 학업적 활용이 인지적, 정서적 참여도 향상과 관련이 있는 것으로 나타남(Guo et al., 2025).
  - 학생들이 생성형 AI를 지식 구성과 확장(knowledge construction and augmentation)의 도구로 활용할 때 보다 높은 수준의 학습이 이루어지는 것으로 확인됨.
  - AI 기반 학습 환경이 학생들의 자신감과 학업적 역량에 대한 신념을 강화하는 역할을 하는 것으로 보고됨(Wang & Fan, 2025).

- **인지 부하 감소**
  - 중국 25개 대학 72,615명의 학부생을 대상으로 생성형 AI 활용은 학습자의 인지적 노력(mental effort)을 감소시키는 효과가 있는 것으로 나타났으나, 이러한 인지 부하 감소가 학습의 깊이에 미치는 영향에 대해서는 추가적인 검토가 필요함(Guo et al., 2025; Stadler et al., 2024).

나 LLM의 활용 유형

- 최근 몇 년 간 대규모 언어 모델(Large Language Models, 이하 LLM)의 발전은 교육 현장에서의 활용 가능성을 획기적으로 넓힘. 특히 LLM은 자연어 이해 및 생성에서 탁월한 성능을 보여주며, 학생 지원, 교수 지원, 적응형 학습 등 다양한 교육 분야에서 새로운 방식의 접근을 가능하게 하고 있음.
- Wang et al. (2024)은 교육에서 LLM의 활용을 기술 중심 관점에서 **학생 지원(study assisting)**, **교사 지원(teach assisting)**, **적응형 학습(adaptive learning)**의 세 가지 핵심 범주로 분류하였음.

〈LLM의 교육적 활용 분류(Wang et al., 2024)〉



1) 학생 지원

- LLM은 사람과 유사한 응답을 생성하는 능력을 바탕으로 학생들에게 실시간 지원을 제공했으며, 어려운 질문을 해결하고 오류를 수정하며 혼란스러운 영역에 대한 설명이나 힌트를 제공함으로써 학습을 지원함
- **질문 해결** : LLM은 수학, 법, 의학, 프로그래밍 등 광범위한 주제에서 강력한 제로샷 문제 해결 능력을 입증했으며, 복잡한 문제 해결을 위해 Chain-of-Thought (CoT) 프롬프팅이나 외부 프로그래밍 도구 사용이 사용되었음
- **오류 수정** : 학습 초기 단계에서 학생들이 저지르는 흔한 문법적 오류나 맞춤법 오류를 즉각적으로 수정하는 데 LLM이 엄청난 잠재력을 보여주었음
- **혼란 도우미** : 정답을 직접 제공하는 대신, 학생 스스로 문제를 해결하도록 돕는 교육적 안내나 힌트를 생성하는 것을 목표로 했으나, LLM이 생성한 힌트가 교사가 작성한 설명보다 덜 효과적이거나 일반적일 수 있다는 결론이 제시되었음

2) 교사 지원

- LLM의 논리적 추론 및 문제 해결 능력 덕분에 교사들이 일상적인 업무 부담을 덜고 교실 수업과 같은

핵심 업무에 집중할 수 있도록 돕는 방향으로 개발되었음.

- **질문 생성 :** LLM은 보충 자료나 교과서 내용을 활용하여 읽기 이해 질문이나 특정 학습 목표에 맞는 객관식 질문을 생성하는 데 활용되었음.
- **자동 채점 :** LLM은 포괄적인 맥락, 명확한 채점 기준 등을 포함한 프롬프트 튜닝을 사용하여 개방형 질문 및 작문 에세이 채점에서 만족스러운 성능을 시연했으며, CoT를 통합하여 점수 외에 자세한 피드백을 제공할 수 있었음.
- **자료 제작 :** LLM은 비동기식 성인 학습 콘텐츠, 외국어 교육 자료 최적화, 대화형 예제 생성 등 고품질 교육 자료 제작을 지원하는 데 큰 잠재력을 보였음

3) 적응형 학습

- 적응형 학습 분야는 지식 추적(KT)과 콘텐츠 개인화(CP)로 분류될 수 있으며, LLM은 연구에 새로운 기회를 제공했음.
- **지식 추적 :** LLM은 질문 텍스트 및 학생 기록 데이터에 대한 보조 정보를 생성하는 데 사용되었으며, 특히 학습 데이터가 제한적인 콜드 스타트 시나리오 해결에 유리함이 입증되었음.
- **콘텐츠 개인화 :** LLM은 학생의 지식 숙달도나 TikTok, NBA와 같은 관심사에 기반하여 맞춤형 학습 경로와 상황에 맞는 질문을 생성하는 데 활용되었음

4) 잠재적 과제

- 교육 분야에서 LLM 구현 시 책임 있는 AI의 기본 요소(공정성, 신뢰성 등)와 더불어 **과도한 의존** 문제가 중요한 우려 사항으로 논의되었음.
- **공정성 및 포괄성 :** LLM 학습 데이터의 편향으로 인해 특정 집단에 대한 편향이 발생할 수 있으며, 특히 비영어권 사용자에게 불평등한 접근 문제가 발생할 수 있음이 보고되었음. 편향을 완화하기 위해 편향되지 않도록 지시하는 프롬프트나, 소수 집단의 관점을 제시하도록 돕는 파이프라인이 연구되었음.
- **신뢰성 및 안전:** LLM의 허위 내용 생성(환각, Hallucinations), 유해한 출력, 응답 불일치 문제가 지적되었음. 검색 증강 생성(RAG)이나 메타인지 전략을 사용하여 LLM의 신뢰성을 개선할 수 있음이 제안되었음.
- **투명성 및 책임성:** LLM이 블랙박스 메커니즘으로 작동하기 때문에 투명성과 책임성에 대한 우려가 있으며, 표절, 시험 부정행위 등 전통적인 평가 방법론에 대한 영향이 논의되었음. 훈련 시 인용을 통합하거나 모델 업데이트를 추적하는 프레임워크가 책임성을 높이는 방안으로 제안되었음.
- **과도한 의존:** 학생들이 비판적 사고나 작문 능력 향상 없이 LLM에 맹목적으로 의존할 위험이 있으며, 따라서 LLM은 인간의 사고 능력을 대체(substitute)하는 것이 아니라 증강(augment)하는 데 사용되어야 한다고 강조되었음

2 생성형 AI의 교육적 효과와 도전 과제

가 개별화 학습 지원

- **적응적 학습 경로 제공**
  - AI 기반 개인화 학습 플랫폼은 학습자의 개별적 요구에 맞게 콘텐츠와 피드백을 조정함으로써 교육적 성과를 최적화할 수 있음(Kasneci et al., 2023).

- 체계적 문헌 연구에 따르면, AI 기술은 적응적 학습 경로, 실시간 피드백, 맞춤형 콘텐츠를 제공함으로써 학생 참여도, 동기, 성취도를 향상시키는 것으로 나타남(Weng et al., 2024).
- 지능형 튜터링 시스템(ITS)과 생성형 AI의 결합은 기존의 사전 정의된 대화 방식의 한계를 극복하고 보다 동적인 콘텐츠와 피드백을 제공할 수 있는 가능성을 제시함(Jeon & Lee, 2023).

● **즉각적 피드백과 상호작용**

- 생성형 AI는 학생들의 질문에 즉각적으로 응답하고, 맥락에 적합한 설명을 제공함으로써 전통적인 교수 모델에서는 어려웠던 개별적 지원을 가능하게 함(Smolansky et al., 2023).
- AI 기반 피드백 시스템은 자기결정이론(Self-Determination Theory)에서 제시하는 자율성, 유능감, 관계성의 심리적 요구를 충족시킴으로써 학습자의 목표 달성, 학업적 자기효능감, 학습 참여도를 향상시킬 수 있음(Derakhshan, & Ghiasvand, 2024; Tariq, 2025).
- 소크라테스식 대화법을 활용한 AI 시스템은 학습자가 스스로 답을 찾아가도록 유도함으로써 고차적 사고를 촉진할 수 있음(Macina et al., 2023).

● **학습 격차 해소**

- 생성형 AI는 학생들의 질문에 즉각적으로 응답하고, 맥락에 적합한 설명을 제공함으로써 전통적인 교수 모델에서는 어려웠던 개별적 지원을 가능하게 함(Smolansky et al., 2023).
- AI 기반 도구는 대규모 수업이나 자원이 제한된 환경에서 개별화된 교수적 지원을 제공함으로써 학습 격차 해소에 기여할 수 있음(Cotton et al., 2023).
- 의료교육에서 AI 기반 개인화 학습 플랫폼을 활용한 무작위 통제 실험 결과, 실험집단이 통제집단에 비해 학습 성과, 학습 만족도, 자기주도학습 능력에서 유의미한 향상을 보임

나 과학학습에서의 효과

● **탐구기반학습(Inquiry-Based Learning) 지원**

- 생성형 AI는 탐구기반학습 환경에서 학생들이 과학 개념에 능동적으로 참여하도록 지원하는 대화형 수업 설계를 가능하게 함(Kotsis, 2024).
- ChatGPT를 과학교육에 통합한 탐색적 연구에서 AI의 응답이 과학교육 연구의 주요 주제와 일치하는 경우가 많았으나, 증거에 기반하지 않은 단일 진리를 가정하는 인식론적 권위의 위험이 지적됨 (Cooper, 2023).
- AI 지원 교수적 실천이 학생들의 오개념에서 과학적 개념으로의 전환을 촉진하는 것으로 보고됨(Chen, 2024).

● **과학적 개념 이해 촉진**

- 규칙 기반 챗봇 시스템을 과학 수업에 활용한 연구에서 학생들의 과학적 개념 이해가 유의미하게 향상됨(Lee et al., 2023).
- AI 로봇 이미지 인식 기술을 활용한 실험집단이 세포 응용 이해에서 통제집단보다 높은 성취를 보임 (Wang & Fan, 2024).
- 가상현실(VR) 기술과 결합된 AI는 다국어 학생들의 과학적 개념 이해를 용이하게 하는 것으로 보고됨 (Ding & Cha, 2024).

- **과학적 역량 및 태도 향상**
  - 학업 수준이 낮은 학생들에게 AI 지원 응용은 학습에 대한 흥미 수준을 높이고, STEM 분야에서 자신을 더 유능하고 지식이 풍부하며 인식이 높은 개인으로 여기게 함(Lee et al., 2023;Xu & Ouyang, 2022).
  - GameGPT를 활용한 학습 집단이 전통적인 게임 기반 학습만을 수행한 집단보다 유의미하게 높은 성취를 보임(Wang & Fan, 2025).
  - AI 기반 학습 환경이 학생들의 자기효능감 인식에 긍정적인 영향을 미치는 것으로 나타남(Ateş, 2024).

다 교수자 역할 및 교수학습 방법의 변화

- **교수자 역할의 변화**
  - 생성형 AI의 통합은 교수자의 역할을 콘텐츠 전달자에서 촉진자 및 안내자로 전환시키며, 학생들이 학습에 능동적으로 참여할 수 있도록 역량을 강화함(Bettayeb et al., 2024).
  - 교수자는 AI 기술의 책임 있고 윤리적인 사용을 안내하는 역할을 담당하게 됨(Chiu, 2024; Chiu et al., 2023).
  - AI 도구가 평가 및 피드백 제공 업무를 지원함으로써 교수자의 업무 부담이 경감될 수 있음(Moorhouse et al., 2023).

- **평가 방식의 변화**
  - 생성형 AI의 등장으로 전통적인 평가 방법이 학생들의 실제 학습 성과를 정확하게 측정하기에 점점 부적절해지고 있음(Chiu, 2024).
  - 고등교육에서의 평가는 학생들의 자기조절학습 기술, 책임 있는 학습, 학문적 무결성을 함양하도록 변환되어야 함(Chiu et al., 2023).

- **학습 설계의 혁신**
  - 생성형 AI의 역량은 학습 설계, 학습 조절, 자동화된 콘텐츠 생성, 피드백 및 평가와 같은 교수학습 실천을 향상시킬 수 있음(Stadler et al., 2024).
  - AI 기술은 시뮬레이션 사례 생성과 개인화된 설명 제공을 통해 교육 자원 개발을 지원할 수 있음(Kasneci et al., 2023).

라 도전과제 및 고려사항

- **학문적 무결성 문제**
  - 생성형 AI가 고품질의 텍스트를 생성할 수 있는 용이성으로 인해 표절 증가에 대한 우려가 제기되고 있으며, 일부 교육기관은 ChatGPT 사용을 제한하는 정책을 채택하였음.
  - 적절한 학업 지원과 부정행위 사이의 경계가 점점 모호해지고 있어 학생과 교수자 모두에게 윤리적, 교육학적 도전을 제기함(Susnjak & McIntosh, 2024).

- 환각 및 정확성 문제
  - 생성형 AI는 사실에 기반하지 않은 내용을 생성하는 '환각' 현상이 발생할 수 있으며, 이는 학습 과정에

서 오개념 형성의 위험을 초래함(Salamin et al., 2023).
- AI가 생성하는 피드백 내용의 과학적 정확성과 교육적 적절성에 대해서는 여전히 비판적 검토가 필요함(Mai et al., 2024).

- **비판적 사고 및 자기조절 역량 저하 우려**
  - 빈번한 생성형 AI 사용이 비판적 사고 능력과 부정적 상관관계를 보이며, 인지적 오프로딩(cognitive offloading)이 매개 요인으로 작용함(Gerlich, 2025).
  - ChatGPT는 학습자가 AI 지원에 의존하게 됨에 따라 메타인지적 나태함을 촉진하여 자신의 메타인지 과정에 대한 참여를 감소시킬 수 있음(Fan et al., 2024).
  - AI의 과도한 활용이 학생들의 자기조절 및 내재적 동기를 약화시킬 수 있다는 우려가 제기됨(Cotton et al., 2023).

- **디지털 격차 및 접근성**
  - AI 도구에 대한 접근성의 차이가 교육 격차를 심화시킬 수 있으며, 공평한 AI 접근을 보장하기 위한 정책적 노력이 필요함(Perkins, 2023).
  - AI 기반 개인화 학습 시스템을 효과적으로 활용하기 위해서는 기존의 자기조절학습 기술이 필요함(Laak et al., 2024).

- **교육적 시사점**
  - 생성형 AI를 단순한 정보 검색 도구가 아닌 학습 촉진과 고차적 사고 함양을 위한 도구로 활용하기 위해서는 증거 기반의 학습 실천에 맞추어 신중하고 사려 깊게 적용되어야 함(Mollick & Mollick, 2023).
  - 학생들에게 AI에 질문하는 방법과 AI 응답을 비판적으로 평가하는 방법에 대한 교육이 필요함.
  - 교수자의 전문성 개발 활동이 AI 기술의 교육적 활용 역량을 강화하는 방향으로 이루어져야 함.
  - RAG(검색증강생성) 기술의 활용, 교육과정 기반 콘텐츠 구축 등을 통해 AI의 환각 효과를 최소화하고 신뢰성 높은 교육적 상호작용을 보장하는 시스템 설계가 필요함.
  - 공교육에서의 생성형 AI 도입은 개인정보보호, 윤리적 사용, 학문적 무결성 등에 대한 종합적인 가이드라인과 정책 수립이 선행되어야 함.

3절 LLM 활용 전략 및 RAG 적용 사례

1 LLM의 개념 및 유형

가 LLM의 개념

- 대규모 언어모델(LLM)은 방대한 텍스트 데이터를 기반으로 학습된 Transformer 기반 언어 모델로, 문맥을 이해해 다음 단어를 예측하는 방식을 통해 자연어 패턴을 학습함. 파라미터 규모가 커질수록 추가 학습 없이도 프롬프트만으로 새로운 과제를 수행하는 In-Context Learning 능력이 발현되었으며, 이를 통해 언어 생성뿐 아니라 추론, 요약, 문제 해결, 코드 작성 등 다양한 과제를 단일 모델로 수행하는 범용성을 갖추게 되었음(Dong et al., 2022).

〈국내외 LLM 주요 모델 비교〉

구분	모델명	개발사/기관	유형	파라미터 규모	주요 특징
해외	GPT-4	OpenAI	비공개	비공개	고품질 추론·요약·대화 능력
	Claude	Anthropic	비공개	비공개	장문 이해·서술 성능 우수
	LLaMA 2/3	Meta	공개	7B ~ 70B	공개 모델 중 최고 수준의 범용성
국내	HyperCLOVA/ CLOVA X	NAVER	공개/비공개	1.5B ~ 204B	한국어 대규모 학습, 국내 서비스 특화
	Midm 2.0	KT	공개/비공개	2.3B ~ 41B	한국 언어 및 문화 맥락 내제, 한국 생태계 특화
	Polyglot-Ko	EleutherAI	공개	1.3B ~ 12.8B	한국어 공개 모델

나 LLM의 유형

- 최근에는 LLM의 언어 이해 능력을 확장하여 이미지·도표·그래프 등 비텍스트 정보를 함께 처리하는 방향으로 연구가 확장되고 있음. 이로 인해 시각 정보 활용이 가능한 VLM(Vision-Language Model)이 등장하였고(Zhang, Huang, Jin, & Lu, 2023), 더 나아가 다양한 모달(modal) 데이터를 통합적으로 처리하는 MLLM(Multi-Modal Large Language Model)으로 발전하고 있음(Yin et al., 2023).

1) VLM (Vision-Language Model)

- 비전-언어 모델은 이미지·도표·그래프 등 시각 정보와 텍스트 정보를 함께 처리하는 모델로, 시각 장면을 이해하고 설명하거나, 이미지 기반 질의응답을 수행할 수 있음.
- 일반적으로 시각 인코더(CNN·ViT 등)와 언어모델을 결합하여 학습하며, 이미지-텍스트 매핑(CLIP 계열(Radford et al., 2021)), 이미지 설명 생성, 시각 질의응답(VQA) 등 작업에 활용됨.

2) MLLM (Multi-modal Large Language Model)

- 멀티모달 대규모 언어모델은 LLM을 중심으로 텍스트 외에 이미지, 음성, 표·차트, 심지어 동영상·센서 데이터 등 다양한 모달 정보를 통합적으로 입력받아 추론하는 모델임.
- 최근 모델(GPT-4V, Gemini 등)은 기존 텍스트 기반 LLM의 언어·추론 능력에 시각 이해 기능을 결합하여, 하나의 모델이 복합적인 판단·해석·설명 작업을 동시에 수행할 수 있도록 발전하고 있음.

2 LLM 모델 개선을 위한 활용 전략

- 대규모 언어모델(LLM)은 범용성을 갖춘 기초 모델로 폭넓게 활용되고 있지만, 특정 업무·도메인·사용자 요구에 그대로 적용하기에는 여전히 여러 한계가 존재함. 예를 들어, 모델이 학습하지 않은 최신 정보를 즉시 반영하기 어렵고, 특정 조직·도메인에서 사용하는 문서 구조나 표현 방식을 자연스럽게 따르지 못함. 또한 사용자 선호·규범·안전 기준을 충분히 충족하지 못하는 경우가 많아, 실사용 단계에서는 별도의 제어·보완 과정이 필수적임. 이러한 이유로 "LLM을 어떤 방식으로 활용하고, 어떤 수준에서 개선할 것인가"는 모델 활용 전략의 핵심 요소가 됨.
- 최근 연구 및 산업 동향을 종합하면, LLM 개선 전략은 크게 세 가지로 정리할 수 있음.
  - 첫째, 파라미터를 바꾸지 않고 입력 프롬프트만으로 모델의 행동을 제어하는 프롬프트 기반 전략
  - 둘째, 외부 문서·지식베이스를 검색해 LLM에 제공함으로써 사실성과 최신성을 보완하는 검색 증강 생성(RAG) 전략
  - 셋째, 감독학습(SFT)·RLHF·DPO 등을 통해 모델 파라미터를 도메인에 맞게 조정하는 후학습(Post-Training) 전략
  - 이 세 가지 전략은 서로 대체 관계가 아니라 상호 보완적인 수단으로, 적용 목적·요구 성능·비용 제약에 따라 가장 적합한 조합을 구성하여 활용할 수 있음.

가 Prompt (In-Context Learning) 방법론

- In-Context Learning(ICL)은 LLM의 파라미터를 변경하지 않고, 입력 프롬프트에 예시·지시·문맥 정보를 제공함으로써 모델의 출력 형태를 조절하는 학습 방식임. LLM은 자기지도학습 과정에서 다양한 패턴을 내재적으로 학습하기 때문에, 추가 학습 없이도 프롬프트에 제시된 양식, 사고 방식, 논리 구조를 모방하거나 새로운 과제를 수행하도록 유도할 수 있음.
- 대표적인 ICL 기법으로는 예시 기반 학습(few-shot prompting)(Brown et al., 2020), 문제해결 과정을 명시하는 Chain-of-Thought(CoT) prompting(Wei et al., 2022), 역할 지시(Role prompting)(Kong et al., 2023) 등이 있음.

1) Few-shot Prompting

- 모델에게 몇 개의 예시 입력-출력 쌍을 제공하여, 그 패턴을 따라 새로운 문제를 해결하도록 유도하는 방식
- 파라미터 업데이트 없이도 예시의 구조, 문체, 논리 진행 방식을 모델이 그대로 모방할 수 있다는 장점이 있다.

2) Chain-of-Thought(CoT) Prompting

- 정답만 요구하지 않고 중간 사고 과정을 프롬프트에 명시하여 모델이 논리적 추론을 하도록 유도하는 방식
- 복잡한 과학 개념 설명, 데이터 해석, 다단계 문제 해결에 효과적임.

3) Role Prompting

- 모델에게 “과학 교사”, “실험 조교”, “평가자” 등 특정 역할을 부여함으로써 발화 방식과 판단 기준을 제어하는 방식
- 역할 설정만으로도 모델의 응답 방식이 크게 달라져 교육 도메인에서 자주 활용됨.

나 RAG (Retrieval-Augmented Generation) 방법론

- RAG는 LLM이 답변을 생성할 때 외부 문서나 지식베이스에서 관련 정보를 검색하여 함께 활용하는 방식으로, 모델이 사전에 학습하지 못한 최신 정보나 세부 도메인 지식을 보완할 수 있는 접근법임(Lewis et al., 2020). 이를 위해 RAG는 보통 검색-증강-생성의 세 단계로 구성됨.
  - **Retrieval(검색)**: 교과서, 실험 자료, 교육과정 문서 등에서 의미적으로 관련된 내용을 벡터 임베딩 기반 검색(semantic retrieval)을 통해 가져옴.
  - **Augmentation(증강)**: 검색된 문서를 LLM 입력에 첨부하거나, 핵심 내용을 정제하여 모델이 참고할 수 있는 문맥으로 제공함.
  - **Generation(생성)**: LLM은 해당 문맥을 기반으로 사실과 근거가 포함된 응답을 생성함.
- 이러한 구조를 통해 RAG는 LLM 단독으로는 확보하기 어려운 사실성(factuality), 도메인 적합성(domain-specificity), 출처 기반 설명(evidence-based explanation)을 강화할 수 있으며, 과학탐구와 같이 근거가 중요한 영역에서 효과적으로 활용됨.

다 Post-Training 방법론

- 프롬프트·RAG만으로는 학생 수준별 적응, 교과 기반 피드백 스타일 통제, 교육 정책·안전 기준 준수와 같은 세밀한 요구사항을 완전히 충족하기 어려움. 이를 보완하기 위해, 도메인과 사용자 요구에 맞게 LLM의 파라미터를 직접 조정하는 후학습(Post-Training) 방법론이 활용됨.

1) 감독 미세조정 (Supervised Fine-Tuning, SFT)

- 교사·전문가가 작성한 “입력(학생 답안·질문)-출력(모범 피드백·설명)” 쌍을 사용해 LLM을 지도학습으로 미세조정하는 방식
- 상대적으로 구현이 단순하고, 도메인 특화 표현·용어·템플릿을 모델에 내재화시키는 데 효과적임.

2) 인간 피드백 기반 강화학습(Reinforcement Learning from Human Feedback, RLHF)

- Ouyang et al.(2022)의 InstructGPT 연구는, 인간이 선호하는 출력에 대한 순위 정보를 활용해 보상모델을 학습하고, 이를 기반으로 RL(PPO 등)로 LLM을 미세조정하는 RLHF 파이프라인을 제안하였음.

- 과학 교육 맥락에서는, “루브릭을 얼마나 잘 반영하는가”, “과학적 정확성·안전성·학생 친화적 표현” 등을 복합적인 보상으로 설계해 학생 피드백 품질을 지속적으로 개선할 수 있음.

3) 선호 최적화(Preference Optimization: DPO 등)

- Direct Preference Optimization(DPO)은 별도의 보상모델·RL 없이, 선호쌍(preferred vs dispreferred 응답)에 대한 단순 분류 손실로 RLHF와 동등한 목적을 최적화하는 효율적인 방법론을 제안함(Rafailov et al., 2023).
- DPO는 학습 파이프라인이 단순하고 안정적이어서, 교육기관·연구소 수준에서도 비교적 구현이 용이하다는 장점이 있음.

3 RAG의 개념 및 의의

가 RAG(Retrieval-Augmented Generation)의 개념

1) 정의

- **LLM(대규모 언어 모델)의 한계와 배경**
  - 대규모 언어 모델(LLM, Large Language Model)은 방대한 데이터 학습을 통해 다양한 작업에 우수한 성능을 보이지만, 여전히 사실에 기반하지 않은 정보를 생성하는 환각(Hallucination) 문제를 완전히 해결하지 못하고 있음
  - 이러한 한계를 극복하기 위해 RAG(Retrieval-Augmented Generation)가 제안됨
- **RAG(Retrieval-Augmented Generation)**
  - 외부 지식 베이스에서 관련 정보를 검색(Retrieval)하고, 이를 모델의 생성(Generation) 과정에 통합하는 기술
  - 모델이 더 신뢰할 수 있고 근거 있는 출력을 생성하도록 도움
  - 단순한 생성 능력의 확장을 넘어, 언어 모델이 외부 지식과 상호작용하며 추론하는 새로운 패러다임으로의 전환을 의미

2) 구조

- **검색기(Retriever)**
  - 사용자의 입력 질의(Question)에 대응하여 외부 지식원으로부터 관련 문서를 검색함
  - 키워드 매칭 기반의 희소 검색(Sparse Retrieval), 임베딩 유사도 기반의 밀집 검색(Dense Retrieval), 두 방식을 결합한 하이브리드 검색(Hybrid Retrieval) 기법이 주로 사용됨
- **생성기(Generator)**
  - 모델의 내부 지식(parametric)과 검색된 문서를 바탕으로 사용자의 입력 질의에 대한 최종 답변을 생성

나 RAG의 의의 및 필요성

1) 확장성 및 유연성

- **고정된 지식의 한계 극복**
  - 파인튜닝(Fine-tuning)된 모델은 학습이 종료된 시점에 지식이 고정되어 정보가 계속 변화하는 상황에서 최신 상태를 유지하기 어려움
  - RAG는 외부 지식 베이스를 실시간으로 검색하고 활용하여 데이터 변동성이 큰 환경에서도 유연한 대응이 가능
- **유지보수 효율성 및 지식 확장성**
  - 모델 파라미터 재학습 없이 외부 데이터베이스 업데이트만으로 즉각적인 최신 정보 반영 가능

2) 신뢰성 및 근거 기반성

- **환각(Hallucination) 문제 완화 및 신뢰성 확보**
  - RAG는 지식 베이스에서 검색된 문서를 바탕으로 답변을 생성하도록 함으로써 언어 모델의 환각 문제 완화
  - 신뢰할 수 있는 문서를 기반으로 답변을 생성하여 근거 문서에 대한 충실성을 담보
- **검증 가능성 및 전문 도메인 활용**
  - 모델은 답변과 함께 참조된 원문 출처를 제공하여 사용자가 사실 관계와 정확성을 직접 검증할 수 있음
  - 의료, 법률, 교육 등 사실적 정확성이 필수적인 도메인에서 모델을 활용하기 위한 핵심 발판 마련

4 RAG의 성능개선 전략 및 활용 사례

가 성능 개선 전략

1) 검색(Retrieval) 단계 최적화

- **데이터 전처리 및 청킹(Chunking) 전략 고도화**
  - 구현이 단순한 길이 기반(Token-level) 청킹이나 처리시간이 긴 의미 기반(Semantic-level) 청킹 대신, 의미 유지와 효율성이 균형 잡힌 문장 단위(Sentence-level) 청킹 사용 (Wang et al., 2024)
  - 작은 크기의 청크 블록을 질의(query)와 매칭하는 데 사용하고, 이 작은 청크를 포함하면서 더 넓은 문맥(Context) 정보를 담은 큰 블록을 함께 반환하는 Small-to-big 기법 사용 (Liu, 2022)
  - 청크 간 일부가 겹치도록 하는 Sliding Window 기법을 적용하여 문맥 손실 방지 (Liu, 2022)
  - 청크 블록에 제목, 키워드, 가상 질문 등의 메타데이터를 추가하여 검색 성능 향상

- **질의(Query) 분석 및 분류**
  - 질의 분류(Query Classification)를 선행하여 모델의 내부 지식(parametric) 정보만으로 충분히 답변 가능한지 판단하고 필요한 경우에만 검색을 수행하여 효율을 높임 (Gao et al., 2023)
  - 사용자의 초기 질의를 LLM을 통해 구체적이고 검색 용이한 형태로 재작성(Query Rewriting) (Ma et al., 2023)

- 하나의 질의를 여러개의 질의로 분해하고 각각의 검색 결과를 통합(TOC) (Kim et al., 2023)
- 사용자 질의(query)를 기반으로 답변 생성에 필요한 가상의 문서를 미리 생성하고 임베딩을 사용해 유사한 문서를 검색(HyDE) (Gao et al., 2023)
- 다중 인덱스 타입 및 하이브리드 검색을 지원하는 벡터 데이터베이스를 통해 대규모 데이터 처리의 확장성을 보장(예: Milvus) (Wang et al., 2021)

- **하이브리드 검색(Hybrid Retrieval) 도입**
  - 키워드 매칭 검색(Sparse Retrieval)과 임베딩 벡터 유사도 기반 검색(Dense Retrieval)을 결합하여 단일 방식의 한계를 보완
  - 전문 용어의 정확한 일치와 문맥적 의미 유사성을 동시에 고려함으로써 검색 성능 향상

- **재순위화(Re-ranking) 기법 적용**
  - 검색된 문서들의 순위를 정밀하게 재조정하여 관련성이 높은 문서를 상위에 배치함으로써 질의와 관련 문서들 간의 유사도 극대화
  - 질의와 문서 간의 관련성 점수를 학습한 신경망 모델을 활용하는 방법(monoT5 (Nogueira et al., 2020), RankLLaMA (Ma et al., 2024))
  - 모든 문서에 대해 각 질의어(token)가 나올 확률을 사전에 계산해두고, 질의어가 들어오면 그 단어들의 확률값을 단순히 합산한 점수를 활용하는 방법(TILDE) (Zhuang & Zuccon, 2021)

- **검색 문서(Retrieved Documents) 후처리 및 요약**
  - LLM의 입력 특성을 고려하여 관련성 점수가 높은 문서가 query와 더 가깝게 위치하도록 Document Repacking 기법 사용 (Liu et al., 2024)
  - 검색된 문서에서 중복 내용을 제거하고 핵심 문맥을 추출하거나 생성하여 요약(Recomp) (Xu et al., 2023)

2) 생성(Generation) 및 통합 단계 최적화

- **생성기 파인튜닝을 통한 노이즈 내성 강화** (Zhang et al., 2024)
  - 관련 문맥과 무작위 비관련 문맥를 함께 섞은 혼합 문맥을 구성하여 생성기를 학습시킴
  - 관련 정보와 무관한 정보를 동시에 처리하는 훈련을 통해, 모델이 검색된 문서 내의 노이즈에 더 강건하게 반응하면서도 관련 문맥을 효과적으로 활용할 수 있음

- **프롬프트 엔지니어링(Prompt Engineering) 및 추론 전략 고도화**
  - 단계적인 추론 과정을 유도하는 CoT(Chain-of-Thought) 프롬프팅을 적용하여 복잡한 질문에 대한 논리적 답변 능력 향상
  - 답변 생성의 예시를 함께 제공하는 퓨샷(Few-shot) 프롬프팅을 통해 생성 형식을 제어하고 도메인 적합성을 높임

- **고급 RAG 파이프라인 구축**
  - 모델이 단순히 검색된 정보를 수동적으로 받아들이지 않고, 자신의 생성 과정과 검색된 정보의 품질을 스스로 평가하고 수정함(Self-RAG) (Asai et al., 2024)
  - 검색기가 부정확한 문서를 검색했을 때, 이를 맹목적으로 신뢰하는 대신 평가자 모델을 통해 검색된 문서의 관련성 점수를 계산하여 모델의 강건함을 유지(CRAG) (Yan et al., 2024)

- LLM의 내부 지식과 검색된 외부 지식 간의 정보 충돌 발생 시, 각 정보원의 신뢰도를 평가하고 적응적으로 통합하여 불완전한 검색 상황에서도 강건함을 유지(ASTUTE RAG) (Wang et al., 2025)

- **부정 응답 억제(Negative Rejection)**
  - 검색된 정보만으로 질의에 답변이 불가능한 경우, 부정확한 답변을 생성하지 않고 거부함
  - 생성된 응답의 신뢰도 점수에 임계값을 설정하여, 낮은 신뢰도의 응답을 필터링함

나 활용 사례

1) 도메인 특화 응용

- **특수 도메인(Domain-specific)에서의 가치 입증**
  - 범용적 지식을 다루는 오픈 도메인(Open-domain)을 넘어, 고도의 전문성과 정확성이 필수적인 분야에서 RAG가 핵심 솔루션으로 자리 잡음
  - 모델의 환각이 치명적인 분야에서 근거 문서에 기반한 신뢰성 있는 답변을 제공

2) 분야별 활용 사례

- **의료 분야**
  - **Med-PaLM 2, Google** (Singhal et al., 2025): 의료 전문 지식 베이스를 학습하여 복잡한 의료 질문에 답하고 진단 및 치료 계획을 제안하는 임상 특화 모델
  - **Apollo 24|7**: Google의 MedLM을 기반으로 임상 지능 엔진(Clinical Intelligence Engine, CIE)을 구축, 의사의 진단 정확도를 지원
  - **MEDRAG** (Zhao et al., 2025): 단순 텍스트 검색을 넘어 지식 그래프 기반의 추론을 결합하여, 의료 질의 응답(QA) 시스템의 사실적 정확성과 신뢰성을 높임
  - **Nuance DAX**: 진료실 내 의사와 환자의 대화를 음성으로 인식하고 RAG를 통해 자동으로 임상 문서(Clinical Documentation)를 생성하여 행정 업무 부담 경감

- **법률 분야**
  - **Harvey AI, CoCounsel, Lexis+** 등: 판례, 법령 등의 방대한 법률 데이터를 RAG로 연결하여 리서치와 문서 작성을 자동화한 상용 서비스
  - **LegalBench-RAG** (Pipitone & Alami, 2024): 위와 같은 상용 서비스에서도 여전히 발생하는 환각 문제와, 법률 도메인 특화 RAG 시스템의 성능을 평가하기 위해 고안된 특화 벤치마크
  - **LawGlance** (Lawglance, n.d.): 법률 조사가 필요한 하위 작업을 스스로 계획하고 수행하는 에이전틱 AI(Agentic AI) 접근 방식의 오픈소스 프로젝트

- **금융 분야**
  - **FinRobot** (Yang et al., 2024): 금융 시장 분석과 보고서 작성을 지원하는 오픈소스 AI 에이전트 플랫폼으로, RAG를 통해 외부 금융 데이터를 내부 파이프라인에 통합하여 보다 정확하고 신뢰성 있는 금융 분석을 수행
  - **AskResearchGPT, Morgan Stanley**: 범용 모델(GPT-4)에 자사의 고유한 금융 데이터를 RAG로 결합하여 애널리스트의 업무를 지원하는 내부 시스템을 구축함
  - **FinRAGBench-V** (Zhao et al., 2025): 텍스트보다 차트나 표에 핵심 정보가 많은 금융 보고서의

특성을 고려하여, 기존 텍스트 기반 RAG가 놓치던 시각적 정보를 처리하고 평가하기 위한 멀티모달 벤치마크

- **교육 분야**
  - **GraphRAG 활용**: 지식 그래프를 탐색하며, 학습 중인 내용과 연관된 개념을 함께 제시하거나 선행 학습이 필요한 개념을 파악하여 체계적인 학습 경로를 제안
  - **NotebookLM, Google** (Tufino, 2025): 사용자 업로드 문서를 기반으로 대화형 인터페이스 및 오디오 개요를 제공하는 RAG 기반 LLM으로, 신뢰성 있는 능동적 학습을 지원함

- **기업 및 지식 관리**
  - **NASA 인력 지능(Workforce Intelligence)** (Tasneem, 2025): GraphRAG 기술을 활용하여 NASA 내부의 방대한 인력 데이터와 프로젝트 이력 간의 복잡한 연결 고리를 분석
  - **Glean** (Glean, n.d.): 기업 내 산재된 데이터 소스를 통합하고, RAG 기반 검색·생성을 통해 사내 지식 탐색을 지원하는 엔터프라이즈 검색 엔진

- **코딩 및 개발**
  - **GitHub Copilot Enterprise**: 기업 내부 리포지토리를 RAG 검색 대상으로 하여, 작성 중인 코드와 연관된 내부 라이브러리 코드를 탐색하고 설명함
  - **Cursor**: 프로젝트 전체를 인덱싱하여 함수 호출 관계, 파일 간 의존성 등 구조 기반 질문에 정확히 답변하는 코드 어시스턴트

제3장 연구수행 내용 및 결과

1절 | 생성형 AI 기반 과학탐구 지원 시스템 개발

2절 | 교사·학생 설문 분석

3절 | 교사·학생 면담 분석

4절 | 시스템 로그데이터 분석

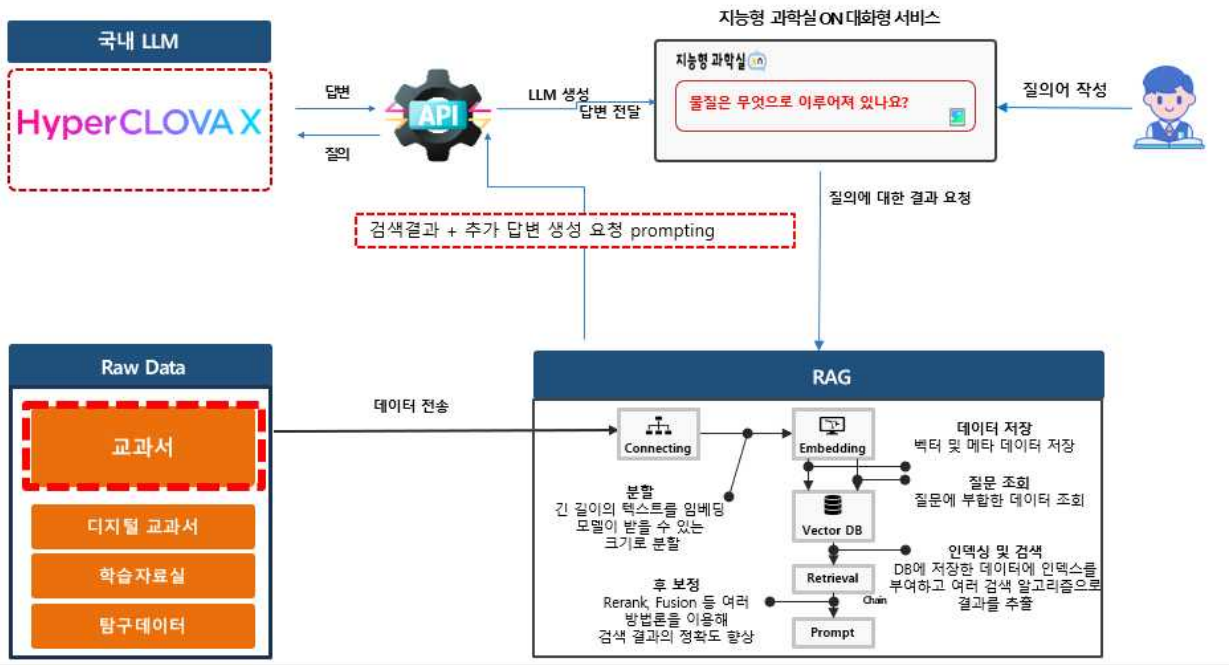
5절 | 시스템 성능 평가

1절 생성형 AI 기반 과학탐구 지원 시스템 개발

1 AI 기반 대화형 서비스

가 대화형 서비스 개요

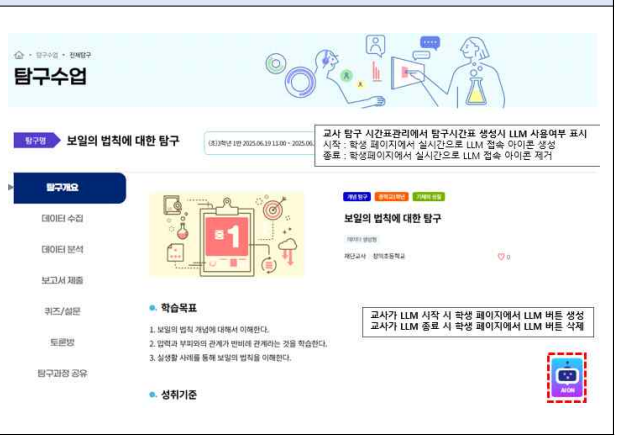
- 대화형 서비스는 교육과정에 맞는 과학탐구 지원 질의응답을 지원하기 위해 RAG 기술을 도입하여 개발하였으며, 질문에 대한 답변을 RAG의 DB에서 검색한 뒤, 검색 결과를 맥락정보로 활용하여 LLM의 추가 검색 결과를 결합하여 답변이 인출되도록 개발하였음. LLM은 개인정보 국외 이전 등의 정보보호 문제를 안정적으로 다루기 위해 국내 LLM인 HyperCLOVA X를 활용하였으며 HCX-007모델을 도입함.



