Vol.
01
2026. 06

MONTHLY EDTECH SOLUTION ANALYSIS REPORT

월간 에듀테크 솔루션 분석 보고서

이번 호 주제 ISSUE TOPIC

LLM 기반 학습 도구

대학에 들어온 범용 생성형 AI: '학습 모드'의 근거와 기관 도입의 조건

대표 분석 솔루션 ChatGPT 학습 모드·ChatGPT Edu · Gemini 가이드 학습·Gemini for Education

키워드 LLM 기반 학습 도구 · 학습 모드 · 고등교육 · 기관 도입 · 데이터 보호

분석 프레임 4-QUADRANT FRAMEWORK

① 기술 구조	② 시장 동향	③ 교육적 적용	④ 한계·논쟁점
Technology	Market	Pedagogy	Limitations

한림대학교 AI 에듀테크 연구과제

글로컬대학 30 한림대학교 AI 에듀테크센터 연구과제「고등교육 AI 에듀테크의 통합적 발전 방안 연구」[연구 2] 산출물

발간 2026. 6. 30.

EXECUTIVE SUMMARY

핵심 요약

이번 호의 질문

범용 LLM 의 '학습 모드'는 일반 대화와 무엇이 다르며 대학생의 학습을 실제로 돕는가? 대학은 어떤 도입 모델과 운영 조건으로 이를 구성원에게 제공해야 하는가?

한 줄 정의

LLM 기반 학습 도구는 범용 대규모 언어 모델(LLM)에 교육용 지시·모델 조정·자료 기반 응답·기관 관리 기능을 더해 학습자에게 설명을 제공하고 연습·이해 점검을 돕는 대화형 도구다.

핵심 발견

1. 학습 모드는 시스템 지시(OpenAI)·학습용 모델 조정(Google)의 두 방식으로 구현되거나 작업 공간·에이전트 안에 배치되지만(Anthropic·Microsoft) 정답 보류를 보장하지 않는다. OpenAI 는 시스템 지시 방식이 대화마다 일관되지 않은 행동과 실수를 낳을 수 있다고 인정했고, 7 개 LLM 독립 평가의 정답 비노출 비율도 모델에 따라 53.12~94.79%로 크게 달랐다(2024 년 모델 수준 평가). 상용 학습 모드의 제품 단위 독립 평가는 ChatGPT 학습 모드를 공개 자료로 평정한 국내 연구(K-12 맥락) 한 건이었다.
2. 메타분석은 g = 1.14(대학생 66 편), g+ = 0.712(실험·준실험 69 편) 등 큰 효과크기를 보고했지만 신중하게 읽어야 한다. 66 편 메타분석에서 이질성(I^2 = 97.62%)과 출판 편향이 확인되었고 69 편 메타분석에 포함된 학업성취 비교 19 건을 감사한 후속 연구(Weidlich 외)에서는 세 가지 기본 기준을 모두 충족한 비교가 4 건(21%)뿐이었다. 상용 학습 모드를 대학생 대상으로 검증한 실험 연구는 연구진이 확인한 범위에서 아직 없다.
3. 대학의 기관 배포는 한 제공사 서비스를 기관 단위로 계약하는 단일 벤더형(서울대 ChatGPT Edu, CSU)과 공급사 플랫폼에서 여러 제공사의 모델을 골라 쓰는 멀티 LLM 통합형(순천향대 조달, 본교(한림대학교) HAI-GPT)으로 나뉜다. 사용량 관리도 CSU 의 무제한 접근과 멀티 LLM 형의 사용자별 월간 한도로 갈렸다. CSU 에서는 자발적 AI 교육 이수율이 학생 0.7%에 그쳤고 학생 대표가 계약 과정에서 협의를 받지 못했다고 밝혀, 기관 도입에는 비용 정당화와 함께 교육 지원·구성원 협의가 필요한 것으로 보인다. 다만 이 논쟁의 근거는 주로 OpenAI 계약 사례(CSU)이고 국내 멀티 LLM 사례는 본교 HAI-GPT 와 같은 공급사의 플랫폼이다.
4. 국내 제도는 투명성과 데이터 보호를 요구한다. AI 기본법 제 31 조의 고지·표시 의무(2026 년 1 월 22 일 시행, 1 년 이상 계도기간)와 개인정보위 생성형 AI 개인정보 처리 안내서(DPA·국외 이전 검토)가 기준점이며, 「대한민국 인공지능행동계획」은 대학 AI 활용 윤리 가이드라인과 학생 개인정보 보호 지침 마련을 권고했다. 2026 년 1~2 월 한국대학교육협의회 총장 설문(140 개교 응답)에서 생성형 AI 활용 정책을 채택한 대학은 40.0%, 제정을 검토 중인 대학도 40.0%였다.

4 분면 요약

① 기술 구조

- ○ 학습 모드는 시스템 지시(OpenAI)·학습용 모델 조정(Google)의 두 방식으로 구현되거나 작업 공간·에이전트 안에 배치된다(Anthropic·Microsoft, 구현 방식 미공개). 정답 보류는 구현 방식만으로 보장되지 않고, 독립 평가(2024 년 모델 수준)에서 정답 비노출 비율은 기반 모델에 따라 크게 달랐다.
- ○ 기관용 상품은 자료 기반 응답·데이터 처리 조건·관리자 기능·LMS 연동에서 차이가 나며 RAG 도 환각을 줄일 뿐 없애지는 못하는 것으로 보인다.

② 시장 동향

- ○ 세계 AI 교육 시장 추정치는 2030 년 58.2 억~410.1 억 달러로 조사 기관에 따라 크게 엇갈려(정의·산정 방법 차이로 추정) 범위로만 읽어야 한다.
- ○ 대학 도입은 단일 벤더형과 멀티 LLM 통합형으로 나뉘며, 사용량 관리 방식(무제한 접근·사용자별 월간 한도)은 다르지만 어느 쪽이든 비용과 데이터의 전달·보관, 구성원 협의가 공통 관리 대상이 될 가능성이 크다.

③ 교육적 적용

○ 메타분석은 큰 정적 효과를 보고하지만 이질성·출판 편향·포함 연구의 설계 문제로 과대 또는 과소 추정 가능성이 있어 신중하게 읽어야 한다.

○ 정답 보류형 튜터가 제한 없는 사용의 학습 저해를 줄인 직접 근거는 K-12 현장실험이고 대학 실험에서는 무제한 사용도 지식 향상에서 뒤지지 않았으며 설계된 튜터의 우위도 일관되지 않았다. 상용 학습 모드의 대학 실험 근거는 아직 확인되지 않았다.

④ 한계·논쟁점

○ 환각·인용 오류, 인지적 외주화와 학습 깊이, 학업 진실성, 형평성, 프라이버시, 편향과 한국어·한국 맥락, 모드 전환, 비용·종속·거버넌스가 남은 쟁점이다.

○ 실험으로 인과가 확인된 위험은 일부이며 대부분은 출처 확인, 정답 보류형 설계와 과제 설계의 병행, 사용 공개, 기관 계약, 비용 관리와 구성원 협의 같은 운영 설계로 관리할 성격으로 보인다.

대표 솔루션 진단 요약

솔루션	사용성 (A)	교육적 특성(B)	AI 신뢰성 (C)	종합	공통 항목 기준	도입 검토 의견
ChatGPT 학습 모드·ChatGPT Edu (OpenAI)	-	26.9	50.0	38.5	38.5	판정 유보
Gemini 가이드 학습·Gemini for Education (Google)	-	42.3	50.0	46.2	46.2	판정 유보

주: 두 제품 모두 자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구로 문서 판정 항목 11 개(필수 13 개 중 5 개)만 판정했고, 실사용 진단이 필요한 AUTO·HYBRID 14 개 항목과 공개 근거를 찾지 못한 A8 은 미측정이다(실사용 진단 결과는 싣지 않음). 필수 항목 측정이 75%에 못 미쳐 도입 검토 의견은 판정 유보이며, 사용성(A)은 측정 항목이 없어 점수를 내지 않았고(-) 종합 점수는 축 B·C 만으로 계산한 참고값이다. 공통 11 개 항목 기준 점수 차 7.7 점은 교수 통합(B6)·표준 연동(B8)의 문서상 LTI 연동 차이에서 나오며 순위가 아니다.

국내 적용 시사점

○ 기관 도입 시 학습 제외·보관·국외 이전과 데이터 처리 부속서(DPA), AI 기본법 제 31 조 고지·표시의 이행 주체를 계약 단계에서 점검하고, 도입 모델에 맞는 사용량 관리와 갱신 조건도 확인할 필요가 있다(부록 4 점검표).

○ 도입한 플랫폼이 학습 모드를 제공하는지 먼저 확인하되, 켜 두는 것만으로는 효과를 기대하기 어렵고 모드 전환의 한계도 있으므로 과정 중심 평가·AI 사용 공개 같은 과제·평가 설계와 함께 운영하는 편이 바람직하다.

○ 교양 AI 기초 교과(대학 AI 기본교육과정 기준)와 연계해 산출물 검증·비판적 수용 역량을 기르고, 통제 비교와 지연 측정을 포함한 학내 효과 측정과 미측정 항목의 실사용 진단을 마련할 필요가 있다.

CONTENTS

목 차

Ⅰ. 배경 및 분석 개요 1

1. 선정 배경 1
2. 분석 범위와 분석 단위 2
3. 분석 대상 솔루션 3
4. 분석 방법 5

Ⅱ. 솔루션군 동향 분석: 4분면 6

1. 기술 구조 6
2. 시장 동향 10
3. 교육적 적용 13
4. 한계·논쟁점 17

Ⅲ. 대표 솔루션 심층 분석 22

1. 솔루션 프로필 22
2. 평가척도 적용 결과 23
3. 비교 종합 26

Ⅳ. 시사점 및 종합 28

1. 국내 적용 시사점 28
2. 본교·연구회 활용 제언 30
3. 종합 32

부록 33

부록 1. 참고 자료 목록 33
부록 2. 평가척도 채점표 38
부록 3. 용어 해설 47
부록 4. 기관 도입 점검표 49

Ⅰ. 배경 및 분석 개요

이 보고서는「월간 에듀테크 솔루션 분석 보고서」의 첫 호로 대규모 언어 모델(LLM) 기반 학습 도구를 다루며, 여기서 정한 분석 틀과 평가 방법은 이후 호의 기준선이 된다. 대상 월은 2026 년 6 월이고 자료 기준일은 발간일과 같은 2026 년 6 월 30 일이다. 기준일 뒤에 발행·공개된 자료와 그 자료에 근거한 내용은 싣지 않았다.

1. 선정 배경

가. 정책 환경: 대학 AI 교육의 제도화

교육부는 2025 년 11 월「AI for All: 모두를 위한 AI 인재양성방안」에서 그간 AI·디지털 교육이 초·중등에 초점이 맞춰졌고 고등·평생 분야 정책은 상대적으로 부족했다고 진단했다(교육부, 2025). 이어 2026 년 대학 AI 기본교육과정 개발 지원사업은 첫해 20 개 대학을 선정해 교당 3 억 원(총 60 억 원)을 지원하며, 선정 대학은 AI 기초 교양 교과를 신입생 필수 이수 과목으로 운영해야 한다(교육부·한국대학교육협의회, 2026).