나 대화형 서비스 활용 방법

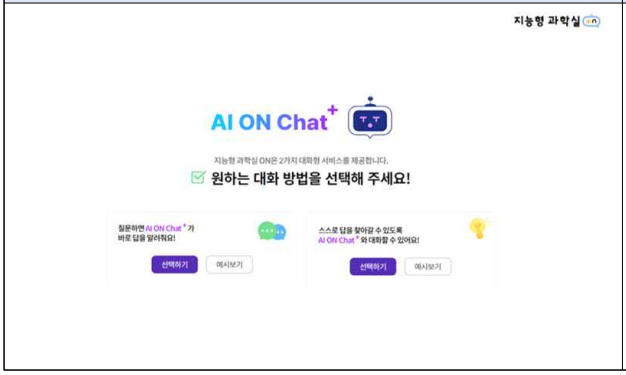
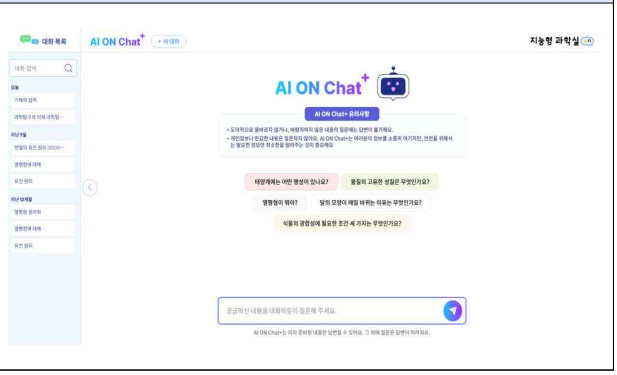
1) 대화형 서비스 활용 준비

- 대화형 서비스의 활용은 일반적인 탐구수업 개설 방법을 통해 탐구시간표를 지정한 뒤 '탐구 시간표 관리' 메뉴에서 교사가 원하는 시간에 LLM 활용 여부를 설정함으로써 가능해짐.
  - LLM 설정은 미사용, 시작, 종료, 모니터링으로 구분되며, 교사가 수업에서 LLM 활용을 '시작'하게 되면 '모니터링' 버튼이 생성되어 학생들의 대화형 서비스 이용 현황을 실시간으로 확인할 수 있음.
  - 교사가 LLM 활용을 '시작'하게 되면 학생들의 탐구수업 화면에 대화형 서비스 접속 창을 열 수 있는 아이콘이 제시됨
  - 학생들은 대화형 서비스 접속 아이콘을 클릭함으로써 'AI ON Chat'이라는 팝업 창에서 AI와의 질의응답을 경험할 수 있음.

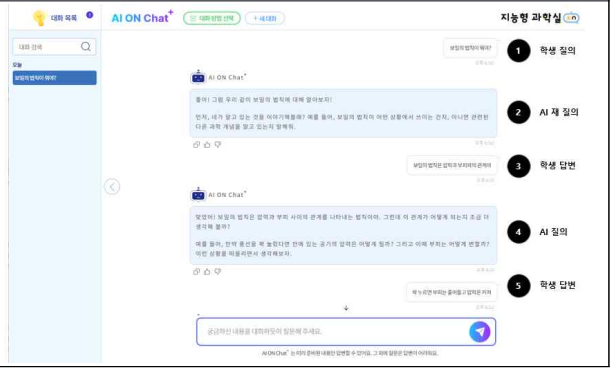
LLM 활용 여부 설정	대화형 서비스 접속 아이콘 생성
	

2) 대화형 서비스 활용

- 대화형 서비스는 학생이 질문하면 즉각 답변을 제공하는 기능과, 학생이 스스로 답을 찾을 수 있도록 AI가 질문하는 소크라테스식 대화법 기능으로 구분됨.
- 대화형 서비스 첫 화면에서 원하는 대화 방식을 선택하고 그에 따라 대화형 서비스를 이용할 수 있음.

LLM 활용 방식 선택	대화형 서비스 활용 화면
	

- 학생들의 선택에 따라 대화형 서비스 활용 방식이 달라짐. 학생이 질문한 결과에 따라 AI가 응답하는 경우 학생들의 질문과 문제를 해결하도록 지원하는 한편, AI가 질문하고 학생이 응답하는 경우 학생들이 자신의 과학적 경험에서 개념적 이해를 이끌어 내도록 사고를 촉진하는 대화가 제공됨.

학생이 질문하는 대화형 서비스 화면	학생이 대답하는 대화형 서비스 화면
	

대화형 UI 주요 기능

No.8	설명
1	사용자 질의 목록
2	질의 내용 변경 및 삭제
3	대화 방법 선택
4	새 대화 추가
5	사용자가 작성한 질의 내용
6	사용자의 질의 내용에 대해 답변한 결과 표시 화면
7	질의 내용 입력 창
8	답변 내용 복사 좋아요, 싫어요
9	질의 목록 닫기

다 대화형 서비스 활용 예시

1) 초등학생

- 과학 개념과 일상 속 실천 방안 등에 대해 질문

학생 질문	생태계 보전을 위해 내가 할 수 있는 걸 알려줘
AI 응답	• 생태계를 보호하기 위해서 우리가 일상 속에서 실천할 수 있는 것들이 많아요! • 먼저 가장 기본적인 것은 '쓰레기를 버리지 않는 것'이에요. 길거리에 쓰레기를 버리면 동물들이 그것을 먹고 아플 수도 있고, 자연 환경이 오염될 수 있어요. 그래서 항상 쓰레기통에 버리도록 해요! • 또한 '재활용'에도 신경 써보세요. 플라스틱 병이나 종이 같은 것을 재활용하면 자원을 절약할 수 있답니다. • 그리고 '물을 아끼는 습관'을 가져보는 것도 중요해요. 양치질하거나 샤워할 때 물을 틀어놓지 않고 꼭 필요할 때만 사용하는 것이 좋아요. 이렇게 하면 물 부족 문제를 예방할 수 있어요. • 마지막으로 '자연을 사랑하는 마음'을 가지는 게 중요하겠죠? 공원이나 산에 가서 나무와 꽃들을 보면서 자연을 느껴보고, 그 소중함을 마음에 새겨보아요. ? • 이렇게 작은 노력들이 모여 큰 변화를 만들 수 있다는 사실 잊지 말아요!

2) 중학생

- 과학 개념 이해를 위한 질문에 더해 사전지식을 연계하여 개념을 심화하기 위해 소크라테스식 대화법을 통한 탐구활동 참여

AI 질문	• 정말 훌륭해! 정확하게 이해하고 있네. ? 네가 말한 것처럼, 손을 사용해 책상을 누를 때 손이 가하는 힘이 책상의 일정 면적에 작용하면서 압력이 생기게 돼. • 마지막으로, 네가 정리한 내용을 기반으로 간단한 문제를 풀어보자. 만약 동일한 힘을 가했을 때, 손바닥보다 손가락 끝으로 책상을 누른다면 압력이 어떻게 될까? 그리고 왜 그렇게 되는지도 함께 설명해 볼래?
학생 응답	• 손가락 끝으로 눌렀을 때는 압력이 손바닥으로 눌렀을 때보다 훨씬 크게 느껴질 거야. 왜냐하면 같은 힘이지만 면적이 작기 때문에 그런 것 같고, 손으로 책상을 누르면 손에 가한 힘이 책상의 일정 면적에 작용하면서 압력이 생기게 돼서 그런 것 같아. **내 지식에 이런 것들이 생겨서 좋고 원래는 아무것도 못 했는데 이제는 좀 할 수 있을 것 같아.**

3) 고등학생

- 교과내용을 넘어 자율탐구를 능동적으로 수행하기 위해 필요한 정보를 질문하고 문제해결을 위해 학생 주도적 과학탐구 수행

학생 질문	PM 2.5, PM10, PM1.0 이 뜻하는 게 뭐야?
AI 응답	• 좋은 질문이에요! PM은 'Particulate Matter'의 약자로, 공기 중에 떠다니는 미세 입자를 의미해요. 숫자들은 그 입자들의 크기를 나타내죠. 이제 하나씩 살펴볼까요? 1. PM 2.5: "2.5 마이크론 이하"의 아주 작은 먼지를 말해요. 이 크기 이하의 미세먼지는 폐 깊숙이 들어가 건강에 나쁜 영향을 미칠 수 있어요. 2. PM 10: "10 마이크론 이하"의 먼지를 가리켜요. 주로 도로에서 자동차가 지나갈 때 발생하는 먼지가 여기에 해당 돼요. 3. PM 1.0: "1 마이크론 이하"의 초미세먼지에요. 가장 작고 위험한 먼지로, 대기 오염의 주범 중 하나로 꼽혀요. • 이제 조금 더 시각적으로 생각해 볼게요. 마이크론은 매우 작은 단위라서 머리카락 굵기의 몇 백분의 일에 해당된답니다. 예를 들어, PM 2.5는 머리카락 굵기의 대략 40분의 1 정도밖에 안 돼요! • 이렇게 작은 먼지들이 왜 위험하냐면, 눈에 보이지 않고 호흡기로 쉽게 들어갈 수 있기 때문이죠. 그래서 이런 미세먼지 농도가 높을 때는 외출 시 마스크를 착용하거나 실내 환기를 잘 관리하는 게 중요해요. 혹시 다른 궁금한 점이 있으면 언제든지 물어보세요!

2 AI 기반 보고서 튜터링 서비스

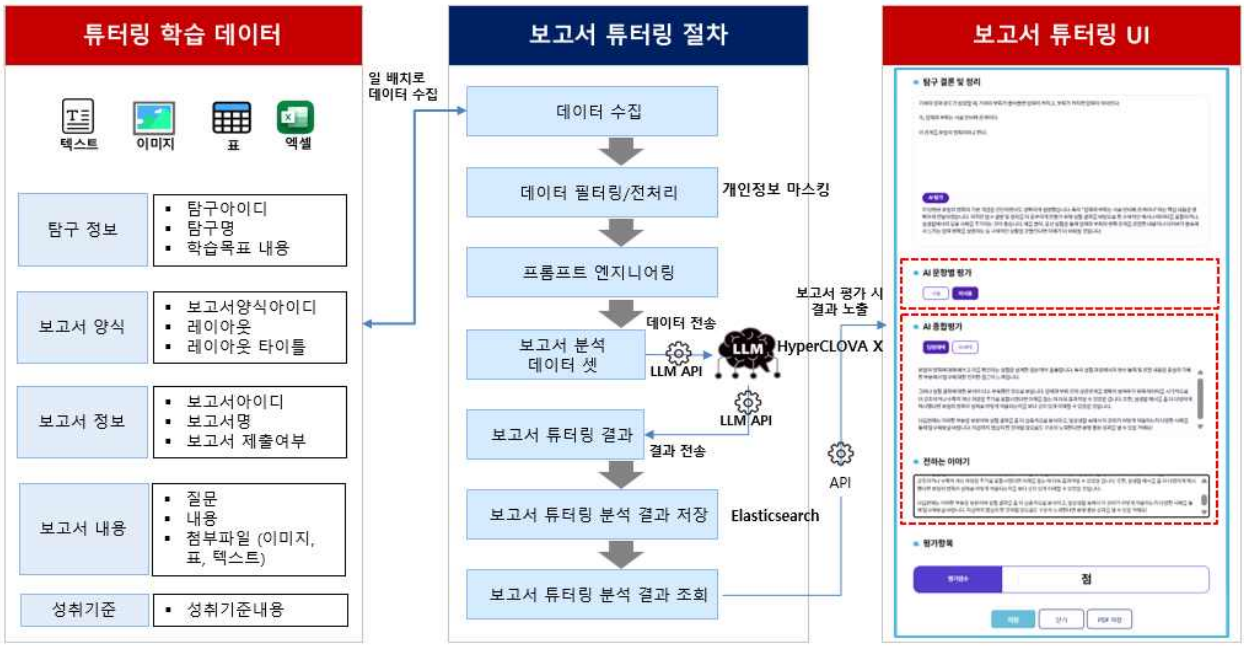
가 보고서 튜터링 서비스 개요

- 보고서 튜터링 서비스는 학습목표와 성취기준, 보고서의 문제에 기반하여 학생들이 작성한 텍스트, 이미지(그래프), 표의 결과를 해석하고 튜터링 피드백을 제공하는 기능으로서 각 보고서 작성 항목별 피드백과 총괄 피드백이 교사의 선택에 의해 제공될 수 있도록 개발되었음. LLM은 개인정보 국외 이전 등의 정보보호 문제를 안정적으로 다루기 위해 국내 LLM인 HyperCLOVA X를 활용하였으며 멀티모달 상호작용이 가능한 HCX-005모델을 도입함.
- 프롬프트 엔지니어링은 1)역할 및 톤 설정, 2)평가방식 설계, 3)전반적 피드백 방식 설계, 4)출력 구조 설계의 프레임워크를 기반으로 이루어졌음.

나 보고서 튜터링 서비스 활용 방법

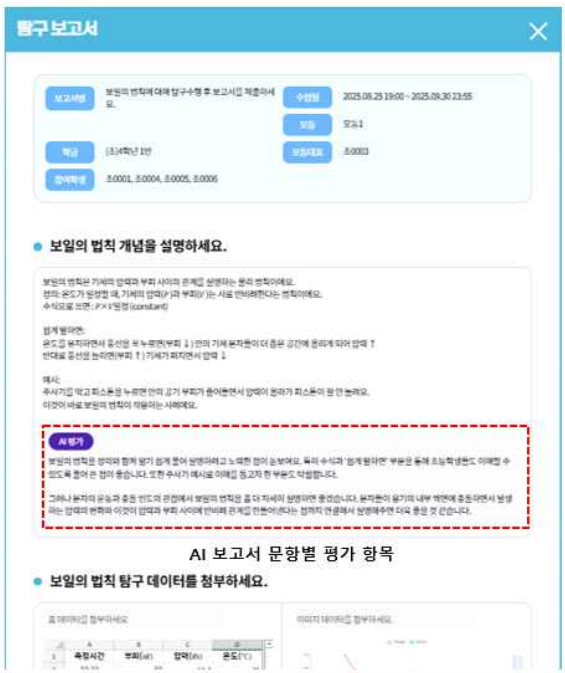
1) 보고서 튜터링 서비스 활용 준비

- 보고서 튜터링 서비스는 별도의 활용 설정이나 준비를 필요로 하지 않음. 학생들이 보고서를 제출하면 일정 시간 안에 자동으로 보고서 튜터링 기능이 제공되고 교사가 AI의 평가 의견을 검토한 뒤 학생들에게 평가 결과를 전송하는 일반적인 '탐구수업' 보고서 평가 방식대로 활용하면 됨.
- 보고서 튜터링 서비스의 작동 시간은 시스템의 정책으로 결정해야 할 사항이며, 본 연구에서는 LLM 사용이 몰리지 않는 밤사이에 보고서 튜터링이 적용되도록 하였음. LLM 활용 토큰 용량이 충분하다면 실시간으로 활용하는 것도 가능하며 현실적인 제약을 검토하여 운영방식을 결정할 필요가 있음.

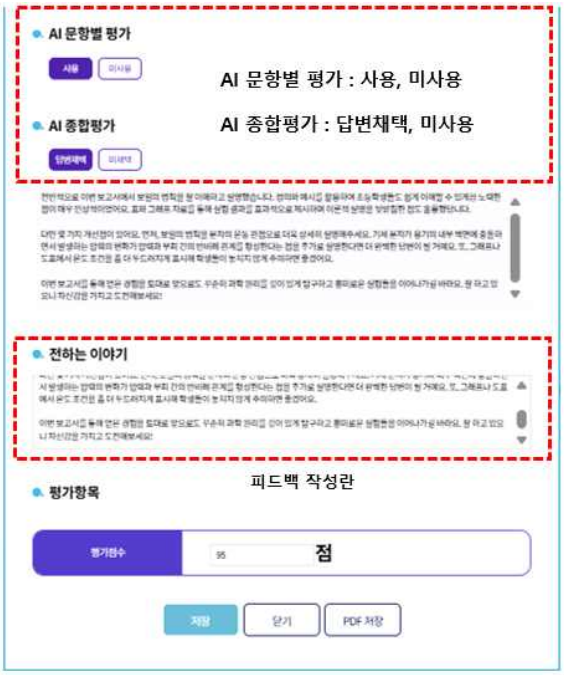


2) 보고서 튜터링 서비스 활용

- 학생들의 보고서가 제출되면 자동으로 'AI 문항별 평가'와 'AI 종합평가' 결과가 반영되고, 교사가 먼저 이를 확인하여 채택을 할 것인지를 선택할 수 있음. AI 문항별 평가 결과를 학생들에게 제공하고자 할 경우 '사용' 버튼을 눌러 전반적인 문항별 피드백을 반영하도록 할 수 있음. AI 종합평가 결과를 활용할 경우에는 '답변 채택' 버튼을 누르면 하단의 '전하는 이야기'란에 해당 내용이 그대로 복사되는데, 교사가 자유롭게 편집, 재구성 하여 종합평가 의견을 수정할 수 있음.
- 교사가 평가점수를 입력하고 평가결과를 저장하게 되면 학생들에게 평가결과가 전송되며 튜터링 결과를 함께 확인할 수 있음

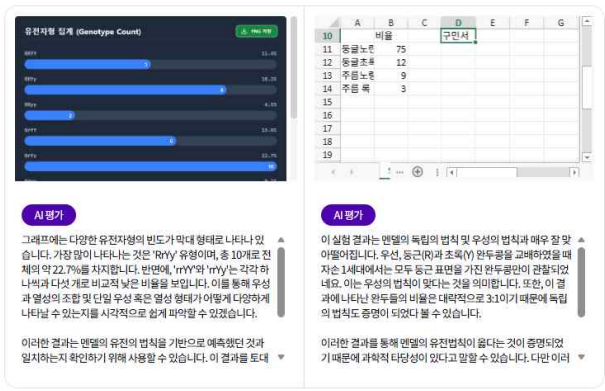


AI 보고서 문항별 평가 항목

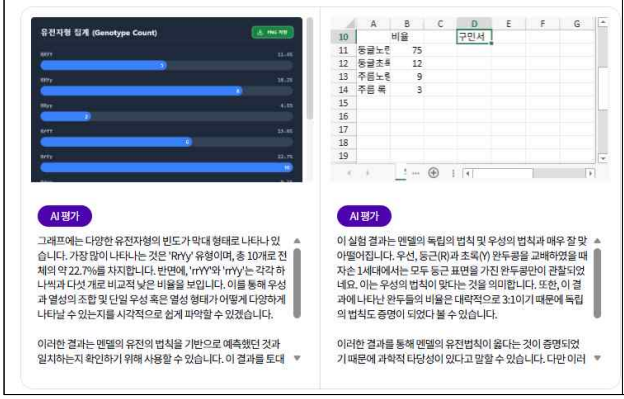
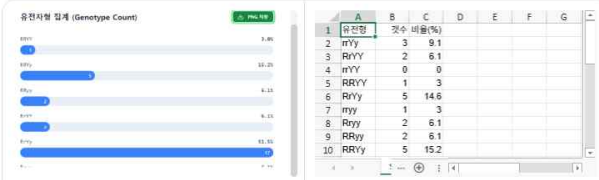


다 보고서 튜터링 서비스 활용 예시

- 학생의 과학탐구 보고서를 제출 후 AI 기반 튜터링 서비스를 통해 보고서 항목별 및 총괄 피드백을 제공하면, 교사의 검토 후 학생에게 결과 전송, 학생이 튜터링 결과를 확인하고 의견을 수렴하여 수정된 과학탐구 보고서를 제출받아 전·후를 비교함
- 학생들은 AI의 보고서 튜터링을 통해 과학 개념 이해를 심화하고, 탐구보고서에 깊이 있는 탐구 결과 및 보다 구체화된 서술 제시

(튜터링 전) 4개 유전형의 비율만 도출	**(튜터링 후)** 10개 유전형의 개수와 비율을 구체적으로 도출
3. [모둠] 마이쥬로 잡종 1대를 자가 수분 시켜 나온 유전형 결과 표와 그래프를 나타내 보세요.  AI 평가 그래프에는 다양한 유전자형의 빈도가 막대 형태로 나타나 있습니다. 가장 많이 나타나는 것은 'RrYy' 유형이며, 총 10개로 전체의 약 22.7%를 차지합니다. 반면에, 'rrYY'와 'rrYy'는 각각 하나씩과 다섯 개로 비교적 낮은 비율을 보입니다. 이를 통해 우성과 열성의 조합 및 단일 우성 혹은 열성 형태가 어떻게 다양하게 나타날 수 있는지를 시각적으로 쉽게 파악할 수 있겠습니다. 이러한 결과는 멘델의 유전의 법칙을 기반으로 예측했던 것과 일치하는지 확인하기 위해 사용할 수 있습니다. 이 결과를 토대 AI 평가 이 실험 결과는 멘델의 독립의 법칙 및 우성의 법칙과 매우 잘 맞아떨어집니다. 우선, 둥근(R)과 초록(Y) 완두콩을 교배하였을 때 자손 1세대에서는 모두 둥근 표면을 가진 완두콩만이 관찰되었네요. 이는 우성의 법칙이 맞다는 것을 의미합니다. 또한, 이 결과에 나타난 완두들의 비율은 대략적으로 3:1이기 때문에 독립의 법칙도 증명이 되었다 볼 수 있습니다. 이러한 결과를 통해 멘델의 유전법칙이 옳다는 것이 증명되었기 때문에 과학적 타당성이 있다고 말할 수 있습니다. 다만 이러	**3. [모둠] 마이쥬로 잡종 1대를 자가 수분 시켜 나온 유전형 결과 표와 그래프를 나타내 보세요.** AI 평가 먼저 이 자료에서는 다양한 유전자형을 나타내고 있습니다. 가장 많이 나타나는 것부터 순서대로 나열하면 'RrYy', 'RrYY' 등의 순으로 나타나고 있네요. 이러한 유형들의 비율을 정확히 알면 학생들이 유전 실험 결과를 이해하는 데 도움이 될 것으로 보입니다. 다만 이 그래프만으로는 어떤 실험인지 명확히 알기 어려우므로, 먼저 이 실험이 어떤 의미를 가지고 있고 또 어떻게 진행되었는지 부가적인 설명을 해주었으면 합니다. 이를 통해 청중들을 AI 평가 이 결과를 보면, 우성과 열성의 부모로부터 자손들이 어떤 비율을 나타내는지 알 수 있어요. 예를 들어, RrYy와 같은 경우 14.6%인 것을 보아 멘델 유전의 분리의 법칙과 독립의 법칙을 잘 이해하고 있는 것 같아요. 또한 이 실험에서는 우열의 관계를 확실히 알 수 있도록 Y가 우성으로 작용한다는 전제를 두고 실험했다는 점도 확인할 수 있습니다. 이러한 결과는 멘델의 분리의 법칙과 독립의 법칙에도 잘 맞아떨어지기 때문에, 학생들이 해당 개념을 정확하게 이해하고 있

- 실험 · 탐구의 도구 및 활용 방법 등에 대한 정보를 습득하여 원활한 탐구수업 참여를 위한 준비 수준 제고

(튜터링 전) 측정 도구에 대한 정보 부족	**(튜터링 후)** 측정 도구에 대한 구체적인 정보 작성
3. [모둠] 마이쥬로 잡종 1대를 자가 수분 시켜 나온 유전형 결과 표와 그래프를 나타내 보세요.  AI 평가 그래프에는 다양한 유전자형의 빈도가 막대 형태로 나타나 있습니다. 가장 많이 나타나는 것은 'RrYy' 유형이며, 총 10개로 전체의 약 22.7%를 차지합니다. 반면에, 'rrYY'와 'rrYy'는 각각 하나씩과 다섯 개로 비교적 낮은 비율을 보입니다. 이를 통해 우성과 열성의 조합 및 단일 우성 혹은 열성 형태가 어떻게 다양하게 나타날 수 있는지를 시각적으로 쉽게 파악할 수 있겠습니다. 이러한 결과는 멘델의 유전의 법칙을 기반으로 예측했던 것과 일치하는지 확인하기 위해 사용할 수 있습니다. 이 결과를 토대 AI 평가 이 실험 결과는 멘델의 독립의 법칙 및 우성의 법칙과 매우 잘 맞아떨어집니다. 우선, 둥근(R)과 초록(Y) 완두콩을 교배하였을 때 자손 1세대에서는 모두 둥근 표면을 가진 완두콩만이 관찰되었네요. 이는 우성의 법칙이 맞다는 것을 의미합니다. 또한, 이 결과에 나타난 완두들의 비율은 대략적으로 3:1이기 때문에 독립의 법칙도 증명이 되었다 볼 수 있습니다. 이러한 결과를 통해 멘델의 유전법칙이 옳다는 것이 증명되었기 때문에 과학적 타당성이 있다고 말할 수 있습니다. 다만 이러	**3. [모둠] 마이쥬로 잡종 1대를 자가 수분 시켜 나온 유전형 결과 표와 그래프를 나타내 보세요.**  AI 평가 먼저 이 자료에서는 다양한 유전자형을 나타내고 있습니다. 가장 많이 나타나는 것부터 순서대로 나열하면 'RrYy', 'RrYY' 등의 순으로 나타나고 있네요. 이러한 유형들의 비율을 정확히 알면 학생들이 유전 실험 결과를 이해하는 데 도움이 될 것으로 보입니다. 다만 이 그래프만으로는 어떤 실험인지 명확히 알기 어려우므로, 먼저 이 실험이 어떤 의미를 가지고 있고 또 어떻게 진행되었는지 부가적인 설명을 해주었으면 합니다. 이를 통해 청중들을 AI 평가 이 결과를 보면, 우성과 열성의 부모로부터 자손들이 어떤 비율을 나타내는지 알 수 있어요. 예를 들어, RrYy와 같은 경우 14.6%인 것을 보아 멘델 유전의 분리의 법칙과 독립의 법칙을 잘 이해하고 있는 것 같아요. 또한 이 실험에서는 우열의 관계를 확실히 알 수 있도록 Y가 우성으로 작용한다는 전제를 두고 실험했다는 점도 확인할 수 있습니다. 이러한 결과는 멘델의 분리의 법칙과 독립의 법칙에도 잘 맞아떨어지기 때문에, 학생들이 해당 개념을 정확하게 이해하고 있

2절 교사 · 학생 설문 분석

1 설문 분석의 개요

- 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』은 학생들의 과학 탐구 과정 전반을 지원하기 위해 설계되었으며, 주요 기능은 **'AI 보고서 튜터링'**과 **'대화형 서비스'** 두 가지로 구성됨.
- AI 보고서 튜터링은 학생들이 작성한 과학탐구 보고서를 분석하여 내용의 논리성, 형식, 그리고 포함된 과학 개념의 정확성 등에 대한 맞춤형 피드백을 실시간으로 제공함.
- 대화형 서비스 (AI 챗봇)은 학생들이 탐구 수업 중 발생하는 과학 개념, 실험 방법, 추가 자료 등에 대한 질문을 던지면, 즉각적으로 답변하고 관련 정보를 탐색할 수 있도록 지원하는 실시간 Q&A 기능을 제공함.
- 본 분석은 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』의 AI 보고서 튜터링 및 대화형 서비스(AI 챗봇)의 **현장 수용도** 및 **교육적 효과**를 다각적으로 진단하기 위해 **초중고 교사 13명과 학생 287명(총 313명)**을 대상으로 수행되었음.
- **정량분석**은 교사 및 학생을 대상으로 한 설문조사 결과를 활용하여 기술 수용도 및 교육적 효과성을 측정하였음.
- **정성분석**은 교사 및 학생의 사전·사후 설문조사 시 제출된 자유 응답형 피드백을 수집·분석하여 사용 행태 및 개선 요구사항을 도출하였음.

2 교사의 인식

가 사전 설문조사

1) 교사의 응답자 특성

- 본 사용성 평가에 참여한 교사 응답자는 총 13명으로, **초등학교(7명, 53.8%)**, **중학교(3명, 23.1%)**, **고등학교(3명, 23.1%)** 교사가 모두 포함되었음.
- 교사 응답자들은 AI 기반 기술에 대한 이해도가 높고, 실제 교육 환경에서 AI 서비스를 적극적으로 활용해 온 것으로 나타났음. 응답자 13명 전원(100%)이 ChatGPT, 문서 요약 AI 등 AI 서비스를 사용해 본 경험이 있다고 응답해 응답 교사 전원이 **AI 기술에 대한 접근성을 이미 확보**한 것으로 나타남.
- 응답자 중 매일 또는 1주일에 3~4회 AI 서비스를 활용하는 교사가 8명으로 과반을 차지함. 이들은 주로 ChatGPT 및 Gemini 를 주로 사용하며, 이는 AI가 이미 **교사의 업무 및 학습 지원 과정에 깊숙이 통합**되어 있다고 판단됨.
- "과학수업 및 학생 지도에서 가장 부담스러운 부분은 무엇입니까?" 라는 질문에 대해 가장 많은 9명의 교사가 보고서(과제) 평가 및 피드백에 부담을 느낀다고 답했으며, 8명의 교사가 개별 학생 맞춤 지도의 어려움을 언급해 교사의 가장 부담스러운 업무는 1) **평가 및 피드백**과 2) **개별 학생 맞춤 지도**인 것으로 나타남. 이는 AI 보고서 튜터링 기능과 대화형 서비스가 단순히 부가적인 편의 기능이 아니라, **교사의 가장 시급한 업무 경감 요구를 충족시킬 수 있는 중요한 수단**이 될 가능성을 시사하고, 따라서 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』 **도입의 정당성을 지지하는 결과**로 볼 수 있음.

2) 서비스 사전 기대에 대한 정량분석

- 정량분석은 “대화형 서비스(AI챗봇)”과 “AI 보고서 튜터링” 서비스에 대한 **교사들의 사전 기대**와 관련된 설문 결과를 보여주고 있음
- **가장 높은 평균 기대치**를 보인 항목은 **'5-1. AI 챗봇이 학생의 과학탐구 역량 강화에 효과적일 것이다'** (M=4.31) 였음. 이는 교사들이 AI 챗봇을 단순한 정보 제공 도구가 아닌, 학생들의 능동적인 탐구 활동을 촉진하고 심화시키는 데 핵심적인 역할을 할 것이라고 기대하고 있음을 시사함. 챗봇의 즉각적인 상호작용 및 개념 지원 기능이 보고서 작성 단계뿐만 아니라 탐구 과정 전반에 걸쳐 유용할 것이라는 초기 인식이 강하게 반영된 것으로 해석됨.
- **가장 낮은 평균 기대치**와 **가장 높은 표준편차**를 기록한 항목은 **'6-4. AI 보고서 튜터링이 학생의 학습 흥미와 동기부여에 효과적일 것이다'** (M=3.69, SD=1.031) 였음. 낮은 평균은 교사들이 AI 보고서 튜터링을 기술적이고 기능적인 지원 도구로 인식할 뿐, 학생들의 내재적 학습 동기나 흥미를 끌어올리는 **정서적/교육 심리적 효과에 대해서는 상대적으로 회의적이거나 낮은 기대**를 가지고 있음을 나타냄. 한편 높은 표준편차는 AI 튜터링이 학생 동기부여에 기여할지에 대해 **교사 집단 내에서 의견 차이가 가장 크다는 것을 시사**함. 일부 교사는 긍정적 효과를 기대하는 반면, 다른 일부는 AI가 제공하는 형식적 피드백이 학생의 흥미를 저해할 수도 있다고 우려하는 것으로 해석될 수 있음.
- 과학 개념 이해 (5-3 vs 6-3)와 관련해, AI 챗봇(M=4.08)과 AI 보고서 튜터링(M=4.23) 모두 **'깊이 있는 이해'에 대해 높은 기대**를 받았음. 이는 교사들이 두 AI 기능 모두 학생의 **개념 학습 심화에 중요한 기여를 할 것으로 예상**한다는 나타냄.
- 기초 학력 향상 (5-2 vs 6-2)와 관련해 두 기능 모두 '기초 학력 향상'에 대해서는 비교적 낮은 기대를 받았음 (M=3.85). 이는 AI가 **기초적인 지식 습득보다는 탐구 역량이나 심화 개념 이해와 같은 고차원적 사고 지원에 더 적합할 것이라는 교사들의 인식을 반영**한다고 해석 될 수도 있음. 또한 이 두 항목은 표준편차도 0.988로 높아, AI가 기초 학력 증진에 효과적일지에 대한 교사들의 합의도가 낮음을 시사함.