국가인공지능전략위원회와 과학기술정보통신부(과기정통부)가 2026 년 2 월 확정한 국가 법정계획인「대한민국 인공지능행동계획(인공지능 기본계획 2026~2028)」은 대학 교육에 관한 권고를 함께 담았다. 이 계획은 과기정통부가 소프트웨어 중심대학 사업을 AI 중심대학 사업으로 전환하고 수행 대학을 확대하는 방안을 수립하도록 권고했다. 또한 교육부가 '(가칭) 대학 AI 활용 윤리 가이드라인'을 2026 년 1 분기까지 개발·배포하고 모든 대학생의 AI 기초 역량을 위해 대학 AI 교육과정 개발·공유·확산을 지원하도록 권고했다(국가인공지능전략위원회, 2026).

지원사업과 행동계획에 비추어 보면 대학 AI 정책의 초점은 교과 개설에서 전 구성원의 AI 활용 관리로 넓어지고, 어떤 AI 도구를 어떤 조건으로 제공할지가 대학의 정책 결정 사항이 될 가능성이 크다.

나. 현장: 보편화된 사용과 '학습 모드' 경쟁

한국교육개발원(KEDI)의 2025 년 대학 교수·학습 실태 조사(참여 대학 자율 신청 방식)에서 일반대학 응답 학생 27,920 명 중 96.6%가 학업에 AI 를 활용하고 있었다(김지하 외, 2025). 영국 고등교육정책연구소(HEPI)와 에듀테크 기업 Kortext 의 민간 조사(2025 년 12 월, 전일제 학부생 1,054 명)에서는 영국 학부생의 95%가 AI 를 한 가지 이상 용도로 쓰고 94%가 평가 과제에 생성형 AI 를 활용한다고 답했다(Stephenson & Armstrong, 2026).

2025 년에는 주요 사업자가 정답 대신 질문·힌트로 학습 과정을 안내하는 대화 모드(이하 학습 모드)를 잇달아 내놓았다. Anthropic 은 2025 년 4 월 2 일 Claude for Education 을 출시하며 Claude 학습 모드(Learning mode)를 도입했다(Anthropic, 2025a). OpenAI 는 2025 년 7 월 29 일 ChatGPT 학습 모드(study mode)를 출시했다(OpenAI, 2025). Google 은 2025 년 8 월 6 일 학습용 모델 LearnLM 의 역량을 Gemini 에 넣은 가이드 학습(Guided Learning)을 출시했다(Heymans, 2025).

KEDI·HEPI 두 조사 결과로 미루어 대학생의 생성형 AI 사용은 보편화 단계에 들어선 것으로 읽힌다. 학습 모드 출시와 함께 대학이 다룰 질문도 생성형 AI 를 쓸 것인가에서 어떤 모드와 조건으로 쓰게 할 것인가로 옮겨 갈 가능성이 크다.

다. 연구: 쌓이는 문헌과 엇갈린 효과 근거

효과 근거는 엇갈린다. 9 개 데이터베이스에서 찾은 5,555 편 가운데 포함 기준을 충족한 대학생 연구 66 편을 종합한 Mo 외(2026)의 메타분석에서 ChatGPT 활용은 대학생 학습성과에 큰 정적 효과(Hedges' g = 1.14)를 보였으나, 저자들은 통합 효과크기가 다소 부풀려졌을 수 있다고 밝혔다. 튀르키예 고등학생(K-12) 대상 현장실험에서는 가드레일(풀이를 바로 주지 않도록 거는 제한) 없는 GPT Base 집단이 AI 없이 치른 시험에서 통제집단보다 성적이 17% 낮았다(Bastani et al., 2025). 한 학술지의 편집인은 2026 년 4 월 메타분석의 불일치에 대한 우려로 ChatGPT 가 학생의 학습 수행에 미치는 효과를 다룬 메타분석 논문을 철회했다(Wang & Fan, 2026).

한 메타분석이 5,000 편이 넘는 검색 문헌에서 포함 연구를 골라낼 만큼 문헌은 쌓였지만, 근거의 일관성과 질은 아직 이를 따라가지 못하는 것으로 보인다. 기관 배포는 국내에서도 한 제공사 서비스의 전 구성원 계약(Ⅰ장 3 절)과 통합 플랫폼 조달(Ⅱ장 2 절 라)로 나뉜다. 이에 이 보고서는 학습 모드의 차이, 기관 배포 방식, 효과와 위험 관리를 묻는 다음 세 핵심 질문으로 범위를 좁힌다.

○ Q1. 범용 LLM 의 학습 모드는 기술적으로 무엇이 다르며 교수 설계 원리를 얼마나 반영하는가?

○ Q2. 대학 기관 배포의 두 모델(단일 벤더형·멀티 LLM 통합형)은 비용·데이터 보호·관리 측면에서 어떻게 다른가?

○ Q3. 학습 효과의 실증 근거는 무엇을 말하며 과의존·학습 저해 위험은 어떻게 관리되는가?

2. 분석 범위와 분석 단위

이 보고서에서 LLM 기반 학습 도구는 범용 LLM 에 교육용 지시·모델 조정·자료 기반 응답·기관 관리 기능을 더해 학습자에게 설명을 제공하고 연습·이해 점검을 돕는 대화형 도구를 말한다. 이번 호의 분석 단위는 '학습자가 범용 LLM 어시스턴트(대학 기관 라이선스 포함)를 학습 보조로 쓰는 경우'이며, 기관 라이선스는 대학이 제공사와 기관 단위로 계약해 구성원에게 제공하는 이용권이다. 학습 모드는 정답을 곧바로 주는 대신 질문·힌트·이해 점검으로 학습 과정을 안내하도록 설계된 범용 LLM 의 대화 모드로 조작적으로 정의하며, 제공사마다 이름(study mode, Guided Learning, Learning mode)이 달라도 이 정의에 해당하면 같은 범주로 비교한다. 다른 호와의 경계는 [표 Ⅰ-1]과 같다.

분석 단위를 좁힌 것은 LLM 의 교육 적용 범위가 넓기 때문이다. 2022 년 11 월~2025 년 3 월 출판된 실증 연구 88 편을 종합한 Shi 외(2026)의 체계적 문헌고찰에 따르면, LLM 은 교육 현장에서 여섯 가지 주요 용도로 활용되며 그중 지능형 튜터링 시스템이 특히 두드러졌다. 교수자 과업으로도 적용이 넓어, Chu 외(2025)의 서베이는 피드백 코멘트 생성·교육과정 설계 등 교육 현장의 핵심 과제에 LLM 에이전트를 적용한 최근 연구를 체계적으로 검토했다. 이번 호는 학습자가 쓰는 기능만 분석한다.

[표 Ⅰ-1] 분석 범위와 경계

구분	이번 호에서 다루는 것	제외·제한(넘기는 호)
사용 주체	학습자	교수자용 강의 설계·자료 생성(Vol.07), 기관의 관리·탐지·조달 심사 도구(Vol.08)
핵심 기능	범용 LLM 과의 대화, 업로드 자료 기반 질의, 학습 모드	교과·강좌 특화 대화형 튜터(Vol.03), 학습 경로·난이도 자동 조정(Vol.02), 계획·점검·성찰 심화 지원(Vol.06)
제공 형태	개인 구독·학생 혜택, 대학 기관 라이선스(단일 벤더형), 멀티 LLM 통합 플랫폼	—

구분	이번 호에서 다루는 것	제외·제한(넘기는 호)
학교급	고등교육	K-12 연구는 참고 근거로만 쓰고 'K-12'로 표시
근거 범위	독립 연구(효과), 제공사 문서(제품 사실), 동료심사 전 자료('동료심사 전' 표시)	제공사 수행·후원 연구는 '자체 보고'로 표시, 제공사 마케팅 성과 수치는 쓰지 않음

자료: 연구진 정리

3. 분석 대상 솔루션

이번 호는 평가척도로 채점할 대표 솔루션으로 OpenAI 의 ChatGPT 학습 모드·ChatGPT Edu 와 Google 의 Gemini 가이드 학습·Gemini for Education 을 골랐다. 선정 기준은 시의성, 국내 고등교육 도입 실적, 국내 이용 가능성, 사업 연속성, 제공사 밖 독립 평가, 실사용 진단 가능성이다(실사용 진단 결과는 이번 호에 싣지 않았다). 시의성과 사업 연속성 면에서 두 학습 모드는 2025 년 하반기에 잇따라 출시되었고(Ⅰ 장 1 절), 두 제공사 모두 고등교육 기관용 상품을 운영한다.

국내 도입 실적으로는 서울대학교가 ChatGPT Edu 를 도입해 2026 년 6 월부터 교수·학생·직원 등 전 구성원에게 정식 서비스를 제공한다(서울대학교, 2026). Gemini for Education 은 기준일 이전 공개 자료에서 국내 대학의 기관 도입을 확인하지 못해 국민 이용 실태를 대리 근거로 삼았다. 만 3 세 이상 국민을 대상으로 한 「2025 인터넷 이용 실태조사」에서는 ChatGPT(41.8%)와 Gemini(9.8%)가 이용한 생성형 AI 서비스 1·2 위였으나(복수 응답), 이는 대학 구성원의 이용 비율과 다른 대리 지표다(과학기술정보통신부·한국지능정보사회진흥원, 2026).

이용 가능성 면에서 OpenAI 는 학습 모드를 로그인 이용자 대상으로 출시했고 ChatGPT Edu 제공은 몇 주 뒤로 예고했다(OpenAI, 2025). Google 은 Gemini for Education 이 Workspace for Education 요금제에 무료로 포함된다고 밝혔다(Kirtikar, 2025). Google 의 학생 대상 Google AI Pro 12 개월 무료 이용(한국 포함)은 신청 기한이 2025 년 10 월 6 일로 끝난 한시 혜택이어서 근거로 삼지 않았다(Pichai, 2025).

제공사 밖 독립 평가는 두 제품에 고르지 않다. ChatGPT 학습 모드는 차덕원·홍후조(2026)가 고교 진로·진학 플랫폼과 같은 틀에서 시스템별 공개 자료를 종합 검토해 가장 보수적인 근거에 따라 점수를 부여하는 방식으로 평정했다(K-12 맥락). Gemini 가이드 학습은 확인된 제품 단위 독립 평가가 없고, ChatGPT·Gemini 각 3 개 버전의 참고문헌 정확도를 비교한 Pastucha 외(2026)처럼 모델 수준 평가만 있다.

같은 시기의 다른 학습 기능은 [표 Ⅰ-2]의 사유로 채점하지 않고 동향 분석 대상으로 다룬다. Microsoft 는 2025 년 10 월 Microsoft 365 Copilot 의 Study and Learn 에이전트를 발표했다(Quarnstrom, 2025). 이 에이전트는 2026 년 5 월 13 일 모든 교육 고객에게 추가 비용 없이 정식 제공되었으나, 발표 시점에는 미국 영어에 최적화되어 있었다(Microsoft Education Team, 2026). Google 은 2025 년 9 월 NotebookLM 에 'Learning Guide' 옵션을 도입했다(Paul & Dhami, 2025).