〈AI기반 서비스 전반에 대한 사전 기대〉

세부 질문 항목	평균	표준편차
대화형 서비스에 대한 기대		
5-1. 학생의 과학탐구 역량 강화에 효과적일 것이다.	4.31	0.632
5-2. 학생의 과학 기초학력 향상에 효과적일 것이다.	3.85	0.988
5-3. 학생의 과학 개념의 깊이 있는 이해에 효과적일 것이다.	4.08	0.613
5-4. 학생의 학습 흥미와 동기부여에 효과적일 것이다.	4.23	0.730
AI 보고서 튜터링에 대한 기대		
6-1. 학생의 과학탐구 역량 강화에 효과적일 것이다.	4.15	0.768
6-2. 학생의 과학 기초학력 향상에 효과적일 것이다.	3.85	0.988
6-3. 학생의 과학 개념의 깊이 있는 이해에 효과적일 것이다.	4.23	0.730
6-4. 학생의 학습 흥미와 동기부여에 효과적일 것이다.	3.69	1.031

3) 서비스 사전 기대에 대한 정성분석

- 정성분석은 “사용성 검토 전, 「 지능형 과학실 ON 」 “대화형 서비스(AI 챗봇)” (“AI보고서 튜터링“)에 대해 실제 수업 현장에서 기대하는 점이나 바라는 점이 있다면 자유롭게 입력해 주세요.” 라는 질문에 대한 교사들의 서술형 응답에 기반함.
- 교사들은 대화형 서비스에 대해 학생의 **자기 주도적 학습 환경 조성**, **미래형 역량 함양**, **개별 맞춤형 지도** 측면에서 가장 크게 기대하는 것으로 나타남

〈대화형 서비스에 대한 사전 기대〉

주요 기대 영역	교사 응답 원문	시사점
자기 주도 학습 및 탐구 지원	“학생들의 과학 학습 – 자기주도 학습이 이어질 수 있는 공간, 기능이 되었으면 좋겠다.”, “학생들의 과학적 개념과 탐구과정의 기회를 확대할 수 있는 대화형 서비스”	교사가 상시적으로 모든 학생의 질문에 대응하기 어려운 환경에서, 챗봇이 학생들의 궁금증을 즉시 해소하고, 과학적 개념을 스스로 검색할 수 있는 환경을 제공함으로써 자기 주도 학습을 촉진할 것으로 기대함.
미래 역량 함양	“이제 AI는 일상이 되었고 학생들은 인공지능과 함께 공부해야되는 시대가 되었습니다. 그에 따라 학생들이 스스로 사고할 수 있는 과정이 필요합니다. 특히 AI에게 질문하는 능력이 요구되는데 이러한 수업을 AI와 함께 적용함으로써 이룰 수 있다고 생각합니다.”	교사들은 단순히 AI가 답을 주는 것을 넘어, 'AI에게 질문하는 능력'과 '스스로 사고하는 과정'을 교육하는 도구로서 AI를 인식함. 이는 AI가 단순 검색 엔진이 아닌, 학생의 탐구 자세를 형성하는 교육적 매개체가 되기를 바라는 기대를 반영함.
개별 맞춤형 지도	“학생들이 맞춤형으로 수업을 진행할 수 있도록 해 주는 좋은 도구가 될 것이라는 기대감이 있습니다. 현장에서 나온 피드백으로 더욱 완성도 높은 도구가 학교에 보급되면 좋겠습니다.”, “수업시간 중의 개인적인 의문점을 해결할 수 있는 맞춤형 도구가 될 수 있을 것 같고, 다른 플랫폼에서는 주제를 필터링 해주지 않는 반면에 지능형 과학살 on 내의 대화형 서비스는 과학 주제에 관한 내용으로 한정 지어 주기에 수업 중 사용을 크게 제한하지 않아도 되리라 기대한다.”	AI가 학생 개개인의 부족한 부분을 파악하고 맞춤형으로 학습을 지원함으로써 교사의 개별 지도 부담을 덜어줄 것으로 기대함.

- 교사들은 AI보고서 튜터링 서비스에 대해 학생의 교사 평가 부담 경감, 개별 맞춤형 피드백 제공, 과정중심 평가 측면에서 가장 크게 기대하는 것으로 나타남

〈보고서 튜터링 서비스에 대한 사전 기대〉

주요 기대 영역	교사 응답 원문	시사점
교사 평가 부담 경감 및 효율화	“학생들을 기준에 따라 평가하고 적용하는데 있어서 개별성 있는 평가가 부담스러웠습니다. AI가 이러한 과정에 대한 부담을 덜어줄 수 있을것이라 생각합니다. 단순한 도구를 떠나서 다양한 피드백을 적용할 수 있을 것 같습니다.”, “모든 학생들의 자료를 선생님이	교사들은 '보고서(과제) 평가 및 피드백'을 가장 큰 부담(9명 선택)으로 꼽았는데, AI 튜터링이 복잡한 과정 중심 평가의 실제적 운영을 지원함으로써 평가 업무 부담을 실질적으로 경감해 주기를 기대함.

	관리하기가 힘들었는데 교사의 업무량도 줄이고 학생들도 맞춤형으로 서비스를 받을 수 있어 기대가 됩니다.”	
개별 맞춤형 피드백 제공	“모든 학생들의 자료를 선생님이 관리하기가 힘들었는데 교사의 업무량도 줄이고 학생들도 맞춤형으로 서비스를 받을 수 있어 기대가 됩니다.”	AI가 학생 개개인에게 개별성 있는 피드백을 제공하고, 교사가 다양한 피드백을 효율적으로 적용할 수 있는 기반을 마련해주기를 기대함. 이는 모든 학생에게 수준 높은 맞춤형 지도를 제공하고자 하는 교사의 전문적인 요구를 반영한다고 볼 수 있음
과정중심 평가	“교사들의 과정중심 평가의 실제적 운영을 지원하는 플랫폼 과정이었으면 좋겠다.”	교사들은 이러한 AI시스템이 오랫동안 논의되어온 과정중심 평가를 실현하는 도구가 될 수 있기를 기대함

나 사후 설문조사

1) 서비스 사용성 정량분석

- 본 분석은 AI 기반 과학탐구 지원 시스템의 사용성 검토에 참여한 교사 13명을 대상으로 사후 설문조사 데이터를 활용하여, 시스템에 대한 최종적인 만족도, 수용 의도, 교육적 가치를 정량적으로 분석함.
- 총 응답자 수는 13명이었으며, 5점 리커트 척도 (5점: 매우 그렇다/만족, 1점: 매우 그렇지 않다/불만족)를 사용하였고, 각 교사별로 유용성, 사용 용이성, 사용 의도, 사용 태도, 신뢰·정확도, 학습효과 및 교육적 가치의 총 6개 사용성 평가 하위 요인에 대해 각 하위 요인 내의 모든 문항 평균을 계산하여 개인 총점을 산출한 후, 13명 교사 전체의 평균(M) 및 표준편차(SD)를 산출하였음.
- “대화형 서비스”와 관련된 사용성 평가 6개 하위 요인 설문조사 결과를 요약하면 아래 표와 같음

〈대화형 서비스 사용성 평가 6개 하위 요인 설문조사 결과〉

하위 요인	문항수	평균	표준편차
유용성	4개	4.62	0.43
사용 용이성	4개	4.60	0.49
사용 의도	3개	4.69	0.67
사용 태도	3개	4.54	0.60
신뢰·정확도	2개	4.35	0.47
학습효과 및 교육적 가치	4개	4.37	0.55

- 대화형 서비스의 **사용 의도(M=4.69)가 가장 높은 점수**를 기록하였고, 이는 교사들이 앞으로도 지속적으로 이러한 시스템을 사용할 의도가 있음을 보여주고 있음.
- 유용성(M=4.62)과 사용 용이성(M=4.50) 또한 5점 만점에 4.5점 이상을 기록하며 매우 높게 나타남. 이는 교사들이 AI 챗봇을 통해 학생들의 학습 목표 달성 및 수업 중 질문 해결이 효율적으로 이루어진다고 판단했으며, 또한 시스템 사용도 쉬웠음을 나타냄.

- 모든 평가 영역 중 신뢰·정확도(M=4.35) 영역이 가장 낮음 점수를 기록함. 비록 4점 이상으로 여전히 긍정적이지만, 사용 의도(M=4.69)와의 격차는 주목할 필요가 있음. 이는 앞으로도 챗봇을 사용하고자 하나 **챗봇이 제공하는 과학적 답변 내용의 질과 정확성에 대해서는 여전히 일부 검토 및 개선의 여지가 있다고 교사들이 인식**하고 있음을 반영함.
- 학습효과 및 교육적 가치(M=4.37)도 4점 이상이나 상대적으로 낮은 점수를 보임. 이는 교사들이 AI 챗봇이 **학생들의 과학탐구 역량 강화나 개념의 깊이 있는 이해에 실질적으로 긍정적인 영향을 미치기 위해서는 더 많은 개선이 필요**하다고 인식함을 보여줌.
- "AI보고서 튜터링"과 관련된 사용성 평가 6개 하위 요인 설문조사 결과를 요약하면 아래 표와 같음.

〈보고서 튜터링 서비스 사용성 평가 6개 하위 요인 설문조사 결과〉

하위 요인	문항수	평균	표준편차
유용성	4개	4.62	0.36
사용 용이성	4개	4.56	0.49
사용 의도	3개	4.62	0.83
사용 태도	3개	4.46	0.60
신뢰·정확도	2개	4.04	0.59
학습효과 및 교육적 가치	4개	4.38	0.59

- AI 보고서 튜터링 기능은 **유용성(M=4.62), 사용 용이성(M=4.56), 사용의도(M=4.62)에서 가장 높은 점수를 보임**. 이는 AI 튜터링이 교사들이 사전적으로 가장 큰 부담으로 꼽았던 학생별 보고서 평가 및 개별 피드백 제공 업무를 실질적으로 경감하고, 보고서의 품질 향상이라는 목표 달성에 크게 기여했고, 시스템의 인터페이스 역시 직관적이고, 사용법 학습에 시간이 거의 들지 않았다는 교사들의 평가를 반영하는 것으로 보임.
- **사용 의도(M=4.62) 및 사용 태도(M=4.46)가 높게** 나타난 것은, AI 보고서 튜터링이 단순한 보조 도구를 넘어 평가 및 지도 활동의 필수적인 도구로 자리매김할 가능성이 매우 높음을 시사함.
- **신뢰·정확도(M=4.04)가 가장 낮은 평균 점수**를 나타냄. 이는 AI 튜터링이 작동 방식이나 효율성 측면에서는 높은 점수를 받았으나, AI가 생성하는 피드백 내용의 과학적 정확성이나 교육적 적절성에는 여전히 교사들의 비판적인 검토와 개선 요구가 존재함을 반영함.

2) 서비스 사용성 정성분석

- "대화형 서비스"에 대한 정성분석은 "「 지능형 과학실 ON 」"대화형 서비스(AI 챗봇)"에서 가장 유용했던 기능 또는 장점이 있다면 자유롭게 입력해 주세요. "와 "「 지능형 과학실 ON 」"대화형 서비스(AI 챗봇)"에서 개선되었으면 하는 점이 있다면 자유롭게 입력해 주세요" 라는 질문에 대한 교사들의 서술형 응답에 기반함.
- 교사들은 대화형 서비스에 대해 학생의 자율적 탐구 활동 지원하는 기능, 과학 개념을 쉽게 설명해 주는 기능, 키워드만 입력해도 관련 개념과 자료를 제공하는 기능, 모둠별 다양한 탐구 활동을 가능하게 하는 기능에 대해 긍정적인 평가를 내림

〈대화형 서비스의 유용한 기능〉

유용한 기능	교사 응답 원문	설명
학생의 자율적 탐구 활동 지원 기능	“학생들이 궁금한 개념에 대해 묻고 답을 찾아갈 수 있어 좋았다.”	학생이 자유롭게 질문하고 개념을 탐색하고, 토론, 실험 활동 중 실시간으로 개념을 검색하며 논의를 확장하는 데 도움이 됨
과학 개념을 쉽게 설명해 주는 기능	“과학 개념 중 모르는 단어, 어휘 등을 설명해주는데 아주 유용했다.”	학생 수준에 맞게 낯선 개념·어휘 설명하고, 교과서 이해 부족을 보완하는 역할을 수행함
키워드만 입력해도 관련 개념과 자료를 제공해 주는 기능	“단어만 입력해도 관련 개념과 교육자료가 나오는 것이 좋았다.”	검색 능력이 약한 초·중등 학생에게 강점을 보이고, 교사의 설명 없이도 핵심 배경지식 확보 가능
모둠별 다양한 탐구 활동 지원 기능	“교사의 획일화된 실험이 아닌 모둠별로 다양한 실험이 나올 수 있어서 좋았다.”	교사의 일률적 지도 방식에서 벗어나, 모둠마다 다른 실험·아이디어를 도출하는 데 기여함

- 교사들은 대화형 서비스에 대해 부정확한 정보 제공, 응답 지연 등 시스템 안정성 문제, 초등학생에게 접근성이 낮은 문제, 정확도 개선 등의 개선 사항을 제시함

〈대화형 서비스에 대한 개선점〉

개선 사항	교사 응답 원문	설명
부정확한 정보 제공	“엉뚱한 답변을 할 때가 많다.”	사실과 다르거나 문맥에 맞지 않는 답변이 종종 발생해 학생이 잘못된 개념을 학습할 가능성 우려가 있음
응답 지연 등 시스템 안정성 문제	“사용 도중 응답하지 않는 경우가 많았다.”	수업 중 챗봇이 멈추고 응답하지 않는 경우가 잦았다는 지적이 있었고, 이는 실시간 활동 중심의 과학 수업에 방해로 작용
초등학생에게 낮은 접근성	“초등학생들에게는 접근성이 좀 떨어진다.”	저학년 학생은 키보드 입력·표현이 어려워, 음성 입력 또는 간단한 인터페이스 요구
정확도 개선	“정확도가 더 올라갔으면 한다.”	과학 개념은 정확성이 매우 중요하기 때문에, 교사 입장에서는 챗봇의 개념 설명 신뢰도를 더 높일 필요가 있음

- “AI보고서 튜터링”에 대한 정성분석은 “「 지능형 과학실 ON 」“AI 보고서 튜터링”에서 가장 유용했던 기능 또는 장점이 있다면 자유롭게 입력해 주세요. “와 ”「 지능형 과학실 ON 」“AI 보고서 튜터링”에서 개선되었으면 하는 점이 있다면 자유롭게 입력해 주세요. “ 라는 질문에 대한 교사들의 서술형 응답에 기반함.
- 교사들은 AI보고서 튜터링에 대해 보고서 방향·내용 정리 및 핵심 개념 보완 기능, 사진자료 기반의 맞춤 피드백 제공 기능, 교사의 피드백 시간 절감 효과, 평가 부담 감소 및 피드백 품질 향상에 대해 긍정적인 평가를 내림

〈보고서 튜터링 서비스의 유용한 기능〉

유용한 기능	교사 응답 원문	설명
보고서 방향·내용 정리 및 핵심 개념 보완 기능	“학생 보고서의 전체 방향과 추가 개념, 보완점 등을 잘 정리해줘서 좋았다.”	AI가 학생 보고서를 읽고 핵심 내용·부족한 개념·보완점을 명확하게 정리해 주는 기능이 매우 유용하다고 평가됨.
사진자료 기반의 맞춤 피드백 제공 기능	“사진 내용을 보고 맞춤 피드백을 해주는 점이 좋았다.”	학생이 제출한 사진·그래프 등 보고서 자료를 직접 분석해 피드백하는 기능을 긍정적으로 평가.
교사의 피드백 시간 절감 효과	“피드백 시간이 매우 줄었고 내용도 탄탄해졌다.”	가장 많이 언급된 장점 중 하나 였으며 반복적인 보고서 확인·피드백 작업의 부담이 획기적으로 줄어들었다는 평가
평가 부담 감소 및 피드백 품질 향상	“적절한 피드백을 추천해줘서 평가 부담이 줄었다.”	일관적인 기준으로 피드백을 제시해 평가 부담을 줄여주고, 학생 개별 특성에 따른 피드백 품질이 높아짐.

- 교사들은 AI보고서 튜터링 서비스에 대해 교육과정·학년 수준과 맞지 않는 용어 사용, 과도한 칭찬 및 부정확한 튜터링 내용, 보고서 내용과 무관한 피드백 등장, 피드백 선택지 제공 필요, 제출된 사진·자료 분석 정확도 향상 등의 개선 사항을 제시함

〈보고서 튜터링 서비스에 대한 개선점〉

개선 사항	교사 응답 원문	설명
교육과정·학년 수준과 맞지 않는 용어 사용	“초등 교육과정 단계에 맞는 용어로 제시되면 좋겠다.”	특히 초등학교 교사들이 지적
과도한 칭찬 및 부정확한 내용	“칭찬이 너무 많고 내용이 부정확할 때가 있다.”	AI가 과하게 칭찬하거나, 흐름과 맞지 않는 피드백을 주는 문제
보고서 내용과 무관한 피드백	“탐구와 관련 없는 내용이 나와 당황스러웠다.”	과학 탐구 보고서인데 전혀 다른 주제(미세먼지 등)를 언급
피드백 선택지	“챗지피티처럼 2가지 중 선택할 수 있는 기능이 있으면 좋겠다.”	한 가지 피드백만 제공하는 것이 아니라, 둘 이상의 대안 제시 요구.
사진·자료 분석 정확도 향상	“그림을 더 정확하게 분석해주면 좋겠다.”	사진 기반 분석 피드백이 유용하지만, 정확성은 아직 부족.

2 학생의 인식

가 학생의 응답자 분석

- 설문 조사에 참여한 287명의 학생은 학교급은 초등학교 91명(31.7%), 중학교 79명(27.5%), 고등학교 117명(40.8%), 성별은 남학생 173명(60.3%), 여학생 114명(39.7%)으로 구성되었음.

〈학생 응답자의 학교급과 성별〉

구분	빈도	백분율
학교급		
고등학교	117	40.8%
중학교	79	27.5%
초등학교	91	31.7%
성별		
남	173	60.3%
여	114	39.7%

- "AI 서비스(예: ChatGPT, 문서 요약 AI 등)를 사용해 본 경험이 있습니까?" 라는 질문에 대해 261명(90.9%)의 학생이 "예"라고 답해 상**당수의 학생들이 이미 AI 서비스를 사용해 본 경험**이 있는 것으로 나타남.

〈AI 서비스를 사용해 본 경험이 있습니까?〉

응답	빈도	백분율
남	261	90.0%
여	26	9.1%

- "AI 서비스를 얼마나 자주 활용하셨나요?" 라는 질문에 대해 209명(72.8%)의 학생이 "1주일에 1~2회" 이상 사용한다고 답해 **상당수의 학생들이 이미 AI 서비스를 자주 사용**하고 있는 것으로 나타남

〈AI 서비스를 얼마나 자주 활용하셨나요?〉

응답	빈도	백분율
매일	31	10.8%
1주일에 3~4회	73	25.4%
1주일에 1~2회	105	36.6%
1~2회	54	18.8%
없다	24	8.4%

- "실제 사용해 본 AI서비스를 선택해 주세요" 라는 질문에 대해 208명(72.4%)의 학생이 **ChatGPT와 Gemini를 사용**한다고 답함.

〈실제 사용해 본 AI 서비스를 선택해 주세요〉

응답	빈도	백분율
Chatgpt(챗지피티)	134	46.7 %
Chatgpt(챗지피티), Gemini (제미나이)	74	25.8 %
없다.	20	7.0 %
Chatgpt(챗지피티), ClovaX (클로바X), Gemini (제미나이)	5	1.7 %
Chatgpt(챗지피티), ClovaX (클로바X), Gemini (제미나이), 퍼블렉시티	5	1.7 %
Chatgpt(챗지피티), Gemini (제미나이), 뤼튼	5	1.7 %
Chatgpt(챗지피티), Gemini (제미나이), 퍼블렉시티	5	1.7 %
Chatgpt(챗지피티), 뤼튼	5	1.7 %
Chatgpt(챗지피티), ClovaX (클로바X)	4	1.4 %
Chatgpt(챗지피티), Gemini (제미나이), 퍼블렉시티, 뤼튼	3	1.0 %
Gemini (제미나이)	3	1.0 %
기타	24	8.0%

나 서비스 사용에 대한 사전-사후 효과성 정량분석

1) 신뢰도 검증

- 사전-사후 효과성 검증을 위해 과학학습정서, 과학관련자아개념, 과학학습동기, 과학관련진로포부, 과학관련태도의 총 5개 하위 요인을 구성하는 문항들에 대해 Cronbach-α를 통해 신뢰도를 검증하였음. 〈표 2-14〉에 제시된 바와 같이 **모든 하위 요인의 Cronbach-α가 0.7 이상으로 양호한 신뢰도**를 보이는 것으로 나타났음.

〈신뢰도 검증〉

하위요인	문항수	Cronbach-α
사전		
과학학습정서	6문항	0.796
과학관련자아개념	6문항	0.878
과학학습동기	10문항	0.851
과학관련진로포부	5문항	0.843
과학관련태도	8문항	0.882
사후		
과학학습정서	6문항	0.831
과학관련자아개념	6문항	0.903
과학학습동기	10문항	0.864
과학관련진로포부	5문항	0.854
과학관련태도	8문항	0.896

2) 대응표본 t-검정을 사용한 효과성 검증

- 사전-사후 효과성을 살펴보기 위해 **대응표본 t-검정**을 사용해 과학학습정서, 과학관련자아개념, 과학학습동기, 과학관련진로포부, 과학관련태도의 총 5개 하위 요인이 사전검사에 비해 사후검사에서 통계적으로 유의미한 평균의 차이를 보이는지 검증함.
- 대응표본 t-검정 결과, 〈표 2-15〉에 나타난 바와 같이 **『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』 사용을 통해 학생들의 과학학습정서, 과학관련자아개념, 과학학습동기, 과학관련진로포부, 과학관련태도 총 5개의 하위 요인 모두 통계적으로 유의미하게 향상**된 것으로 나타남.

〈사전-사후 효과성 검증을 위한 대응표본 t-검정 결과〉

하위 요인	사전 평균	사전 표준편차	사후 평균	사후 표준편차	평균차이	p-value	효과크기(d)
과학학습정서	4.06	0.90	4.25	0.77	0.19	〈.001	0.25
과학관련자아개념	3.37	1.01	3.62	1.03	0.26	〈.001	0.30
과학학습동기	3.79	0.74	3.95	0.72	0.15	〈.001	0.25
과학관련진로포부	3.71	0.95	3.88	0.93	0.18	〈.001	0.22
과학관련태도	3.99	0.79	4.15	0.76	0.16	〈.001	0.24

3) 서비스 사용성 정량분석

- 본 분석은 AI 기반 과학탐구 지원 시스템의 사용성 검토에 참여한 학생을 대상으로 사후 설문조사 데이터를 활용하여, 시스템에 대한 최종적인 만족도, 수용 의도, 교육적 가치를 정량적으로 분석함.
- 총 응답자 수는 287명이었으며, 5점 리커트 척도 (5점: 매우 그렇다/만족, 1점: 매우 그렇지 않다/불만족)를 사용하였고, 각 학생별로 유용성, 사용 용이성, 사용 의도, 사용 태도, 신뢰·정확도, 학습효과 및 교육적 가치의 총 6개 사용성 평가 하위 요인에 대해 각 하위 요인 내의 모든 문항 평균을 계산하여 개인 총점을 산출한 후, 287명 학생 전체의 평균(M) 및 표준편차(SD)를 산출하였음.
- “대화형 서비스”와 관련된 사용성 평가 6개 하위 요인 설문조사 결과를 요약하면 아래 표와 같음.

〈대화형 서비스에 대한 사용성 평가 6개 하위 요인 설문조사 결과〉

하위 요인	문항수	평균	표준편차
유용성	4개	3.93	0.85
사용 용이성	4개	3.97	0.87
사용 의도	3개	3.90	0.92
사용 태도	3개	3.96	0.84
신뢰·정확도	2개	3.82	0.86
학습효과 및 교육적 가치	4개	3.93	0.79

- 사용 용이성은 6개 하위 요인 중 가장 높은 평균(M=3.97)을 보였으며, 이는 학생들이 대체로 대화형 서비스를 쉽게 이해하고 사용할 수 있었다고 인식하고 있음을 보여줌. 인터페이스 접근성이나 조작성에 대한 부담이 크지 않아 학습 과정에서 서비스 활용이 비교적 수월하게 이루어졌음을 의미함.

- 유용성(M=3.93)과 학습효과 및 교육적 가치(M=3.93)는 모두 평균 4점에 근접한 수치를 기록함. 이는 학생들이 대화형 서비스를 통해 학습 목표 달성, 정보 탐색, 이해 증진 등에서 도움을 받았다고 느꼈음을 의미함. 특히 학습효과 및 교육적 가치 요인의 점수는 대화형 서비스가 학생들의 학습을 실질적으로 지원하는 보조 도구로 기능할 수 있음을 시사함.
- 사용 태도는 3.96으로 높은 편에 속하며, 학생들이 대화형 서비스를 사용하는 것에 대해 전반적으로 긍정적인 정서와 인식을 가지고 있음을 알 수 있음. 이는 학습 과정에서 AI를 활용하는 것에 대한 거부감이 낮고, 기대·흥미·호감 같은 정서적 요소도 비교적 긍정적임을 반영한다고 볼 수 있음.
- 사용 의도는 3.90으로 전체 평균 대비 약간 낮은 편임. 이는 학생들이 현재 서비스 사용에 대해서는 긍정적이나, 앞으로 지속적이고 적극적으로 활용하려는 의지는 상대적으로 덜 형성된 상태임을 의미함.
- 전체 요인 중 가장 낮은 평균(M=3.82)을 보인 영역은 신뢰·정확도였음. 이는 사용자가 대화형 서비스의 응답 품질에 대해 여전히 내용의 정확성·신뢰성에 대한 의구심을 일부 갖고 있음을 시사함. 특히 AI 기반 서비스의 일반적 한계(오류 가능성, 사실성 판단 어려움 등)가 학생 인식에도 반영된 것으로 볼 수 있음.
- “AI보고서 튜터링” 서비스와 관련된 사용성 평가 6개 하위 요인 설문조사 결과를 요약하면 아래 표와 같음.

〈보고서 튜터링에 대한 사용성 평가 6개 하위 요인 설문조사 결과〉

하위 요인	문항수	평균	표준편차
유용성	4개	3.91	0.84
사용 용이성	4개	3.95	0.84
사용 의도	3개	3.88	0.92
사용 태도	3개	3.96	0.84
신뢰·정확도	2개	3.84	0.86
학습효과 및 교육적 가치	4개	3.91	0.80

- 사용 태도(M = 3.96)가 6개 하위 요인 중 가장 높은 점수를 나타냈으며, 이는 학생들이 AI보고서 튜터링 기능을 전반적으로 긍정적으로 받아들이고 있고, 서비스 사용 자체에 대한 심리적 저항이 적으며 친숙함을 느끼고 있음을 시사함.
- 사용 용이성(M = 3.95) 역시 높은 수준을 보였으며, 학생들은 보고서 자동 생성·수정 지원 등 기능을 사용하는 과정에서 크게 어려움을 느끼지 않았고 인터페이스 및 기능 흐름을 비교적 쉽게 이해한 것으로 보임. 이는 시스템의 조작성 및 접근성이 학생 사용자에게 적절히 설계되었음을 반영함.
- 유용성(M = 3.91)은 4점 가까운 점수를 기록하여, 학생들이 보고서 작성 과정에서 AI 튜터링 기능이 실제로 도움을 제공한다고 인식했음을 의미함. 특히 글 흐름 개선, 문장 수정, 요약 등의 기능이 학습 효율성 향상에 기여했을 가능성이 큼.
- 학습효과 및 교육적 가치(M = 3.91)도 비교적 높은 점수로 나타나, AI보고서 튜터링이 학생들의 글쓰기 역량 향상이나 학습 과정 전반에 긍정적인 교육적 효과를 제공한다고 학생들이 평가했음을 보여줌. 이는 튜터링 기능이 단순 기술적 편의를 넘어 교육적 역할도 수행하고 있음을 시사함.
- 사용 의도(M = 3.88)는 다소 낮은 편이지만 여전히 중간 이상 수준으로, 현재 서비스 경험이 긍정적이긴 하나 학생들이 향후에도 지속적으로 해당 서비스를 사용하려는 의지는 상대적으로 덜 확립된 상태임을 의미함. 이는 서비스 기능에 대한 기대 수준, 필요한 시점의 제한, 또는 개인적 선호 차이 등이 영향을 미쳤을 수 있음.
- 신뢰·정확도(M = 3.84)는 모든 하위 요인 중 가장 낮은 점수를 보였으며, 학생들은 AI보고서 튜터링이 제공하는 문장 수정·내용 보완 등의 기능이 유용하다고 느끼면서도 생성된 문서의 정확성·타당성·내용 신뢰성에 대해 일부 우려를 가지고 있는 것으로 해석됨. 이는 AI 기반 문서 생성 도구에 공통적으로

나타나는 오답 가능성, 논리적 일관성 부족 등의 특성과 연결될 수 있음.

- 전반적으로 6개 모든 요인이 3.8~4.0 사이의 안정적인 긍정적 평가를 나타내고 있으며, 학생들은 AI보고서 튜터링 기능을 학습 보조 도구로 유용하게 인식하고 있으나 정보의 정확성에 대한 신뢰는 상대적으로 낮아, 교육 현장에서 실질적 효과를 높이기 위해서는 정확성 강화·출처 제시·검증 로직 개선 등 품질 향상이 필요함을 시사함.

다 서비스 사용성에 대한 교사-학생 정량분석 결과 비교

- 서비스 사용성에 대한 교사-학생 정량분석 결과는 두 가지 점에서 차이를 보임. 첫째, **교사의 평균이 학생의 평균보다 높음**. 둘째, **교사의 표준편차가 학생의 표준편차보다 작음**.

1) 교사의 사용성 평가가 학생의 평가보다 높은 현상

- 교사의 사용성 평균이 학생보다 높은 결과는『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』이 학생보다 교사에게 더 긍정적으로 받아들여지고 있음을 시사하고, 이는 여러 관점에서 해석 가능함.
- 교사는 교육 현장에서 AI 도구를 학습 설계·수업 운영·평가 지원 등 실제 업무에 직접적으로 활용할 수 있어 서비스의 기능적 가치를 더 잘 이해하기때문에 AI 서비스의 기능적 장점을 학생보다 훨씬 더 명확하게 체감했을 가능성이 있음.
- 교사는 직무 특성상 ICT·AI 도구를 지속적으로 사용하며 학생보다 디지털 리터러시 수준이 높음. 이런 특성이 사용성 평가에서 더 높은 신뢰·용이성·유용성 인식으로 이어졌을 가능성이 있음.
- 학생들은 AI 서비스를 학습 보조도구로 사용하지만, 학업 수준 차이, 학습 동기 차이, AI 사용 목적 차이, 정보의 신뢰성 판단 능력 차이로 인해 AI를 반드시 “도움이 되는 도구”로 느끼지 않을 수도 있어 서비스의 학습효과를 체험하는 방식이 더 제한적일 수 있음. 반면 교사는 서비스 사용이 “수업 효율화”, “학생 질문 대응”, “업무 경감” 등 직접적인 이점을 제공하므로 평가가 높았을 가능성이 있음.
- 교사는 AI 기반 서비스가 학생 참여 확대, 개별 피드백 제공, 질문 대응 자동화, 수업 준비 시간 절감 등 교육적 효과를 가져올 수 있다는 통찰이 있어, 학생보다 더 긍정적·전략적·실용적 관점에서 서비스의 가치를 평가할 가능성이 있음.
- 학생은 AI의 오답이나 부정확성을 “학습 오류”로 경험하기 때문에 사용성 점수를 낮게 매기기 쉬움. 반면 교사는 AI의 한계를 이해하고 해석할 수 있어 오류가 있어도 서비스 전체에 대해 보다 관대하고 정교하게 평가할 가능성이 있음.

2) 학생 사용성 평가의 편차가 매우 큰 현상

- 학생 사용성 평가의 표준편차가 교사 사용성 평가의 표준편차보다 매우 큰 현상은 학생들의 응답이 더 다양하고 이질적임을 의미함.
- 교사들은 직무 특성상 AI·디지털 도구를 일정 수준 이상 활용해본 경험이 있기 때문에 사용성 문항에 대한 반응이 상대적으로 더 균질하게 모일 가능성이 있음. 반면, 학생들은 디지털 리터러시 수준, AI 서비스 사용 경험, 학년·학교급(초·중·고) 차이, 학습 동기나 흥미, 개인별 태도 등에서 이질성이 매우 크기 때문에 다양한 편차를 보였을 가능성이 있음.
- 교사는 오답이나 부정확한 답변이 나와도 분석적인 관점에서 평가할 수 있지만, 학생은 그 오차를 더 민감하게 받아들일 수도 있고, 반대로 오답을 오히려 문제로 인식하지 못할 수도 있음. 이 인지적 편차가 더 큰 표준편차로 나타날 수 있음

라 서비스 사용에 대한 정성분석

1) 대화형 서비스

- 대화형 서비스에 대한 학생들의 긍정적 응답은 주로 정보 접근성, 학습 이해 촉진, 편의성 등으로 요약될 수 있음. 학생들은 AI 챗봇이 모르는 내용을 쉽고 빠르게 해결해주는 점에 높은 만족도를 보였음.

〈대화형 서비스의 유용한 기능〉

유용한 기능	학생 응답 원문	설명
과학 개념의 쉽고 빠른 설명	“Ai가 내가 모르는 것을 쉽고 빠르게 알려줘서 좋았다”, “예시까지 들어 설명해주었다.”, “세세하게 풀어서 설명하는 것이 좋았다”	모호하거나 어려운 과학 개념을 학생 수준에 맞게, 혹은 예시와 함께 즉각적으로 풀어서 설명해주는 능력.
정보 검색 및 질문 해결의 편리성	“우리가 힘들게 검색해야하는걸 즉답해준다”, “모르는 것을 질문해서 해결”, “과학관련 질문을 여러 답변으로 쉽게 알려줌”	검색 엔진을 따로 이용할 필요 없이 과학 관련 질문에 대해 빠르고 정확하게 답을 제공하는 점
높은 과학적 정보 정확도 및 신뢰성	“챗gpt랑 다르게 거짓말을 안한다.”, “과학에 대한건 완벽하게 알고 있는것 같아서 좋았다”	일반 AI에 비해 과학 지식에 특화되어 있어 답변의 정확도가 높고 신뢰할 수 있다고 평가함.
학습 유도 및 흥미 유발	“질문을 통해서 답을 유도해주는 부분이 생각할 힘을 길러주기에 좋은 요소인 것 같다.”, “챗봇이랑 대화하면서 과학 내용을 알게 되니 신기했다.”	단순 답변을 넘어 스스로 답을 찾아가도록 유도하거나, 새로운 방식으로 학습 흥미를 유발하는 점
사용의 간편성	“편하다”, “이용하기 편리하고 사용법이 간편하다”	별도 로그인 없이 바로 시작할 수 있거나, 인터페이스가 직관적이어서 사용하기 쉬운 점

- 대화형 서비스의 주요 단점은 성능 및 기술적 제약, 답변의 질적 문제, 사용자 상호작용의 불편함의 세 영역에서 집중적으로 제기되었음.

〈대화형 서비스에 대한 개선점〉

개선 사항	학생 응답 원문	설명
느린 응답속도 및 서버 문제	“답변 몇 개를 받으면 다음부터 너무 느리고 대답을 해줄 수 없다는 듯이 말을 하여 더 이상 사용하지 못 하였다.” “서버가 인터파크 마냥 작아서 많은 학생들이 사용하면 응답이 불가능해진다.”	동시 접속자가 많을 경우 로딩 시간이 길어지거나 (렉, 서버 지연), 아예 사용이 중단되는 기술적 안정성 문제가 가장 큰 불만 요인으로 지적됨.

반복 질문 제약 및 기억 능력 부족	“전에 보낸 내용과 같은 내용을 다시 물어보면 답변을 해주지 않는 것”, “일단 같은 질문을 하게 해줘야 하며 뭔가 서비스가 느린 감이 있어 답답했다”	이전 질문 내용을 기억하지 못하거나, 같은 질문에 대해 반복적으로 답변을 거부하는 정책/기능상의 제약에 대한 불편함.
답변의 질적 문제 (장황함, 요약 부족)	“내용을 조금만 요약해서 알려주면 좋겠다”, “요약을 해주지않고 답변이 너무 길며 답변속도가 좀 느리다”	제공되는 답변의 내용이 너무 길거나 복잡하여 (말이 너무 많다), 핵심 내용을 파악하기 어렵다는 지적.
질문 이해 및 답변 정확도 문제	“질문하는 사람의 의도를 잘 파악 했으면 한다.” “정확하지 않는 답변을 해주어서 그 점이 개선되었으면 좋겠다.”	가끔 질문의 의도를 제대로 파악하지 못하거나, 부정확한 정보를 제공하는 경우가 발생함.

2) 보고서 튜터링 서비스

- 학생들은 보고서 튜터링의 피드백의 구체성, 보고서 작성 능력 향상, 빠른 검토라는 측면을 긍정적으로 판단함.