[표 Ⅰ-2] 분석 대상 솔루션 개요

구분	솔루션(제공사)	학습 기능·출시	기관용 상품	국내 도입·이용 근거	이번 호 처리
대표	ChatGPT 학습 모드·ChatGPT Edu(OpenAI)	study mode(2025. 7. 29.)	ChatGPT Edu	서울대 전 구성원 도입(2026. 6.~), 국민 이용 41.8%(1위)	채점(문서 판정)
대표	Gemini 가이드 학습·Gemini for Education(Google)	Guided Learning(2025. 8. 6.)	Gemini for Education(Workspace for Education 포함)	국내 대학 기관 도입 확인되지 않음, 국민 이용 9.8%(2 위)	채점(문서 판정)
동향	Claude 학습 모드·Claude for Education(Anthropic)	Learning mode(2025. 4. 2.)	Claude for Education	국내 기관 도입 확인되지 않음	동향 분석(이해관계로 채점 제외)
동향	Study and Learn(Microsoft)	2025. 10. 발표, 2026. 5. 13. 정식 제공(교육 고객 추가 비용 없음)	Microsoft 365 Copilot	국내 고등교육 도입 확인되지 않음	동향 분석(교육 기관 고객용 에이전트)
동향	NotebookLM Learning Guide(Google)	2025. 9. 도입	—	—	동향 분석(업로드 자료 기반 학습, Vol.03·07 과 경계)
기관 도입 모델	국내 멀티 LLM 통합 플랫폼(순천향대 조달)	—	크레딧 기반 구독형	대학 조달 공고(2025)	도입 모델로 분석(이해관계로 채점 제외, Ⅱ 장 2 절·Ⅳ 장 1 절)
사례	본교 HAI-GPT	—	FactChat(마인드로직) 기반, 본교 AI 에듀테크센터 운영	—	Ⅳ장 2 절 사례(평가하지 않음)

자료: 각 사 공식 발표·도움말, 서울대학교(2026), 과학기술정보통신부·한국지능정보사회진흥원(2026), 대학 조달 공고를 바탕으로 연구진 정리

자료 수집·분석과 보고서·채점 초안 작성에 활용한 생성형 AI(Anthropic Claude 등)는 분석 대상 사업자와 경쟁 관계에 있는 제품이므로, Anthropic 제품은 채점에서 제외했고 두 대표 솔루션의 판정 근거는 부록 2 에 공개해 검증할 수 있게 했다. 대표 솔루션은 같은 평가척도로 심층 분석하고 동향 분석 대상은 같은 비교 항목으로 서술했으나, Claude for Education 은 채점에서 빠져 이 보고서가 제시한 결함 근거에 노출되지 않았으므로 이를 상대적 우위로 해석해서는 안 된다. 국내 대학의 생성형 AI 도입은 주로 여러 LLM 을 묶는 통합 플랫폼 형태인데, 본교(한림대학교)도 같은 유형의 플랫폼(FactChat 기반 HAI-GPT)을 운영하고 있어 이해관계가 있으므로 국내 플랫폼은 채점에서 제외하고 Ⅱ 장 동향과 Ⅳ장 사례로 다루었다. Ⅱ 장 2 절 라의 순천향대 조달 건도 FactChat 기반임을 연구진이 대학 공지로 확인했다. HAI-GPT 는 이 보고서를 산출한 본교 AI 에듀테크센터가 운영하므로 평가하지 않으며, 뤼튼 등 국내 범용 서비스는 학습 모드 제공 여부를 확인하지 못해 다루지 않았다.

4. 분석 방법

분석은 [표 Ⅰ-3]의 다섯 단계로 진행했고 자료 목록은 부록 1 에 실었다. 근거 자료는 기준일까지 발행·공개된 것만 썼다. DOI 가 있는 논문은 OpenAlex 와 Crossref 로 철회·정정 여부를 확인했고, 철회된 논문은 효과 근거로 쓰지 않았다.

이번 호의 평가는 자료 기준일까지 공개된 문서, 곧 제공사의 발표문·도움말·정책 문서와 정부 자료, 독립 연구에 근거하며 실사용 진단 결과는 싣지 않았다. 그래서 평가척도 가운데 실사용 진단이 있어야 판정할 수 있는 항목은 문서만으로 판정하지 않고 미측정으로 두었다. 자료 수집·분석과 초안·채점 초안 작성에는 생성형 AI 를 보조로 활용했으며, 판정과 원고는 연구진이 근거 문서와 대조해 확정했다.

평가척도 v1 은 축 A 사용성(8 개 항목), 축 B 교육적 특성(10 개), 축 C AI 신뢰성·윤리·거버넌스(8 개)의 26 개 항목으로 이루어진다. 항목마다 측정 방식이 정해져 있어, AUTO 항목은 도구로 실측하고 HYBRID 항목은 실측 뒤 검토자가 확인하며 DOC 항목은 문서로 판정한다. 항목은 ●(충족, 2 점)·◐(부분충족, 1 점)·○(미충족, 0 점)으로 판정하고 가중(필수 ×3, 권장 ×2, 보조 ×1)을 곱한다. 축 점수는 측정 항목 만점에 대한 득점 비율을 100 점으로 환산하며, 해당 없는 항목과 측정하지 못한 항목(미측정)은 분자와 분모에서 모두 뺀다. 측정 항목이 없는 축은 점수를 내지 않는다.

종합 점수는 점수가 산출된 축에 이번 호 가중치(A:B:C = 30:35:35)를 적용해 계산하고, 두 제품의 비교에는 두 제품 모두 측정된 공통 항목(이번 호는 11 개)의 점수만 쓴다. 도입 검토 의견은 게이트 항목(C1 타당성·신뢰성과 환각 관리, C2 투명성·설명가능성, C4 프라이버시·데이터 거버넌스)과 종합 점수로 정한다. 적극 검토는 게이트 모두 ●·필수 항목에 ○ 없음·종합 80 이상, 조건부 검토는 적극 검토에 해당하지 않으면서 게이트 ◐ 이상·종합 60 이상, 신중 검토는 게이트에 ○가 있거나 종합 60 미만인 경우다. 다만 필수 항목 13 개 가운데 측정 항목이 75%(10 개)에 못 미치면 측정 범위가 부족한 것으로 보고 경계값을 적용하지 않는 '판정 유보'로 둔다. 판정 유보는 평가척도 v1 에 없던 범주로 이번 호에서 보완했으며, 평가·배치를 결정하는 도구에만 적용하는 고부담 게이트는 이번 호에 해당하지 않는다.

[표 Ⅰ-3] 분석 방법 요약

단계	내용	자료·도구	산출
① 자료 수집·선별	학술 문헌·정책·조달·제공사 문서 수집(자료 기준일까지), 철회·정정 확인	OpenAlex·Crossref	부록 1
② 4 분면 분석	기술 구조, 시장 동향, 교육적 적용, 한계·논쟁점	수집 자료	[표 Ⅱ-1]~[표 Ⅱ-4]
③ 효과 근거 등급화	설계·대상·표본·효과크기·사용 조건·근거 수준 기록, '자체 보고·동료심사 전·K-12' 표시	메타분석·실험 연구	[표 Ⅱ-3]
④ 평가척도 채점	26 개 항목 중 문서 판정(DOC) 항목 판정·가중, AUTO·HYBRID 항목은 미측정, 게이트 C1·C2·C4, 도입 검토 의견	평가척도 v1, 가중치 A:B:C = 30:35:35, 공식 문서·정부 자료·독립 연구	[표 Ⅲ-3], 부록 2
⑤ 국내 적용 해석	법제·정책·조달 문서 대조	법령·지침·조달 공고	[표 Ⅳ-1]·[표 Ⅳ-2], 부록 4

자료: 연구진 작성

Ⅱ. 솔루션군 동향 분석: 4 분면

이 장은 LLM 기반 학습 도구 솔루션군을 기술 구조, 시장 동향, 교육적 적용, 한계·논쟁점의 네 분면으로 나누어 분석한다.

1. 기술 구조

가. 학습 모드의 구현 방식과 배치 형태

범용 LLM 의 기본 행동이 학습 지원과 어긋난다는 진단은 제공사 안팎에서 나왔다. Google 연구팀의 기술보고서(제공사 자체 보고, 동료심사 전)인 LearnLM 보고서는 현재의 생성형 AI 가 기본적으로 정보 제시 위주로 조정되어 인간 튜터처럼 학습을 이끌지 못한다고 진단했다(LearnLM Team, 2024). 제공사와 독립된 Maurya 외(2025)의 평가에서도 GPT-4 는 응답의 약 47%에서 정답을 노출해 실행 가능성이 떨어졌다(Ⅱ 장 1 절 마).

학습 모드의 구현 방식은 두 가지이고, 이와 별도로 학습 기능을 작업 공간이나 에이전트 안에 두는 배치 형태가 있다(사업자별 비교는 [표 Ⅱ-1]). 첫째 구현 방식은 시스템 지시다. OpenAI 의 ChatGPT 학습 모드는 맞춤형 시스템 지시로 작동하며, 정답을 바로 주는 대신 소크라테스식 질문·힌트·자기성찰 프롬프트를 결합해 이해를 안내한다. OpenAI 는 이 방식이 대화마다 일관되지 않은 행동과 실수를 낳을 수 있다고 인정하면서, 학습 모드의 행동을 향후 주력 모델에 직접 학습시킬 계획이라고 밝혔다(OpenAI, 2025).

둘째 구현 방식은 학습용 모델 조정이다. Google 은 학습용으로 미세조정한 모델군 LearnLM 을 개발했다고 밝히면서, 프롬프트 개선만으로는 의미 있는 학습 도구를 만들기에 충분하지 않았다는 점을 그 배경으로 들었다. Gemini 가이드 학습은 LearnLM 의 역량을 Gemini 에 넣은 기능으로, 탐색적·개방형 질문으로 참여를 유도하고 문제를 단계별로 나누어 설명을 학습자에 맞춘다(Heymans, 2025). 다만 같은 제공사의 LearnLM 보고서는 한계에도 불구하고 프롬프트가 교육 제품 개발자가 행동을 지정하는 최선의 방법으로 남을 것으로 보고, 교수법 행동 주입을 시스템 수준 지시를 따르는 '교수법 지시 따르기' 문제로 재정의했다(LearnLM Team, 2024).

배치형의 예로 Anthropic 은 Claude for Education 의 학습 모드가 답을 주기보다 학생의 추론 과정을 안내한다고 소개했으며 이 모드는 과제·주제별로 대화를 묶어 저장하는 Projects 기능 안에서 작동한다(Anthropic, 2025a). Microsoft 365 Copilot 의 Study and Learn 에이전트(2026 년 5 월 정식 제공, Ⅰ 장 3 절)는 개념 이해·연습과 플래시카드·퀴즈 같은 내장 활동을 제공한다(Quarnstrom, 2025). 두 기능이 지시와 모델 조정 가운데 어떤 방식으로 구현되었는지는 공개 문서에서 확인되지 않았다.

모델 조정은 지시를 대체하기보다 모델이 교수법 지시를 더 잘 따르게 하는 방향이고 OpenAI 도 모델 수준 학습을 계획하고 있어, 두 구현 방식은 서로 배타적이지 않은 것으로 볼 수 있다. 시스템 지시만 쓰는 방식은 재학습 없이 지시를 고쳐 빠르게 개선할 수 있지만 대화마다 행동이 흔들릴 수 있다. 모델 조정은 일관성에 유리할 수 있으나, 이 보고서가 확인한 범위에서 그 이점을 보인 평가는 제공사 자체 평가뿐이다.

나. 근거 생성: 자료 기반 응답과 환각 완화

검색 증강 생성(RAG)은 관련 자료를 먼저 찾아 그 내용을 바탕으로 응답하는 방식이다. Swacha 와 Gracel(2025)은 교육 목적 RAG 챗봇을 다룬 논문 47 편을 분석해 교육용 LLM 챗봇 도입의 주된 장벽을

환각으로 보고 RAG를 그 대응책으로 제시했다. 하버드대 CS50 개발팀은 LLM을 미세조정하거나 재학습하지 않고 강의 자막 기반 RAG를 적용해 강의 관련 답변의 환각 가능성을 줄일 수 있다고 보고했다. 다만 이는 Microsoft·OpenAI 등의 지원을 밝힌 개발팀의 자체 배치 보고이며, 2023년 가을 학기 답변 180건 중 직원 승인을 받은 답변은 70건으로 단순 계산 정확도는 39%에 그쳤다(Liu et al., 2024).

제품 수준에서는 Google의 NotebookLM에서 플래시카드·퀴즈가 사용자가 올린 자료에만 근거해 생성되며, 설명 기능은 원자료로 연결되는 인용을 함께 제시한다(Paul & Dhami, 2025). 유료 상품인 Microsoft 365 Copilot은 문서·이메일·회의 같은 사용자 데이터와 웹을 결합해 출처와 함께 응답한다고 소개된다(Quarnstrom, 2025).

그러나 RAG를 적용해도 환각이 사라지는 것은 아니다. 법률 분야에서 평가 계획을 미리 등록하고 수행한 Magesh 외(2025)의 평가에서 RAG 기반 법률 AI 도구는 범용 챗봇(GPT-4)보다 환각이 줄었지만 각각 17~33%의 비율로 환각을 일으켰다. 이 평가는 202개 질의로 한 시점의 결과만 포착했다.

CS50의 단순 계산 값과 법률 분야의 수치를 학습 도구에 그대로 옮길 수는 없지만, 두 근거는 RAG가 환각을 줄이는 수단일 뿐 없애는 장치는 아님을 시사한다. 학습 도구의 신뢰성은 자료 기반 응답, 원자료 출처 표시, 학습자가 출처를 직접 확인하게 하는 설계를 함께 보아 판단하는 것이 필요하다.

다. 데이터 처리와 관리자 기능

OpenAI는 2024년 5월 ChatGPT Edu를 발표하며 이 상품이 엔터프라이즈 수준의 보안과 통제를 갖추었고, ChatGPT Edu의 대화와 데이터를 OpenAI 모델 학습에 쓰지 않는다고 밝혔다(OpenAI, 2024). 다만 기관 고객용 개인정보 문서는 ChatGPT Edu 등 기관 상품의 데이터를 '기본값으로' 모델 학습에 쓰지 않는다고 적는다. 이 문서에 따르면 관리자가 데이터 보관 기간을 정하고 삭제된 대화는 법적 보존 의무가 없으면 30일 안에 OpenAI 시스템에서 제거되며, 관리자는 Enterprise Compliance API로 대화와 GPT의 감사 로그에 접근할 수 있다(OpenAI, 2026c). ChatGPT Edu는 전용 워크스페이스와 관리 콘솔, 그룹 권한·GPT 관리 기능을 제공한다(OpenAI, 2026a).

Google은 Gemini for Education 발표에서 이 서비스가 Workspace 핵심 서비스로서 엔터프라이즈급 데이터 보호와 관리자 통제를 갖추며 모든 교육 사용자의 대화 데이터를 AI 모델 개선에 쓰지 않는다고 밝혔다. 관리자는 관리 콘솔에서 Gemini 앱 접근을 관리하고 Vault로 도메인 내 대화를 검색하며 사용 보고서를 볼 수 있다(Kirtikar, 2025). Anthropic도 2025년 7월 Claude for Education의 대화가 기본적으로 비공개이며 기본적으로 AI 학습에서 제외되고, 기관의 데이터 요청에는 공식 승인이 필요하며 셀프서비스 데이터 내보내기는 기본적으로 제한된다고 밝혔다(Anthropic, 2025b).

개인정보보호위원회(개인정보위)의 생성형 AI 개인정보 처리 안내서의 비교표에서 개인용 라이선스는 기본적으로 AI 학습에 활용(opt-out 가능)되고 기업용 라이선스는 기본값으로 학습에 활용되지 않는다(개인정보보호위원회, 2025). 이 비교표는 2025년 8월 3일 기준 OpenAI ChatGPT 중심으로 의료기관 사례 맥락에서 정리한 것이어서, 대학 기관 계약에는 유추 근거로만 쓰고 기관판 조건은 각 사의 공식 문서로 다시 확인했다.

기관판의 '학습 제외' 문구는 기본값 여부와 대상 표현('모델 학습', 'AI 모델 개선', 'AI 학습')에서 사업자마다, 때로는 같은 사업자의 문서끼리도 다르게 표현되어 있다. 따라서 '모델 학습에 쓰지 않는다'는 문장만으로 조건이 같다고 보기 어렵고, 기관 도입 시에는 학습 제외 범위, 보관·삭제 주기,

사람의 검토 여부, 감사 로그 접근 범위를 계약 문서에서 점검하는 것이 바람직하다(IV 장 1 절 나, 개인 계정은 II장 4 절 마).

라. LMS 연동과 에이전트·확장 기능

1EdTech 컨소시엄의 학습 도구 상호운용성(LTI)은 학습관리시스템(LMS) 연동의 기술 표준으로, 학습 도구를 도구별 로그인 없이 기관의 학습 환경에 연결한다. LTI Advantage 의 과제·성적 서비스(AGS)는 외부 도구가 성적부 열을 만들고 성적을 게시하게 한다(1EdTech Consortium, 2026).

Google 은 2024 년 12 월 9 일 타사 학습관리시스템(LMS)에 통합되는 AI 도우미 Gemini LTI 를 발표했고, Gemini LTI 는 처음에 Canvas 와 Schoology 에 연동되었다(Google, 2024). 2025 년 6 월에는 Gemini LTI 개선으로 교사가 수업 자료 기반 Gem 과 노트북을 Schoology·Canvas 에서도 학생에게 배정할 수 있게 된다고 밝혔다(Kirtikar, 2025). 2026 년 5 월에는 Gemini LTI 의 지원 범위를 기존 Canvas·Schoology 에 더해 Moodle 로 넓혔다(Google, 2026a).

Microsoft 는 2025 년 10 월 Copilot Chat 이 엔터프라이즈 데이터 보호와 IT 통제를 갖췄다고 소개했다. 이 Copilot Chat 은 Microsoft 365 LTI 로 Canvas·Schoology·Brightspace·Blackboard·Moodle 에 연결될 예정(12 월 프리뷰)이었다(Quarnstrom, 2025). Anthropic 은 2025 년 7 월 9 일 Claude 가 Canvas LTI 를 지원해 학생이 Canvas 강좌 안에서 Claude 를 바로 쓸 수 있게 되었다고 발표하면서, Wiley 와 Panopto 는 미리 만든 MCP 서버로 Claude 에 연결될 예정이라고 밝혔다(Anthropic, 2025b).

OpenAI 쪽에서는 공식 쿡북의 Canvas 연동 레시피(2024 년 9 월)가 맞춤형 GPT 에 Custom Action 을 붙여 Canvas REST API 로 Canvas 와 연결하는 방법을 안내했다(Schule & Casson, 2024). Instructure 와 OpenAI 는 2025 년 7 월 23 일 교수자가 Canvas 안에서 맞춤형 GPT 와 비슷한 학습 경험을 만드는 LLM 활용 과제를 도입하는 협력을 발표하고, 학습자 정보는 OpenAI 와 공유되지 않는다고 밝혔다(Instructure, 2025). 앞의 방식은 Canvas API 를 직접 호출하는 비표준 연동이고 LLM 활용 과제는 Canvas 의 기능이어서, 둘 다 ChatGPT Edu 자체의 LTI 연동은 아니다.

학습 기능은 대화창에서 LMS·에이전트로 넓어지고 있으며(일부는 발표 단계), 표준 연동 여부가 새로운 도입 변수가 될 가능성이 있다. 공개 문서에서 ChatGPT Edu 의 LTI 지원은 확인되지 않았고, Copilot Chat 과 Claude 의 연동은 발표 뒤 실제 제공 여부를 확인하지 않았으며, Claude LTI 로 학습 모드를 강좌 단위로 배정하는 방법도 발표문에 없다. 계약 전에는 LTI 지원과 발표된 연동의 실제 제공 여부를 확인할 필요가 있다.

마. 독립 평가가 보여 준 '정답 보류'의 한계

정답 보류 전략은 LLM 튜터 독립 평가에서 한계를 보였다. Maurya 외(2025)는 수학 오류 교정 대화를 대상으로 7 개 LLM·인간 튜터의 192 개 대화와 1,596 개 응답에 8 개 교수 차원별 주석을 단 MRBench 를 공개했다. 인간 평가에서 정답을 노출하지 않은 응답의 비율은 모델에 따라 53.12%에서 94.79%까지 크게 달랐고, 저자들은 최신 LLM 이 아직 AI 튜터로서 충분히 좋지 않으며 개선 여지가 크다고 결론지었다.

실제 강좌에 배치된 도구는 프롬프트 지시 외에 별도의 방어 장치를 두었다. CS50 개발팀은 GPT-4 를 속여 숙제 정답 전체를 받아내려는 프롬프트 주입 공격을 막기 위해 학생 요청을 점검하는 '가드' 기능을 두었다(Liu et al., 2024). 대형 프로그래밍 강좌에 배치된 CodeAid 는 모델이 생성한 코드를 보여 주지 않고 별도 LLM 함수로 의사코드로 바꿔 표시했지만, 37 건(2%)에서 짧은 코드 정답이 생성되었다(Kazemitabaar et al., 2024).

MRBench 는 2024 년 당시 모델 수준의 평가여서 그 결과를 상용 학습 모드에 그대로 적용하기는 어렵다. 그럼에도 정답 보류 수준이 기반 모델에 따라 크게 달랐고 강좌 배치 도구마저 요청 점검·출력 변환 장치를 따로 두었다는 점에서, 정답 보류는 프롬프트 지시만으로는 견고하지 않을 수 있다. 이런 장치의 효과는 비교 측정되지 않았다. 상용 학습 모드가 실제로 정답을 보류하는지는 실사용 진단으로 확인할 항목이며, 이번 호는 이를 측정하지 않았다(Ⅲ 장 2 절).

[표 Ⅱ-1] 학습 모드·기관용 상품의 기술 구조 비교

구분	ChatGPT 학습 모드 ·Edu(OpenAI)	Gemini 가이드 학습 ·for Education(Google)	Claude 학습 모드·for Education(Anthropic)	Study and Learn·Copilot(Microsoft)
구현 방식·배치 형태	구현: 맞춤형 시스템 지시(모델 학습 계획)	구현: 교수법 지시를 따르도록 조정한 모델(LearnLM)	배치: Projects 안의 기능(구현 방식 미확인)	배치: Copilot 에이전트(구현 방식 미확인)
학습 지원 방식	소크라테스식 질문·힌트·자기성찰 프롬프트	탐색적·개방형 질문, 단계별 설명	답 대신 학생의 추론 과정 안내	개념 이해·연습·퀴즈 등 내장 활동
자료 기반 응답	미확인	NotebookLM: 업로드 자료 근거 생성·원자료 인용	미확인	Microsoft 365 Copilot(유료): 사용자 데이터·웹 결합 응답과 출처 제시(제공사 표현)
데이터 보호(기관판)	기본값 학습 미사용, 삭제 대화 30 일 내 제거	교육 사용자 대화 AI 모델 개선 미사용, 엔터프라이즈급 데이터 보호	대화 기본 비공개·학습 기본 제외(2025. 7. 발표)	'엔터프라이즈 데이터 보호'(Copilot Chat, 제공사 표현)
관리자 기능	관리 콘솔, 그룹 권한·GPT 관리, 감사 로그(Compliance API)	관리 콘솔(Gemini 앱 접근 관리), Vault 대화 검색, 사용 보고서	기관 데이터 요청 공식 승인, 셀프서비스 내보내기 기본 제한(2025. 7. 발표)	'IT 통제'(Copilot Chat, 제공사 표현)
LMS 연동	LTI 지원 미확인, 쿡북의 Canvas API 연동 안내(2024. 9.), Instructure 와 Canvas 안 LLM 활용 과제 협력(2025. 7.)	Gemini LTI(Canvas·Schoology 2024. 12.~, Moodle 2026. 5.~)	Canvas LTI 지원(2025. 7. 발표, 학습 모드 강좌 배정 미확인)	Copilot Chat 의 Microsoft 365 LTI(2025. 10. 발표 당시 12. 프리뷰 예정)
확장·에이전트	과목·과제용 맞춤형 GPT 워크스페이스 공유, Custom Action(쿡북 안내)	수업 자료 기반 Gem·노트북 배정(2025. 6. 발표)	Wiley·Panopto MCP 연결(2025. 7. 발표 당시 예정)	Study and Learn 에이전트(2026. 5. 정식 제공)

자료: OpenAI·Google·Anthropic·Microsoft·Instructure 공식 발표·도움말 문서(2024~2026)를 바탕으로 연구진 정리

주: '미확인'은 이 보고서가 근거로 삼은 공식 문서에서 해당 내용을 확인하지 못했다는 뜻이며, 기능이 없다는 뜻은 아니다. '예정'으로 적은 항목은 발표 당시의 계획이며 이후 제공 여부는 확인하지 않았다.