〈보고서 튜터링 서비스의 유용한 기능〉

유용한 기능	교사 응답 원문	설명
구체적이고 정확한 문제점 지적	“내 보고서에서 부족한 부분을 정확히 잡아서 이런 부분은 이런 말로 고치면 좋을거 같다”, “내가 실수하는 부분을 잘 잡아주었다.”	학생 보고서의 부족한 부분을 명확히 짚어주고, 구체적인 개선 방향을 제시해주는 점.
보고서 품질 향상 및 자신감 증진	“부족했던 부분을 알려주면서 어떤 식으로 수정하면 훨씬 더 좋을지 알려줘서 보고서 작성이 편했던 점이 좋았다.” “AI 보고서 튜터링을 읽고 나의 보고서에서 부족한 점이 무엇인지... 확신이 생기고 자신감이 생긴다.”	피드백을 통해 보고서의 품질이 향상되었고, 이 과정에서 작성 능력 및 자신감이 높아졌다고 평가.
신속하고 객관적인 평가	“간편함과 내 보고서의 문제점을 제대로 짚어주어 이해하기 쉬었다.”, “자동으로 보고서 튜터링을 해주어 직접 틀린 이유를 찾아보지 않아도 되는 점이 가장 유용했습니다.”	교사 대신 AI가 신속하고 객관적인 평가를 제공하여, 학생들의 부담을 줄이고 편리성을 제공함.
다양한 보완 정보 제공	“부족한부분 용어를 알려준다 혹시 모를수있는 상황에 대해 알려주어서 결과를 더 정확히 쓸수있게 한다”	보고서 내용 외에도 관련 용어나 추가적인 상황 설명을 제공하여 보고서의 완성도를 높이는 데 도움을 줌.

- 보고서 튜터링의 개선 필요 사항은 주로 피드백의 난이도 및 구체성, 기능의 제약이라는 항목에서 제기되었음.

〈보고서 튜터링 서비스에 대한 개선점〉

개선 사항	교사 응답 원문	설명
피드백 내용의 이해 난이도	“보고서 AI 평가 내용이 조금 어려운 내용도 있었다.”, “조금 내용이 길어서 어려웠다.”	제공된 피드백의 내용이나 용어가 학생들의 수준에 비해 어려워 이해하기 힘들었다는 지적.
구체적인 감점/가점 기준 설명 요구	“어떤 문제에서 몇 점이 감소되었는지 자세히 알려주면 좋을 것 같다.”	단순히 피드백이나 점수만 제공할 것이 아니라, 평가 항목별 상세한 점수 변동 이유에 대한 구체적인 정보를 요구.
반복적인 형식적 요청 및 특수 문자 처리	“뭔가 아래부분에 자꾸 뭔 느낀점을 쓰라는데 보고서의 마지막에 들어가야 하는거 같은걸 자꾸 제시해서 불편하다”, “특수문자 같은 경우에는 너그럽게 넘어가도 될 것 같다”	보고서 형식과 관련하여 불필요하거나 반복적인 느낀점 작성 요청이나, 특수문자 입력에 대한 불합리한 지적.
구체적인 수정 제안 형식 필요	“근데 정확하게 지적해 주지는 않는다”, “어떤식으로 바꿔야 하는지도 알려줬으면 좋겠다”	문제점을 알려주는 것을 넘어, 문장 수준에서 구체적인 수정 예시나 '이렇게 바꿔라'와 같은 명확한 가이드를 필요로 함.

3절 교사 · 학생 면담 분석

1 교사 면담

가 생성형 AI 기반 과학탐구 지원 시스템의 효과

- 본 연구를 통해 개발한 생성형 AI 기반 과학탐구 지원 시스템을 수업에 활용한 초등교사 7인, 중등교사 6인 총 13인을 대상으로 FGI 면담을 실시한 결과 4가지 주제로 시스템 활용의 효과가 나타남.
- 연구에 참여한 교사들은 생성형 AI 기반 과학탐구 지원 시스템으로서 대화형 서비스와 보고서 튜터링 서비스를 학생에게 적용하였으며, 학생들이 자유롭게 대화형 서비스를 활용하여 과학탐구를 추진할 수 있도록 하였음. 학생들이 보고서를 제출하면 보고서 튜터링이 적용되도록 하였으며, 교사의 검토를 통해 평가 결과가 제시된 후 보고서를 수정·보완할 수 있는 기회를 제공하였음. 보고서의 개선을 위해서도 대화형 서비스를 활용할 수 있었음

1) 지능형 과학실 ON 연계 AI 기반 과학 수업 운영 결과 학생 주도적 탐구가 이루어졌다.

- 연구에 참여한 교사들은 지능형 과학실 ON에서의 AI 기반 과학 수업 운영 결과 학생들의 주도성과 참여도가 높아졌으며, 스스로 탐구 문제를 해결하면서 깊이 있는 과학탐구에 참여하는 경우가 많았다고 하였음. 특히 과학탐구의 설계부터 학생 스스로 진행하여 자연스럽게 교과 수업 내에서 자율탐구의 형식이 도입될 수 있었고, 학생이 실험 준비물, 실험 방법 등을 자신의 관심사에 맞게 구성하면서 일상 생활에서의 문제를 과학적으로 해결해보는 과정을 주도적으로 이끌어나갈 수 있었음.
- 과학 실험·탐구 수업에서 학생들은 교사가 정해준 실험 절차에 따라 수동적으로 수행하는 경우가 많았으나, 탐구활동에 대한 학생의 권한이 높여주고, 학생의 부족함을 채워줄 수 있는 지원 도구가 마련되니 학생의 참여도가 높아지고, 다양한 방식의 탐구가 추진됨.
  - “제가 평소에 어떤 말을 하면은 애들이 우르르 다 그쪽으로 따라가서 비슷하게 활동을 하니까 이렇게 안 하려고 했고요. 그래서 모둠별로 AI를 사용해서 탐구활동 계획을 세워봐라, 계획을 세울 때 어떤 재료를 사용하는 지 등을 AI에 질문하면서 결정해서 실험해 보라고 했어요. 이렇게 하니까 각자 실험 과정이 다 달라지는 거에요. 보통 한 2~3차시에 실험 끝내고 보고서 작성할 수 있지만, 이렇게 하니 몇 일 걸려서 계획 세우고 준비물도 가져오고 하면서 학생이 훨씬 주도적으로 되었죠”(교사1)

- 대화형 서비스는 실험 전 개념학습, 실험 중 탐구 설계, 보고서 작성 과정에서 활용하는 경우가 많았음. 개념학습 활동을 과제로 수행하고 학교에서는 협력적 프로젝트 수행 방법을 제안하는 플립트 러닝(flipped learning)이 강조하는 바처럼 교실 수업이 지식 이해 위주 수업보다는 탐구와 문제해결 활동을 강화하는 방향으로 변화함. 실험·탐구의 앞뒤로 AI와의 상호작용을 통해 실험·탐구의 맥락을 학생 주도적으로 파악하도록 하고, 실험의 결과를 해석하였으며, 실제 실험하지 못한 결과에 대한 예측과 호기심 해소에도 도움을 주었음.
  - “초등학생들은 pH에 대한 개념을 모르잖아요. 제가 학생들에게 개념을 알려주지 않고, 학생들이 대화형 서비스를 직접 이용하면서 개념을 익히게 했어요. 그렇게 개념을 이해하고 실험을 했는데, 학생들이 스스로 산성과 염기성을 분류해보면서 원래 준비한 것 외에도 추가적으로 샴푸, 린스 등 다른 것들의 pH도 궁금해하더라고요. 실험으로 다 못한 것은 먼저 산성인지 염기성인지 예측을 해보게 하고 챗봇에 질문했는데 정말 잘 알려주더라고요...(중략) 학생이 AI에 각자 궁금한 것을 물어보면서 더 능동적으로

참여하는구나 느꼈습니다."(교사2)

- 연구에 참여한 교사들은 과학수업에서의 학생 간 소통이 증대하는 등 수업의 활기가 높아졌다고 보고하였음. 모둠별 활동이 아닌 경우에도 개별적인 AI와의 상호작용 결과를 주변 친구들과 공유하면서 대화가 많아졌고, 전반적인 교실 속 상호작용도 증가하는 현상이 나타났다고 하였음. 또한 학생 누구나 보고서 작성을 완수할 수 있다는 자신감 속에서 수업에 참여하니 참여도가 높아졌음.
  - "저는 평상시보다 더 수업에 활기가 있어서 좋았어요. 실험은 정해진 답이 없고 데이터를 통해서 저희가 답을 찾아서 결론을 이끌어내야 되잖아요. 애들이 챗봇을 이용하니까 의지할 수 있는 대상이 있다는 생각 때문인지 보고서를 끝까지 쓸 수 있다는 자신감이 생겨서 그런지 다들 적극적으로 챗봇 활용해서 질문을 하고 답을 찾아나가려고 하더라고요. 그래서 평상시보다 더 많이 수업에 참여했었고 애들이 자신의 개별보고서 쓴 게 맞는지 주변 친구들이랑도 소통하면서 비교도 좀 해보더라고요. 그래서 저는 전반적으로 수업을 재미나게 했습니다."(교사11)

2) 지능형 과학실 ON 연계 AI 기반 과학 수업 운영 결과 개별화 학습이 이루어졌다.

- 교사들은 지능형 과학실 ON에 AI 기능이 도입되어 학생의 사전지식 수준과 관심사에 맞는 개별화 학습이 가능해졌다는 것을 중요한 교육적 효과로 인식하였음, 학생의 사전지식에 따라 대상별로 차별화된 교실 수업을 운영하기 어려운 현실에서 학생 스스로 자신의 미흡한 부분을 찾아 메우면서 교실 수업이 매끄럽게 운영되는 효과나 나타났음, 학생들이 과학 교과 지식 중 부족한 부분이 서로 다를 수 있고, 이러한 지식의 부재를 적시에 해소하지 못함으로 인해 점진적으로 교육 격차가 벌어질 수 있는데 사전지식이 부족한 학생도 학습목표를 성취할 수 있도록 학습환경을 제공하여 만족도가 높음.
  - "대화형 서비스에서 제가 생각했을 때 학생들과 수업했을 때 가장 긍정적인 부분은 학생들이 스스로 AI 기능을 활용하면서 개별화 학습을 할 수 있다는 게 굉장히 긍정적인 부분이었던 것 같아요. 어떠한 미생물들이 어떤 특징을 갖는지라든지 잘 아는 학생들이 있고 못 따라가는 학생들이 있는데 AI를 활용해서 대부분 잘 따라가려고 노력을 하더라고요. 친구들에게 물어보지 않더라도 교사에게 질문하지 않더라도 지능형 과학실 ON의 AI 기능을 활용 하는 것만으로도 자신감을 불어넣어줄 수 있는 부분이 가장 긍정적인 부분이었고요."(교사9)

- 과학 교과는 개념에 대한 이해만큼이나 오개념에 대한 해소가 중요한 과제 중의 하나인데, 학교 교육에서 별도의 도구나 지원 없이 개별 학생들의 오개념 해소에 노력하는 것은 쉽지 않음. 그러나 지능형 과학실 ON에서의 AI 기능은 공교육에 도입되기 위한 안전성에 우선하여 개발하였고, 현장 적용 과정에서 환각에 대한 문제가 발견되지 않았으므로 교사들은 안심하고 학생들의 개별화된 과학 개념 점검에 이를 활용함.
  - "학생들이 이렇게 기존에 가지고 있던 오개념이나 이런 것들에 대해서 이제 점검을 하고 수업에 진행을 하는데 이제 자기가 아는 지식에 대해서 물어보면서 또 교정도 될 수도 있고 하는 도구로 활용할 수 있어서 좀 좋았다라고 볼 수가 있고요."(교사6)

- 교사들은 보고서 튜터링 기능에 대한 만족도가 높았는데, AI의 기능의 시범운영이 필요하다면 보고서 튜터링 기능을 더 먼저 학교 현장에 적용해 달라는 요구가 있기도 하였음(교사10). 이러한 이유는 보고서 튜터링 서비스를 통해 학생들에게 피드백하기 어려웠던 구체적인 과정지식, 탐구수행 결과, 데이터 해석 결과 등에 피드백을 주게 됨으로써 학생들의 과학탐구 역량 향상에 도움을 주었다고 인식하였기 때문임. 따라서 보고서 튜터링 기능이 도입되지 않던 시기와 비교하여 수혜 학생과 비수혜 학생의 과학탐구 경험의 차이에 대해 아쉬움을 표현하기도 하였음.
  - "AI 튜터링이 이제 들어오는데 어찌 보면 좀 아쉬운 게 이 기능을 쓰지 않았을 때는 애들한테 개별적으로 어떤 지도라든가 이런 것들이 거의 불가능했어요."(교사3)

3) 지능형 과학실 ON 연계 AI 기반 과학 수업은 교사의 역할 변화를 촉진한다.

- 지능형 과학실 ON 연계 AI 기반 과학 수업을 운영한 교사들은 새로운 테크놀로지의 도입으로 인해 교실 수업의 변화를 체감하였고, 이러한 변화는 교사의 역할 변화로 이어질 수 있다고 인식하였음, 먼저 학생들의 능동적 과학탐구 참여는 교사에 대한 의존도를 낮추었고, 교사는 한정된 차시 수업 시간 활용을 보다 의미있게 활용하는 방법에 고민하게 되었음, 대화형 서비스를 활용한 개별화된 과학탐구 과정에서 교사의 도움이 더 필요한 학생에게 집중적인 교수적 처치를 제공하는 등 교사의 개입이 더 다양해질 수 있음을 경험함.
  - "아이들이 교사에게 의존도가 굉장히 떨어지더라고요. 무슨 말이냐 하면 AI 챗봇을 통해서 질문을 하고 답을 얻으면서 그 시간에 교사는 더 손이 필요한 학생들에게 많이 가게 되더라고요. 굉장히 교사의 지도 시간이 많이 확보되는 것 같아요. 그래서 지능형 과학실 ON의 AI가 순수하게 과학 튜터의 개념으로도 이해됩니다. 그런 면에서 도움을 많이 받았습니다."(교사4)

- 교사는 보고서 튜터링 서비스가 주는 교사의 평가 업무 경감에 고무적이었는데, 이것이 단순히 업무가 축소되었기 때문에 느끼는 반가움이 아니라, 교사가 학생에게 제공하기 어려운 피드백을 적은 노력으로 제공한다는 것에 대한 만족과 피드백의 질적인 수준에 대한 만족, 그리고 교육적으로 의미 있는 피드백을 평소에 주지 못하는 아쉬움에 대한 해결이 복합적으로 작용한 결과로 볼 수 있음. 이러한 업무 경감은 교사의 시간을 어떻게 할애할지, 어떠한 역할에 집중할지에 대한 고민으로 연결됨. 또한 디지털 리터러시가 높은 교사의 경우 다른 도구를 활용하여 AI를 활용한 튜터링을 수업에 적용해보기도 하였는데 이제 별도의 시스템을 활용하지 않고도 지능형 과학실 ON 플랫폼 내에서 일반 과학 교사들이 편하게 활용할 수 있도록 도구가 제공된 데에 대한 업무 경감의 측면도 포함하고 있었음.
  - "보고서 튜터링 같은 경우에는 이전에는 제가 별도의 방법으로 수업에 적용해 본 적이 있습니다. 이게 지능형 과학실 ON 플랫폼 안으로 들어와서 매우 좋았고 업무 경감에 효과적이었다고 생각합니다. 학생들이 그래프 분석하는 거 하고 통계 분석 오차 분석 이거 하는 게 아이들이 생각보다 되게 힘들어하는데 AI 튜터링 결과를 보여주니까 '아 이렇게 숫자를 활용하는구나.', '오차가 있을 땐 이렇게 답변을 하는구나.' 그런 것에 대한 일종의 학습 효과가 있었다고 생각합니다."(교사10)
  - "AI가 튜터링한 내용을 보니 뭐라 할까요. 제가 피드백 할 수 있다고 감히 생각하지도 못한 내용들을 제시해 주더라고요. 다른 선생님이 설명하신 것처럼 지난 사용성 검토 때보다 AI 수준이 향상되었다는 생각이 들었습니다."(교사2)
  - "보고서 평가하는 것이 솔직히 말하면 시간이 많이 듭니다. 그런데 이번에 보고서 튜터링이 이런 측면에서 도움이 많이 되었고 교사가 해주고 싶은 말인데 그런 말들을 많이 해 준 것이 장점이라는 생각이 들었습니다."(교사12)

- 앞서 언급한 교사의 역할 변화와 업무 경감 시간을 어떻게 활용할지에 대한 고민 등을 통해 교사들은 교사에게 더 강화되어야 할 역할과 노력에 대해 생각하였음. 연구에 참여한 교사들은 "전문적 수업 설계자"로서의 교사 역할을 강조함. 수업 설계가 체계적이어야 그러한 정보를 토대로 목적에 맞는 정교한 보고서 튜터링이 이루어질 수 있다고 인식함.
  - "튜터링이 좀 더 좋아지기 위해서는 일단 제일 중요한 것은 교사가 탐구 설계를 잘해야 될 것 같아요. 지능형 과학실 ON에 탐구 개설부터 잘해야 된다고 생각하고 AI 역시 이 수업에 참여하는 학생이라 생각하고 실험을 개발하면 그러한 정보를 바탕으로 튜터링이 잘되지 않을까 싶습니다. 탐구 목표, 과학 개념에 적합한 데이터를 잘 선정해야 하고 보고서 양식을 잘 만들어야 AI가 튜터링을 잘해 줄 수 있다는 생각이 듭니다."(교사10)

4) 지능형 과학실 ON 연계 AI 기반 과학 수업을 통해 교사와 학생의 효능감이 제고되었다.

- 교사들은 지능형 과학실 ON에서 AI 기반 과학 수업을 하면서 학생의 자율성이 높아지고 개별화 학습이 가능해졌다고 하면서도, 수업 중 적재적소에 교사의 개입을 실천하면서 학생을 지원할 수 있게 됨으로써 수업에 대한 장악력도 높아졌다고 생각하였음. 또한 AI 활용 수업을 통해 학생이 성장하는 데 도움을 줄 수 있었다는 인식이 교사의 효능감 제고로 이어질 수 있었음.
  - "대규모 학급에서 굉장히 빛을 발할 것 같아요. 교사 효능감은 당연히 좀 올라간 것 같아요." (교사4)
  - "항상 다수의 학생에게 제가 보고서에 대한 피드백을 상세하게 해 줄 수 없는 부분이 있었는데 아이들의 답변에 대해서 상세하게 피드백을 제공할 수 있다는 부분이 교사로서 만족스럽고 제가 무언가 했어야 하는 일을 못하고 있었는데 할 수 있게 되었다는 것이 저의 교사 효능감을 향상 시키지 않았나 생각했습니다."(교사8)

- 교사들은 지능형 과학실 ON 연계 AI 기반 과학탐구 수업에 참여한 학생들에게서도 효능감의 향상이 나타났다고 인식하였음. 학생이 수업에서 적시에 모르는 것을 질문할 수 있어 수업 참여를 잘 할 수 있게 된다는 것과 더 깊이 있는 이해를 하고 탐구 문제를 해소할 수 있다는 점에서 효능감의 향상이 이루어질 수 있다고 생각함. 대화형 서비스에는 학생이 한 질문에 AI가 응답하는 서비스와 함께 AI가 질문을 하고 학생이 답변하는 소크라테스식 서비스를 선택적으로 활용할 수 있는데, 소크라테스식 대화를 통해 학생의 과학적 사고가 심화될 수 있었다는 의견도 제시됨.
  - "아이들이 실험하는 과정 중간에 모르는 것은 지능형 과학실 ON의 채팅 기능을 이용해서 선생님에게 물어보는데 아이들도 그렇고 선생님들도 그렇듯이 '이걸 물어봐도 되나. 너무 작은 질문인 것 같은데'라고 느끼는 학생들이 있어요."(교사7)
  - "수업에 참여했던 친구들도 한 클래스에서 적어도 한 10명의 친구들은 내가 무언가를 지금 하고 있고 과학탐구 활동을 할 때 AI의 도움을 받아서 좀 더 깊이 있는 개념 이해를 할 수 있었다는 측면에서 효능감이 높아진 것 같아 의미 있었던 활동 같습니다."(교사8)
  - "대화형 서비스 중 소크라테스식 질문형으로 스스로 답을 찾아가게 하는 부분도 학생들에게 써보게 했을 때 왜라는 질문과 어떻게 라는 질문 그러니까 실험 과정에서 어떤 식으로 답을 찾아가야 되는지에 대해서 사고의 과정을 도와서 자연스럽고 매끄럽게 질문을 해준다는 느낌도 받았습니다."(교사9)

나 생성형 AI 기반 과학탐구 지원 시스템의 개선점

- 교사 FGI 면담 분석 결과 시스템의 정교화, 기능 확대, 그리고 체계화 등을 고려하여 다음의 세 가지 주제의 시스템 개선점을 도출함.

1) 학습 지원 최적화를 위한 AI 페르소나 개선이 필요하다.

- 대화형 서비스나 보고서 튜터링 서비스에서 학교급에 맞는 용어를 사용할 수 있도록 조정하는 것이 필요하다는 응답이 있었는데, 대다수 초등 교사들에게서 교육과정에 맞는 어휘 사용의 필요성이 지적되었으며, 중등 교사들은 교육과정에 맞게 잘 답변이 도출되고 있다고 하였음. 따라서 초·중·고 수준에 맞는 AI 응답 개선이 필요하되, 초등학생들이 활용하는 경우 더 쉽게 상호작용이 이루어질 수 있도록 초등학생의 페르소나로 개선할 필요가 있음.
  - "AI 답변 수준을 초등학생, 중학생, 고등학생 수준으로 선택을 할 수 있다면 초등학생들에게 AI가 답변해 줄 때 초등학교 아이들의 시선과 언어로 해석할 수 있을 것 같아요." (교사4)
  - "초등학생들이 이해하기 어려운 용어로 AI가 답변을 해주니까 이게 어떤 의미인지 선생님한테 물어보는 현상도 있더라고요."(교사6)
  - "다른 선생님도 얘기하셨는데 성취기준 그리고 그 학습 목표 이상의 용어와 단어를 얘기해 주더라고

요.”(교사7)

- “제가 생각했을 때는 아까 이야기 하신 중학교 선생님 말씀과 같이 고등학교 교육과정 안에서도 굉장히 교육과정에 맞춰서 잘 답변해 준다는 생각을 좀 많이 했습니다.”(교사9)

- 간혹 대화형 서비스나 보고서 튜터링의 피드백 내용이 길어 학생들이 충분히 숙고하고 읽기 어려운 경우가 있기 때문에 장황한 설명보다 간결한 표현으로 응답이 제시되도록 조정할 필요가 있음. 또한 학생의 학업성취 수준에 따라 피드백을 받아들이는 수준과 태도에도 차이가 있기 때문에 모든 학생들이 받아들일 수 있는 정도로 응답이 제공되어야 한다는 의견이 제시됨.
  - “예를 들어 교수님과 대화하면 어려움을 느끼듯이 초등학교 6학년 학생들인데 길게 설명하면 어려워해서 초등학생들한테는 개조식으로 설명하든지 짧게 대답을 해주면 좋겠다는 생각을 했습니다.”(교사7)
  - “이거를 꼼꼼하게 읽는 친구들은 정말 상위권의 그런 모범적인 친구들이에요. 그런 친구들은 꼼꼼하게 다 읽지만 다른 친구들 같은 경우는 사실 의미를 생각하면서 다 읽지는 않을 수도 있어요,”(교사8)

- 과도한 칭찬은 청소년들에게 와닿지 않을 수 있으므로 칭찬을 줄이는 등의 AI의 말투, 어조의 수정이 필요함. 같은 AI의 피드백에 대해서도 학생마다 반응이 다를 수 있으므로 하나의 방식으로 모든 학생을 만족시키기는 어려울 수 있지만 많은 학생들이 만족할 수 있도록 보다 차분하게 상호작용이 이루어질 필요가 있음. 교사들은 초등학생의 경우 칭찬이 도움이 되는 면이 있지만 중등학생들은 칭찬에 그다지 반응하지 않으므로 과학탐구와 관련된 피드백 위주로 언급이 되는 것이 낫다고 판단하였음.
  - “AI의 똑같은 피드백에 대해서도 아이들마다 반응이 다른 거예요. 예를 들면 AI가 아이들의 작은 답에도 칭찬해 주고 격려해 주는 의미의 말도하고 감탄사도 쓰는데 어떤 친구들은 ‘그런 말이 저를 좀 으쌰으쌰 해 주는 것 같아서 기분이 좋았어요’ 하는 친구들도 있었는데 ‘선생님 부담스러워요. 너무 필요 없는 말을 과하게 칭찬해 주는 게 좀 읽기가 조금 부담스럽네요’라고 하는 친구도 있었어요.”(교사8)
  - “너무 칭찬이 많은데 우리 고등학생 애들이 그렇게 칭찬 많다고 해서 그렇게 좋아하거나 그러지는 않거든요.”(교사12)

2) 이미지, 음성 인식 등 과학탐구 지원을 위한 AI 기능 확대가 필요하다.

- 교사들은 이미지 데이터를 활용한 과학탐구 수업의 용이성을 위해, 혹은 학생들의 대화형 상호작용의 직관성을 높이기 위해, 그리고 과학탐구 활동의 흥미 요소와 창의적 활동을 위해 대화형 서비스에서 이미지 파일 입력과 AI의 이미지 생성이 가능하기를 바라기도 하였음. 일반적인 생성형 AI 활용 경험에 비추어 이미지의 입출력을 동반한 AI와의 상호작용이 일상화 되었으므로 지능형 과학실 ON에서도 그러한 경험이 이어지기를 요구함.
  - “예산 지원을 더 받아서 이미지 API를 활용하면 좋을 것 같아요. 그래서 아이들이 실험 설계부터 결론 도출에 이르기까지 질문이 있을 때 ‘지능형 과학실 ON 채팅’은 순수하게 학생과 교사 간에 텍스트로 상호작용을 하지만 ‘AI 대화형 서비스’는 질문 프롬프트에 따라 답변이 나올 때 거기에 이미지화도 되면 좋을 것 같아요.”(교사4)
  - “학생들이 이미지를 대화형 서비스에 업로드를 많이 하고 싶어 하더라고요. 그런데 이제 그 부분이 안 됐던 점이 조금 아쉬웠습니다.”(교사9)
  - “애들이 그림으로 좀 나왔으면 좋겠다 이걸 원하더라고요. 그리고 마지막에 실험·탐구를 빨리한 친구들은 재미난 이미지 만들기도 좀 했으면 좋겠는데 보고서 작성 다 하고 나니까 좀 지루한 느낌이 든 친구도 몇 명 있었어요. 뭔가 재미적인 요소를 담아 창의적으로 할 수 있는 활동까지 지원하면 좋겠다는 것들이 애들의 요구 사항이었던 것 같습니다.”(교사11)
- 음성 인식을 통한 대화형 서비스 이용이 가능하다면 타자 치는 것이 어려운 초등학생들이나, 실험 과정이 복잡해서 AI과 밀접한 상호작용이 필요한 중고등학생들에게도 도움이 될 것이라는 의견이 개선점으로

제시되었음.

- “초등에서는 타자 치는 게 굉장히 어렵습니다. 그래서 STT 기반으로 학생들이 이야기를 하면 AI가 인식하여 피드백을 주는 기능이 있으면 오히려 더 잘 쓸 수 있겠다는 생각이 많이 들더라고요. 저희 반 같은 경우는 마이크로소프트 팀즈 안에 있는 리딩 프로그래스를 쓰는데 한 반에 22명의 친구가 있더라도 이어폰을 끼고 마이크를 껴서 얘기를 하면 인식을 잘합니다. 그렇게 지능형 과학실 ON도 개선되면 초등에서 더 잘 활용할 수 있을 것 같습니다.”(교사6)
- “문자적인 것뿐만 아니라 음성적인 것도 한번 해봤으면 어떨까 그런 생각도 한번 해봤거든요. 그래서 실험할 때 바로 음성으로 질문해서 그 결과들이 나오는 것들도 포함시키면 어떨까. 문제는 비용적인 부분이 가장 크게 작용할 텐데 그런 부분도 고민했으면 어떨까 싶었습니다.”(교사12)

3) 지능형 과학실 ON에서의 과학탐구를 종합적으로 고려한 체계적인 AI 연계 방안이 필요하다.

- 공동탐구 등 지능형 과학실 ON 플랫폼 전반에서 AI 기능을 활용할 수 있도록 기능 확대의 필요성이 제기되었음. 연구에 참여한 교사들은 새로운 시도나 수업개선 활동에 상대적으로 적극적인 교사 집단이므로 일반적인 교사들이 지능형 과학실 ON을 잘 활용할 수 있도록 하는 유인책을 강구하는 것에도 관심이 있었음. 그러한 측면에서 접근성이 좋은 ‘공동탐구’ 메뉴에도 AI 기능이 활용되는 것을 전략으로 인식함.
  - “저는 선생님들이 지능형 과학실 ON을 가장 쉽게 접근하는 방법은 공동탐구라고 생각합니다. 공동탐구에서도 AI를 곳곳에 활용해서 평가도 해준다고 하면 여기에 유입되는 선생님들도 많아지고 활용의 필요성도 더 많이 느낄 수 있을 것이라는 생각이 들었습니다.”(교사12)

- 본 연구에서는 연구에 참여한 학생들에게 각 수업에서 동일한 보고서 양식 2개를 순차적으로 제공하여 보고서 튜터링 결과를 받은 후 보고서를 개선하여 다시 제출하도록 수업을 구성하였음. 이는 보고서 튜터링의 결과가 일방향적으로 전달되는 것에 그치지 않고, 학생들에게 자신의 과학탐구 성과물을 개선할 기회를 주고, 얼마나 개선될 수 있는지를 확인하기 위함이었음. 연구 참여 교사들은 이러한 보고서 작성 절차를 긍정적으로 평가하였으며, 탐구수업 메뉴에서 보고서 작성 시 ‘보고서 제출-AI튜터링 적용-보고서 개선-재제출’의 절차가 시스템 상에서 이루어질 수 있도록 기능 개선이 필요하다고 인식하였음.
  - “보고서에 수정할 게 있으면 자기가 썼던 보고서에서 AI 피드백을 보고 다시 한 번 수정할 수 있는 기회를 주는데, 밑에 댓글 형식 등으로 원본을 그대로 두면서 수정할 수 있게 해주면 애들이 자기가 한 거를 또 한번 볼 수 있는 기회가 되지 않을까 이런 생각을 했습니다.”(교사11)
  - “이번에는 보고서 1차, 2차로 따로 냈었는데 애들이 똑같은 거 또 써요? 이 얘기를 꼭 하더라고요. 그래서 1차 냈던 거에서 바로바로 수정할 수 있었으면 하는 생각을 했습니다.”(교사13)

다 생성형 AI 기반 과학탐구 지원 시스템 활용을 위한 교육적 요구

- 본 연구에서의 교사 면담을 통해 생성형 AI 기반 과학탐구 지원 시스템 활용의 효과와 개선점을 파악하고자 하였으며, 이러한 과정에서 시스템의 교육 현장 적용을 위한 요구사항을 확인하였음.

1) 학생을 대상으로 질문하는 방법에 대한 교육이 필요하다.

- 과학은 일상에 대한 관찰과 호기심이 중요한 학문이므로 질문이 중요함에도 교육 현장에서는 탐구 질문을 충분히 수용하는 수업을 운영하는 것에 제한점이 있음. 따라서 AI를 활용하여 이를 해소하려는 노력을 추진하고자 하지만 학생들이 질문을 하는 방법에 대해 잘 알지 못해 AI 기능을 잘 활용하지 못하는 경우도 나타남. 특히 AI에게 질문을 한다는 것은 맞춤법에 대한 이해, 어휘력, 문해력 등 기초 역량이 요구됨과 함께 질문의 목적과 기대하는 답변에 대한 질문자의 인식을 글로 정리하는 활동이므로 과학적 소양과

더불어 지능정보사회에서 키워야할 중요한 역량 중의 하나임. 그러므로 학생에게 별도의 교육을 통해 질문을 잘 하는 방법을 익히게 할 필요가 있음.

- “앞으로 우리 선생님들도 함께 지능형 과학실 ON의 AI 기능의 방향성에 대해서 좀 고민을 해야 될 부분인 것 같은데, 아이들이 AI에게 도대체 무엇을 어떻게 질문하는지 그 질문하는 방향과 방법을 잘 모르더라고요. 어찌 보면 독서 교육으로 회귀가 필요하기도 한데 평상시에 선생님들과 더 이러한 취지의 과학 수업을 통해 과학 개념과 과학실험에 대해 근본적으로 질문하는 힘이 좀 선행되어야 AI와 채팅을 100% 활용할 수 있지 않을까 하는 생각이 들었거든요.”(교사4)
- “공부를 잘하는 애들은 질문을 막 세 줄 네 줄씩 쓰더라고요. 근데 좀 문장력이 좀 부족하거나 학업 능력이 좀 부족한 아이는 한 줄씩 이렇게 물어보는데... 오히려 공부 잘하는 애들이 한번에 너무 많은 걸 물어보니까 답변이 더 정확하게 좀 안 나오는 그런 것도 좀 있는 것 같아요.”(교사1)
- “챗봇을 사용할 때는 질문을 잘해야만 챗봇이 양질의 대답을 내놓잖아요. 우리 아이들은 제가 한 2주 정도 질문하는 방법과 어떻게 하면 챗봇이 TMI를 적게 하고 우리가 원하는 답변을 내놓을 수 있는지에 대해서 제가 교육을 조금 시켰습니다. 그다음에 우리는 답변을 받는 챗봇보다 대화용 챗봇을 제가 아이들에게 주로 사용하라고 했습니다.(교사10)

2) 교사의 전문성 향상을 위한 가이드 제공이 필요하다.

- AI 기반 과학탐구 지원 시스템이 교육 현장에서 효과적으로 운영되기 위해서는 교사들이 새로운 디지털 도구를 수업에 적용할 수 있도록 구체적인 안내가 필요하다고 인식하였음. AI이 도입되지 않더라도 지능형 과학실 ON에서 데이터 기반 과학탐구 수업을 운영하기 위해서는 학생들에게 데이터를 친숙하게 하고 어떻게 해석할 수 있을지를 잘 가르칠 수 있도록 교사도 준비가 필요하다고 하였음. 또한 AI 기능을 교사의 수업에 적합하게 조절하기 위해서는 수업 사례 및 가이드 등이 필요하다는 요구가 나타남.
  - “지능형 과학실 ON을 사용하면서 학생들이 그래프와 표를 활용해서 정량적인 데이터를 기반으로 해석하는 능력이 좀 향상되었다는 것을 느꼈습니다. 원래는 과학 시간에 따로 가르치지는 않았거든요. 그러나 지능형 과학실 ON을 이용하다 보니까 자연스럽게 센서와 연동해서 데이터를 해석하는 방법을 가르칠 수밖에 없겠더라고요. 학생들이 표와 그래프에 대해서 더 관심을 갖게되고, 데이터를 해석하는 문장을 보고서에 작성하는 게 인상 깊었습니다.”(교사2)
  - “AI 기능이 당연히 좋은데 이걸 쓸 수 있는 교사의 조절 능력이 필요해요. 어떨 때 AI 챗봇을 쓰고 어쩔 때 AI 보고서 튜터링을 쓸지에 대한 예시나 수업 사례, 혹은 매뉴얼이 잘 정리돼서 선생님들께 안내되면 좋겠다라고 생각했습니다.”(교사7)
  - “AI에게 질문해서 학생이 원하는 답변을 얻지 못하면 교사에게 다시 질문을 하지 않습니까? 그런 상황에서 오히려 선생님에게 더 많은 지식과 공부와 전문성이 필요하지 않겠나. 반대로 생각해서 AI보다 더 큰 전문성이 필요하지 않겠냐 그런 생각도 해봤고요.”(교사9)

2 학생 면담

가 생성형 AI 기반 과학탐구 지원 시스템의 효과

- 본 연구를 통해 개발한 생성형 AI 기반 과학탐구 지원 시스템 활용 수업에 참여한 학생 중 초등학생 5인, 중학생 3인, 고등학생 4인 등 총 12인의 학생들 대상으로 FGI 면담을 실시한 결과 3가지 주제로 시스템 활용의 효과가 나타남. 과학에 대한 흥미와 자신감, 참여도가 높아졌다는 것은 교사들의 면담에서도 드러난 부분이며, 학생들도 교사가 느끼는 바와 같이 과학 학습에 대한 정의적 성취가 높아졌음을

언급하였음. 학생들의 면담을 통해 학생들이 AI를 활용하면서 어떻게 학습 방법이 변화하였는지를 알 수 있었음.

1) 지능형 과학실 ON의 AI 기능을 활용하고 과학에 대한 흥미와 자신감이 커졌다.