소결 학습 모드는 시스템 지시와 학습용 모델 조정의 두 방식으로 구현되고, 작업 공간·에이전트 안에 배치되기도 한다. 다만 정답 보류는 구현 방식만으로 보장되지 않으며, 독립 평가(2024 년 모델 수준)에

서는 그 수준이 기반 모델에 따라 크게 달랐다(상용 학습 모드는 이번 호에서 측정하지 않음). 대학이 도입 전에 확인할 기술 요소는 구현 방식과 일관성 검증 자료, 업로드 자료 기반 응답과 출처 표시, 학습 제외·보관·인적 검토에 관한 데이터 처리 문구, 관리자 통제와 감사 로그, LTI 같은 표준 연동의 다섯 가지로 정리할 수 있다. 이는 핵심 질문 Q1 의 답으로 IV장 3 절에서 종합한다.

2. 시장 동향

가. 시장 규모: 추정치의 큰 편차와 국내 기준선

세계 AI 교육 시장의 규모는 추정 기관에 따라 크게 다르다. MarketsandMarkets(2024)는 세계 AI 교육 시장이 2024 년 22.1 억 달러에서 2030 년 58.2 억 달러로 연평균 17.5% 성장한다고 추정했다. 반면 Mordor Intelligence(2025)는 같은 이름의 시장을 2025 년 69.0 억 달러, 2030 년 410.1 억 달러(연평균 42.83%)로 추정했으며, 이 추정에서 2024 년 AI 교육 수요의 45.00%를 고등교육이 차지했다.

국내에서 참고할 수 있는 기준선은 국가승인통계인 이러닝산업 실태조사다. 2024 년 국내 이러닝 매출액은 5 조 8,521 억 원으로 2023 년(5 조 5,946 억 원)보다 4.6% 늘었고, 학교급별 이러닝 지출 비중은 4 년제 대학교가 29.4%(1,007 억 원)로 가장 높았다(산업통상자원부·소프트웨어정책연구소, 2025). 다만 이 통계는 콘텐츠·솔루션·서비스 부문을 아우르는 이러닝 산업 전체를 다루므로, LLM 기반 학습 도구 시장과는 정의가 다른 맥락 지표로 읽어야 한다.

두 추정치는 유료 보고서의 공개 요약에서 확인한 값이어서 산정 방법과 표본은 확인할 수 없었고, 2030 년 전망치가 크게 벌어지는 것은 시장 정의나 산정 방법의 차이 때문으로 보여 두 값을 견주거나 평균을 내기는 어렵다. 이 보고서는 시장 수치를 정의와 함께 범위로만 제시하며([표 II-2]), 대학 단위의 비용 규모는 II장 2 절 다·라의 조달 문서로 확인한다.

나. 공급 구조와 가격: 기관 라이선스·개인 구독·학생 혜택

공급 방식은 기관 라이선스, 개인 구독, 학생 대상 혜택으로 나누어 볼 수 있다. OpenAI 는 2024 년 5 월 30 일 대학이 학생·교수·연구자·캠퍼스 운영 부서에 AI 를 배포하도록 만든 ChatGPT Edu 를 발표하면서 교육기관이 감당할 수 있는 가격이라고 소개했다(OpenAI, 2024). Google 은 Gemini for Education 을 전 세계 교육기관 학생·교육자에게 추가 비용 없이 제공한다고 밝혔다(Pichai, 2025). II 장 1 절 라에서 본 Microsoft 의 Copilot Chat 은 Microsoft 365 에 추가 비용 없이 포함된다. 유료 상품인 Microsoft 365 Copilot 의 교육용 학술 요금은 교직원과 13 세 이상 학생 사용자당 월 18 달러(2025 년 12 월부터, 지역별 가격은 다름)다(Quarnstrom, 2025).

학생 개인을 겨냥한 한시 혜택(I 장 3 절)도 있었지만 개인 유료 구독은 아직 일부에 그친다. 2025 인터넷 이용 실태조사에서 생성형 AI 를 유료로 구독하는 비율은 7.9%였고 챗지피티(ChatGPT, 7.3%)가 가장 높았다(과학기술정보통신부·한국지능정보사회진흥원, 2026). 이 수치는 대학생이 아니라 전 국민을 대상으로 한 대리 지표다.

이처럼 시장은 기관 라이선스와 개인 구독의 이원 구조를 띤다고 볼 수 있다. 이런 학생 대상 혜택은 신청 기한과 대상 국가가 정해진 한시적 조건이어서, 개인 차원의 접근은 제공사의 조건 변경에 따라 달라질 수 있다. 또한 제공사마다 가격 공개 방식이 달라 공개 자료만으로 기관용 상품의 단가를 견주기는 어려운 것으로 보인다.

다. 단일 벤더형 기관 계약: 구조와 쟁점

단일 벤더형은 대학이 한 제공사의 서비스를 직접 또는 국내 공급사를 거쳐 기관 단위로 계약해 전 구성원에게 제공하는 방식이다. 미국 캘리포니아주립대(CSU)는 2025 년 OpenAI 에 1,700 만 달러를 지불하고 전 캠퍼스에 교육용 ChatGPT 무제한 접근을 제공했다(Corzo, 2026). CSU 는 2026 년 6 월 말 만료되는 이 18 개월 계약을 갱신해 학생 47 만여 명과 교직원 6 만 3,000 명의 체제 전체 접근을 위해 3 년간 연 1,300 만 달러를 지불한다(Rix, 2026). 계약 당사자인 CSU 의 설문(응답 9 만 4,060 명)에서 응답자의 ChatGPT 사용률은 학생 84%, 교수 87%, 직원 89%였으나, 이 설문은 자발적 참여라 전체 구성원을 대표하지 않을 수 있다(California State University, 2026).

국내에서는 서울대가 같은 방식으로 ChatGPT Edu 를 전 구성원에게 도입했는데, 이는 대학 차원의 통합 지원 체계를 마련하고 높은 구독 비용으로 고급 AI 서비스를 쓸 수 없는 학생에게도 무료로 제공하려는 사업이었다(서울대학교, 2026). 제안요청서에 실린 2026 년 3 월 구성원 설문에서 도입 희망 AI 서비스는 ChatGPT 30.4%, Gemini 31.7%, Claude 32.7%, LLM 플랫폼(종합 제공) 2.2%로, 세 제품에 대한 선호가 비슷하게 나뉘었다. 제안요청서는 학생·교원·연구원 약 3 만 9,000 명을 대상으로 구성원 선호를 반영한 상용 생성형 AI 서비스를 제공하는 사업에 총사업비 3,288,986,000 원(부가가치세 포함)을 잡았다. 이용권은 ChatGPT Edu, Google AI Pro for Education, Claude for Education 가운데 하나 이상을 제공하도록 요구했다(서울대학교 정보화본부, 2026).

다른 제공사 제품을 전교 단위로 제공한 사례도 있다. 이후 운영 여부는 확인되지 않았으나, 제공사 발표에 따르면 Anthropic 은 2025 년 4 월 Claude for Education 을 출시하면서 노스이스턴대·런던정치경제대(LSE)·챔플레인칼리지와 전교 단위 이용 계약을 맺었다(Anthropic, 2025a).

단일 벤더형은 대규모 일괄 계약으로 구성원 사이의 접근 격차를 줄이려는 선택으로 볼 수 있다. 그러나 서울대 설문처럼 구성원의 선호가 여러 제품으로 나뉘는 상황에서 한 제공사를 고르면 구성원 선호와의 불일치, 특정 사업자 종속, 비용 정당화 문제가 함께 따를 수 있다. 다만 여러 모델을 묶은 LLM 플랫폼은 같은 설문에서 희망 응답이 2.2%에 그쳐, 선호가 나뉜 것이 통합 플랫폼 선호를 뜻하지는 않는다. 따라서 구성원 선호, 이용 규모와 비용, 갱신 조건을 검토하는 것이 바람직하다(IV 장 1 절 라, CSU 논쟁은 II장 4 절 아).

라. 멀티 LLM 통합 플랫폼: 구조와 쟁점

멀티 LLM 통합형은 대학이 공급사의 플랫폼을 구독하고 구성원이 그 안에서 여러 제공사의 모델을 골라 쓰게 하는 방식이다. 순천향대 규격서는 고비용 LLM 계정을 개별 제공하면 대규모 예산이 들고 개인별 편차로 낭비가 크다는 점을 통합 플랫폼 도입 배경으로 들고, 하나의 크레딧으로 여러 LLM 을 고르고 모델별 토큰 소모량에 따라 크레딧이 차감되는 과금 구조를 요구했다. 사업예산은 60,000,000 원(부가가치세 포함)이었다(순천향대학교, 2025). 이 금액은 공고에 적힌 예산이며, 최종 계약 금액은 확인되지 않았다.

마인드로직의 FactChat 은 공급사 자체 문서인 개인정보처리방침(2026 년 5 월 28 일 시행본)에서 이용자가 글로벌 LLM 모델을 골라 질의하는 시점에 데이터를 암호화해 전송한다고 밝힌다(마인드로직, 2026). 본교 AI 에듀테크센터가 운영하는 HAI-GPT(이해관계는 IV장 2 절 가)도 FactChat 기반으로 ChatGPT·Gemini·Claude 등을 한 홈페이지에서 제공하며, 재학생에게 월 5,000 크레딧, 교직원에게 월 2 만 크레딧이 무료로 제공된다(강지오, 2026). 사용량은 이처럼 크레딧 한도로 관리되며, 순천향대 규격서도 그룹·개별 사용자별 월간 사용 한도를 설정해 정해진 예산 안에서 운영하는 관리자 기능을 요구했다(순천향대학교, 2025).

멀티 LLM 통합형은 크레딧 과금과 사용자별 한도로 비용을 관리하면서 모델 선택의 폭을 넓히는 방식이라 할 수 있다. 다만 질의 데이터가 공급사를 거쳐 해외 모델 제공사로 전달되는 흐름은 별도의 관리 과제가 될 수 있다. 제공사 앱의 학습 모드 같은 기능이 통합 플랫폼에서도 같은 방식으로 제공되는지도 도입 전에 확인해야 한다. 이 소절의 국내 사례는 조달 문서 한 건, 학내 보도 한 건, 공급사 처리방침에 기대고, 조달 건의 플랫폼도 HAI-GPT 와 같은 FactChat 기반(연구진이 대학 공지로 확인)이어서 사실상 한 공급사에 몰려 있다. 따라서 유형 전체로 일반화하기 어렵고, 중개 공급사 의존은 단일 벤더형의 특정 사업자 종속과 같은 성격의 위험이 될 수 있다(두 모델 비교는 IV장 1 절 라).

마. 이용 실태와 기관 지원의 간극

Ⅰ장 1 절에서 본 영국 HEPI 조사에서 소속 기관이 AI 도구를 제공한다는 학생 응답은 2024 년 9%, 2025 년 23%, 2026 년 38%로 늘었지만 기관이 제공해야 한다고 보는 학생(50%)에는 못 미쳤다 (Stephenson & Armstrong, 2026). 국내에서는 KEDI 조사(2025 년 1 학기)에서 대학으로부터 AI 도구 사용 교육·지원을 받은 경험이 있다는 학생은 37.7%, 대학에 명확한 AI 지침·정책이 있다는 학생은 32.1%에 그쳤다(김지하 외, 2025). 두 조사는 국가와 대상, 시점, 문항이 달라 수치를 직접 비교하지 않는다.

기관의 도구 제공과 교육·지침은 Ⅰ장 1 절에서 본 학생 이용 수준에 아직 못 미치는 것으로 해석된다 (도구 제공과 교육적 활용의 관계는 Ⅱ장 4 절 아).

[표 Ⅱ-2] 시장 지표(수치·기준 시점·출처·정의)

지표	수치	기준 시점	출처	정의·범위·주의
세계 AI 교육 시장(추정 ①)	22.1 억 → 58.2 억 달러(연평균 17.5%)	2024 → 2030	MarketsandMarkets	LMS·적응형 학습 플랫폼·챗봇·표절 탐지 도구 등 소프트웨어 포함
세계 AI 교육 시장(추정 ②)	69.0 억 → 410.1 억 달러(연평균 42.83%)	2025 → 2030	Mordor Intelligence	2024 년 수요의 45.00%가 고등교육. 정의·산정 방법을 확인할 수 없어 ①과 직접 비교 불가
국내 이러닝 매출액	5 조 8,521 억 원(전년 대비 4.6% 증가)	2024	이러닝산업 실태조사(국가승인통계)	이러닝 산업 전체로 AI 교육 시장과 정의가 다름
정규교육기관 이러닝 지출액	3,428 억 원(전년 대비 1.9% 증가, 4 년제 대학교 1,007 억 원·29.4%)	2024	이러닝산업 실태조사	학교급별 비중 1 위는 4 년제 대학교
생성형 AI 유료 구독 비율	7.9%(ChatGPT 7.3%)	2025	인터넷 이용 실태조사	만 3 세 이상 전 국민 대상 대리 지표
대학 체제 계약(CSU)	연 1,300 만 달러 × 3 년(직전 18 개월 1,700 만 달러)	2026 갱신	언론 보도(Rix, 2026)	학생 47 만여 명·교직원 6 만 3,000 명 대상
대학 단일 서비스(서울대)	총사업비 3,288,986,000 원(약 32.9 억 원)	2026 공고	제안요청서	약 3 만 9,000 명 대상, 이용권 1 종 이상 요구. 계약 금액과 다를 수 있음
멀티 LLM 플랫폼(순천향대)	60,000,000 원(0.6 억 원, 최대 1 만 계정·4 개월)	2025 공고	입찰공고	공고 예산, 계약 금액 미확인

자료: MarketsandMarkets(2024), Mordor Intelligence(2025), 산업통상자원부·소프트웨어정책연구소(2025), 과학기술정보통신부·한국지능정보사회진흥원(2026), 각 대학 조달 문서와 언론 보도를 바탕으로 연구진 정리

소결 시장 규모 추정치는 조사 기관에 따라 크게 달라 범위로 읽어야 하며, 국내 통계도 LLM 기반 학습 도구만의 규모는 보여 주지 않는다. 학생 대상 혜택은 한시적일 수 있어 기관 차원의 접근 보장 여부가 대학의 판단 사항이 되는 것으로 보인다. 대학 도입은 단일 벤더형과 멀티 LLM 통합형으로 나뉘며, 어느 쪽이든 비용·사용량과 데이터의 전달·보관이 핵심 관리 대상이 될 가능성이 크다. 기관의 도구 제공·교육·지침은 학생 이용 수준에 아직 못 미친다. 핵심 질문 Q2 는 IV장 1 절 라에서 다시 검토한다.

3. 교육적 적용

가. 대학 내 활용 유형

Ravšelj 외(2025)는 109 개 국가·지역의 대학생 23,218 명을 조사했고, 학생들은 주로 아이디어 발상·텍스트 요약·논문 찾기에 ChatGPT 를 썼다. 다만 편의표집과 자기보고에 의존한 조사여서 일반화에 주의가 필요하다. 영국 전일제 학부생 1,054 명 조사(2025 년 12 월)에서는 가장 흔한 AI 용도가 개념 설명, 학술 자료 요약, 아이디어 구조화였고, 학부생의 약 3 분의 2 가 평가 과제 준비에 AI 로 개념 설명을 받았다(Stephenson & Armstrong, 2026). 국내에서는 한국직업능력연구원이 생성형 AI 이용 경험이 있는 대학 재학생 726 명(2024 년 1 학기 기준)을 조사한 결과 응답 대학생의 77.8%가 글쓰기 과제에 AI 를 활용한 경험이 있었으나, 표본이 이용 경험자로 한정되어 전체 대학생의 이용률을 뜻하지는 않는다(윤혜준 외, 2025).

활용 방식은 전공과 학습자 수준에 따라서도 달랐다. 서울 소재 4 년제 사립대학 재학생 573 명을 조사한 연구에서 이공계는 코딩 해결, 인문·사회계는 에세이 작성·요약에 AI 를 주로 활용해 전공별 활용 양상과 표절 인식이 달랐다(박돈하·이효숙, 2026). 대학 글쓰기 강좌 수강생 40 명의 사례 연구에서는 ChatGPT 답변을 원고에 실제 반영한 비율이 48.8%였고 하위 집단(73.2%)이 상위 집단(42.6%)보다 훨씬 높았으며, 활용 유형의 74.1%는 답변을 의미적으로 변형하지 않았다. 상위 필자는 미흡한 답변을 폐기한 반면 하위 필자는 미흡성을 인지하지 못하고 수용하는 경향이 높았다. 단일 대학 40 명의 사례여서 일반화에는 한계가 있다(이윤빈, 2024).

국내외 조사와 사례 연구로 미루어 보면 대학생의 활용이 아이디어 발상·개념 설명·요약·글쓰기·코딩과 평가 과제 준비에 걸쳐 있고, 답변을 거의 그대로 옮기는 사용도 적지 않은 것으로 읽힌다. 학습 모드는 이 가운데 개념 이해를 돕는 용도를 겨냥한 기능으로 볼 수 있어, 효과와 위험도 활용 유형별로 나누어 살펴야 한다.

나. 메타분석이 보여 준 효과의 범위와 조건

메타분석은 대체로 큰 정적 효과를 보고한다. 대학생 연구 66 편을 종합한 Mo 외(2026)의 메타분석(g = 1.14)에서 ChatGPT 를 처음 접한 참여자 연구의 평균 효과크기는 g = 1.73 으로 특히 컸다. Deng 외(2025)는 2022~2024 년 실험·준실험 연구 69 편(대학 84.06%, 초·중등 14.49%)을 분석해 ChatGPT 개입이 학업성취(g+ = 0.712)와 고차사고 성향(g+ = 0.703)을 높였다고 보고했다.

Chen 과 Cheung(2025)은 57 편 연구·97 개 추정치를 종합해 생성형 AI 가 대학생 학습성과에 큰 효과(g+ = 0.804)를 보였다고 보고했다. 다만 생성형 AI 는 메타인지에는 유의한 효과를 보이지 않았고, 신입생과 언어 학습자, 검사로 측정한 결과, 4~12 주 개입, 학술지 게재 연구에서 효과크기가 더 컸다. 백다은 외(2024)가 학령기(K-12)와 성인기를 포함한 국내 연구 24 편을 분석한 결과 효과크기는 통제-처치 설계가 고정효과모형에서 0.4262, 사전-사후 설계가 무선효과모형에서 0.7646 이었다.

그러나 66 편 메타분석에서는 효과크기 간 이질성이 매우 컸고(I^2 = 97.62%) Egger 검정에서 출판 편향의 증거가 확인되었으며, 표본이 작은 연구일수록 불균형하게 큰 정적 효과를 보고했다(Mo et al., 2026). 69 편 메타분석에서도 검정력 분석을 수행한 연구는 62 편 중 5 편(8.06%)뿐이었다(Deng et al., 2025). Weidlich 외(2025)는 이 메타분석에 포함된 학업성취 비교 19 건이 처치, 통제집단, 결과 측정에 관한 세 가지 기본 고려사항을 충족하는지 감사했는데, 세 기준을 모두 충족한 비교는 4 건(21%)뿐이었다. 저자들은 현재로서는 관찰된 이득을 ChatGPT 에 확신 있게 귀속할 수 없고 메타분석 효과크기는 효과를 과대 또는 과소 추정할 수 있다고 보았다.

메타분석의 큰 효과크기는 신중하게 읽어야 한다. 66 편 메타분석은 출판 편향과 소표본 효과로 평균 효과를 과대 추정했을 수 있지만, 포함 연구의 설계 문제는 효과를 어느 쪽으로도 왜곡할 수 있어 방향을 단정하기에는 무리가 있다. 처음 접한 참여자의 큰 효과에는 신기성 효과가 섞였을 수 있고, 국내 메타분석의 설계별 차이도 추정 모형이 달라 직접 비교하기 어렵지만 설계에 따라 효과가 커 보일 수 있음을 시사한다. 따라서 효과크기를 도입 근거로 삼을 때에는 통제집단이 있는지, AI 없이 측정했는지, 일정 기간 뒤에도 측정했는지를 먼저 확인해야 한다.

다. 설계된 튜터와 제한 없는 사용: 실험 연구

제한 없는 사용과 설계된 튜터를 직접 비교한 실험은 연습 중의 성과와 실제 학습이 서로 다를 수 있음을 보여 준다. 튀르키예 고등학생(K-12)을 대상으로 한 Bastani 외(2025)의 대규모 무작위 통제 실험(RCT)에서는 문제 풀이 중 GPT-4 를 쓰게 하자 GPT Base 와 GPT Tutor 집단 모두 성적이 크게 올랐다. 그러나 가드레일 없는 GPT Base 집단은 AI 없이 치른 시험에서 통제집단보다 성적이 17% 낮았고, 교사가 설계한 힌트를 주도록 지시한 GPT Tutor 집단의 시험 성적은 통제집단과 통계적으로 구별되지 않았다. 이 연구는 수학 단일 주제에 한정되었고 단기 성과만 측정했다. 대학생 대상으로는 Bassner 외(2026)가 독일 TUM 의 입문 프로그래밍(CS1) 강좌 학생 275 명에게 3 집단 무작위 통제 실험을 수행했다. 보정된 힌트를 주되 완전한 해답은 보류하는 스캐폴딩 튜터(Iris) 집단과 완전한 해답까지 주는 무제한 ChatGPT 집단은 모두 통제집단보다 실습 점수가 크게 높았지만, 두 조건 모두 사전-사후 지식 향상이나 코드 이해에서 우위를 보이지 않았다.