- 연구에 참여한 학생들은 초·중·고 학생 모두 과학 수업이 흥미로워지고 과학에 대한 자신감이 제고되었다고 응답함. 과학탐구는 정답이 없고, 문제를 해결하는 과정이 비구조화되어 있어 학생들이 부담을 느끼는 순간들이 있지만 AI와 상호작용하면서 문제를 해결하고 실험·탐구를 수행하는데 유능함을 느끼게 됨. 특히 과학탐구 보고서를 작성하는 데에 있어 초등학생부터 중학생까지는 어려움이 있다고 인식하는 학생들이 더러 있었는데 명확한 사실에 기반하여 보고서를 작성할 수 있고, 보고서를 보완할 수 있는 기회를 통해 자신의 성과물에 대한 자신감을 가질 수 있었던 것으로 보여짐.
- 학생 사전·사후 설문조사를 통해 과학긍정경험이 유의하게 상승하였고, 하위 5가지 지표 모두 각각 유의하게 상승하였기 때문에 지능형 과학실 ON의 AI 기능이 학생의 정의적 성취에 전방위적으로 긍정적으로 작용한다고 할 수 있음. 학생 대상 면담 결과에서도 이러한 결과가 일관성 있게 나타남.
  - “지능형 과학실 ON에 AI 기능이 새로 추가되었는데 답변이 잘 나와서 좋았어요, 더 확실한 답을 보고서에 쓸 수 있게 되었어요. 애들이 수업을 재미있어했던 것 같아요. 저도 AI 기능 쓰고 과학탐구가 훨씬 즐거워졌어요. 보고서도 원래 한두 줄밖엔 못 썼는데 이제는 많이 쓸 수 있게 되어서 자신감이 들어요.”(학생1)
  - “지능형 과학실 ON을 이용하면서 AI를 이용해서 궁금했던 정보나 단어 같은 거를 구체적이고 정확하게 알 수 있게 되었고요. 제가 보고서 작성한 거를 AI가 평가하고 조언을 해주면서 보고서를 더 잘 쓰게 되고 자세하게 쓸 수 있게 된 것 같아요.”(학생6)

- 중등 학생들은 온라인 상에서 보고서를 작성하는 편리함을 인식하고 있었는데, 이에 더해 보고서 작성 과정을 지원받고, 이러한 지원이 신뢰롭다 생각하기 때문에 지능형 과학실 ON에서의 AI 활용에 대해 만족하고 있었음. 학생들은 새로운 교구나 디지털 도구를 도입하는 등의 수업의 변화에 민감하게 반응하고 있음을 언급하였고, 이러한 변화가 학생들의 참여도나 주도성을 제고해주는 데 도움이 된다고 인식함.
  - “저희가 지능형 과학실 ON을 활용하니까 A4 용지에다 줄글로 보고서를 쓰는 것보다 훨씬 더 간편하고 쉽게 보고서를 작성할 수 있게 되어서 보고서 작성에 대한 두려움이랄까요? 그런 게 좀 줄어드는 게 있고요. 그리고 이제 AI를 이용하다 보니까 저희가 모르는 정보를 엄청나게 정확하게 알려주니 좀 더 지능형 과학실 ON의 AI를 신뢰할 수 있게 되는 것 같습니다.”(학생7)
  - 우선 수업을 할 때 무언가를 활용해서 한다. 그것만으로도 좀 더 참여성을 독려하고 약간 자기 주도적으로 할 수 있는 기회를 만들어 준다고 생각합니다. 그 미세한 게 별게 아닌 것 같지만 학생들 입장에서는 뭔가 새로운 거를 시도하는 게 어렵지만 좀 더 흥미롭게 느껴지기 때문에 그 점이 좋은 것 같습니다.(학생9)

2) 지능형 과학실 ON의 AI 기능을 통해 학업 수준에 상관없이 과학 수업의 참여가 높아졌다.

- 학생들은 지능형 과학실 ON의 AI 기능을 활용하면서 과학탐구를 계획하고 수행하며, 결론을 도출하는 전 과정에서의 수업 참여가 고르게 높아졌다고 인식하였음. 과학탐구 과정에서의 학생들의 권한이 높아지고, 사전지식에 대한 요구와 과학적 의사결정에 대한 부담이 완화됨으로써 학생들의 참여도가 높아질 수 있었음. 또한 어렵게 느껴졌던 과학탐구에 편리하고 쉽게 접근할 수 있게 되었다는 점에서 학생들이 과학탐구에 참여하는 심리적 장벽이 완화되었다고 볼 수 있음.
  - “질문도 저희가 직접 할 수 있고, 질문 방법도 저희가 정할 수 있어서 AI가 들어오고 저희가 직접 하는 것이 많아졌어요.”(학생3)
  - “더 편리하게 과학탐구를 할 수 있으니까 친구들도 더 열심히 하는 계기가 되었던 것 같아요.”(학생5)

- 학생들은 지능형 과학실 ON을 활용한 수업이 본연적으로 학생의 실험·탐구 중심으로 운영되기 때문에 강의식 수업보다 참여도가 높다고 인식하였고, 이에 더해 AI와의 상호작용까지 가능하기 때문에 학생들이 보다 긴밀하게 지능형 과학실 ON에서의 상호작용과 탐구활동에 참여하게 된다고 인식하였음. 이러한 참여를 통해 성취감을 느끼기도 하였음.
  - "대체로 많은 학생들이 자기가 직접 뭔가를 해보면 그 이론을 받아들이는 게 더 빠르고 그 이론을 안다는 것 자체에도 정말 성취감을 느끼면서 더 좋아하는 게 맞는 것 같아요."(학생7)
  - "활용하지 않을 때는 그냥 강의식으로 수업을 듣는 거다 보니까 가끔 토론을 하긴 하는데 그래도 강의식이면 애들이 좀 졸려하고 참여도가 떨어지는 경우가 많습니다. 그래서 저는 활용을 하는 게 확실히 도움이 된다고 생각합니다."(학생9)

- 특히, 대화형 서비스를 통해서 학생들은 자신의 수준에 맞게 질문하고 답변을 받을 수 있다는 점을 장점으로 인식하였음. 지능형 과학실 ON에 연계된 생성형 AI는 교육적 목적으로 개발되어 목적에 어긋난 대화는 지원하지 않도록 개발되었기 때문에 좋은 방향으로 질문을 유도한다는 것을 이해하고 있었음. 또한 텍스트 기반의 질의응답을 통해 충분히 생각하고 질문을 이어갈 수 있다는 점을 긍정적으로 생각함.
  - "대화형 서비스에서 대해서 생각을 해보았는데 일단은 제가 보기에 좋았던 점은 학생의 수준에 맞게 분석을 해주고 설명해 주는 것이 좋다고 생각했고 또 제 주변 학생들이 AI에게 좀 안 좋은 의도로 질문을 했을 때 좋은 방향으로 물어봐 달라고 하면서 학생들이 좋게 질문을 하게 의도하는 그런 점이 좋았다 생각을 했어요."(학생10)
  - "대화형 서비스를 사용하면서 나의 수준에 맞춰서 질문할 수 있고 그에 맞는 답변을 받을 수 있어서 좋았고, 그리고 질문을 생각해 보고 질문할 수 있어서 좋았습니다."(학생11)

- 학생별로 수준과 역량에 맞게 보고서 작성을 지원해주는 것을 AI 활용 수업의 차이로 인식하기도 하였음. 학생들이 느끼기에 독서를 열심히 하지 않은 동료들의 경우 문맥에 맞지 않는 보고서를 작성하는 경우도 많은데 AI 기능을 활용하면 훨씬 자연스럽고 문맥에 맞는 보고서를 작성할 수 있다고 생각하고 있었음. 즉, AI의 종합적인 피드백을 통해 학생 역량 향상에 도움을 줄 수 있다고 생각함. 또한 개선된 보고서 혹은 보고서에 대한 피드백을 봄으로써 좋은 사례를 확인하게 되어 학생의 성과물 수준이 높아질 수 있다고 인식함.
  - "살짝 결과물의 수준이 개인의 역량에 맞게 높아지는 것 같은데 그런 면에서 AI가 활용되고 안 되고 수업에 차이가 있었던 것 같습니다. 한 번 잘된 걸 보면 나도 활용해볼까 하고 생각할 것 같아요. 저도 예전에는 'AI가 내 것을 어떻게 만들어주지' 하고 살짝 마이너스적인 생각을 했는데 한 번 두 번 써보니까 틀린 말도 잘 고쳐주고 해서 꽤 사용하기 좋은 AI인 것 같아요."(학생11)

- 연구에 참여한 학생들의 경우 대다수 과학탐구에 적극적으로 참여하는 학생이었기 때문에, 그렇지 않은 주변의 학생의 경우 참여수준이 어떠한지 질문한 결과, 공부에 관심이 적은 학생들이 AI 기능 활용 이전의 수업시간에 책에 낙서를 하는 등 참여하지 않는 모습을 보인 반면 AI를 통해 기초 지식도 물어볼 수 있기 때문에 과학수업에 적극적으로 참여하는 모습을 확인하였다고 응답 하였음.
  - "제가 저번에 지능형 과학실 ON으로 수업을 했을 때 수행평가로 AI를 활용한 탐구를 진행하였는데요. 이 수업을 원래 열심히 듣지 않던 학생들도 한두 번씩 다 AI한테 물어봐서 결과를 도출하는 등 활동에 참여하는 모습을 보였습니다."(학생12)

3) 지능형 과학실 ON의 AI 기능을 활용하고 과학 학습의 방법이 달라졌다.

- 학생들이 과학탐구를 위한 정보가 필요할 때 여러 가지 방법으로 정보를 찾지 않고, 지능형 과학실 ON 플랫폼 내에서 AI을 활용하여 학습이 확장되기 때문에 인터넷 상의 정보의 홍수 속에서 불필요한 정보를 걸러내는 작업 없이 직관적인 질의응답 활동을 통해 정보를 습득할 수 있게 된다고 인식함. 과학적 지식을

습득하는 방법이 간편하고 쉽게, 더 적은 시간으로 가능해짐. 정보 획득에 드는 시간을 단축하여 수업 차시 내에서 보고서 작성까지 완성할 수 있었다고 보고함. 또한 스스로 질문을 하면서 심화탐구를 할 수 있어 학습 방법과 함께 수준도 변화할 수 있다고 인식함.

- "옛날에는 인터넷을 죄다 찾아서 정보를 찾아봤는데 대화형 서비스가 생기니까 한 번에 질문이 가능하고, 원하는 답만 나오니까 좋아요."(학생2)
- "예전엔 한 차시에 실험하고 데이터 수집하면서 힘들었는데, AI 기능이 들어오고 나서는 학생들이 모르는 것을 따로 알아보거나 선생님한테 일일히 질문해야 하는 상황이 아니니까 시간을 아낄 수 있다는 장점이 있었던 것 같아요."(학생7)
- "궁금한 점이 있어도 살짝 질문하기 어려워하는 학생들이 있을 텐데 AI가 있으면 혼자서 심화 탐구를 해볼 수 있는 기회가 주어져서 저는 이 점이 다르고 좋다라고 생각하고 있습니다."(학생10)

- 학생들은 보고서 제출 후 과학탐구가 종료되는 것이 아니라 AI의 피드백을 받아 자신의 과학탐구 결과를 성찰할 수 있게 되었다는 점을 긍정적으로 인식함. AI 기반 과학탐구 지원 시스템 활용에 대한 학생의 인식을 통해 과학 지식 이해 위주의 수업과 과학탐구 수업이 분리되어 운영되는 교육 현장의 실태를 개선할 수 있는 실마리를 찾을 수 있음. 입시 위주의 교육을 위해 학교에서 과학 실험·탐구가 제한적으로 운영되는 경향이 있지만 연구 참여 학생의 발언은 과학탐구 수업을 더 강화함으로써 깊이있는 지식 이해도 이루어질 수 있음을 보여줌.
  - "온도계와 지능형 과학실 ON을 연계해서 표나 그래프 등 객관적인 자료를 뽑아낼 수 있었다는 점이 좋았고요. 표나 그래프 첨가가 쉽다는 온라인의 이점을 이용해서 학생들이 쉽게 탐구 결과를 보고서로 작성하도록 하는 것과 그 보고서를 제출했을 때 AI가 피드백을 줘서 스스로 학생들이 성찰할 수 있었다는 점이 좋았던 것 같습니다. AI 피드백을 통해 제가 '이런 부분을 몰랐었구나' 하면서 그 부분을 다시 복습했어요. 그때가 시험 기간이었거든요. 그래서 시험 치는 데 더 유리해진 것 같습니다."(학생8)

- 주목할 점은 학생들이 AI를 수동적으로 활용하는 것에 그치지 않고, 소크라테스식 대화법을 지원하는 대화형 서비스를 이용하여 AI의 질문에 답변해 보고 탐구를 심화하는 학습 방법을 채택한 경우도 있었음. AI을 활용하면서 학생들은 질문을 더 잘하는 방법에 대해 고민하거나, AI의 답변을 검증하기 위해 노력하면서 학습 방법의 변화를 경험하였고, 그러한 과정에서 AI에 대한 비판적 사고도 함께 키울 수 있었다고 응답함.
  - 질문하는 AI 있잖아요. 거기서 제가 답을 하면서 실제 선생님과 대화를 한다는 느낌을 받았는데요. 이게 뒤처지는 애를 방지하는 중요한 역할을 할 것 같습니다. 제가 모르는 키워드를 여러 개 던져주면서 이게 왜 그런지 제가 상상을 한번 해보게끔 하고 그 상상을 통해서 제가 답을 작성해 가지고 AI한테 맞는지 물어보는 과정을 통해서 단백질이 왜 손상이 됐는지 그런 것을 깨달았는데요. 이런 게 사고력 논리력을 키워줘서 나중에 고등학교에 갔을 때 어려운 개념들을 나왔을 때 잘 헤쳐나가는 데 도움이 될 것 같습니다."(학생8)
  - "AI를 쓰면서 질문을 더 잘 만드는 연습이 되었고 답변을 검증하는 과정을 통해 학습 방식이 조금은 달라지기는 했던 것 같아요. 그렇지만 핵심 과제는 결국 제가 판단해야 한다는 생각도 들었습니다. AI에는 신뢰성 문제도 있다고 하잖아요. 뭔가 신뢰성을 한 번 더 검증해 줘야 될 것 같아서 다시 한 번 더 찾아보게 되고 더 머리에 깊게 박히게 되는 것 같아요. 만약 그 틀린 부분이 있으면 정확하게 수정해서 이해할 수 있게 하다 보니 학습 방법이 수정되었습니다."(학생11)

나 생성형 AI 기반 과학탐구 지원 시스템의 개선점

- 학생들의 면담 결과를 종합하여 다음의 세 가지 개선점을 도출하였음.

1) 지능형 과학실 ON의 AI 기능을 과학 분야를 넘어선 활용에 허용하면 융합탐구에 도움이 될 것이다.

- 지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템은 공교육에서의 교사와 학생에게 도입되는 생성형 AI으로서 안전성 확보와 과학탐구 활성화를 주요한 가치로 인식하고 개발을 하였음. 과학 분야에 대한 질문이 아닌 경우 응답할 수 없도록 제한하였는데 이러한 점이 실생활 기반 문제해결형 과학탐구나, 환경문제 탐구 등 융합 탐구를 수행하는 것에는 불편함을 준다고 응답하였음.
  - "과학 이외에 다른 영역에서는 답변이 아직 충분하지 않다고 느꼈는데 이제 이 방면을 개선해서 학생들이 과학과 또 다른 영역을 묶어서 또 다른 데이터를 수집할 수 있게 도와주는 쪽이 학생들의 자율적인 탐구를 수행하는데 도움을 주는 거라고 저는 그렇게 생각을 하고 있어요."(학생10)
  - "과학적 지식에서 벗어나는 질문을 했을 때는 답변이 풍부하지 않거나 제대로 된 답변을 받지 못한다는 단점이 있었습니다."(학생11)

- 과학분야에 대한 질의응답 외에도 지능형 과학실 ON에서의 AI 기능을 다른 교과목의 수업에도 활용한다면 학생들의 학습에 도움이 될 수 있다는 인식이 나타남. 보고서 작성을 지원하여 학생들의 보고서 작성 능력을 키워준다는 생각이나 본 시스템을 신뢰한다는 생각을 통해 본 시스템이 과학 분야를 넘어서서 활용될 때에도 이익이 있다고 인식하도록 하였음.
  - "보고서를 평가해 주는 AI의 경우에는 저희가 다른 과목과 연계했을 때도 꽤 큰 도움을 받을 수 있을 것 같다는 생각이 들었어요. 과학이 아니라 국어에서 보고서를 작성해야 하는 상황에서도 이 AI를 이용해서 저희가 보고서를 작성하는 능력을 기를 수 있기 때문에 국어 시간에 보고서를 작성할 때도 더 잘 적을 수 있을 것 같아요."(학생6)
  - "저희가 챗GPT 많이 사용하잖아요. GPT한테 과학 질문을 하다 보면 GPT가 약간 헛소리 비스무리한 걸 할 때가 많거든요. 근데 이게 과학 전용 AI다 보니까 오히려 더 정확하게 용어 정리도 해주고 개념 설명도 해주더라고요. 그러니까 그런 면을 보면 다른 부분에서 추후에 활용될 가능성도 높다고 생각합니다."(학생8)

2) 지능형 과학실 ON의 AI가 시각적 인식·표현 능력이 높아진다면 과학탐구에 더 도움이 될 것이다.

- 대화형 서비스 응답의 문단 구분, 불릿 기호 등을 통한 직관적 설명, 폰트 크기, 정렬 방식 등 텍스트의 시각적 표현 방식에 대해서도 개선의 필요성이 언급되었음. 이러한 부분은 프롬프트 엔지니어링을 통해 해결하기는 어려운 내용으로서 본 시스템이 활용하고 있는 LLM인 HyperCLOVA X의 발전을 통해 해결할 수 있음. 그러나 학생들이 문서의 시각적 표현에 대해서도 요구가 있음을 확인할 수 있었음.
  - "AI를 사용해 본 경험으로 비교를 해보자면 저희가 잘 아는 챗GPT는 설명을 해 줄 때 나눠서 설명을 해 주는 경우가 있는데, 이 AI를 사용했을 때는 한번에 글로 써져 있어서 보기에도 좀 불편하고 어디가 중요한지 확인하기도 어려워요. 이런 방면에서도 개선을 해서 AI가 자료를 시각화해 주는 기술을 지원을 해주면 학생들의 탐구 역량을 함양할 수 있다고 생각합니다."(학생10)

- 보고서 튜터링의 경우 생물 이미지 사진을 첨부한 학생이 AI가 사진 인식을 정확히 못하였던 경험을 토대로 이미지 인식률을 높여야 함을 개선 의견으로 제시하였음. 이 또한 본 시스템이 활용하고 있는 LLM인 HyperCLOVA X의 발전을 통해 해결할 수 있음. 과학 분야에 따라 LLM의 이미지 인식률이 더 중유한 분야가 있을 수 있음을 확인하였음.
  - 튜터링 서비스의 단점은 사진을 첨부하면 AI가 사진을 제대로 인식을 못해서 좋은 답변이 나오지

않았던 점이 있었어요.(학생11)

3) 대화형 서비스가 나를 더 잘 이해하면 과학탐구에 도움이 될 것이다.

- 학생들은 대화형 서비스가 학생들이 앞서 한 질문의 내용을 토대로 학생의 후속 질문의 맥락과 의도를 더 잘 파악해주기를 원하고 있었음. 멀티턴(multi-turn) 대화란, 대화가 여러 차례 걸쳐 진행되는 과정에서 챗봇이 과거의 질문과 답변을 기억하고 반영하여 대화를 이어나가는 경우를 말함. 학생들의 요구는 멀티턴 기능의 향상을 의미한다고 볼 수 있음. 본 시스템은 학생의 정보를 기억하거나 수집하는 것을 최소화하는 방향으로 프로토타입을 개발하였으나 보다 자연스러운 대화를 위해 앞선 대화를 맥락 정보로 활용하는 기술적 고려가 필요함.
  - “대화형 서비스에서 개선점은 제가 글을 정리해 달라고 했는데 잘 정리를 못 해주더라고요. 예를 들어 챗GPT는 500자 내외로 요약해 달라고 하면 잘 요약해 주거든요.”(학생11)
  - “저는 AI를 사용할 때 층층이 연관 개념들을 올려가면서 질문하는 편입니다. 예를 들어 공학에서의 활용성 이런 거에 대해 질문할 때 먼저 공학이 뭐야, 기계 공학이 뭐야, 여기서 어떻게 활용돼 이런 식으로 층층이 쌓아가면서 질문을 하는 것이에요. 그런데 지능형 과학실 ON에서는 층층이 질문하지 못하고 각각 따로따로 분리되는 것처럼 느꼈습니다.”(학생12)

4절 시스템 로그데이터 분석

1 대화형 서비스

가 로그데이터 기술 통계

- 대화형 서비스 로그 데이터 분석을 위해 13개 초·중·고등학교로부터 수집된 총 2,845건의 이벤트(event) 로그를 사용하였음.

〈학교급별 대화형 서비스 로그 이벤트 수(비율)〉

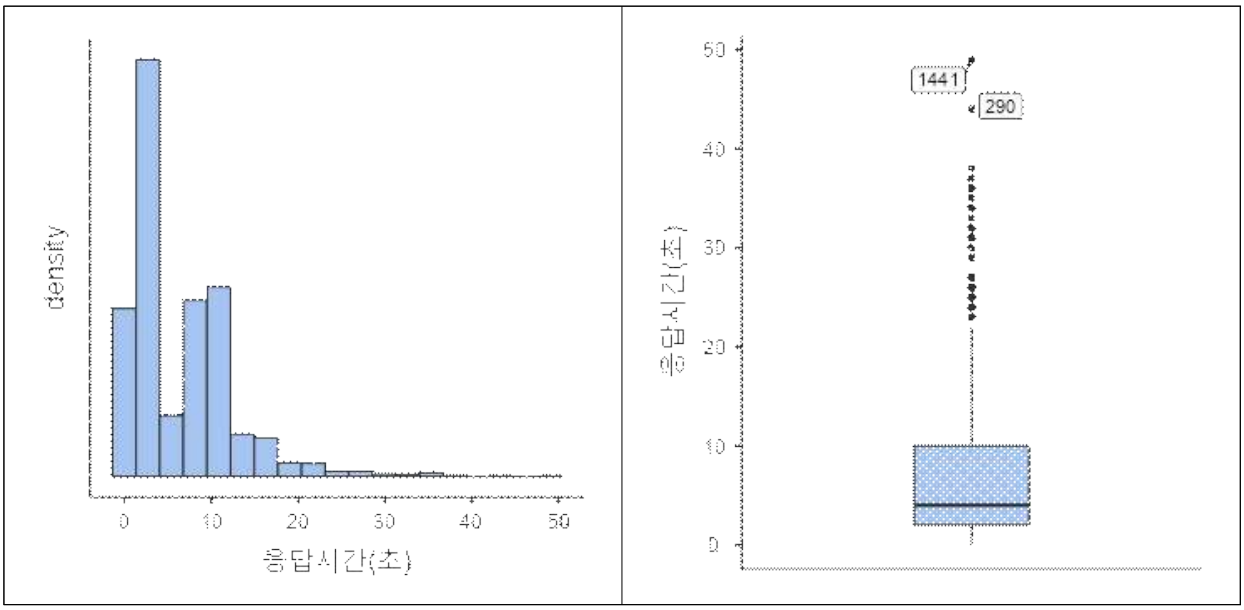
학교급	이벤트수	전체 비율
고등학교	1,136	39.9%
중학교	1,086	38.2%
초등학교	623	21.9%
총계	2,845	100%

〈대화형 서비스 로그 이벤트 수(비율)〉

하위 요인	이벤트수	전체 비율
생식과 유전	498	17.5 %
기체의 성질	363	12.8 %
변화와 다양성	362	12.7 %
온도와 열	353	12.4 %
열과 우리생활	234	8.2 %
과학과 미래 사회	163	5.7 %
생물의 구조와 기능	158	5.6 %
열과 우리 생활	92	3.2 %
다양한 생물과 우리 생활	67	2.4 %
식물의 구조와 기능	34	1.2 %
용해와 용액	33	1.2 %
산과 염기	18	0.6 %
여러 가지 기체	17	0.6 %

- 대화형 서비스의 평균 응답시간(초)은 6.49초였으며 표준편차는 5.84초 였음. 히스토그램을 통해 파악된 응답시간의 분포는 일반적인 **응답시간 전형적인 분포인 오른쪽으로 편포된 분포**를 보임. 또한 박스플롯에서 알 수 있듯이, **이상치를 포함해 매우 긴 응답시간의 경우도 다수 발견**되었음. 이는 학생들이 서술형 답변을 통해 지적한 느린 응답속도와 관련된 경우로 볼 수 있으며, 향후 **응답시간과 관련해 균일한 수준의 서비스를 구현할 수 있도록 개선될 필요**가 있음.

〈시스템의 평균 응답 시간 분포(초)〉



- 학교급별 평균 응답시간의 경우, 초등학교가 가장 큰 평균 응답시간(7.35초)과 표준편차(6.85초)를 보이는 것으로 나타남. 이는 **초등학교의 경우 가장 응답이 느리고, 또한 초등학교들 간 편차가 크다는 것**을 나타냄. 향후 **학교급별 정보통신 인프라와 관련해서도 균일한 수준의 서비스가 제공될 수 있도록 점검이 필요**함.

〈학교급별 평균 응답시간〉

학교급	평균(초)	표준편차(초)
고등학교	5.85	4.55
중학교	6.66	6.31
초등학교	7.35	6.85

나 질문 유형 분석

- 학생들의 **질문 내용이 한 단어로 매우 짧은 경우도 많고**, 교과와 관련 없는 의미 없는 내용도 많아, 토픽 모델링(Topic modeling)과 같은 **텍스트 마이닝 기법을 본격적으로 적용하기는 어려움이 있음**.
- 이에 학생들의 질문 내용을 검토해 **정의, 원리, 실천, 기타**의 네 가지 분류 체계를 도출하였음.

〈학생들 질문 내용으로부터 도출된 분류 체계〉

범주	설명
정의	특정 용어, 개념, 사실, 특징 등에 대한 단순한 정의를 요청하는 질문
원리	특정 현상의 작동 메커니즘, 발생 과정, 원인과 결과의 논리적 연결, 또는 과학적 이론을 바탕으로 설명할 것을 요청하는 질문
실천	특정 목표 달성을 위한 구체적인 방법, 개인적/사회적 행동지침, 문제 해결책, 또는 실험/실천 계획의 설계를 요청하는 질문
기타	위 세 가지 분류에 해당하지 않는 의미 없는 질문

- 위와 같은 분류 체계를 기반으로 **프롬프트(prompt)를 작성해 LLM(ChatGPT)을 통해 총 2845건의 학생 질문을 정의, 원리, 실천, 기타로 분류(classification)**하는 작업을 수행하였음. 이러한 LLM기반 분류 작업의 결과, 학생들의 질문은 아래와 같이 분류되었음.

〈학생들의 질문 내용 분류〉

범주	빈도	비율
정의	1968	69.2%
원리	281	9.9%
실천	77	2.7%
기타	519	18.2%
총계	2845	100%

- LLM기반 분류 작업의 결과에 따르면, 학생들은 **정의에 대한 설명을 가장 많이 요청하였고 (1968건, 69.2%)**, 그 다음으로 원리에 대한 설명을 많이 요청하였고 (281건, 9.9%), 마지막으로 **실천에 대한 설명 (77건, 2.7%)은 가장 적게 요청**하였음. 이러한 결과는 학생들이 대화형 서비스를 **자기주도적으로 원리를 탐구하는 고차원적인 학습을 위한 도구보다는 검색엔진을 대체하는 도구로 인식하고 사용할 가능성을 시사**한다고도 해석할 수 있어, 향후 심층 면접 등을 통한 연구를 통해, 학생들의 사용 의도를 보다 정확하고 심도있게 파악하고, 이러한 AI 도구들이 단순한 검색엔진의 대체물이 아닌 보다 고차원적인 학습을 위한 도구로 사용되도록 **체계적인 질문 과정을 교육시키는 사전 개입**이 이루어질 필요가 있음.
- 한편, LLM기반 분류 작업의 결과에 따르면, 의미 없는 기호 등 정의, 원리, 실천으로 분류될 수 없는 질문인 기**타에 해당하는 질문도 519건(18.2%)**이나 되었음. 이러한 불필요한 입력을 줄이기 위한 사전 교육도 필요할 것으로 판단됨.
- 학교급에 따라 LLM기반 분류 작업의 결과를 분류한 결과는 다음과 같음. **학교급별 질문 유형의 분포는 대체로 유사**하였으나, **초등학생의 경우 중학생이나 고등학생에 비해 실천에 해당하는 질문(7.2%)을 많이 하는 것으로 나타남**.

〈학교급별 학생들의 질문 내용 분류〉

학교급	구분	질문내용 분류				전체
		기타	실천	원리	정의	
고등학교	관찰값	235	26	97	778	1136
	행 비율	**20.7 %**	**2.3 %**	**8.5 %**	**68.5 %**	100.0 %
	열 비율	45.3 %	33.8 %	34.5 %	39.5 %	39.9 %
중학교	관찰값	182	6	124	774	1086
	행 비율	**16.8 %**	**0.6 %**	**11.4 %**	**71.3 %**	100.0 %
	열 비율	35.1 %	7.8 %	44.1 %	39.3 %	38.2 %
초등학교	관찰값	102	45	60	416	623
	행 비율	**16.4 %**	**7.2 %**	**9.6 %**	**66.8 %**	100.0 %
	열 비율	19.7 %	58.4 %	21.4 %	21.1 %	21.9 %
전체	관찰값	519	77	281	1968	2845
	행 비율	18.2 %	2.7 %	9.9 %	69.2 %	100.0 %
	열 비율	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %

2 보고서 튜터링 서비스 성과 비교

가 연구 방법

1) 분석 환경 및 도구

- 분석 데이터: 연구 참여 교사의 평가 계획에 따라 구안된 평가 루브릭을 기준으로 교사가 튜터링 적용 전과 후의 학생 보고서 2건을 평가한 점수
- 프로그래밍 언어: Python 3.13 기반
- 핵심 라이브러리:
  - 데이터 처리: Pandas를 활용하여 CSV 데이터 로드, 결측치 처리 및 파생 변수 생성.
  - 통계 분석: SciPy 및 Statsmodels를 사용하여 T-검정 및 분산분석(ANOVA) 수행.
  - 시각화: Matplotlib 및 Seaborn을 활용하여 데이터 분포 및 통계적 차이를 시각화.

2) 데이터 전처리

- 유효 데이터 추출: 사전 점수와 사후 점수 중 결측값(Null)이 포함된 데이터는 분석의 정확도를 위해 제외(dropna)함.
- 성취도 향상 폭 계산: 각 학생 별 피드백 효과를 정량화하기 위해 사후 점수에서 사전 점수를 뺀 Score_Diff(점수 차이) 변수를 생성하여 분석에 활용함.

3) 통계적 검증 방법

- 대응표본 t-검정 (Paired T-test)
  - 목적: 동일 집단(학생)의 1차(사전) 및 2차(사후) 점수 간의 평균 차이가 통계적으로 유의미한지 검증.
  - 적용: 전체 학생 대상 검정 및 각 학교별 개별 검정을 수행하여 집단별 효과성을 세분화하여 분석함.
- 일원분산분석 (One-way ANOVA)
  - 목적: 학교(School)라는 독립변수가 성취도 향상 폭(Score_Diff)에 미치는 영향이 유의미한지 검증.
  - 적용: 학교별로 피드백의 효과 차이가 통계적으로 존재하는지 확인하기 위해 실시함.
  - 성취도 향상 폭 계산: 각 학생별 피드백 효과를 정량화하기 위해 사후 점수에서 사전 점수를 뺀 Score_Diff(점수 차이) 변수를 생성하여 분석에 활용함.

나 지능형 과학실 ON 튜터링 서비스에 따른 학생 성과 변화

1) 전체 효과성 분석

- 전반적 학업 성취도 향상: 전체 학생(N=250)을 대상으로 대응표본 t-검정(Paired t-test)을 실시한 결과, 사전 점수 평균 85.73점에서 사후 점수 평균 89.38점으로 3.66점 상승함.
- 통계적 유의성 확보:T-통계량은 6.539, 유의확률(p-value)은 0.000으로 나타나, 지능형 과학실의 탐구 피드백 처치가 학생들의 성취도 향상에 통계적으로 매우 유의미한 효과($p<.001$)가 있음을 확인함.

2) 학교별 효과 차이 및 그룹별 특성

- 학교별 사전 수준과 점수 향상 패턴에 따라 다음과 같이 3가지 그룹으로 유형화할 수 있음.

구분 (Group)	표본 수	사전 평균	사후 평균	점수 차이	T값	유의확률	유의성
전체 (Overall)	250	85.73	89.38	+3.66	6.539	0	***
E 초등학교	5	84	100	+16.00	4	0.0161	*
A 초등학교	16	92.5	97	+4.50	34.857	0	***
F 초등학교	22	92.73	92.95	+0.23	0.165	0.8703	ns
C 초등학교	26	100	100	0	–	–	–
G 초등학교	14	94.79	100	+5.21	45.818	0	***
B 초등학교	3	70	95	+25.00	8.66	0.0131	*
D 초등학교	5	80	90	+10.00	inf	0	***
I 중학교	23	91.96	100	+8.04	6.903	0	***
H 중학교	24	90	90.62	+0.62	0.175	0.8624	ns
K 고등학교	60	100	100	0	–	–	–
L 고등학교	24	20	19.75	−0.25	−1.813	0.083	ns
M 고등학교	28	78.57	91.61	+13.04	6.319	0	***

- 유의미한 성취도 향상 그룹 (피드백 효과가 뚜렷한 학교)
  - 대상 학교: A초, G초, M고, E초, B초, D초, I중
  - 특징:이들 학교는 피드백 제공 후 사후 점수가 통계적으로 유의미하게 상승함($p<.05$).
  - 해석:특히 A초(+4.5점), G초(+5.21점), M고(+13.04점) 등은 피드백을 통해 보고서의 완성도가 크게 개선됨. B초의 경우 표본 수는 적으나 평균 25점의 가장 큰 상승 폭을 보임.
- 천장 효과(Ceiling Effect) 그룹 (사전 성취도가 이미 높은 학교)
  - 대상 학교: C초, K고
  - 특징:사전 점수와 사후 점수가 모두 만점(100점)으로 동일하여 통계적 유의성 검정이 불가능함.
  - 해석:학생들이 이미 해당 과제에 대해 높은 수준의 이해도를 갖추고 있어, 기본 피드백 외에 심화 탐구 등의 추가적인 도전 과제가 필요할 것으로 판단됨.
- 성취도 유지 그룹 (통계적 차이가 없는 학교)
  - 대상 학교: F초, H중, L고
  - 특징:사전-사후 점수 간의 차이가 통계적으로 유의미하지 않음(ns).
  - 해석: F초, H중의 경우, 사전 점수가 이미 90점대로 매우 높아, 추가적인 점수 상승의 여지가 적었음(상위권 유지). L고는 사전 점수가 약 20점으로 타 학교 대비 낮으며, 피드백 후에도 유의미한 변화가 나타나지 않음. 이는 과제 난이도나 학생들의 학습 동기 등 외적 요인에 대한 추가적인 질적 분석이 요구됨.

3) 종합 결론

- 지능형 과학실 ON의 AI 및 교사 피드백 시스템은 전반적으로 학생들의 과학적 탐구 보고서 작성 능력 향상에 긍정적인 영향을 미침.
- 다만, 학교급 및 학생들의 기초학력 수준(Baseline)에 따라 피드백의 효과가 상이하므로, 이미 성취도가 높은 집단에게는 심화 피드백을, 성취도가 낮은 집단(예: L고등학교 사례)에게는 기초 개념 보정 피드백을 강화하는 등 수준별 맞춤형 피드백 전략이 필요함.

다 지능형 과학실 ON 튜터링 서비스에 따른 학교별 및 집단 특성

1) 전체 집단에 대한 효과성 검증 (Overall Paired T-test)

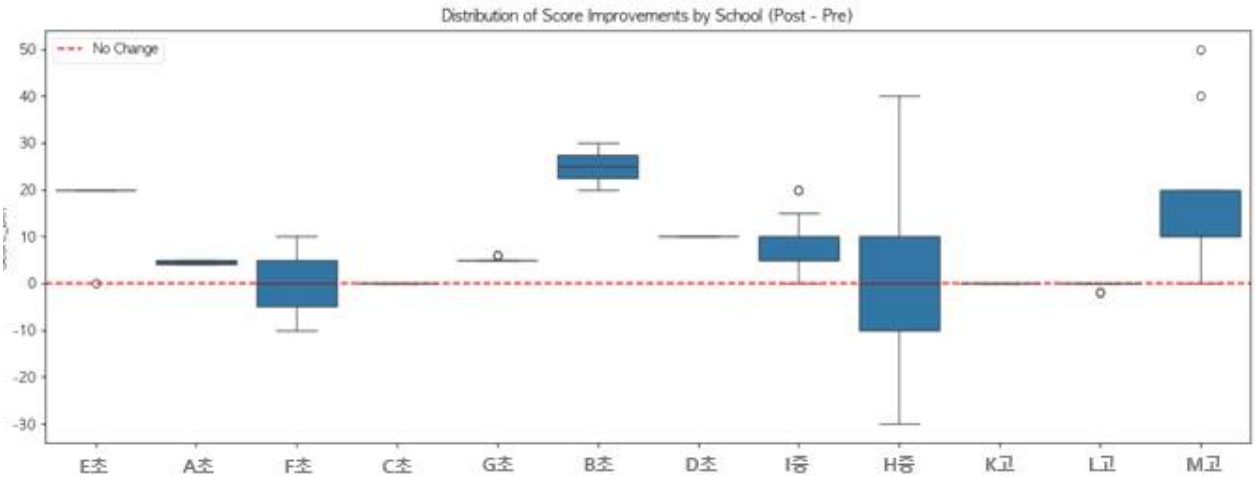
- 통계적 수치: 전체 데이터를 대상으로 분석한 결과, T-통계량(T-statistic)은 6.539, 유의확률(p-value)은 약 3.48×10^{-10}으로 산출됨.
  - 결과 해석: 일반적으로 p<0.05일 때 유의미하다고 판단하나, 본 결과는 0에 가까운 매우 낮은 값을 보임.
  - 이는 지능형 과학실 ON의 탐구 피드백 처치가 우연이 아닌, 실질적으로 학생들의 학업 성취도를 향상시키는 요인으로 작용했음을 강력하게 시사함.

2) 학교별 차이 검증 (ANOVA Analysis)

- 통계적 수치: 학교 변수에 따른 점수 향상 폭(Difference)의 차이를 일원분산분석(ANOVA)으로 검정한 결과, F-통계량은 12.92, 유의확률(p-value)은 약 3.68×10^{-19}로 나타남.
- 결과 해석
  - F-값이 12.92로 매우 높게 나타난 것은 학교 간(Group-between) 변동이 학교 내(Group-within) 변동보다 훨씬 크다는 것을 의미함.
  - 즉, 피드백의 효과는 모든 학교에서 동일하게 나타나는 것이 아니며, 학교별 특성(기초학력, 참여도 등)에 따라 성취도 향상의 정도가 유의미하게 달라짐을 통계적으로 입증함.

3) 데이터 시각화 분석

- 박스 플롯(Boxplot) 분석
  - 점수 분포 이동: 'Pre vs Post' 그래프에서 사후 점수(Post)의 박스(Box)가 사전 점수(Pre) 대비 전체적으로 상향 이동한 패턴이 관찰됨.
  - 학교별 편차 확인: 'Score Difference' 그래프에서 A초, G초 등은 중앙값이 0보다 훨씬 높은 위치에 형성된 반면, 일부 학교는 0 부근에 머무르거나 변화폭이 좁게 나타남.
- 박스 플롯(Boxplot) 분석
  - D초와 같이 모든 학생의 점수가 일률적으로 상승한 경우 분산이 0에 수렴하여 시각적으로 매우 좁은 분포를 보임.
  - 이는 해당 학교의 학생들이 피드백을 매우 균일하게 수용했거나, 특정 수업 방식이 일관된 효과를 냈을 가능성을 보여줌.



5절 시스템 성능 평가

1 교사의 평가 루브릭과 보고서 튜터링 간의 유사도

- 본 실험은 실제 교사가 작성한 보고서 평가 루브릭과 AI 튜터가 생성한 피드백(튜터링) 간의 유사도를 정량·정성적으로 비교하였음. 유사도 평가는 Macina et al.(2023)에서 사용한 방식을 차용하여 전통적 유사도 평가 지표(KF1), sacreBLEU(Post, 2018), BERTScore(Zhang, Kishore, Wu, Weinberger, & Artzi, 2020)를 구성하였으며, 대규모 언어모델을 활용한 정성평가(G-Score)를 병행하여 수행함.
- 특히 최근 연구(Zheng et al., 2023)에서 LLM의 평가 결과가 인간 평가자와 높은 상관을 보이는 것으로 보고됨에 따라, 본 연구에서도 이를 활용하여 교사의 평가 루브릭과 AI 피드백 간의 정성적 유사도 및 품질을 측정하였음. 이를 통해 단순 문장 표면 유사도뿐 아니라, 내용 정렬도, 개념 정합성, 표현 수준 차이 등 다양한 관점에서 교사-AI 피드백 간 관계를 분석하고자 함.

가 유사도 평가 지표

- 본 연구에서 사용한 유사도 평가 지표는 다음과 같음.
  - **KF1**(Token-level F1 Score): 교사의 루브릭과 AI 피드백 간에 공유되는 핵심 단어의 정밀도·재현율 기반 유사도를 측정함
  - **sacreBLEU**: 교사 루브릭과 AI 피드백간 n-gram 일치도를 통해, 표면적 문장 유사도와 표현 일치도를 평가함.
  - **BERTScore**: 교사 루브릭과 AI 피드백간 의미적 유사도를 임베딩 기반으로 측정하여, 표현이 달라도 의미가 유사할 경우 높은 점수를 부여함.
  - **G-Score**(GPT-based Scoring): GPT를 활용한 정성·의미 기반 유사도 평가로, 내용 구성, 개념 정합성, 설명 목적의 일치 여부 등을 종합적으로 판단함
- 전통적 유사도 평가 지표(KF1, BLEU, BERTScore)는 표면적 어휘·구문 유사도, G-Score는 내용적·개념적 정렬 정도를 반영한다는 점에서 상호 보완적인 성격을 가짐.

나 유사도 분석

- 유사도 분석은 피드백 전(before-tutoring)과 피드백 후(after-tutoring) 조건을 각각 평가하여, AI 튜터링이 교사 루브릭과의 정렬에 어떤 영향을 미쳤는지 확인하는 방식으로 이루어짐.

1) 전통적 텍스트 유사도 지표기반 정량 평가

- 교사가 제시한 평가 루브릭은 상·중·하의 세 수준으로 기술되어 있으며, 본 연구에서는 AI 튜터가 생성한 피드백과 각 루브릭 수준 간의 텍스트 유사도를 개별적으로 산출한 뒤, 가장 높은 유사도를 보이는 루브릭 수준을 AI 피드백의 대응 기준으로 선택하였음.
- 유사도 평가는 두 관점에서 수행되었는데, 개별 문항 단위의 피드백을 기준으로 하는 Item Level, 그리고 보고서 전체를 요약한 AI의 종합 평가(Overall)를 기준으로 하는 방식임. 또한 단순 표면적 문장 비교를

피하고 의미적 비교의 정확도를 높이기 위해, 각 텍스트에서 핵심 내용어(nouns)를 추출하여 유사도 계산에 활용하였음. 이러한 전처리 과정을 거친 뒤, Token 단위 F1 score, sacreBLEU, BERTScore 등 전통적 텍스트 유사도 지표를 적용하여 AI 피드백과 교사 루브릭 간의 정량적 유사도를 평가하였음.

〈교사 평가 루브릭과 AI 튜터링 간 유사도 평가 결과〉

구분		KF1	sacreBLEU	BERTScore
피드백 전	Item Level	14.12	71.25	68.60
	Overall	18.22	74.63	69.00
피드백 후	Item Level	14.26	70.48	68.70
	Overall	18.22	79.11	69.11

- 위 표에 나타난 바와 같이, 전통적 지표들은 전반적으로 중간 정도의 유사도를 보였으며, 지표별 특성이 반영된 차이가 관찰되었음. 먼저 KF1은 교사 루브릭과 AI 피드백이 사용하는 핵심 어휘가 서로 다르기 때문에 특히 낮은 값을 보였음. 반면 sacreBLEU와 BERTScore는 표현 방식과 문장 구조 차이에 더 민감하여, 교사의 루브릭이 상위 개념 중심의 요약적 평가 문장으로 구성된 반면 AI 피드백은 세부 요소를 짚어주는 설명 중심 서술에 가까워 표면적 유사도가 상대적으로 낮게 나타났음. 이러한 차이는 두 텍스트가 지향하는 목적과 서술 방식이 본질적으로 다르다는 점에서 기인한 것으로 해석됨.

2) GPT 기반 정성 평가

- KF1, sacreBLEU, BERTScore와 같은 전통적 텍스트 유사도 지표는 어휘 일치나 문장 구조의 표면적 유사성에는 민감하지만, 피드백의 질·의도·교육적 타당성과 같이 정성적 판단이 필요한 요소를 충분히 반영하기 어려움. 이러한 한계를 보완하기 위해 GPT 기반 정성 평가 방식인 G-Score를 활용하여, AI 피드백이 교사 루브릭과 의도적으로 얼마나 정렬되는지, 그리고 단순 유사도를 넘어 확장성·표현의 구체성·실행 가능성 등 실제 교육적 품질 요소를 추가적으로 평가하였음.
- 본 평가에서는 Item Level과 Overall을 구분하지 않고 하나의 피드백으로 통합하여 분석하였고, 교사가 작성한 상·중·하 루브릭 전체를 제시한 뒤 GPT가 다음 다섯 기준(① 유사도(G-Score), ② 루브릭 준수도(Compliance), ③ 루브릭 확장성(Enrichment), ④ 어휘·표현 구체성(Lexical), ⑤ 작업 가능성(Actionability))에 대해 1~5점 리커트 척도로 점수를 부여하도록 하였음. 이를 통해 전통적 지표만으로는 파악하기 어려운 피드백의 교육적 완성도와 질적 차이를 정교하게 분석하였음.

〈GPT 기반 정성 평가 결과〉

구분	G-Score	Compliance	Enrichment	Lexical	Actionability
피드백 전	3.79	3.80	4.27	3.69	3.72
피드백 후	3.84	3.86	4.22	3.76	3.69

- 실험 결과에 따르면, 전체적으로 평균 점수는 3점대 중반에서 후반에 분포하며, 피드백 전·후 모두 큰 폭의 변화 없이 안정적인 수준을 유지하는 경향을 보임. 우선 유사도(G-Score), 루브릭 준수도(Compliance), 어휘·표현 구체성(Lexical Specificity), 작업 가능성(Actionability) 항목은 모두 3.6~3.8 수준으로 비교적 균일하게 나타났으며, 이는 AI 피드백이 교사 루브릭의 핵심 요소를 어느 정도 반영하면서도, 특정 항목에서 매우 우수하거나 매우 부족한 사례가 많지 않았음을 의미함.
- 특히 주목할 점은 점수 분포가 전반적으로 3점대에 집중되었다는 것임. 이는 평가 과정에서 1점(부적합)이나 5점(탁월)에 해당하는 극단적 사례가 적었다는 의미로 해석할 수 있음. 루브릭의 구조적 특성상, AI 피드백은 대체로 루브릭의 큰 방향성에는 부합하지만, 동시에 교사 루브릭이 요구하는 구체적

기준을 완전히 충족하거나, 반대로 전혀 맞지 않는 형태도 드물기 때문에, 점수가 자연스럽게 중간-상위 정도의 범위에 수렴하는 경향을 보인 것으로 볼 수 있음.

- 또한 루브릭 확장성(Enrichment) 항목은 상대적으로 높은 수준(4점대 초반)을 유지하고 있음. 이는 AI 피드백이 교사 루브릭에 명시된 요소 외에도 특정 개념 설명, 예시, 추가 조언 등을 포함하여 확장성 측면에서 일관되게 높은 평가를 받은 것으로 해석됨.

다 키워드 분석

- 전통적 텍스트 유사도 지표(KF1, sacreBLEU, BERTScore)와 G-Score 기반 정성 평가를 통해 교사 루브릭과 AI 피드백 간의 유사도를 분석하였으나, 이러한 지표들은 텍스트가 어떤 어휘를 중심으로 구성되어 있으며 개념 수준이나 설명 방식이 어떻게 다른지를 충분히 드러내기 어려움. 실제로 교사 루브릭은 상·중·하 수준을 기술하기 위해 상위 개념 중심의 요약적 문장으로 구성되는 반면, AI 피드백은 학생의 답변 맥락에 기반하여 세부 요소를 짚어주는 설명형 서술이 많아, 표면적 유사도나 정성 평가만으로는 두 텍스트의 개념적 차이를 명확히 파악하기 어려움.
- 본 연구에서는 루브릭과 AI 피드백 각각에서 핵심 명사·개념어를 추출하여 비교하는 키워드 분석을 추가적으로 수행하였음. 특히 문서 수준의 특성을 정량적으로 파악하기 위해 TF-IDF 기반 상위 10개 단어를 도출하여 루브릭과 AI 피드백이 강조하는 개념의 범위, 세부성 수준, 어휘 다양성의 차이를 비교하였음. 이를 통해 두 텍스트가 강조하는 개념 수준의 차이(상위 개념 vs 세부 개념), 사용 어휘의 범위와 다양성의 차이를 확인하고자 하였음.

〈TF-IDF 기반 Rubric-Tutor 키워드 비교 (대표 3개 보고서)〉

보고서	교사 루브릭 상위 TF-IDF 10개	AI 튜터링 상위 TF-IDF 10개	해석
2	멘델, 유전, 미흡하다, 연구, 대화, 잡종, 자가, 시뮬레이터, 수분, 지능	법칙, 유전, 멘델, 우성, 형질, 우열, 열성, 결과, 비율, 분리	Rubric은 평가적·요약적(파랑), AI 튜터링은 구체 개념(빨강) 중심
3	병원체, 표본, 분류, 비교, 개념, 경우, 관찰, 활용, 서술, 모호하다	세포, 미생물, 사진, 현미경, 관찰, 바이러스, 염색, 크기, 특징, 형태	Rubric은 상위 개념(파랑), AI 튜터링은 실제 관찰 대상을 구체적으로 기술(빨강)
7	기체, 이동, 현상, 방법, 측정, 관찰, 도구로, 센서, 자다, 구체	공기, 대류, 온도, 기체, 열의, 이동, 현상, 변화, 위로, 자다	Rubric은 개념·절차 중심(파랑), AI 튜터링은 현상 설명 중심(빨강)

- TF-IDF 분석 결과, 교사 루브릭과 AI 튜터 피드백은 공통 개념을 공유하면서도 강조하는 어휘의 수준과 구체성에서 뚜렷한 차이를 보임. 루브릭은 '기체 이동', '멘델 유전', '병원체 분류'와 같이 상위 개념(high-level concepts) 을 중심으로 기술되어 있으며, 평가 기준을 요약적으로 표현하는 경향이 나타났음. 반면 튜터 피드백은 '대류', '온도 변화', '우성 열성 형질', '현미경 관찰', '크기 특징 형태', '세포 미생물 바이러스' 등 구체적 현상·절차·사례에 해당하는 세부 개념(low-level concepts)이 높은 TF-IDF 값을 보였음. 이러한 차이는 두 텍스트가 수행하는 역할의 본질적 차이에서 기인함.

1. **루브릭은 평가 기준을 정의하는 문서**로서, **핵심 개념을 추상화**하여 제시함.
2. **AI 튜터 피드백**은 학생의 답변을 개선시키기 위한 **설명 중심 발화**이므로, **개념을 더 세분화**하고 **구체적 어휘**를 활용하는 경향이 큼.

- 따라서 전통적 텍스트 유사도 지표에서 낮은 값이 나타난 이유는 내용의 부정합 때문이 아니라, 동일한 개념군을 서로 다른 추상도에서 서술한 결과로 해석할 수 있음.
- TF-IDF 기반 키워드 분석 결과, 교사 루브릭과 AI 피드백은 핵심 개념어 차원에서는 일정 수준의 연관성을 유지했으나, 실제 사용되는 어휘의 구체성·세부 수준에서 뚜렷한 차이를 보였음. 이러한 결과는 전통적 텍스트 유사도 지표와 G-Score에서 낮은 유사도가 관찰된 이유가 단순 성능 저하가 아니라, 두 텍스트가 서로 다른 개념 수준(상위 개념 vs. 세부 개념)과 설명 방식을 사용하는 데 기인함을 뒷받침함. 다시 말해, 낮은 유사도 점수는 의미적 불일치라기보다는 서술 단위의 차이에서 발생한 것으로 해석할 수 있으며, 본 분석은 이러한 구조적 차이를 명확히 확인해 주는 근거가 됨.

2 RAG 성능평가

가 평가 개요 및 구성

1) 평가 목적

- 서비스의 정량적 신뢰성 검증
  - 대화형 서비스에 적용된 RAG 시스템이 사용자 질문에 대해 신뢰할 수 있는 정확한 정보를 제공하는지 정량적으로 검증함
  - 외부 지식 베이스 활용 능력에 기반한 시스템의 실질적 성능을 측정함
- 검색(Retrieval) 단계의 문서 검색 정확도 평가
  - 시스템이 검색한 문서 집합 내에 사용자 질문 해결에 필요한 핵심 정보가 정확히 포함되어 있는지 검증함
  - 불필요한 노이즈(Noise) 정보 없이 정답 문서를 상위에 노출시키는 검색 과정의 신뢰성을 평가함
- 생성(Generation) 단계의 정답 복원력 및 정확성 평가
  - 검색된 문서를 바탕으로 생성된 답변이 정답 답변과 의미적으로 얼마나 일치하는지 측정함
  - 모델이 사용자의 질의 의도를 정확히 파악하고, 왜곡 없이 올바른 정보를 전달하는지 검증함
- 근거 기반성 확보 및 환각 방지
  - 생성된 답변이 실제 검색된 문서의 내용을 논리적 근거로 삼아 작성되었는지 평가함
  - 사실이 아닌 정보를 생성하는 환각 현상을 억제하여, 교육용 서비스로서 필수적인 사실적 정확성을 확보함

2) 평가 항목 정의

- 시스템이 출력한 데이터(Evaluable Outputs)
  - question(질의): 사용자가 질문한 내용
  - retrievedDocs(검색된 문서): 사용자의 question에 대해 RAG 시스템의 검색기(Retriever)가 추출한 관련 문서 집합
  - answer(생성된 답변): 검색된 retrievedDocs를 바탕으로 LLM이 생성한 실제 답변
- 정답 기준 데이터 (Ground Truth)
  - Gold_Context(정답 문서): 입력된 question에 대해 시스템이 찾아야 하는 이상적인 맥락 정보
  - Gold_Answer(정답 답변): 질문에 대한 정답 응답

3) 핵심 평가 지표 (Gan et al., 2025)

- **Context Relevance** (retrievedDocs ↔ question)
  - 사용자가 입력한 질의(question)의 정보 요구에 대해, 검색된 문서(retrievedDocs)가 얼마나 적합한 정보를 담고 있는지 평가
  - 검색 단계에서 질의와 무관한 문서(Noise)를 배제하고 유의미한 문서를 검색했는지 측정
- **Answer Relevance** (answer ↔ question)
  - 생성된 답변(answer)이 초기 사용자의 질의(question)의 의도를 정확하게 파악하고 적절히 답변했는지 평가
  - 질의에 대한 응답으로서의 적절성을 검증
- **Faithfulness** (answer ↔ retrievedDocs)
  - 생성된 답변(answer)이 검색된 문서(retrievedDocs)의 내용에 충실히 기반하였는지 평가
  - RAG 시스템의 핵심 목표인 환각(Hallucination) 방지 여부를 판단
- **Accuracy** (retrievedDocs ↔ Gold_Context)
  - 시스템이 검색한 문서(retrievedDocs) 집합 내에 실제 정답 문서(Gold_Context)가 정확히 포함되어 있는지 평가
  - 검색기(Retriever)가 정답 문서(Gold_Context)를 누락하지 않고 상위에 노출시켰는지 정량적으로 측정
- **Correctness** (answer ↔ Gold_Answer)
  - 생성된 답변(answer)이 정답 답변(Gold_Answer)과 정보적, 논리적으로 얼마나 일치하는지 평가
  - 최종적으로 사용자가 얻게 되는 정보의 정확성을 전문가 기준(Ground Truth)과 비교하여 검증

나 검색(Retrieval) 단계 성능평가

1) 평가 메커니즘

- 정답 문서 포함 여부 검증
  - 시스템이 실제 검색을 통해 추출한 문서(retrievedDocs) 내에 정답 문서(Gold_Context)가 포함되어 있는지 확인
  - 시스템이 필요한 정보를 누락하지 않고 찾아냈는지(Recall)와 불필요한 문서를 얼마나 배제했는지(Precision) 판단
- 순위 기반 품질 분석
  - 단순 검색 여부 뿐만 아니라 검색된 문서(retrievedDocs)의 관련성 순위를 평가
  - 정답 문서(Gold_Context)가 검색 문서(retrievedDocs)의 앞순위에 배치될수록 높은 점수를 부여함
- 검색된 Top-K 문서 평가
  - 전체 검색 결과가 아닌 상위 K개의 문서(retrievedDocs@K)를 대상으로 정답 문서(Gold_Context)와의 일치 여부를 판별
  - 실제 RAG 시스템이 제한된 컨텍스트 윈도우(Context Window) 내에서 LLM에게 전달할 수 있는 문서의 수가 한정적이라는 현실적인 제약을 반영

2) 주요 평가 지표

- 비순위 기반(Non-Rank) 지표

- **Accuracy**: 전체 문서 중에서 올바르게 분류된 문서(관련 문서를 관련으로(TP), 비관련 문서를 비관련으로(TN) 분류한 경우)의 비율

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TotalNumber}}$$

- **Precision@K**: 검색된 상위 k개의 문서 중 실제 정답 문서가 차지하는 비율

$$\text{Precision@k} = \frac{|\text{Retrieved}_{1:k} \cap \text{Relevant}|}{k}$$

- **Recall@K**: 실제 정답 문서 집합 중에서 상위 k개의 검색 결과에 포함된 문서의 비율

$$\text{Recall@k} = \frac{|\text{Retrieved}_{1:k} \cap \text{Relevant}|}{|\text{Relevant}|}$$

- 순위 기반(Rank-Based) 지표
  - **MRR**(Mean Reciprocal Rank): 전체 질의 집합(Q)의 각 질의(q)에 대해 첫 번째 관련 문서가 등장하는 순위 역수의 평균

$$\text{MRR} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \frac{1}{\text{rank}_q}$$

  - **nDCG**(Normalized Discounted Cumulative Gain): 문서의 순위가 낮을수록 문서의 관련성 점수가 감쇠(discount)되도록 조정한 DCG를 최대 누적 이득(IDCG)에 대해 정규화하여 0-1 범위로 스케일링한 값

$$\text{DCG@k} = \sum_{i=1}^{k} \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)}, \quad \text{nDCG@k} = \frac{\text{DCG@k}}{\text{IDCG@k}}$$

  - **MAP**(Mean Average Precision): 각 질의(q)마다 관련 문서가 등장하는 순서를 고려하여 평균 정밀도(Average Precision, AP)를 계산하고, 이를 전체 질의에 대해 평균(mean)한 값

$$\text{MAP} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \frac{\sum_{k=1}^{N} \text{Precision@k}(q) \cdot rel(k)}{|\text{Relevant}_q|}$$

다 생성(Generation) 단계 성능평가

1) 평가 메커니즘

- 정답 유사도 비교 분석
  - 생성된 답변(answer)과 정답 답변(Gold_Answer) 간에 단어 또는 구(phrase)의 중첩 정도를 n-gram 기반으로 측정하여 텍스트의 표면적 일치도(lexical overlap) 평가
  - 생성된 답변(answer)과 정답 답변(Gold_Answer) 텍스트 임베딩 간의 코사인 유사도(Cosine Similarity)를 측정하여 텍스트의 의미적 유사도(Semantic Similarity) 평가
- 근거 기반성 및 환각 검증
  - 생성된 답변(answer)이 검색된 문서(retrievedDocs)의 내용에 논리적으로 기반하는지 검증

2) 주요 평가 지표

- 표면적 유사도(Lexical Similarity) 기반 지표
  - **BLEU**(Bilingual Evaluation Understudy) (Papineni et al., 2002): 생성된 답변(answer)과 정답 답변(Gold_Answer) 간의 n-gram 정밀도(P_n)를 기반으로 표면적 유사도를 측정하며, 생성 문장이 지나치게 짧을 때 정밀도가 과대평가되는 문제를 보정하는 짧은 문장 페널티(Brevity Penalty, BP)를 적용함

$$\text{BLEU} = \text{BP} \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

  - **ROUGE**(Recall-Oriented Understudy for Gisting Evaluation) (Lin, 2004): 생성된 답변(answer)과 정답 답변(Gold_Answer) 간의 n-gram 단어 중첩을 기반으로 한 재현율 중심 평가지표로, 주로 ROUGE-N과 ROUGE-L이 사용되며, ROUGE-L은 문장의 최장 공통 부분수열(Longest Common Subsequence, LCS)을 기반으로 생성 문장 기준 LCS Precision(P_{LCS})과 참조 문장 기준 LCS Recall(R_{LCS})의 조합으로 측정

$$\text{ROUGE-L} = \frac{(1+\beta^2) \times \text{P}_{\text{LCS}} \times \text{R}_{\text{LCS}}}{\text{P}_{\text{LCS}} + \beta^2 \times \text{R}_{\text{LCS}}}$$

  - **METEOR** (Banerjee & Lavie, 2005): 어휘 수준의 동의어(synonym)와 어형 변화(stemming)를 반영하여 의미적 유사도 측면을 개선한 n-gram 단어 중첩 기반 평가지표이며, 매칭된 단어들의 불연속 구간(chunks) 수에 따라 순서 불일치 패널티를 적용함

$$\text{METEOR} = F_{\text{mean}} \times (1 - \text{Penalty}),$$
$$F_{\text{mean}} = \frac{10 \times \text{Precision} \times \text{Recall}}{\text{Recall} + 9 \times \text{Precision}}$$

- 의미적 유사도(Semantic Similarity) 기반 지표
  - **BERTScore** (Zhang et al., 2019): 사전학습 언어모델(BERT 계열)의 문맥 임베딩(Contextual

Embedding)을 활용하여 생성된 답변(answer)과 정답 답변(Gold_Answer) 간의 의미적 유사도를 측정하는 평가지표로, 두 문장 토큰 임베딩의 정밀도(P_{BERT})와 재현율(R_{BERT})의 조화평균으로 계산됨

$$R_{\text{BERT}} = \frac{1}{|x|}\sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} \cos(\mathbf{x}_i, \hat{\mathbf{x}}_j), \quad P_{\text{BERT}} = \frac{1}{|\hat{x}|}\sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} \cos(\mathbf{x}_i, \hat{\mathbf{x}}_j)$$

$$F_{\text{BERT}} = \frac{2\,P_{\text{BERT}}\,R_{\text{BERT}}}{P_{\text{BERT}} + R_{\text{BERT}}}$$

- 근거 기반성(Faithfulness) 및 적합성(Relevance) 지표
  - **FActScore** (Min et al., 2023): 생성된 답변(answer)을 검증 가능한 최소 단위인 원자적 사실(Atomic Fact)의 집합(A_y)으로 분해하여 평가하며, 개별 답변(y)에 대한 점수($f(y)$)를 생성된 원자적 사실 중 지식 소스(C)에 의해 지지되는(supported) 사실의 비율로 정의하고, 언어 모델(M)에 대한 최종 FActScore는 프롬프트 집합(X)에 대한 기대값으로 계산함

$$f(y) = \frac{1}{|\mathcal{A}_y|}\sum_{a \in \mathcal{A}_y} \mathbb{I}[a \text{ is supported by } \mathcal{C}]$$

$$\text{FACTSCORE}(\mathcal{M}) = \mathbb{E}_{x \in \mathcal{X}}[f(\mathcal{M}_x) | \mathcal{M}_x \text{ responds}]$$

  - **AlignScore** (Zha et al., 2023): 생성된 답변(answer)이 검색된 문서(retrievedDocs)의 내용과 얼마나 의미적으로 정렬(align)되어 있는지를 평가하며, 검색된 문서와 생성된 답변을 청크 단위로 분할한 뒤 각 생성 문장 청크(l_i')가 가장 잘 지지되는 근거 문서 청크(o_i')를 찾아 NLI(Natural Language Inference) 기반 정렬 모델이 산출하는 Aligned 클래스 예측 확률의 최대값을 문장별로 계산하고, 이를 전체 문장에 대해 평균을 내어 최종 AlignScore를 계산함

$$\text{AlignScore}(o, l) = \operatorname*{mean}_j \max_i \text{alignment}(o_i', l_j')$$

라 LLM 기반 통합 평가 (LLM-as-a-Judge)

1) 평가 메커니즘

- 정성적 평가의 자동화 및 고도화
  - 인간과 유사한 추론 능력을 가진 고성능 LLM(GPT-4 등)을 '평가자(Judge)'로 활용하여 단순 통계적 수치로 평가하기 힘든 답변의 논리적 일관성, 문맥 적절성, 사실 관계 등을 심층적으로 평가
  - 단순한 점수 부여를 넘어, 평가의 근거와 개선을 위한 서술형 피드백을 제공하여 평가 결과의 해석 가능성 확보
- 평가 프로세스 구조
  - 입력(Input): 사용자 질문(question), 검색된 문서 집합(retrievedDocs), 생성된 답변(answer), (선택적) 정답 문서와 정답 답변(Ground Truth)
  - 평가 프롬프트(Prompt): 평가 기준(Rubric), 점수 척도(1~5점), 평가 예시(Few-shot)가 포함된 시스템 프롬프트 적용

- 출력(Output): 항목별 정량적 점수 및 점수 부여 사유(Chain-of-Thought) 도출

2) 주요 평가 지표 및 프레임워크

- **ARES**(An Automated RAG Evaluation System) (Saad-Falcon et al., 2024), **RAGAS**(Retrieval-Augmented Generation Assessment Suite) (Es et al., 2024), **SCARF** (Rengo et al., 2025)
  - **Context Relevance**: 사용자가 입력한 질의(question)의 정보 요구에 대해, 검색된 문서(retrievedDocs)가 얼마나 적합한 정보를 담고 있는지 평가하여 검색기의 잡음(Noise) 포함 여부를 판단
  - **Faithfulness**: 생성된 답변(answer)이 검색된 문서(retrievedDocs)의 내용에 충실히 기반하였는지 평가하여 환각(Hallucination) 환각 여부를 판단
  - **Answer Relevance**: 생성된 답변(answer)이 초기 사용자의 질의(question)의 의도를 정확하게 파악하고 적절히 답변했는지 평가하여 질의에 대한 답변으로서의 적절성을 판단

마 RAG 성능평가

- 개요
  - LLM-as-a-Judge 방식(RAGAS)으로 시범 평가를 수행
  - 평가 데이터셋은 전문가(현장 교사)의 검토가 완료되어 정답(Gold_Answer)이 확보된 질문-응답(QA) 100개를 바탕으로 함
  - 입력 데이터는 사용자 질의(question), 검색 문서 집합(retrievedDocs), 생성된 답변(answer), 정답 답변(Gold_Answer), 정답 문서(Gold_Context)를 포함

1) LLM-as-a-Judge(RAGAS) 평가

- 평가 전제
  - 정답 문서(Gold_Context)와 정답 답변(Gold_Answer)을 기준으로 평가를 진행하며, 정답 답변(Gold_Answer)의 정답 여부가 'X'인 경우 수정된 정답 답변(Gold_Answer)을 사용함
  - 검색된 문서(retrievedDocs)는 실제 RAG 검색을 통해 추출된 문서로 생성된 답변(answer)의 기반이 됨
- RAGAS 주요 평가 지표
  - **Context Relevance**: 사용자가 입력한 질의(question)의 정보 요구에 대해, 검색된 문서(retrievedDocs)가 얼마나 적합한 정보를 담고 있는지 평가하여 검색기의 잡음(Noise) 포함 여부를 판단
  - **Faithfulness**: 생성된 답변(answer)이 검색된 문서(retrievedDocs)의 내용에 충실히 기반하였는지 평가하여 환각(Hallucination) 환각 여부를 판단
  - **Answer Relevance**: 생성된 답변(answer)이 초기 사용자의 질의(question)의 의도를 정확하게 파악하고 적절히 답변했는지 평가하여 질의에 대한 답변으로서의 적절성을 판단

〈RAGAS 기반 평가 결과 요약〉

구분	Answer Relevancy	Answer Correctness	Answer Similarity	Faithfulness	Context Precision	Context Recall
점수	0.6804	0.4275	0.4965	0.6346	0.2276	0.8

2) 평가 결과 해석

- 데이터셋의 구조적 특징이 성능에 미치는 영향
  - 정답 문서(Gold_Context), 정답 답변(Gold_Answer)은 한 두 문장 정도의 짧은 단답형으로 이루어져 있으며, 검색 문서(retrievedDocs), 생성 응답(answer)은 양이 방대하면서도 응답이 자세하고 구체적임
  - Ground Truths와 시스템 출력 간의 정보량 및 길이의 불균형이 주요 평가 지표의 왜곡 또는 저하를 유발
- 검색 성능: Needle in the Haystack
  - 정답 문서(Gold_Context)는 한 두 문장 수준으로 매우 짧기 때문에, 시스템이 검색한 방대한 문서 집합(retrievedDocs)에서 정답 문서(Gold_Context)가 차지하는 비율(Context Precision)이 매우 낮게 나타남(0.2276)
  - 검색 결과의 대부분이 정답과 직접 관련이 없는 Noise로 작용하고 있음
  - 반면, 정답 문서(Gold_Context)가 시스템이 검색한 문서(retrievedDocs)에 포함될 확률(Context Recall)은 상대적으로 높게 나옴(0.8)
- 검색 성능: 답변 스타일 불일치
  - 단답형, 단문형으로 이루어진 정답 답변(Gold_Answer)과 길고 자세한 설명으로 이루어진 생성 답변(answer) 사이에 서술 구조 및 정보량 차이가 커서 의미적 유사도(Answer Similarity)와 일치도(Answer Correctness)가 낮게 산출됨 (각 0.4965, 0.4275)
- 답변 품질 및 신뢰성: 정보 과잉으로 인한 희석
  - 생성된 답변(answer)은 필요 이상의 세부 설명을 포함하면서 핵심 내용의 비중이 낮아지고, 이에 따라 질문과의 적합성(Answer Relevancy)이 제한적임(0.6804).
  - 검색된 문서(retrievedDocs) 중 비관련 문서(Noise)의 비중이 높아, 모델이 답변 생성시 근거를 충분히 반영하지 못하거나, 비관련 정보를 응답에 반영함으로써 근거 기반성(Faithfulness)이 약화됨(0.6346).

3) 성능 향상을 위한 개선 방향

- Ground Truth 확장
  - 단답형 정답 답변(Gold_Answer)을 문장형·설명형으로 확장하여 검색 정밀도(Precision)와 근거 기반성(Faithfulness) 향상을 유도
- 평가 지표의 다양화
  - RAG 시스템의 다면적 특성을 평가하기 위해 다양한 평가 지표를 활용

제4장 시사점 및 제언

1절 | 결론 및 시사점

2절 | 후속 연구 제언

1절 결론 및 시사점

1 시스템 개발 측면

가 개인화 학습환경 조성

- 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』을 통해 학생들은 1) 데이터를 수집·분석·시각화 하고 보고서를 작성하는 데이터 기반 과학탐구 전반에서 AI와의 대화형 상호작용을 통해 지원을 받을 수 있고, 2) 학생들의 탐구 산출물에 대한 구체적 피드백을 항목별로 제공받을 수 있다는 점에서 본 시스템 개발은 '과학탐구 맥락에서의 개인화 학습환경 조성'의 노력으로 볼 수 있음. 특히, 데이터 기반 과학탐구와 같이 정답이 없고, 비구조화된 탐구 영역에서 학생들이 즉각적으로 질의응답을 통해 탐구를 이어가고 학생의 개별화된 탐구목적에 따라 과학탐구를 심화할 수 있도록 지원한다는 점에서 의미가 있음. 이것은 생성형 AI가 학생들의 질문에 즉각 응답하고 맥락화된 설명을 제공함으로써 전통적 교수모델에서 제공하기 어려웠던 개별화 지원이 가능해졌다는 선행연구와 맥을 함께한다고 볼 수 있으며, 나아가 공교육에 도입될 경우 학습 격차 해소에 기여할 수 있다는 가능성을 확인하게 해줌(Cotton et al., 2023; Smolansky et al., 2023; Wang et al., 2024).
- 시스템 활용 전 교사들의 기대를 확인하고자 실시한 설문에서 '과학수업 및 학생 지도에서 가장 부담스러운 부분'에 대한 질문에 대해 대다수의 교사가 보고서(과제) 평가 및 피드백에 부담을 느낀다고 답했으며, 개별 학생 맞춤 지도의 어려움을 언급해 교사들이 평가·피드백 및 학생 맞춤 지도에 어려움을 겪고 있는 것으로 확인됨. 이러한 점은 교사들의 AI기반 보고서 튜터링 만족도를 높이는 요인이 되었을 수 있음. 교사들은 보고서 튜터링 서비스를 통해 학생들에게 피드백하기 어려웠던 구체적인 과정지식, 탐구수행 결과, 데이터 해석 결과 등에 피드백이 가능해짐으로써 학생들의 과학탐구 역량 향상에 도움을 주었다고 인식하였기 때문임. 이것은 AI기반 개인화 학습 플랫폼이 제공하는 실시간, 맞춤형 피드백은 교육적 성과를 최적화하는 것을 도울 뿐 아니라 학생의 참여도와 동기를 향상시킨다는 보고들과 관련이 있음(Kasneci et al., 2023; Weng et al., 2024).
- 교사 면담 결과에서도 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』의 중요한 효과로 개별화 학습이 가능해졌다는 공유된 인식이 도출되었는데, 교사들은 시스템을 활용하면서 학생들의 사전지식 수준과 관심사에 맞는 개별화 학습을 긍정적으로 인식함. 이러한 학습 환경 변화는 교수자의 역할을 콘텐츠 전달자에서 촉진자 및 안내자로 전환시키며, 새로운 역할에서의 교수자 역량 강화를 요구함(Bettayeb et al., 2024).
- 다만, 균등한 맞춤형 교육의 실천을 위해 초·중·고 학교급별 성취기준과 언어 활용 수준에 맞게 차별화된 프롬프트의 필요성이 제기되었음. 초등교사들을 위주로 AI의 답변이나 피드백에 초등 수준을 넘어서는 용어가 제시되는 경우가 있다는 의견과 함께 학생들의 연령대에 따라 과도한 칭찬이 학습 분위기에 도움이 되지 않을 수도 있다는 인식이 나타났음. 초등 학생과의 자연스러운 상호작용을 위한 보다 쉬운 언어 사용, 중등 학생에게는 차분한 어조가 활용되도록 개발이 필요하며, 이는 초중등 학교급에 따라 보다 세분화된 프롬프트가 필요함을 의미함.
- 또한 학생의 성취도 수준 차이에 따른 튜터링 차별화가 학생들의 개인화 학습 만족도를 높일 수 있는 설계 방안이 될 수 있음. 학업능력이 높은 학생의 경우 상대적으로 구체적인 AI의 피드백이 학생 성장을 견인할 수 있지만, 성취도가 낮거나 기초 학업 능력이 뒷받침되지 못하는 학생의 경우 텍스트를 읽고 이해하는 것 자체에 심적 부담을 느낄 수 있음. 따라서 학생 수준별로 AI의 반응을 다양한 버전으로 개발하는 것을 고려할 수 있음. 예를 들어, 보고서 튜터링 시 상위권 학생의 경우 학생에게 도전적인 내용을 사전지식에 해당할 과학 개념을 포함하여 적절한 분량으로 제시하고, 하위권 학생의 경우 과학에

대한 친밀감을 느낄 수 있도록 일상생활과 연계한 내용을 쉽고 간략하게 제시할 수 있음.