H.-Y. Lee 외 연구팀의 소표본 실험 두 건에서는 힌트나 질문으로 안내하는 설계가 일반 ChatGPT 보다 지식 구성이나 인지적 참여에서 나은 결과를 보였다. 대만 한 대학의 학부생 61 명을 대상으로 블렌디드 기초화학 강좌에서 안내형 학습 보조(GCLA)와 일반 ChatGPT 사용을 비교한 무작위 통제 실험에서는 GCLA 집단이 지식 검사의 사후·지연검사 점수에서 앞섰고, 자기보고로 측정한 자기조절학습과 고차사고도 높았다. 다만 한 대학의 편의표본이고 지연검사 시점 등 일부 보고가 논문 안에서 서로 맞지 않는다(H.-Y. Lee et al., 2024). 석사과정생 72 명 대상 H.-Y. Lee 외(2026)의 12 주 무작위 시험에서 일반 ChatGPT 는 즉각 지원으로 행동적 참여와 정서적 편안함을 높였고, 소크라테스식 시스템은 분석적 사고와 지식 구성 같은 인지적 참여를 더 효과적으로 촉진했다. 결과는 성취도가 아닌 참여·인식 지표였다.

학생의 선호와 학습 효과가 어긋날 수 있다는 점도 확인되었다. 앞의 프로그래밍 실험에서 학생들은 ChatGPT 를 더 쉽고 도움이 된다고 평가했으며, 저자들은 해답을 무제한 제공하는 AI 가 학생 선호와 교수적 효과가 어긋나는 '편안함의 함정(comfort trap)'을 낳는다고 결론지었다(Bassner et al., 2026). 대학생 글쓰기 실험에서도 ChatGPT 집단은 에세이 점수 향상에서만 앞섰는데, 저자들은 이 점수 우위가 글 작성을 제한하는 프롬프트에도 학습자가 지시를 우회해 완성 글을 생성했기 때문일 수 있다고 보았다(Fan et al., 2025).

효과를 좌우하는 요인으로는 도구 자체보다 사용 방식이 지목된다. Weidlich 외(2025)는 ChatGPT 를 어떻게 사용했는지가 효과의 주된 결정요인이라는 초기 증거가 있다고 보았다. 다만 대학생 66 편 메타분석에서 안내(guidance) 제공 여부의 조절 효과는 하위 집단 신뢰구간이 넓어 확정하기 어려웠다 (Mo et al., 2026).

이 실험들의 튜터는 각 연구팀이 설계한 것으로 상용 학습 모드를 직접 시험한 것은 아니다. 정답 보류 설계가 제한 없는 사용의 학습 저해를 줄인다는 직접 근거는 고등학생 대상 현장실험(K-12)이고, 대학생 대상 프로그래밍 실험에서는 무제한 ChatGPT 집단도 지식 향상에서 뒤지지 않았고 강좌 안에서 운영한 설계된 튜터의 우위도 일관되지 않았다. 과제·평가 연계는 어느 실험에서도 변인으로 검증되지 않았으므로, 학습 모드를 켜는 것만으로 효과를 기대하기는 어렵고 과제·평가 연계의 효과는 학내 측정으로 확인할 과제로 보인다.

라. 상용 학습 모드 자체의 근거: 대학 실험 연구의 부재

연구진이 OpenAlex·Crossref·arXiv 에서 주요 사업자의 학습 모드 제품명으로 검색한 범위에서는 기준일까지 상용 학습 모드의 학습 효과를 대학생 대상으로 검증한 실험·준실험 연구를 찾지 못했다. 색인되지 않았거나 이후 발표될 연구는 있을 수 있다.

이 보고서가 검토한 독립 연구 가운데 학습 모드를 직접 다룬 연구는 한 건이며, 이 연구도 학습 효과가 아니라 설계와 기능을 평가했다. 고교 진로·진학 플랫폼과 같은 틀로 공개 자료를 평정한 국내 연구 (K-12 맥락)에서는 7 개 플랫폼 중 ChatGPT 학습 모드만 상호작용·탐색 영역에서 만점을 받았지만 정보 신뢰성·근거는 2/8 점, 학교 맥락·교사 연계는 0/4 점이었다. 저자들은 사용 방식에 따라 기능이 달라지는 범용 생성형 AI 는 학생·교사 대상 인터뷰·관찰·실험 연구로 보완해야 한다고 밝혔다(차덕원·홍후조, 2026).

제공사가 직접 수행한 실험도 있다. LearnLM Team 과 Fab AI(2026)는 시에라리온 중학교에서 사전등록된 2 집단 무작위 통제 실험을 수행해, Gemini 가이드 학습을 사용한 학급의 수학 성취가 의도처치(ITT) 기준 +0.258 표준편차 향상되었다고 보고했다. 그러나 이는 Google 연구팀이 협력 기관과 수행한 제공사 자체 보고이자 교사가 주간 수학 수업의 절반에 가이드 학습을 넣도록 지시받은 수업과 결합된 K-12 결과이며, 보고서는 동료심사를 거치지 않은 프리프린트라고 밝혔다(LearnLM Team & Fab AI, 2026). 같은 제공사의 LearnLM 기술보고서도 평가 기준의 원리가 근거 기반이어도 그 결과가 실제 학습성과 향상으로 이어지는지는 불분명하다고 밝혔다(LearnLM Team, 2024).

제공사 실험은 대상·맥락·검증 수준이 대학과 달라 이 보고서의 핵심 근거로 삼지 않았다. 대학 대상 근거가 쌓이기 전까지는 학습 모드를 도입하는 대학이 통제 비교와 지연 측정을 포함한 자체 효과 측정을 함께 설계하는 편이 바람직하다.

마. 국내 실험 연구

이 보고서가 검토한 범위에서 국내 대학생을 대상으로 한 무작위 실험은 많지 않았다. Sok 과 Shin(2025)은 서울 소재 대학 영어 강좌 학부생 80 명을 통제집단과 두 ChatGPT 처치집단에 무작위 배정했다. ChatGPT 와 구두 요약 과제를 수행한 두 처치집단은 전반적 말하기 과제 수행에서 통제집단보다 유의하게 나은 성과를 보였지만, 전달력과 언어 사용 등 다른 하위 요소에서는 유의한 효과가 없었다. 개입은 50 분 세 차례로 짧았고 저자들은 더 긴 처치 기간의 효과를 후속 연구 과제로 제시했다.

국내 근거는 표본이 작고 처치 기간이 짧아 아직 일반화하기 어렵다. 한편 Ⅱ장 3 절 가의 글쓰기 사례에서 하위 필자는 답변의 미흡성을 인지하지 못한 채 수용하는 경향이 높았다(이윤빈, 2024). 학습자 수준에 따라 같은 도구가 학습을 돕기도 하고 해치기도 할 가능성이 있다는 뜻이다. 국내 대학은 도구 도입과 병행해 답변 검증과 비판적 수용을 가르치는 수준별 지원을 마련하는 것이 바람직하다. Ⅱ 장 3 절에서 다룬 효과 실증 연구는 [표 Ⅱ-3]에 정리했다.

[표 Ⅱ-3] 효과 실증 연구 요약

연구	설계	대상	표본	효과크기·주요 결과	근거 수준
Mo 외(2026)	메타분석	대학 학부생	66 편	g = 1.14, 첫 노출 g = 1.73, I^2 = 97.62%, 출판 편향 확인	메타
Deng 외 (2025)	실험·준실험 메타분석	대학 84.06%, 초·중등 14.49%	69 편(메타분석 62 편)	학업성취 g+ = 0.712, I^2 = 91.789%, 검정력 분석 5/62 편	메타
Weidlich 외 (2025)	메타분석 포함 연구 감사	Deng 외의 학업성취 비교	19 건	세 기준 모두 충족 4 건 (21%)	메타(감사)
Chen 과 Cheung(2025)	메타분석	대학생	57 편, 5,389 명	g+ = 0.804, 메타인지 효과 비유의	메타
백다은 외 (2024)	국내 메타분석	국내 학령기·성인기 학습자 (K-12 포함)	24 편	통제-처치 0.4262(고정), 상관 0.5417(무선), 사전-사후 0.7646(무선)	메타(국내)
Bastani 외 (2025)	현장 RCT(3 집단)	튀르키예 고등학생(9~11 학년)	약 1,000 명	연습 중 성적 +48% (Base)·+127%(Tutor), AI 없는 시험에서 Base 17% 낮음, Tutor 차이 없음	상(K-12)
Bassner 외 (2026)	RCT(3 집단)	독일 TUM CS1 수강생	275 명	실습 점수 향상, 지식 향상·코드 이해 우위 없음	상
Barcaui(2025)	RCT(2 집단)	AI 를 배우는 학부생	무작위 배정 120 명(분석 표본 미확인)	45 일 후 유지 시험 57.5% 대 68.5%, d = 0.68	중
Fan 외(2025)	무작위 실험 (4 집단)	대학생(학부생 55%)	117 명	에세이 점수 향상, 지식 습득·전이 차이 없음	중
Stadler 외 (2024)	무작위 배정 실험	대학생	91 명	인지부하 감소, 추론·논증 품질 저하	중
Sok 과 Shin(2025)	무작위 배정 3 집단 실험	서울 소재 대학 학부생	80 명	말하기 과제 전반 유의 향상, 하위 요소 효과 없음	중(국내)
H.-Y. Lee 외 (2026)	12 주 무작위 시험(3 조건)	석사과정생	72 명	소크라테스식 조건의 인지적 참여 촉진(성취도 미보고)	중

연구	설계	대상	표본	효과크기·주요 결과	근거 수준
H.-Y. Lee 외 (2024)	RCT(2 집단)	대만 학부생	61 명	지식 검사 사후 79.7 대 74.7, 지연 75.9 대 69.3	중(보고 불일치)
LearnLM Team 과 Fab AI(2026)	사전등록 2 집단 RCT	시에라리온 중학생(7·8 학년)	1,763 명	수학 성취 ITT +0.258 표준편차	참고(제공사 자체 보고·동료심사 전 ·K-12)

자료: 각 연구 원문을 바탕으로 연구진 정리

주: 근거 수준은 상 = 동료심사 RCT·표본 250 명 이상, 중 = 동료심사 무작위 실험·표본 250 명 미만, 메타 = 메타분석과 그 포함 연구 감사, 참고 = 제공사 자체 보고·동료심사 전이다. 보고 불일치는 원문 안에서 지연검사 시점 등 일부 보고가 서로 맞지 않는다는 뜻이다. 인지부하와 장기 유지를 다룬 실험 2 건은 Ⅱ장 4 절 나에서 다룬다.

소결 교육적 적용의 근거는 '효과가 있다' 또는 '없다'로 정리하기 어렵다. 메타분석의 큰 정적 효과는 이질성·출판 편향·포함 연구의 설계 문제를 고려해 신중하게 받아들여야 한다. 실험에서는 연습 중 성과 향상이 AI 없는 시험이나 지식 검사의 우위로 이어지지 않은 사례(대학 2 건, K-12 1 건)와 45 일 후 유지 시험에서 ChatGPT 집단이 낮았던 사례(Ⅱ장 4 절 나)가 있었다. 정답 보류 설계가 학습 저해를 줄인 직접 근거는 K-12 현장실험이고, 대학 실험에서는 안내형 설계가 지연검사에서 앞선 소표본 실험도 있었지만 설계된 튜터의 지식 향상 우위는 일관되지 않았다. 상용 학습 모드의 대학생 학습 효과는 연구진이 확인한 범위에서 아직 실험으로 검증되지 않아, 핵심 질문 Q3 전반부의 답은 효과가 설계·측정 시점·사용 방식에 따라 달라진다는 조건부 형태가 될 것으로 보인다.