나 시스템 성능 개선

- 본 연구의 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』을 사용해 본 교사들은 대화형 서비스와 보고서 튜터링 서비스에 모두 높은 사용 의도를 가지고 있었으며, 유용성과 사용 용이성 측면에서 높은 점수를 주었으므로 과학탐구 활성화를 위한 AI의 첫 도입은 안정적으로 이루어졌음을 확인하였음.
- LLM의 학생 활용 연령 제한이 공교육에서의 생성형 AI 활용을 억제하고, 일반적인 포털 검색으로 정보를 찾는 과정에는 시간적, 인지적 노력이 추가적으로 소요되는 반면, 지능형 과학실 ON에 LLM이 연계되어 보다 직관적으로 원하는 정보에 접근할 수 있고, 과학탐구 맥락에서 벗어나지 않고 AI와 상호작용 할 수 있음이 긍정적으로 평가되었음.
- 학생들은 자신이 모르는 개념에 대해 질문하였을 때, 정확한 답변을 받았던 경험과 과학에서 벗어난 질문이거나 학습과 관계없는 질문일 경우 답변을 제한하는 기능으로 인해 본 시스템을 신뢰한다고 응답하였음. 특히, LLM의 문제로 지적되는 환각 현상이 나타나지 않아 교육적으로 활용하는 것에 긍정적인 인식을 주었음. 본 연구에서 LLM의 교육적 활용의 신뢰성 확보를 위해 고려할 수 있는 기술로 제안된 RAG 기술을 적용한 것은 시스템의 안정성 측면에서 바람직한 접근이라고 볼 수 있음(Wang et al., 2024).
- 그러나 모든 평가 영역 중 교사와 학생 모두 신뢰·정확도 영역이 가장 낮은 점수를 기록하였음. 교사의 경우 여전히 4.35점/5점으로 긍정적 인식을 바탕에 두고 있음을 확인하였으나 대화형 서비스의 과학적 답변 내용의 질과 정확성을 개선하기 위한 노력을 통해 시스템의 사용성을 제고할 수 있음. 생성형 AI는 사실에 기반하지 않은 내용이 생성될 수 있다는 가능성을 염두에 두고 교육적 활용을 고민해야 하며, 생성형 AI의 활용이 학생들의 오개념에 영향을 미치는지, 혹은 교육적 적절성에 부합하는지에 대한 비판적 검토가 뒷받침 되어야 함(Mai et al., 2024; Salamin et al., 2023).
- 답변의 정확도, 응답속도 개선을 위해 RAG의 벡터 DB의 자료 질 관리 및 확대가 필요하며, LLM과 RAG 간에 질문 내용이 반복적 순환되지 않도록 합리적인 개발 아키텍쳐를 구축할 필요가 있음. 튜터링이나 대화형 서비스에서 정답과 풀이를 분리하여 두괄식으로 서술함으로써 글 읽기 활동의 선호나 수준이 다른 학생들이 앞부분만 읽고도 핵심 내용을 파악할 수 있도록 개발하는 것도 하나의 방법이 될 수 있음.
- 본 연구에서는 RAG에 대한 성능평가를 위해 'LLM as a judge'의 방법을 적용한 RAGAS 지표를 활용하여 단순 통계적 수치로 평가하기 힘든 답변의 논리적 일관성, 문맥 적절성, 사실 관계 등을 심층적으로 분석하였음(Es et al., 2024). 분석 결과 지능형 과학실 ON 대화형 서비스에서 도출한 정답문서가 시스템이 검색한 문서에 포함될 확률이 상대적으로 높게 도출됨. 다만 생성된 답변이 필요 이상의 세부 설명을 포함하면서 핵심 내용의 비중이 낮아질 수 있으므로 질문과의 적합성을 높이려는 개발 노력이 요구됨. 또한 본 연구의 성능평가에서는 100개의 데이터 셋을 활용하였지만, 보다 큰 데이터 셋을 구성하여 RAGAS외의 다양한 지표로 검토해볼 필요가 있음. LLM의 판단 결과에 대한 신뢰성 제고를 위해 교육전문가와 자동평가 간의 일치도 확인이 중요한 역할을 수행할 것으로 기대함.

다 시스템 성능 확장

- 본 연구의 참여자들을 통해 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』의 확장에 대한 요구가 도출되었으며, 이미지 인식, 음성 인식 및 파일 업로드 등의 멀티모달 측면에서 기능적 확장이 이루어지기를 바람. 학생들의 대다수는 이미 상용 LLM을 활용 중에 있었으며, 기존의 LLM 활용 경험이 본 시스템에 대한 기대를 높이는 기준이 되었을 수 있음. 연구 참여자들은 대화형 서비스를 활용하여 실험 과정 및 측정·관찰 결과 사진을 업로드하여 피드백을 받거나, 타자를 치기 어려운 실험 중 음성으로

지원을 받을 수 있기를 희망하였고, 보다 시각적으로 구조화된 텍스트 제시를 기대하였음.

- 멀티모달 AI의 지원이 학습에 유의미한 효과를 제공함을 입증한 연구들을 통해 이러한 연구참여자의 요구 수용의 필요성을 뒷받침 해줌. AI 로봇 이미지 인식 기술을 활용한 실험 집단의 성취가 통제집단보다 높게 나타난 연구나, 가상현실(VR) 기술과 결합된 AI의 활용이 다국어 학생들의 과학 개념이해를 제고했다는 연구들은 생성형 AI의 다차원적인 지원이 학습을 용이하게 함을 지지해줌(Ding & Cha, 2024; Wang & Fan, 2024).
- 한편, 시스템의 확장 측면에서 융합 탐구를 위해 과학 분야를 넘어서는 질문을 수용하거나, 본 시스템의 강점을 과학 교과 외에서도 활용할 수 있기를 기대하는 학생들의 요구가 발생하기도 하였음. 공교육 차원에서의 생성형 AI 활용에 대한 가이드, 안전성 확보 등에 대한 체계 및 이해관계자의 인식이 조성된 후에 순차적으로 시스템의 기능 확장, 타 시스템간의 연계 등을 고려할 수 있을 것으로 사료됨.

2 교육적 측면

가 교사 교육

- 면담 결과, 본『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』이 교사의 업무경감에 탁월하게 도움을 준 것으로 교사들이 일관되게 인식하고 있었음. 또한 확보된 시간을 학생 지원의 우선순위에 따라 맞춤형 개입을 할 수 있어, 교실 수업 운영의 효율성을 높일 수 있었음을 보고하였음. 교사가 생성형 AI을 활용하여 수업을 운영하게 되면, 기존의 교수학습 맥락과는 다른 개입과 조절이 이루어지게 되므로 교사의 AI기반 과학교육 역량 강화 전략이 필요해짐(Bettayeb et al., 2024). Wang et al. (2024)은 교육에서 LLM의 활용을 기술 중심 관점에서 학생 지원(study assisting), 교사 지원(teach assisting), 적응형 학습(adaptive learning)의 세 가지 핵심 범주로 분류하였는데, LLM이 동시적으로 교사와 학생을 지원하고 있음에 주목해야 함. 맞춤형 학습 준비, 학습자 유형 분석, AI·디지털 도구가 제공하는 여러 특성의 학습자 유형화 결과물을 어떻게 해석하고 활용해야 하는지에 대한 숙제가 교사에게 발생하는 것이므로, 자동화, 개인화 등의 기술적 효과를 교실 수업 맥락에서 어떻게 총체적으로 이해하여 반영할 지에 대한 교사의 고도화된 전략과 스킬이 요구됨(Bettayeb et al., 2024).
- 교사들은 학생들의 질문하기가 학생들의 수준과 관심, 탐구하는 방법에 따라 다름을 확인하였고, 과학탐구 맥락에서 탐구를 심화·지속하고, 질문을 잘하기 위해 질문 하는 법에 대한 교수법 노하우가 공유되어야 할 필요성을 제기함. 과학교사가 과학적 탐색과 발견, 탐구의 절차 계획·실행을 안내해야 할 의무가 있지만 개인화된 학습 환경에서 생성형 AI에게 질문을 잘 하는 법을 안내하는 것에는 다양한 어려움이 존재함. 질문을 잘 하기 위해서는 문해력, 논증적 글쓰기에 대한 기초적 이해가 바탕이 되어야 하고, 체계적 질문을 위한 논리적 사고와 AI을 효율적으로 활용할 수 있는 기술적 이해가 고루 뒷받침되어야 함.
- Marzano와 Simms(2012)는 학생들의 창의적 사고와 깊이있는 학습을 위해 질문연속체(questioning sequence theory) 개념을 통해 모든 영역에서 질문을 효과적으로 하기 위한 절차로 세부 내용, 범주, 정교화, 근거 제시의 절차로 질문을 도출할 것을 제안한 바 있음. 촉진자 및 안내자로서의 교사 역할 강화 및 학생들의 과학탐구 활성화를 위해 과학교육의 접근 방향에 새로운 변화가 필요하며, 교육적 목적의 AI와의 상호작용을 학습 방법의 한 차원으로 인정하고, 체계적인 교수법을 마련해야 할 필요가 있음. 또한 교사의 역할에 대한 인식 변화를 이끌어내고 새로운 교육방법을 교육현장에 안착시키기 위한 교사연수와 프로그램 개발이 병행되어야 할 필요가 있음.

나 학생 교육

- 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』 사용을 통해 학생들의 과학학습정서, 과학관련자아개념, 과학학습동기, 과학관련진로포부, 과학관련태도 등 과학긍정경험 총 5개의 하위 요인 모두 통계적으로 유의미하게 향상된 것으로 나타남. 정량적 수치 외에도 학생들은 면담을 통해 학생 스스로

과학탐구에 주도적으로 참여하였고, 학습 동기가 높아졌다는 인식을 제시하였음. 특히 학생들은 그간 과학탐구 과정에서 소외되었던 학생들도 자신의 수준에 맞는 AI와의 질의응답을 통해 보고서 작성하기 활동에 참여할 수 있었음을 AI 기반 과학탐구 지원 시스템의 효과로 인식하였음. 이것은 학업 수준이 낮은 학생들에게 AI 지원 응용이 학습에 대한 흥미를 높여주고, STEM 분야에서 자신을 더 유능하고 지식이 풍부하며 인식이 높은 개인으로 여기게 한다는 기존 연구의 결과를 지지하는 결과가 도출된 것으로 볼 수 있음(Lee et al., 2023; Xu & Ouyang, 2022). 학생들이 자신의 개념 이해를 점검하는 데에 AI 대화형 서비스를 활용했던 것이 시험 준비에 도움이 되었다는 학생의 응답은 본 과학탐구 지원 시스템이 학생들의 정의적 영역뿐만 아니라 인지적 성취에도 즉각적인 도움을 줄 수 있다는 가능성을 내포하고 있으며, 학생 주도적인 학습 방법 변화를 위한 촉매제가 될 수 있음을 시사함.

- 또한 본 연구에서 AI 기반 보고서 튜터링을 제공한 학생들이 사전 대비 사후에 통계적으로 유의한 성취 향상을 보여주어 보고서 튜터링이 학생의 성장에 의미 있는 영향을 미칠 수 있음을 실증적으로 제시하였음. 학생들이 초·중·고 학창 시절 동안 AI가 제공한 구체적인 수준으로 피드백을 받기 어려운 교육 현실을 고려한다면, AI 보고서 튜터링의 적용이 교사는 부족해지는 한편, 과학탐구 역량 강화에 대한 교육적 요구 기준은 높아지고 있는 교육의 난제를 해결하기 위한 대안이 될 수 있음을 보여줌.
- 다만, 학생들이 본 AI 기반 대화형 서비스를 활용한 질의응답 기록을 분석한 결과는 학생들의 학습 방식의 질적인 개선이 필요함을 나타냄. 질문의 내용을 '정의', '원리', '실천'으로 구분하여 분석한 결과 질문 중 대다수가 과학적 개념 이해를 목적으로 하는 '정의'에 머물러 있었고, 과학적 탐구 실행을 위한 '원리'와 '실천'에 대한 질문은 제한적으로 나타났음. 따라서 학생들의 고차적 사고를 증진하고, 과학탐구의 깊이를 심화할 수 있도록 '원리'와 '실천'의 질문을 위한 개입 프로그램이 마련될 필요가 제기됨. 이것은 앞서 교사 측면의 전문성 강화를 위해 제안된 과학적 질문에 대한 교육과 함께 학생 측면의 질문하기 교육 프로그램이 함께 마련되어야 할 필요성을 뒷받침함. 또한 시스템의 설계 차원에서도 학생의 과학탐구 활성화 전략을 채택할 수 있음. 예를 들어, 학생에게 정답을 말해주는 대신 학생의 이해 상태를 예측하여 학생이 적절한 이해에 도달할 때까지 정답을 알려주지 않고, 바람직한 답에 근접하게 다가가도록 프롬프트를 지원하는 교육적 설계를 제안함.

2절 후속 연구 제언

가 과학탐구 활성화를 위한 맞춤형 교육 설계 기반 마련

- 맞춤형 과학탐구 학습을 위한 이론적 기반이 마련될 필요가 있음. ITS(Intellgent Tutoring System)을 활용한 학습환경과 비 ITS 학습환경에서의 학습 결과를 비교한 연구들을 메타분석한 연구에서는 ITS의 활용 유형과 수준, 교과 영역에 상관없이 유의미한 양의 평균 효과 크기가 도출되었음(Ma et al., 2014). 효과적인 ITS 구성을 위해서 본 시스템 맥락에 맞는 학습자 모델, 도메인 모델, 튜터 모델이 요구됨.
- 학습자 모델을 구축하여 학생의 사전지식과 평가 모델에 맞게 개인화된 피드백을 제시할 수 있음. 또한 학생에게 제공된 대화형 서비스의 응답이나 보고서 튜터링 결과, 과학 콘텐츠 추천 결과를 학생들이 평가하도록 하고(예: 좋아요 등) 이러한 학습자 반응에 대한 데이터 베이스를 별도로 구축하여 학생들의 반응과 성취 수준을 연계한 피드백을 제공할 수도 있음. 이를 위해서는 학습자 분석이 선행되어야 함.
- 그리고 학생은 과학과 관련하여 여러 가지 개념(multiple concept)을 가지고 있을 수 있는데, 틀린 답이 도출될 가능성을 낮추고, 옳은 개념을 이끌어 내는 '과학탐구 지원' 시스템으로서의 맞춤형 전략을 모색하는 연구가 추진될 필요가 있음. 이러한 연구를 위해서는 시스템 메타데이터를 확대하고 지식 맵을 구축하는 등의 노력이 병행되어야 함.

나 생성형 AI 기반 과학탐구 지원을 위한 수업모델 및 교육 프로그램 개발

- 과학교육 맥락에서의 생성형 AI 활용 수업 개선을 위해 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』 활용 과학탐구 수업을 안내하는 수업모델을 개발하고 실제 교실 수업에서의 적용 사례를 안내하는 것이 현장을 지원하는 중요한 지원이 될 수 있음. 생성형 AI의 활용에 대한 현장의 이해와 창의적인 접근을 촉진할 필요가 있으며, AI 활용 윤리와 과학탐구의 맥락을 결합하여 교육 프로그램을 개발한다면 2022 개정 교육과정의 주안점과, AI인재 양성을 위한 정책적 기조에 부합하는 실천적 연구가 될 수 있음.
- 또한, AI가 답을 얻기 위한 도구를 넘어 학생들의 과학적 사고력을 강화하기 위해 어떻게 질문을 유도해야 하는지에 대한 후속 연구결과를 '교사가 어떻게 발문해야 하는지'와 연계하여 수업모델에 포함하는 한편, 교육적 설계와 공학적 설계의 일관성을 위해 시스템 개발의 방향성을 함께 고민해야 함.

다 학습자 이력 추적을 통한 과학탐구 지원 시스템의 중장기 효과 분석

- 지능형 과학실 ON에 학생들이 가입을 하게 되면 가입 이후의 초·중·고 학습 과정에서 산출된 탐구 산출물(데이터 수집·분석 결과, 탐구 보고서)이 포트폴리오에 누적 관리되어 학생들의 자기이해 및 진로 모색에 활용할 수 있도록 지원하고 있음. 또한 해당 학생을 지도하는 교사들은 학생의 포트폴리오에 접근할 수 있기 때문에 학생의 과학탐구 참여 수준과 활동 내용을 파악하여 지도에 활용할 수 있음. 이러한 지능형 과학실 ON의 장점과 AI 기반 과학탐구 지원 시스템 활용의 연계성을 고려하여 학습자 이력 추적을 통한 과학탐구 시스템 활용의 중장기 효과를 분석할 수 있음.
- 학생들의 과학탐구 수준을 질적으로 평가하고, 개선의 방향, 학생 역량과 인식 변화 등의 데이터를 추가적으로 확보하여 『지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템』의 효과를 다면적으로 확인할 수 있음. 이러한 접근은 시스템 개발의 방향성을 제안해줄뿐만 아니라 근거 기반의 과학교육 정책 추진의 강력한 원동력이 될 수 있을 것임. 미래사회가 요구하는 과학분야 인재양성 문제와 인력 수급의 불균형 문제를 고려한다면, 중장기 관점에서 학생들의 변화를 관찰하고, 과학교육의 근본적인 변화를 모색하는 것이 의미있는 시도가 될 수 있음.

제5장 참고문헌 및 부록

1절 | 참고문헌

2절 | 부록

1절 참고문헌

Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2024). Self-rag: Learning to retrieve, generate, and critique through self-reflection.

Ateş, H. (2024). Integrating augmented reality into intelligent tutoring systems to enhance science education outcomes. Education and Information Technologies, 29(17), 22617-22651. https://doi.org/10.1007/s10639-024-12970-y

Azoulay, A. (2023). UNESCO issues urgent call for appropriate use of technology in education. UNESCO.

Baidoo-Anu, David and Owusu Ansah, Leticia. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. http://dx.doi.org/10.2139/ssrn.4337484

Ball, A. (2012). Review of data management lifecycle models. University of Bath.

Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization (pp. 65-72).

Bettayeb, A. M., Abu Talib, M., Sobhe Altayasinah, A. Z., & Dakalbab, F. (2024). Exploring the impact of ChatGPT: Conversational AI in education. Frontiers in Education, 9, 1379796.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

Calzada, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. Libri, 63(2), 123-134.

Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. Portal: Libraries and the Academy, 11(2), 629-657.

Chen, X. (2024). AI-assisted instructional practices facilitating conceptual change in science education. Science Education Research.

Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. Educational Technology Research and Development, 72, 1329-1353. https://doi.org/10.1007/s11423-024-10366-w

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence, 4, 100118.

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology, 32(3), 444-452.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring

academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 60(3), 1–12.

Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A., & De Laat, C. (2012). Addressing big data challenges for scientific data infrastructure. In 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings (pp. 614–617). IEEE.

Deng, Y., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224.

Derakhshan, A., & Ghiasvand, F. (2024). Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of research-active EFL teachers' perceptions. International Journal of Applied Linguistics, 34(4), 1246–1264. https://doi.org/10.1111/ijal.12561

Ding, A. E., & Cha, E. E. (2024). The integration of virtual reality-enhanced multimodal meaning-making improves knowledge acquisition and disciplinary literacy development in science classrooms. Learning & Instruction, 94, 101999.

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., ... & Sui, Z. (2024, November). A survey on in-context learning. In Proceedings of the 2024 conference on empirical methods in natural language processing (pp. 1107–1128).

Es, S., James, J., Anke, L. E., & Schockaert, S. (2024, March). Ragas: Automated evaluation of retrieval augmented generation. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations (pp. 150–158).

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544

Fataliyev, T. K., & Mehdiyev, S. A. (2019). Integration of cyber-physical systems in e-science environment: State-of-the-art, problems and effective solutions. International Journal of Modern Education and Computer Science, 11(9), 35–48.

Faundeen, J., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., ... & Zolly, L. S. (2014). The United States geological survey science data lifecycle model (No. 2013-1265). US Geological Survey.

Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. Science, 331(6018), 705–708.

Gan, A., Yu, H., Zhang, K., Liu, Q., Yan, W., Huang, Z., ... & Hu, G. (2025). Retrieval Augmented Generation Evaluation in the Era of Large Language Models: A Comprehensive Survey. arXiv preprint arXiv:2504.14891.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey.arXiv preprint arXiv:2312.10997, 2(1).

Gao, L., Ma, X., Lin, J., & Callan, J. (2023, July). Precise zero-shot dense retrieval without relevance labels. InProceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1762-1777).

Gebre, E. (2022). Conceptions and perspectives of data literacy in secondary education.

British Journal of Educational Technology, 53(5), 1080–1095.

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

Glean. (n.d.). Glean: Work AI that Works. https://www.glean.com/

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. Teachers College Record, 117(4), 1–22.

Guo, S., et al. (2024). Whether and when could generative AI improve college student learning engagement? Behavioral Sciences, 15(8), 1011.

Hey, T., Tansley, S., & Tolle, K. M. (Eds.). (2009). The fourth paradigm: Data-intensive scientific discovery. Microsoft Research.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Education and Information Technologies, 28(12), 15873–15892.

Jerry Liu. (11. 2022). LlamaIndex. https://github.com/jerryjliu/llama_index.

Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering, 29(10), 2318–2331.

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: A new paradigm for biodiversity studies. BioScience, 59(7), 613–620.

Kim, G., Kim, S., Jeon, B., Park, J., & Kang, J. (2023, December). Tree of clarifications: Answering ambiguous questions with retrieval-augmented large language models. InProceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 996–1009).

Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? Studies in Educational Evaluation, 56, 21–31.

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. CBE—Life Sciences Education, 18(2), es2.

Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R., ... & Dong, X. (2024). Better zero-shot reasoning with role-play prompting. arXiv. Retrieved 2024-05-18, from http://arxiv.org/abs/2308.07702.

Kotsis, K. T. (2024). Integrating ChatGPT into the inquiry-based science curriculum for primary education. European Journal of Education and Pedagogy, 5(6), 95–104. https://doi.org/10.24018/ejedu.2024.5.6.891

Laak, K. J., Abdelghani, R., & Aru, J. (2024). Personalisation is not guaranteed: The challenges of using generative AI for personalised learning. In International Conference on Innovative Technologies and Learning (pp. 40–49). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65881-5_5

Lawglance. (n.d.). lawglance. https://github.com/lawglance/lawglance

Lee, J., An, T., Chu, H.-E., Hong, H.-G., & Martin, S. N. (2023). Improving science conceptual understanding and attitudes in elementary science classes through the development and application of a rule-based AI chatbot. Asia-Pacific Science Education, 9(2), 627-667. https://doi.org/10.1163/23641177-bja10070

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474.

Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts.Transactions of the Association for Computational Linguistics, 12, 157-173.

Ma, X., Gong, Y., He, P., Zhao, H., & Duan, N. (2023, December). Query rewriting in retrieval-augmented large language models. InProceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 5303-5315).

Ma, X., Wang, L., Yang, N., Wei, F., & Lin, J. (2024, July). Fine-tuning llama for multi-stage text retrieval. InProceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2421-2425).

Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. Frontiers in Education, 9, 1328769.

Macina, J., Daheim, N., Chowdhury, S., Sinha, T., Kapur, M., Gurevych, I., & Sachan, M. (2023, December). Mathdial: A dialogue tutoring dataset with rich pedagogical properties grounded in math reasoning problems. In Findings of the Association for Computational Linguistics: EMNLP 2023 (pp. 5602-5621).

Marzano, R. J., & Simms, J. A. (2012). Questioning sequences in the classroom. Solution Tree Press.

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. Trends in Ecology & Evolution, 27(2), 85-93.

Min, S., Krishna, K., Lyu, X., Lewis, M., Yih, W. T., Koh, P., ... & Hajishirzi, H. (2023, December). Factscore: Fine-grained atomic evaluation of factual precision in long form text generation. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 12076-12100).

Mollick, E., & Mollick, L. (2023). Assigning AI: Seven approaches for students, with prompts. arXiv preprint, arXiv:2306.10052. https://arxiv.org/abs/2306.10052

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. Computers and Education Open, 5, 100151.

Mustafee, N., Bessis, N., Taylor, S. J., Hou, J., & Matthew, P. (2020). Co-citation analysis of literature in e-science and e-infrastructures. Concurrency and Computation: Practice and Experience, 32(9), e5620.

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. The National Academies Press.

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. The

National Academies Press.

Nogueira, R., Jiang, Z., Pradeep, R., & Lin, J. (2020, November). Document ranking with a pretrained sequence-to-sequence model. InFindings of the association for computational linguistics: EMNLP 2020 (pp. 708-718).

OECD. (2019). PISA 2018 Assessment and Analytical Framework. OECD Publishing. https://doi.org/10.1787/b25efab8-en

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35, 27730-27744.

Pangrazio, L., & Selwyn, N. (2019). 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data. New Media & Society, 21(2), 419-437.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching and Learning Practice, 20(2), Article 7. https://doi.org/10.53761/1.20.02.07

Pipitone, N., & Alami, G. H. (2024). Legalbench-rag: A benchmark for retrieval-augmented generation in the legal domain. arXiv preprint arXiv:2408.10343.

Post, M. (2018). A call for clarity in reporting BLEU scores. arXiv preprint arXiv:1804.08771.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big Data, 1(1), 51-59.

Qiao, C., Chen, Y., Guo, Q., Wang, X., Luo, H., & Shen, Y. (2024). Understanding science data literacy: A conceptual framework and assessment tool for college students majoring in STEM. International Journal of STEM Education, 11, Article 25.

Qin, J., & D'ignazio, J. (2010). The central role of metadata in a science data literacy course. Journal of Library Metadata, 10(2-3), 188-204.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PmLR.

Rafailov, R., Sharma, A., Mitchell, E., Manning, C. D., Ermon, S., & Finn, C. (2023). Direct preference optimization: Your language model is secretly a reward model. Advances in neural information processing systems, 36, 53728-53741.

Rengo, M., Beadini, S., Alfano, D., & Abbruzzese, R. (2025). A system for comprehensive assessment of rag frameworks. arXiv preprint arXiv:2504.07803.

Saad-Falcon, J., Khattab, O., Potts, C., & Zaharia, M. (2024, June). Ares: An automated evaluation framework for retrieval-augmented generation systems. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 338-354).

Salamin, A.-D., Russo, D., & Rueger, D. (2023). ChatGPT, an excellent liar: How conversational agents' hallucinations impact learning and teaching. In Proceedings

of the 7th International Conference on Teaching, Learning and Education (pp. 10368). Diamond Scientific Publishing. https://www.dpublication.com/abstract-of-7th-ictle/e-10368/

Samala, A. D., et al. (2024). Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects. Education and Information Technologies, 30(3), 1–40.

Sander, I. (2020). What is critical big data literacy and how can it be implemented? Internet Policy Review, 9(2), 1–22.

Schultheis, E. H., & Kjelvik, M. K. (2020). Using messy, authentic data to promote data literacy & reveal the nature of science. The American Biology Teacher, 82(7), 439–446.

Shaffer, J. F., Ferguson, J., & Denaro, K. (2019). Use of the test of scientific literacy skills reveals that fundamental literacy is an important contributor to scientific literacy. CBE—Life Sciences Education, 18(3), ar31.

Shields, M. (2004). Information literacy, statistical literacy, data literacy. IASSIST Quarterly, 28(2/3), 6–11.

Sholikah, L., & Pertiwi, F. N. (2021). Analysis of science literacy ability of junior high school students based on Programme for International Student Assessment (PISA). INSECTA: Integrative Science Education and Teaching Activity Journal, 2(1), 95–104.

Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Amin, M., ... & Natarajan, V. (2025). Toward expert-level medical question answering with large language models.Nature Medicine, 31(3), 943–950.

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In Proceedings of the Tenth ACM Conference on Learning @ Scale (pp. 378–382). Association for Computing Machinery. https://doi.org/10.1145/3573051.3596191

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Computers in Human Behavior, 160, 108386.

Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity? Education Sciences, 14(6), 656. https://doi.org/10.3390/educsci14060656

Tansley, S., & Tolle, K. M. (2009). The fourth paradigm: Data-intensive scientific discovery (Vol. 1). Microsoft Research.

Tariq, M. U. (2025). AI-driven personalized learning-enhancing student engagement and outcomes. In Exploring the Intersection of AI and Human-Computer Interaction (pp. 151–178). IGI Global. https://doi.org/10.4018/979-8-3693-7195-4.ch007

Tasneem, S. (2025, August 7). 4 real-world success stories where GraphRAG beats standa rd RAG. Memgraph. https://memgraph.com/blog/graphrag-vs-standard-rag-success-stories

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. PLoS ONE, 6(6), e21101.

Trantham, P. S., Sikorski, J., de Ayala, R. J., & Doll, B. (2021). An item response theory

and Rasch analysis of the NUDKS: A data literacy scale. Educational Assessment, Evaluation and Accountability, 35, 155–177.

Tufino, E. (2025). NotebookLM: An LLM with RAG for active learning and collaborative tutoring. arXiv preprint arXiv:2504.09720.

Wang, J., & Fan, X. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. Humanities & Social Sciences Communications, 12, 621.

Wang, J., Yi, X., Guo, R., Jin, H., Xu, P., Li, S., ... & Xie, C. (2021, June). Milvus: A purpose-built vector data management system. InProceedings of the 2021 international conference on management of data (pp. 2614–2627).

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., ... & Wen, Q. (2024). Large language models for education: A survey and outlook. arXiv preprint arXiv:2403.18105.

Wang, X., Wang, Z., Gao, X., Zhang, F., Wu, Y., Xu, Z., ... & Huang, X. J. (2024, November). Searching for best practices in retrieval-augmented generation. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing(pp. 17716–17736).

Wang, F., Wan, X., Sun, R., Chen, J., & Arik, S. O. (2025, July). Astute rag: Overcoming imperfect retrieval augmentation and knowledge conflicts for large language models. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 30553–30571).

Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. F. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. Australasian Journal of Educational Technology, 40(6), 37–65.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824–24837.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1–9.

Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. The Journal of Community Informatics, 12(3), 9–26.

Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. International Journal of STEM Education, 9, Article 59. https://doi.org/10.1186/s40594-022-00377-5

Xu, F., Shi, W., & Choi, E. (2023). Recomp: Improving retrieval-augmented lms with compression and selective augmentation. arXiv preprint arXiv:2310.04408.

Yan, S. Q., Gu, J. C., Zhu, Y., & Ling, Z. H. (2024). Corrective retrieval augmented generation.

Yang, H., Zhang, B., Wang, N., Guo, C., Zhang, X., Lin, L., ... & Wang, C. D. (2024). Finrobot: An open-source ai agent platform for financial applications using large language models. arXiv preprint arXiv:2405.14767.

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). A survey on multimodal large language models. National Science Review, 11(12), nwae403.

Zha, Y., Yang, Y., Li, R., & Hu, Z. (2023). AlignScore: Evaluating factual consistency with a unified alignment function. arXiv preprint arXiv:2305.16739.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. arXiv preprint arXiv:1904.09675.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. arXiv preprint arXiv:1904.09675.

Zhang, J., Huang, J., Jin, S., & Lu, S. (2024). Vision-language models for vision tasks: A survey. IEEE transactions on pattern analysis and machine intelligence, 46(8), 5625-5644.

Zhang, T., Patil, S. G., Jain, N., Shen, S., Zaharia, M., Stoica, I., & Gonzalez, J. E. (2024). Raft: Adapting language model to domain specific rag. arXiv preprint arXiv:2403.10131.

Zhao, S., Jin, Z., Li, S., & Gao, J. (2025). FinRAGBench-V: A Benchmark for Multimodal RAG with Visual Citation in the Financial Domain. arXiv preprint arXiv:2505.17471.

Zhao, X., Liu, S., Yang, S. Y., & Miao, C. (2025, April). Medrag: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot. In Proceedings of the ACM on Web Conference 2025 (pp. 4442-4457).

Zheng, L., Chiang, W. L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., ... & Stoica, I. (2023). Judging llm-as-a-judge with mt-bench and chatbot arena. Advances in neural information processing systems, 36, 46595-46623.

Zhuang, S., & Zuccon, G. (2021, July). TILDE: Term independent likelihood moDEl for passage re-ranking. InProceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1483-1492).

2절 부록

1 설문지

가 교사용

1) 사전설문지

안녕하세요. 지능형 과학실 ON입니다.

'AI 기반 「지능형 과학실 ON」 과학탐구 지원 시스템 학교 현장 적용'을 위한 사용성 검토 실시 전, 해당 서비스에 대한 기대 및 AI 서비스 인식 및 경험에 대하여 사전 설문조사를 진행합니다.

이번 설문조사의 목적은
AI 기반 지능형 과학실 ON 과학탐구 지원 시스템 적용을 위한 연구 자료로만 활용됩니다.

□ 참여자 정보

1. 선생님의 성함을 입력해 주세요.(필수답변)

2. 선생님의 소속 학교 지역을 선택해 주세요. (필수답변)

○ 서울 ○ 부산 ○ 대구 ○ 인천 ○ 광주 ○ 대전 ○ 울산 ○ 세종 ○ 경기 ○ 강원 ○ 충북 ○ 충남
○ 전북 ○ 전남 ○ 경북 ○ 경남 ○ 제주

3. 선생님 소속 학교의 학교급을 선택해 주세요 (필수답변)

① 초등학교
② 중학교
③ 고등학교

4. 선생님의 소속 학교명을 입력해 주세요. * 공식 명칭으로 작성해주세요.(필수답변)

창의중(X), 00사대부속고(X), 서울창의중학교(O)

5. 선생님의 이메일 주소를 입력해 주세요

*사용성 검토 안내를 위한 선생님의 메일 주소를 입력해 주세요. (필수답변)

6. 선생님의 휴대폰 번호를 입력해 주세요.

*사용성 검토 안내를 교사님의 정확한 휴대폰 번호를 입력해 주세요. (필수답변)

7. 사용성 검토 참여 및 혜택 관리를 위한 개인정보 수집에 동의하십니까? (필수답변)

사용성 검토 참여 및 혜택을 위한 개인정보 수집 이용 동의

사용성 검토를 위해 귀사가 본인의 개인정보를 수집 이용하고자 하는 경우, 개인정보보호법 제15조, 제22조에 따라 동의를 얻어야 합니다. 이에, 이벤트 응모 및 경품 지급을 위하여 아래와 같이 개인정보를 수집 이용 및 제공하고자 합니다. 내용을 자세히 읽어 보신 후 동의 여부를 결정하여 주시기를 바랍니다.

[개인정보 수집 및 이용 동의]

수집기관	한국과학창의재단, ㈜인튜브
수집 및 이용 목적	AI 기반 지능형 과학실 ON 과학탐구 지원 시스템 사용성 검토 운영 및 참여자 경품 발송
수집하는 항목	**이름, 휴대전화 번호(경품 발송용),** 사용성 검토 참여 기록
보유기간	개인정보는 사용성 검토 후 참여자 경품 수령이 완료되면 즉시 파기됨
동의를 거부할 권리	귀하는 개인정보 제공 동의를 거부할 권리가 있으며, 동의를 거부하게 되면 경품 지급이 제한될 수 있음

① 예, 동의합니다.
② 아니오. 동의하지 않습니다.

☐ AI 경험

8. AI 서비스(예: ChatGPT, 문서 요약 AI 등)를 사용해 본 경험이 있습니까? 필수답변)

① 예
② 아니오

9. AI 서비스를 얼마나 자주 활용하셨나요? (필수답변)

① 없다.
② 1~2회
③ 1주일에 1~2회
④ 1주일에 3~4회
⑤ 매일

10. 실제 사용해 본 AI 서비스를 선택해 주세요. (팔수답변) (중복선택 가능)

① 없다.
② Chatgpt(챗지피티)
③ ClovaX (클로바X)
④ Gemini (제미나이)
⑤ 퍼블렉시티
⑥ 기타 ( )

☐ 사전 설문 문항

11. 과학수업 및 학생 지도에서 가장 부담스러운 부분은 무엇입니까?
(팔수답변) (중복선택 가능)

① 자료 검색

② 수업 자료 제작
③ 개별 학생 맞춤 지도
④ 보고서(과제) 평가 및 피드백
⑤ 기타 ( )

12. 사용성 검토 전, 지능형 과학실 ON "대화형 서비스(AI 챗봇)"에 대해 기대하는 정도를 선택해 주세요. (필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
12-1. 학생의 "과학탐구 역량 강화"에 효과적일 것이다.	①	②	③	④	⑤
12-2. 학생의 "과학 기초학력 향상"에 효과적일 것이다.	①	②	③	④	⑤
12-3. 학생의 "과학 개념의 깊이 있는 이해"에 효과적일 것이다.	①	②	③	④	⑤
12-4. 학생의 "학습 흥미와 동기부여"에 효과적일 것이다.	①	②	③	④	⑤

13. 사용성 검토 전, 지능형 과학실 ON "AI 보고서 튜터링"에 대해 기대하는 정도를 선택해 주세요. (필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
13-1. 학생의 "과학탐구 역량 강화"에 효과적일 것이다.	①	②	③	④	⑤
13-2. 학생의 "과학 기초학력 향상"에 효과적일 것이다.	①	②	③	④	⑤
13-3. 학생의 "과학 개념의 깊이 있는 이해"에 효과적일 것이다.	①	②	③	④	⑤
13-4. 학생의 "학습 흥미와 동기부여"에 효과적일 것이다.	①	②	③	④	⑤

14. 사용성 검토 전, 「 지능형 과학실 ON 」 "대화형 서비스(AI 챗봇)"에 대해 실제 수업 현장에서 기대하는 점이나 바라는 점이 있다면 자유롭게 입력해 주세요.

15. 사용성 검토 전, 「 지능형 과학실 ON 」 "AI 보고서 튜터링"에 대해 실제 수업 현장에서 기대하는 점이나 바라는 점이 있다면 자유롭게 입력해 주세요.

2) 사후설문지

안녕하세요. 지능형 과학실 ON입니다.

'AI 기반 「지능형 과학실 ON」 과학탐구 지원 시스템 학교 현장 적용'을 위한 사용성 검토 설문조사를 진행합니다.

이번 설문조사의 목적은 AI 기반 지능형 과학실 ON 과학탐구 지원 시스템 학교 현장 적용을 위한 연구 자료로만 활용됩니다.

※ 지능형 과학실 ON의 대화형 서비스(AI챗봇)와 AI 보고서 튜터링을 실제 수업에 활용하여 사용성 검토 후 결과보고서 제출과 설문조사를 진행 부탁드립니다.

[사용성 검토 진행 순서]
1. 지능형 과학실 ON AI활용 과학탐구 수업계획 수립 및 제출
1. 지능형 과학실 ON 탐구수업 전 사전 설문조사 참여 완료 (교사/학생)
2. 지능형 과학실 ON 탐구수업을 통한 AI 시스템 사용성 검토 진행 (교사/학생)
3. 사용성 검토 후 보고서 제출 (교사)
4. 사용성 검토 설문조사 참여 완료 (교사/학생)

□ 참여자 정보

1. 선생님의 성함을 입력해 주세요.(필수답변)

2. 선생님의 소속 학교 지역을 선택해 주세요. (필수답변)

○ 서울 ○ 부산 ○ 대구 ○ 인천 ○ 광주 ○ 대전 ○ 울산 ○ 세종 ○ 경기 ○ 강원 ○ 충북 ○ 충남
○ 전북 ○ 전남 ○ 경북 ○ 경남 ○ 제주

3. 선생님 소속 학교의 학교급을 선택해 주세요 (필수답변)

① 초등학교
② 중학교
③ 고등학교

4. 선생님의 소속 학교명을 입력해 주세요. * 공식 명칭으로 작성해 주세요.(필수답변)

창의중(X), 00사대부속고(X), 서울창의중학교(O)

5. 선생님의 이메일 주소를 입력해 주세요.

*****사용성 검토 안내를 위한 선생님의 메일 주소를 입력해 주세요. (필수답변)

6. 선생님의 휴대폰 번호를 입력해 주세요.

*****사용성 검토 안내를 교사님의 정확한 휴대폰 번호를 입력해 주세요. (필수답변)

7. 사용성 검토 참여 및 혜택 관리를 위한 개인정보 수집에 동의하십니까? (필수답변)

사용성 검토 참여 및 혜택을 위한 개인정보 수집 이용 동의

사용성 검토를 위해 귀사가 본인의 개인정보를 수집 이용하고자 하는 경우, 개인정보보호법 제15조, 제22조에 따라 동의를 얻어야 합니다. 이에, 이벤트 응모 및 경품 지급을 위하여 아래와 같이 개인정보를 수집 이용 및 제공하고자 합니다. 내용을 자세히 읽어 보신 후 동의 여부를 결정하여 주시기를 바랍니다.

[개인정보 수집 및 이용 동의]

수집기관	한국과학창의재단, ㈜인튜브
수집 및 이용 목적	AI 기반 지능형 과학실 ON 과학탐구 지원 시스템 사용성 검토 운영 및 참여자 경품 발송
수집하는 항목	**이름, 휴대전화 번호(경품 발송용),** 사용성 검토 참여 기록
보유기간	개인정보는 사용성 검토 후 참여자 경품 수령이 완료되면 즉시 파기됨
동의를 거부할 권리	귀하는 개인정보 제공 동의를 거부할 권리가 있으며, 동의를 거부하게 되면 경품 지급이 제한될 수 있음

① 예, 동의합니다.
② 아니오. 동의하지 않습니다.

□ 사용성 검토 설문 문항

〈유용성〉

8. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 유용성에 대해 각 항목별 응답해 주세요. (필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
8-1. “대화형 서비스(AI 챗봇)”는 학생의 학습 성과 향상에 도움이 된다.	①	②	③	④	⑤
8-2. “대화형 서비스(AI 챗봇)”는 학생의 자기주도 학습에 도움이 된다.	①	②	③	④	⑤
8-3. “대화형 서비스(AI 챗봇)”는 수업 중 학생의 질문 해결에 효과적이다.	①	②	③	④	⑤
8-4. “대화형 서비스(AI 챗봇)”는 탐구수업 운영 효율을 높이고 학생 참여를 촉진한다.	①	②	③	④	⑤
8-5. “AI 보고서 튜터링”은 학생의 과학탐구 보고서 작성 능력 향상에 도움이 된다.	①	②	③	④	⑤
8-6. “AI 보고서 튜터링”을 사용하면 학생의 과학탐구 보고서를 더 빠르게 평가할 수 있다.	①	②	③	④	⑤
8-7. “AI 보고서 튜터링”을 통해 학생의 과학탐구 보고서의 품질을 향상시킬 수 있다.	①	②	③	④	⑤
8-8. “AI 보고서 튜터링”을 사용함으로써 과정중심 평가를 실현할 수 있다.	①	②	③	④	⑤

〈사용 용이성〉

9. AI 기반 「 지능형 과학실 ON 」 과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 사용 용이성에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
9-1. “대화형 서비스(AI 챗봇)”를 사용하는 것은 나에게 쉬운 일이다.	①	②	③	④	⑤
9-2. “대화형 서비스(AI 챗봇)” 사용법을 배우는 데 시간이 오래 걸리지 않았다.	①	②	③	④	⑤
9-3. “대화형 서비스(AI 챗봇)”는 인터페이스가 직관적이라 사용하는 데 불편함이 없다.	①	②	③	④	⑤
9-4. 도움 없이도 “대화형 서비스(AI 챗봇)”를 잘 사용할 수 있다.	①	②	③	④	⑤
9-5. “AI 보고서 튜터링”을 사용하는 것은 나에게 쉬운 일이다.	①	②	③	④	⑤
9-6. “AI 보고서 튜터링” 사용법을 배우는 데 시간이 오래 걸리지 않았다.	①	②	③	④	⑤
9-7. “AI 보고서 튜터링”은 인터페이스가 직관적이라 사용하는 데 불편함이 없다.	①	②	③	④	⑤
9-8. 도움 없이도 “AI 보고서 튜터링”을 잘 사용할 수 있다.	①	②	③	④	⑤

〈사용 의도〉

10. AI 기반 「 지능형 과학실 ON 」 과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 사용 의도에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
10-1. “대화형 서비스(AI 챗봇)”가 지능형 과학실 ON에 상용화된다면 이 서비스를 지속적으로 사용할 의향이 있다.	①	②	③	④	⑤
10-2. “대화형 서비스(AI 챗봇)”가 지능형 과학실 ON에 상용화된다면 이 서비스를 다른 사람에게 추천하고 싶다.	①	②	③	④	⑤
10-3. “대화형 서비스(AI 챗봇)”는 향후 탐구수업 활동에 있어 필수적인 도구가 될 수 있다고 생각한다.	①	②	③	④	⑤
10-4. “AI 보고서 튜터링”이 지능형 과학실 ON에 상용화된다면 이 서비스를 지속적으로 사용할 의향이 있다.	①	②	③	④	⑤
10-5. “AI 보고서 튜터링”이 지능형 과학실 ON에 상용화된다면 이 서비스를 다른 사람에게 추천하고 싶다.	①	②	③	④	⑤

10-6. “AI 보고서 튜터링”은 향후 과학탐구 보고서 작성 및 평가에 있어 필수적인 도구가 될 수 있다고 생각한다.	①	②	③	④	⑤

〈사용 태도〉

11. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 사용 태도에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
11-1. “대화형 서비스(AI 챗봇)”를 사용하는 것은 긍정적인 경험이었다.	①	②	③	④	⑤
11-2. 나는 “대화형 서비스(AI 챗봇)” 사용에 대해 전반적으로 만족한다.	①	②	③	④	⑤
11-3. 지능형 과학실 ON에서의 “대화형 서비스(AI 챗봇)”는 교육적으로 유익하다고 생각한다.	①	②	③	④	⑤
11-4. “AI 보고서 튜터링”을 사용하는 것은 긍정적인 경험이었다.	①	②	③	④	⑤
11-5. 나는 “AI 보고서 튜터링” 사용에 대해 전반적으로 만족한다.	①	②	③	④	⑤
11-6. 지능형 과학실 ON에서의 “AI 보고서 튜터링”은 교육적으로 유익하다고 생각한다.	①	②	③	④	⑤

〈신뢰·정확도〉

12. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 신뢰 및 정확도에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
12-1. 지능형 과학실 ON에서의 “대화형 서비스(AI 챗봇)”를 신뢰할 수 있다고 느낀다.	①	②	③	④	⑤
12-2. “대화형 서비스(AI 챗봇)”의 답변이 정확하다고 생각한다.	①	②	③	④	⑤
12-3. 지능형 과학실 ON에서의 “AI 보고서 튜터링”을 신뢰할 수 있다고 느낀다.	①	②	③	④	⑤
12-4. “AI 보고서 튜터링” 결과값이 정확하다고 생각한다.	①	②	③	④	⑤

〈학습효과 및 교육적 가치〉

13. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 학습효과 및 교육적 가치에 대해 각 항목별 응답해 주세요.

(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
13-1. “대화형 서비스(AI 챗봇)”는 학생의 “과학탐구 역량 강화”에 효과적이다.	①	②	③	④	⑤
13-2. “대화형 서비스(AI 챗봇)”는 학생의 “과학 기초학력 향상”에 효과적이다.	①	②	③	④	⑤
13-3. “대화형 서비스(AI 챗봇)”는 학생의 “과학 개념의 깊이 있는 이해”에 효과적이다.	①	②	③	④	⑤
13-4. “대화형 서비스(AI 챗봇)”는 학생의 “학습 흥미와 동기부여”에 효과적이다.	①	②	③	④	⑤
13-5. “AI 보고서 튜터링”은 학생의 “과학탐구 역량 강화”에 효과적이다.	①	②	③	④	⑤
13-6. “AI 보고서 튜터링”은 학생의 “과학 기초학력 향상”에 효과적이다.	①	②	③	④	⑤
13-7. “AI 보고서 튜터링”은 학생의 “과학 개념의 깊이 있는 이해”에 효과적이다.	①	②	③	④	⑤
13-8. “AI 보고서 튜터링”은 학생의 “학습 흥미와 동기부여”에 효과적이다.	①	②	③	④	⑤

14. 「 지능형 과학실 ON 」“대화형 서비스(AI 챗봇)”에서 가장 유용했던 기능 또는 장점이 있다면 자유롭게 입력해 주세요.

15. 「 지능형 과학실 ON 」“대화형 서비스(AI 챗봇)”에서 개선되었으면 하는 점이 있다면 자유롭게 입력해 주세요.

16. 「 지능형 과학실 ON 」“AI 보고서 튜터링”에서 가장 유용했던 기능 또는 장점이 있다면 자유롭게 입력해 주세요.

17. 「 지능형 과학실 ON 」 “AI 보고서 튜터링”에서 개선되었으면 하는 점이 있다면 자유롭게 입력해 주세요.

나 학생용

1) 사전설문지

안녕하세요. 지능형 과학실 ON입니다.

'AI 기반 「지능형 과학실 ON」 과학탐구 지원 시스템 학교 현장 적용'을 위한 사용성 검토 실시 전 과학 긍정경험에 대한 사전 설문조사를 진행합니다.

이번 설문조사의 목적은 과학 공부와 관련된 긍정적인 경험을 늘리기 위한 방법을 찾는데 도움이 되는 정보를 얻으려는 것입니다. 학생의 개인정보나 모든 응답 내용은 공개되지 않으며. AI 기반 지능형 과학실 ON 과학탐구 지원 시스템 적용을 위한 연구 자료로만 활용됩니다.

- 이 설문에서는 이번 학기(3월부터 현재까지) 과학수업 시간(온라인 과학 수업 포함)의 경험과 관련된 여러분의 생각과 느낌에 대해 물어볼 것입니다.
- 각 질문을 주의 깊게 읽고, 자신의 생각과 가장 맞는 것을 골라주시기 바랍니다.
- 이 설문에는 '맞는 답'이나 '틀린 답'이 없으므로, 여러분의 생각이나 느낌을 가장 잘 나타내는 것에 표시하면 됩니다.
- 설문의 질문 내용이 잘 이해되지 않으면 선생님께 여쭈어봐도 됩니다.

□ 참여자 정보 [공통]

1. 법적 성별을 선택해 주세요.

① 남 ② 여

2. 이름을 입력해 주세요.

3. 거주지역을 선택해 주세요.

○ 서울 ○ 부산 ○ 대구 ○ 인천 ○ 광주 ○ 대전 ○ 울산 ○ 세종 ○ 경기 ○ 강원 ○ 충북 ○ 충남
○ 전북 ○ 전남 ○ 경북 ○ 경남 ○ 제주 ○ 해외

4. 현재 소속 중인 학교 이름을 입력해 주세요. * 공식 이름으로 작성해 주세요. (필수답변)

창의중(X), 00사대부속고(X), 서울창의중학교(O)

5. 학교급을 선택해 주세요.

① 초등학교
② 중학교
③ 고등학교

6. 학년을 선택해 주세요.

① 1학년
② 2학년
③ 3학년
④ 4학년

⑤ 5학년
⑥ 6학년

□ 설문 문항: 과학 긍정경험 지표 검사(사전)

〈과학학습정서〉

7. 이번 학기(3월부터 현재까지) 동안의 과학수업 경험에 대한 질문입니다. 이번 학기(3월부터 현재까지) 과학수업(온라인 수업 포함)을 생각해볼 때, 과학을 배울 때 느낀 나의 감정과 관련된 아래의 각 내용에 대해 얼마나 자주 경험하였습니까?

문항	전혀 없음	가끔	자주	매우 자주
7-1. 나는 과학 수업이 즐거웠다.	①	②	③	④
7-2. 나는 과학 수업이 만족스러웠다.	①	②	③	④
7-3. 나는 과학 수업이 재미있었다.	①	②	③	④
7-4. 나는 과학 수업이 지루하였다.	①	②	③	④
7-5. 나는 과학 수업이 짜증나거나 귀찮았다.	①	②	③	④
7-6. 나는 과학 수업 시간에 불안하거나 초조하였다.	①	②	③	④

〈과학관련 자아개념〉

8. 이번 학기(3월부터 현재까지) 과학수업 시간의 경험에 대한 질문입니다. 이번 학기(3월부터 현재까지) 과학수업(온라인 수업 포함)을 생각해볼 때, 과학수업에서 자신에게 느낀 점과 관련된 아래의 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
8-1. 나는 과학 수업 시간에 주어진 과제 및 활동을 잘 해결할 수 있다.	①	②	③	④
8-2. 나에게 과학 공부는 쉬운 일이다.	①	②	③	④
8-3. 과학은 내가 잘하는 과목 중 하나이다.	①	②	③	④
8-4. 나는 과학 수업 시간에 선생님과 친구들로부터 인정받고 있다.	①	②	③	④
8-5. 과학 수업을 통해 나는 스스로를 필요한 사람이라고 느낀다.	①	②	③	④
8-6. 과학 수업은 내가 스스로에게 만족할 수 있게 해준다.	①	②	③	④

〈과학학습동기〉

9. 이번 학기(3월부터 현재까지) 과학수업 시간의 경험에 대한 질문입니다. 이번 학기(3월부터 현재까지) 과학수업(온라인 수업 포함)을 생각해볼 때, 과학수업에 대한 여러분의 태도 및 생각과 관련된 아래의 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
9-1. 나는 과학 수업 시간에 최선을 다하려고 노력한다.	①	②	③	④

9-2. 나는 과학 공부에 많은 시간을 들인다.	①	②	③	④
9-3. 나는 과학 수업 시간에 적극적으로 참여한다.	①	②	③	④
9-4. 나는 과학 수업 시간에 특히 질문을 많이 한다. (온라인 수업 시 댓글, 커뮤니티 활동 포함)	①	②	③	④
9-5. 나는 과학 수업 시간에 딴 짓을 하지 않고 집중한다.	①	②	③	④
9-6. 나는 과학 수업 시간에 종종 수업과 관련이 없는 딴 생각을 한다.	①	②	③	④
9-7. 과학 수업 시간에 배운 내용은 나와 관련이 있다.	①	②	③	④
9-8. 과학 수업 시간에 배운 내용은 일상생활에서 사용할 수 있다.	①	②	③	④
9-9. 나에게는 과학 수업 시간에 배운 내용을 제대로 이해하는 것이 중요하다.	①	②	③	④
9-10. 나에게는 과학 수업 시간에 주어진 과제나 활동을 성공적으로 마치는 것이 중요하다.	①	②	③	④

〈과학관련 진로포부〉

10. 이번 학기(3월부터 현재까지) 과학수업 시간의 경험에 대한 질문입니다. 이번 학기(3월부터 현재까지) 과학수업(온라인 수업 포함)을 생각해볼 때, 과학 관련 직업 희망과 관련된 아래의 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
10-1. 과학 관련 직업에 대하여 알게 되었다.	①	②	③	④
10-2. 나는 과학 관련 직업이 나에게 스스로 더 배우고 발전할 수 있는 기회를 준다고 생각한다.	①	②	③	④
10-3. 과학 관련 직업은 우리 사회에 많은 영향을 끼친다.	①	②	③	④
10-4. 과학 관련 직업에 흥미를 가지게 되었다.	①	②	③	④
10-5. 미래에 과학과 관련된 직업을 가지고 싶다.	①	②	③	④

〈과학관련 태도〉

11. 이번 학기(3월부터 현재까지) 과학수업 시간의 경험에 대한 질문입니다. 이번 학기(3월부터 현재까지) 과학수업(온라인 수업 포함)을 생각해볼 때, 과학에 대한 태도와 관련된 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
11-1. 과학은 이 세상을 더욱 살기 좋은 곳으로 만드는 데 도움이 된다.	①	②	③	④
11-2. 과학은 공부할 만한 가치가 있다.	①	②	③	④
11-3. 과학은 학교를 졸업한 후에도 쓸모가 있다.	①	②	③	④

11-4. 과학의 발전은 환경, 기술 및 사회의 발전과 영향을 주고받는다.	①	②	③	④
11-5. 학교에서 과학 수업 시간이 늘어나는 것은 바람직하다.	①	②	③	④
11-6. 과학자는 합리적으로 생각하고 판단한다.	①	②	③	④
11-7. 나는 과학에 대하여 더 알고 싶다.	①	②	③	④
11-8. 나는 과학과 관련된 경험(독서, 견학, 과학 영상 시청 등)을 좋아한다.	①	②	③	④

□ AI 경험

12. AI 서비스(예: ChatGPT, 문서 요약 AI 등)를 사용해 본 경험이 있습니까? 필수답변)

① 예
② 아니오

13. AI 서비스를 얼마나 자주 활용하셨나요? (필수답변)

① 없다.
② 1~2회
③ 1주일에 1~2회
④ 1주일에 3~4회
⑤ 매일

14. 실제 사용해 본 AI 서비스를 선택해 주세요. (필수답변) (중복선택 가능)

① 없다.
② Chatgpt(챗지피티)
③ ClovaX (클로바X)
④ Gemini (제미나이)
⑤ 퍼블렉시티
⑥ 기타 ( )

2) 사후설문지

안녕하세요. 지능형 과학실 ON입니다.

'AI 기반 「지능형 과학실 ON」 과학탐구 지원 시스템 학교 현장 적용'을 위한 사용성 검토 실시 후, 사용성 검토 및 과학 긍정경험에 대한 사후 설문조사를 진행합니다.

이번 설문조사의 목적은 과학 공부와 관련된 긍정적인 경험을 늘리기 위한 방법을 찾는데 도움이 되는 정보를 얻으려는 것입니다. 학생의 개인정보나 모든 응답 내용은 공개되지 않으며. AI 기반 지능형 과학실 ON 과학탐구 지원 시스템 적용을 위한 연구 자료로만 활용됩니다.

◦ 이 설문에서는 사전 설문조사 참여 및 사용성 검토 진행 이후부터 과학수업 시간(온라인 과학 수업 포함)의 경험과 관련된 여러분의 생각과 느낌에 대해 물어볼 것입니다.
◦ 각 질문을 주의 깊게 읽고, 자신의 생각과 가장 맞는 것을 골라주시기 바랍니다.
◦ 이 설문에는 '맞는 답'이나 '틀린 답'이 없으므로, 여러분의 생각이나 느낌을 가장 잘 나타내는 것에 표시하면 됩니다.
◦ 설문의 질문 내용이 잘 이해되지 않으면 선생님께 여쭈어봐도 됩니다.

※ 지능형 과학실 ON의 대화형 서비스와 AI 보고서 튜터링을 실제 수업에 활용하여 사용성 검토 후 설문조사를 진행 부탁드립니다.

[사용성 검토 진행 순서]
1. 지능형 과학실 ON 탐구수업 전 사전 설문조사 참여 완료
2. 선생님의 지도하에 지능형 과학실 ON 탐구수업을 통한 AI 시스템 사용성 검토 진행
3. 사용성 검토 설문조사 참여 완료

□ 참여자 정보 [공통]

1. 법적 성별을 선택해 주세요.

① 남 ② 여

2. 이름을 입력해 주세요.

3. 거주지역을 선택해 주세요.

○ 서울 ○ 부산 ○ 대구 ○ 인천 ○ 광주 ○ 대전 ○ 울산 ○ 세종 ○ 경기 ○ 강원 ○ 충북 ○ 충남
○ 전북 ○ 전남 ○ 경북 ○ 경남 ○ 제주 ○ 해외

4. 현재 소속 중인 학교 이름을 입력해 주세요. * 공식 이름으로 작성해 주세요. (필수답변)

창의중(X), 00사대부속고(X), 서울창의중학교(O)

5. 학교급을 선택해 주세요.

① 초등학교
② 중학교
③ 고등학교

6. 학년을 선택해 주세요.

① 1학년
② 2학년
③ 3학년
④ 4학년
⑤ 5학년
⑥ 6학년

□ 사후 설문 문항: 과학 긍정경험 지표 검사(사후)

〈과학학습정서〉

7. 사용성 검토 진행 이후부터 지금까지의 과학수업 경험에 대한 질문입니다. 사용성 검토 진행 이후부터 지금까지 과학수업(온라인 수업 포함)을 생각해볼 때, 과학을 배울 때 느낀 나의 감정과 관련된 아래의 각 내용에 대해 얼마나 자주 경험하였습니까?

문항	전혀 없음	가끔	자주	매우 자주
7-1. 나는 과학 수업이 즐거웠다.	①	②	③	④
7-2. 나는 과학 수업이 만족스러웠다.	①	②	③	④
7-3. 나는 과학 수업이 재미있었다.	①	②	③	④
7-4. 나는 과학 수업이 지루하였다.	①	②	③	④
7-5. 나는 과학 수업이 짜증나거나 귀찮았다.	①	②	③	④
7-6. 나는 과학 수업 시간에 불안하거나 초조하였다.	①	②	③	④

〈과학관련 자아개념〉

8. 사용성 검토 진행 이후부터 지금까지의 과학수업 시간의 경험에 대한 질문입니다. 사용성 검토 진행 이후부터 지금까지 과학수업(온라인 수업 포함)을 생각해볼 때, 과학수업에서 자신에게 느낀 점과 관련된 아래의 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
8-1. 나는 과학 수업 시간에 주어진 과제 및 활동을 잘 해결할 수 있다.	①	②	③	④
8-2. 나에게 과학 공부는 쉬운 일이다.	①	②	③	④
8-3. 과학은 내가 잘하는 과목 중 하나이다.	①	②	③	④
8-4. 나는 과학 수업 시간에 선생님과 친구들로부터 인정받고 있다.	①	②	③	④
8-5. 과학 수업을 통해 나는 스스로를 필요한 사람이라고 느낀다.	①	②	③	④
8-6. 과학 수업은 내가 스스로에게 만족할 수 있게 해준다.	①	②	③	④

〈과학학습동기〉

9. 사용성 검토 진행 이후부터 지금까지의 과학수업 시간의 경험에 대한 질문입니다. 사용성 검토 진행 이후부터 지금까지 과학수업(온라인 수업 포함)을 생각해볼 때, 과학수업에 대한 여러분의 태도 및 생각과 관련된 아래의

각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
9-1. 나는 과학 수업 시간에 최선을 다하려고 노력한다.	①	②	③	④
9-2. 나는 과학 공부에 많은 시간을 들인다.	①	②	③	④
9-3. 나는 과학 수업 시간에 적극적으로 참여한다.	①	②	③	④
9-4. 나는 과학 수업 시간에 특히 질문을 많이 한다. (온라인 수업 시 댓글, 커뮤니티 활동 포함)	①	②	③	④
9-5. 나는 과학 수업 시간에 딴 짓을 하지 않고 집중한다.	①	②	③	④
9-6. 나는 과학 수업 시간에 종종 수업과 관련이 없는 딴 생각을 한다.	①	②	③	④
9-7. 과학 수업 시간에 배운 내용은 나와 관련이 있다.	①	②	③	④
9-8. 과학 수업 시간에 배운 내용은 일상생활에서 사용할 수 있다.	①	②	③	④
9-9. 나에게는 과학 수업 시간에 배운 내용을 제대로 이해하는 것이 중요하다.	①	②	③	④
9-10. 나에게는 과학 수업 시간에 주어진 과제나 활동을 성공적으로 마치는 것이 중요하다.	①	②	③	④

〈과학관련 진로포부〉

10. 사용성 검토 진행 이후부터 지금까지의 과학수업 시간의 경험에 대한 질문입니다. 사용성 검토 진행 이후부터 지금까지 과학수업(온라인 수업 포함)을 생각해볼 때, 과학 관련 직업 희망과 관련된 아래의 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
10-1. 과학 관련 직업에 대하여 알게 되었다.	①	②	③	④
10-2. 나는 과학 관련 직업이 나에게 스스로 더 배우고 발전할 수 있는 기회를 준다고 생각한다.	①	②	③	④
10-3. 과학 관련 직업은 우리 사회에 많은 영향을 끼친다.	①	②	③	④
10-4. 과학 관련 직업에 흥미를 가지게 되었다.	①	②	③	④
10-5. 미래에 과학과 관련된 직업을 가지고 싶다.	①	②	③	④

〈과학관련 태도〉

11. 사용성 검토 진행 이후부터 지금까지의 과학수업 시간의 경험에 대한 질문입니다. 사용성 검토 진행 이후부터 지금까지 과학수업(온라인 수업 포함)을 생각해볼 때, 과학에 대한 태도와 관련된 각 내용에 대해 어느 정도 동의합니까?

문항	전혀 그렇지 않다	그렇지 않다	그렇다	매우 그렇다
11-1. 과학은 이 세상을 더욱 살기 좋은 곳으로 만드는 데 도움이 된다.	①	②	③	④
11-2. 과학은 공부할 만한 가치가 있다.	①	②	③	④
11-3. 과학은 학교를 졸업한 후에도 쓸모가 있다.	①	②	③	④
11-4. 과학의 발전은 환경, 기술 및 사회의 발전과 영향을 주고받는다.	①	②	③	④
11-5. 학교에서 과학 수업 시간이 늘어나는 것은 바람직하다.	①	②	③	④
11-6. 과학자는 합리적으로 생각하고 판단한다.	①	②	③	④
11-7. 나는 과학에 대하여 더 알고 싶다.	①	②	③	④
11-8. 나는 과학과 관련된 경험(독서, 견학, 과학 영상 시청 등)을 좋아한다.	①	②	③	④

□ 사용성 검토 설문 문항

〈유용성〉

12. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 유용성에 대해 각 항목별 응답해 주세요. (필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
12-1. “대화형 서비스(AI 챗봇)”는 나의 학습 성과를 향상시키는 데 도움이 된다.	①	②	③	④	⑤
12-2. “대화형 서비스(AI 챗봇)”는 나의 자기주도 학습에 도움이 된다.	①	②	③	④	⑤
12-3. “대화형 서비스(AI 챗봇)”는 수업 중 질문 해결에 효과적이다.	①	②	③	④	⑤
12-4. “대화형 서비스(AI 챗봇)”는 내가 이해하지 못한 개념을 쉽게 설명해준다.	①	②	③	④	⑤
12-5. “AI 보고서 튜터링”은 나의 과학탐구 보고서 작성 능력 향상에 도움이 된다.	①	②	③	④	⑤
12-6. “AI 보고서 튜터링”을 사용하면 과학탐구 보고서를 더 빠르게 작성할 수 있다.	①	②	③	④	⑤
12-7. “AI 보고서 튜터링”을 통해 나의 과학탐구 보고서의 품질을 향상시킬 수 있다.	①	②	③	④	⑤
12-8. “AI 보고서 튜터링”을 사용함으로써 과정중심 평가를 실현할 수 있다.	①	②	③	④	⑤

〈사용 용이성〉

13. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 사용 용이성에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
13-1. “대화형 서비스(AI 챗봇)”를 사용하는 것은 나에게 쉬운 일이다.	①	②	③	④	⑤
13-2. “대화형 서비스(AI 챗봇)” 사용법을 배우는 데 시간이 오래 걸리지 않았다.	①	②	③	④	⑤
13-3. “대화형 서비스(AI 챗봇)”는 인터페이스가 직관적이라 사용하는 데 불편함이 없다.	①	②	③	④	⑤
13-4. 도움 없이도 “대화형 서비스(AI 챗봇)”를 잘 사용할 수 있다.	①	②	③	④	⑤
13-5. “AI 보고서 튜터링”을 사용하는 것은 나에게 쉬운 일이다.	①	②	③	④	⑤
13-6. “AI 보고서 튜터링” 사용법을 배우는 데 시간이 오래 걸리지 않았다.	①	②	③	④	⑤
13-7. “AI 보고서 튜터링”은 인터페이스가 직관적이라 사용하는 데 불편함이 없다.	①	②	③	④	⑤
13-8. 도움 없이도 “AI 보고서 튜터링”을 잘 사용할 수 있다.	①	②	③	④	⑤

〈사용 의도〉

14. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 사용 의도에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
14-1. “대화형 서비스(AI 챗봇)”가 지능형 과학실 ON에 상용화된다면 이 서비스를 지속적으로 사용할 의향이 있다.	①	②	③	④	⑤
14-2. “대화형 서비스(AI 챗봇)”가 지능형 과학실 ON에 상용화된다면 이 서비스를 다른 사람에게 추천하고 싶다.	①	②	③	④	⑤
14-3. “대화형 서비스(AI 챗봇)”는 향후 탐구수업 활동에 있어 필수적인 도구가 될 수 있다고 생각한다.	①	②	③	④	⑤
14-4. “AI 보고서 튜터링”이 지능형 과학실 ON에 상용화된다면 이 서비스를 지속적으로 사용할 의향이 있다.	①	②	③	④	⑤
14-5. “AI 보고서 튜터링”이 지능형 과학실 ON에 상용화된다면 이 서비스를 다른 사람에게 추천하고 싶다.	①	②	③	④	⑤

14-6. “AI 보고서 튜터링”은 향후 과학탐구 보고서 작성 및 평가에 있어 필수적인 도구가 될 수 있다고 생각한다.	①	②	③	④	⑤

〈사용 태도〉

15. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 사용 태도에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
15-1. “대화형 서비스(AI 챗봇)”를 사용하는 것은 긍정적인 경험이었다.	①	②	③	④	⑤
15-2. 나는 “대화형 서비스(AI 챗봇)” 사용에 대해 전반적으로 만족한다.	①	②	③	④	⑤
15-3. 지능형 과학실 ON에서의 “대화형 서비스(AI 챗봇)”는 교육적으로 유익하다고 생각한다.	①	②	③	④	⑤
15-4. “AI 보고서 튜터링”을 사용하는 것은 긍정적인 경험이었다.	①	②	③	④	⑤
15-5. 나는 “AI 보고서 튜터링” 사용에 대해 전반적으로 만족한다.	①	②	③	④	⑤
15-6. 지능형 과학실 ON에서의 “AI 보고서 튜터링”은 교육적으로 유익하다고 생각한다.	①	②	③	④	⑤

〈신뢰·정확도〉

16. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 신뢰 및 정확도에 대해 각 항목별 응답해 주세요.(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
16-1. 지능형 과학실 ON에서의 “대화형 서비스(AI 챗봇)”를 신뢰할 수 있다고 느낀다.	①	②	③	④	⑤
16-2. “대화형 서비스(AI 챗봇)”의 답변이 정확하다고 생각한다.	①	②	③	④	⑤
16-3. 지능형 과학실 ON에서의 “AI 보고서 튜터링”을 신뢰할 수 있다고 느낀다.	①	②	③	④	⑤
16-4. “AI 보고서 튜터링” 결과값이 정확하다고 생각한다.	①	②	③	④	⑤

〈학습효과 및 교육적 가치〉

17. AI 기반「 지능형 과학실 ON 」과학탐구 지원 시스템 (대화형 서비스, AI 보고서 튜터링)의 학습효과 및 교육적 가치에 대해 각 항목별 응답해 주세요.

(필수답변)

문항	매우 그렇다	그렇다	보통이다	그렇지 않다	전혀 그렇지 않다
17-1. “대화형 서비스(AI 챗봇)”는 학생의 “과학탐구 역량 강화”에 효과적이다.	①	②	③	④	⑤
17-2. “대화형 서비스(AI 챗봇)”는 학생의 “과학 기초학력 향상”에 효과적이다.	①	②	③	④	⑤
17-3. “대화형 서비스(AI 챗봇)”는 학생의 “과학 개념의 깊이 있는 이해”에 효과적이다.	①	②	③	④	⑤
17-4. “대화형 서비스(AI 챗봇)”는 학생의 “학습 흥미와 동기부여”에 효과적이다.	①	②	③	④	⑤
17-5. “AI 보고서 튜터링”은 학생의 “과학탐구 역량 강화”에 효과적이다.	①	②	③	④	⑤
17-6. “AI 보고서 튜터링”은 학생의 “과학 기초학력 향상”에 효과적이다.	①	②	③	④	⑤
17-7. “AI 보고서 튜터링”은 학생의 “과학 개념의 깊이 있는 이해”에 효과적이다.	①	②	③	④	⑤
17-8. “AI 보고서 튜터링”은 학생의 “학습 흥미와 동기부여”에 효과적이다.	①	②	③	④	⑤

18.「 지능형 과학실 ON 」“대화형 서비스(AI 챗봇)”에서 가장 유용했던 기능 또는 장점이 있다면 자유롭게 입력해 주세요.

19.「 지능형 과학실 ON 」“대화형 서비스(AI 챗봇)”에서 개선되었으면 하는 점이 있다면 자유롭게 입력해 주세요.

20.「 지능형 과학실 ON 」“AI 보고서 튜터링”에서 가장 유용했던 기능 또는 장점이 있다면 자유롭게 입력해 주세요.

21.「 지능형 과학실 ON 」 “AI 보고서 튜터링”에서 개선되었으면 하는 점이 있다면 자유롭게 입력해 주세요.

지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템 개발 연구
(A Study on the Development of Generative AI-Based Science Inquiry Support System Integrated with the 'Intelligence Science Lab ON' Platform)

연 구 주 관 기 관 **한국과학창의재단**

연 구 담 당 조수경 (과학교육실 선임연구원)

연 구 수 행 기 관 **한국과학창의재단**

연 구 수 행 조수경 (과학교육실 선임연구원 / 연구책임자)

김학수 (건국대학교 교수 / 공동연구원)

이선복 (이화여자대학교 교수 / 공동연구원)

조헌국 (단국대학교 교수 / 공동연구원)

박찬준 (숭실대학교 교수 / 공동연구원)

선주오 (건국대학교 박사과정 / 연구조원)

※ 이 보고서는 2025년도 교육부의 재원으로 한국과학창의재단에서 수행한 성과물임

지능형 과학실 ON 연계 생성형 AI 기반 과학탐구 지원 시스템 개발 연구

주 의 문

- 본 연구의 주장이나 제언은 연구진의 견해이며, 한국과학창의재단의 공식 입장이 아닙니다.
- 이 보고서는 한국과학창의재단에서 시행한 정책연구과제의 결과로, 인용하실 때에는 반드시 출처를 표기하여 주시기 바랍니다.

한국과학창의재단
Korea Foundation for Science and Creativity

06130 서울특별시 강남구 테헤란로 7길 22(역삼동, 과학기술회관 2관)
TEL 02-555-0701 **FAX** 02-555-0702 www.kosac.re.kr