4. 한계·논쟁점

효과 근거의 질(이질성·출판 편향·논문 철회)은 Ⅰ장 1 절과 Ⅱ장 3 절에서, 규제는 Ⅳ장 1 절 가에서 다루므로 [표 Ⅱ-4]에만 요약한다.

가. 환각과 인용 오류

ChatGPT(GPT-3.5·GPT-4)로 42 개 주제의 짧은 문헌 리뷰 84 편을 만들고 참고문헌 636 건을 분석한 연구에서 GPT-3.5 참고문헌의 55%, GPT-4 참고문헌의 18%가 조작된 것이었다(Walters & Wilder, 2023). 이후 버전을 비교한 독립 평가에서도 완전히 허위인 참고문헌은 ChatGPT 버전에서 1~22%, Gemini 버전에서 0~26%였다(Pastucha et al., 2026). 다만 두 연구는 각각 2023 년 4 월에 생성한 문헌 리뷰와 이비인후과 지침 주제를 다루어 학습 모드에 바로 옮겨 읽기는 어렵다.

RAG 기반 법률 연구 도구를 평가한 연구는 답변이 길수록 모든 명제와 인용을 독립적으로 검증해야 해 확인에 더 많은 시간이 든다고 지적했다(Magesh et al., 2025). 국내 일반대학 학생들은 생성형 AI 사용 시 가장 큰 우려로 제공 정보의 신뢰성(48.8%)을 꼽았다(김지하 외, 2025).

모델이 개선되어도 인용 오류가 남고 확인 부담이 학습자에게 넘어간다는 점이 앞의 연구와 조사 결과에서 읽히므로, 학습 모드를 쓰더라도 출처 확인을 과제 절차의 한 단계로 넣는 방안을 검토할 수 있다.

나. 인지적 외주화와 학습 깊이: 상관과 실험의 구분

상관 연구에서는 AI 도구 빈번 사용과 비판적 사고 능력 간에 유의미한 부적 상관이 있고 인지적 외주화(cognitive offloading)가 이를 매개하는 것으로 나타났다(Gerlich, 2025). Microsoft Research 소속 저자가 다수인 연구팀의 지식노동자 319 명 조사에서는 생성형 AI 에 대한 신뢰가 높을수록

비판적 사고가 덜 나타났다(H.-P. Lee et al., 2025). 생성형 AI 이용 경험이 있는 국내 대학생 조사에서도 'AI 에 의존하고 나면 사고력이 낮아질까 봐 두렵다'는 항목의 긍정 응답이 60.2%로 가장 높았다(윤혜준 외, 2025). 세 연구는 설문·자기보고에 기댄 상관·인식 조사여서 인과를 단정할 수 없고, 앞의 두 연구는 각각 여러 연령층과 지식노동자를 조사해 대학생 표본이 아니다.

실험 연구는 인과 추론에 더 가깝지만 결과는 과제와 측정 방식에 따라 다르다. 대학생 91 명을 ChatGPT-3.5 또는 Google 검색으로 자외선차단제 나노입자 쟁점을 조사하도록 무작위 배정한 실험에서 LLM 을 쓴 학생들은 인지부하가 유의하게 낮았지만 최종 권고의 추론·논증 품질은 검색엔진 집단보다 낮았다(Stadler et al., 2024). AI 를 배우는 학부생 대상 무작위 통제 실험(n = 120)에서는 45 일 후 유지 시험 정답률이 ChatGPT 집단 57.5%로 전통 학습 집단 68.5%보다 유의하게 낮았고 효과크기는 Cohen's d = 0.68 이었다(Barcaui, 2025).

Melumad 와 Yun(2025)은 온라인·실험실 실험 7 건(n = 10,462)에서 LLM 요약으로 학습한 참가자가 웹 검색 대비 더 얕은 지식을 형성했다고 보고했다. 다만 참가자는 주로 일반 성인이며, 효과는 타인에게 조언하려고 주제를 학습하는 특정 과제 맥락에서 자기보고 학습 깊이 등으로 측정되었다. Fan 외(2025)의 무작위 실험 연구도 ChatGPT 같은 AI 기술이 기술 의존을 부추기고 학습자가 메타인지 부담을 덜고 학습 과제와 메타인지 과정을 덜 연결하는 '메타인지적 게으름'을 유발할 수 있다고 지적했다.

II장 3 절 다에서 본 튀르키예 고등학교 수학 수업의 무작위 통제 실험(K-12)에서 가드레일 없는 GPT Base 집단은 시험 성적이 낮았는데도 자신이 덜 배웠다고 인식하지 않았다(Bastani et al., 2025). 영양 의사결정 과제의 세 실험 가운데서는 AI 설명만 제공하고 학습자가 스스로 결론에 도달하게 한 조건(실험 3)에서만 더 정확한 결정과 학습 이득(learning gains)이 함께 나타났다(Gajos & Mamykina, 2022).

요컨대 일부 실험은 LLM 에 의존한 학습이 검색이나 전통 학습보다 학습의 깊이와 지속성을 낮출 수 있음을 시사한다. 정답 보류·설명 중심 설계의 효과를 보인 근거는 의사결정 과제 실험과 K-12 연구가 중심이고, 대학생 대상으로는 보고 불일치가 있는 소표본 실험 1 건(안내형 설계 우위)과 우위가 없었던 실험 1 건이 있어 결과가 일관되지 않다(II 장 3 절 다). 따라서 이런 설계와 체감한 이해가 실제 학습과 다를 수 있다는 안내는 대학 수업에서 효과를 확인하며 도입할 과제로 보인다.

다. 학업 진실성

영국 학부생 대상 민간 조사에서 AI 생성 문장을 평가 과제에 직접 넣었다는 학생은 2024 년 3%, 2025 년 8%에서 2026 년 12%로 늘었다(Stephenson & Armstrong, 2026). II 장 3 절 가에서 본 단일 대학 설문에서는 AI 를 활용하는 학생의 87.5%가 표절 여부에 대한 윤리적 불안감을 느꼈지만, 대학 단계의 표절 예방 교육 경험은 42.7%로 중등 단계보다 낮았다(박돈하·이효숙, 2026).

국내 대학 총장들도 생성형 AI 환경에서 학생 평가 신뢰도를 높일 우선 과제로 '평가 기준 명확화'와 '과정 중심 평가'를 공동 1 순위로 꼽았다(한국대학교육협의회, 2026).

직접 삽입은 늘고 대학 단계의 예방 교육은 부족하므로, 학업 진실성 관리는 AI 사용 사실의 공개와 과정 중심 평가로 옮겨 갈 가능성이 있고 윤리 교육 보강도 요구된다(과제·평가 설계는 IV장 2 절, 탐지 도구는 Vol.08).

라. 형평성: 유료 기능 격차와 접근

미국 대학생 1,001 명 설문에서 4 년제 대학 재학생은 커뮤니티·기술 칼리지 재학생보다 ChatGPT 를 더 자주 사용했고, 비용·연결 문제 등 접근성 우려를 표한 학생들은 가구소득 최저 구간(0~5 만 달러)

출신이었다(Baek et al., 2024). 영국 학부생 조사 보고서는 AI 도구 접근의 형평성 문제가 지속되지만 무료 ChatGPT 가 있어 비용만이 원인은 아닐 수 있다고 보았다(Stephenson & Armstrong, 2026).

2023 년 8 월 기준 GPT-3.5 는 무료였고 GPT-4 는 월 20 달러 유료 구독자에게만 제공되었으므로, II 장 4 절 가에서 본 조작 비율 차이는 무료·유료 모델의 차이이기도 했다(Walters & Wilder, 2023). 109 개 국가·지역 대학생 조사의 저자들은 최신 ChatGPT 버전의 유료화가 저소득 지역·농촌 학생을 구형 도구에 머물게 할 위험을 지적하고, 격차 해소책으로 접근 보조금, 기관 라이선스, 소외 지역 디지털 리터러시 프로그램을 제안했다(Ravšelj et al., 2025).

국내 일반대학 학생의 생성형 AI 활용 수준은 수도권 대학(3.29 점)이 비수도권 대학(3.06 점)보다, 국공립(법인)대학(3.17 점)이 사립대학(3.10 점)보다 유의하게 높았다(김지하 외, 2025).

유료 기능의 품질 차이가 학습 격차로 이어질 위험은 있으나, 횡단 설문의 차이는 인과를 보여 주지 않고 비용 외 요인도 지적되므로 원인을 비용 하나로 단정하기에는 무리가 있다. 기관 라이선스는 대응 수단이 될 수 있지만 사용 지원이 함께 필요하다.

마. 프라이버시: 개인 계정과 기관 계약의 차이

개인정보보호위원회(2024)의 사전 실태점검에서 개선권고 대상 LLM 사업자는 OpenAI·Google·Microsoft·Meta·NAVER·뤼튼이었으며, Anthropic 은 보도자료에 없어 점검 여부를 알 수 없다. 사업자들은 검토 인력이 이용자 질문과 AI 답변을 직접 열람·수정해 만든 데이터셋을 모델 학습과 서비스 개선에 활용했고, 이용자는 자신의 입력 데이터가 인적 검토를 거친다는 사실을 알기 어려웠다(개인정보보호위원회, 2024).

Google 의 한국 거주자 대상 개인정보 추가 정보는 Google 이 모델의 개발·학습·미세 조정·개선 등을 위해 Gemini 앱 같은 AI 모델·기술과의 상호작용을 사용한다고 밝힌다(Google, 2026b). OpenAI ChatGPT 를 기준으로 한 개인정보위 안내서의 비교표(2025 년 8 월 3 일 기준)에서 개인용 라이선스는 기본적으로 AI 학습에 활용(opt-out 가능)되고 기업용 라이선스는 기본값으로 학습에 활용되지 않는다(개인정보보호위원회, 2025). Anthropic 은 2025 년 8 월 28 일 소비자 약관을 바꿔 학습 활용 허용 여부를 이용자가 선택하게 했고, 허용하면 보관 기간이 5 년으로 늘어나고 허용하지 않으면 30 일이 유지되며, Claude for Education 등 상용 약관 서비스에는 적용되지 않는다(Anthropic, 2025c).

40 개 대학의 정책·지침 분석에서도 대학 정책에는 데이터 프라이버시 대응과 공평한 접근 보장에서 공백이 남아 있었다(Y. Jin et al., 2025).

개인 계정의 데이터 처리 조건은 같은 제공사의 기관 계약(II 장 1 절 다)과 크게 다르고 사업자마다도 달라 서비스별 확인이 필요하다. 개인 계정을 전제로 한 과제는 개인정보와 비공개 자료의 입력 금지를 안내하는 편이 바람직하다(기관 계약 조건은 IV장 1 절 나). 이번 호는 Microsoft 등 다른 사업자의 개인 계정 조건을 검토하지 않았다.

바. 편향과 한국어·한국 맥락

Weissburg 외(2025)의 실험에서 GPT-4o, Gemini 1.5 Pro, Llama 3.1 405B 는 인구학적 집단에 맞춰 교육 콘텐츠를 생성·선택하는 방식에서 유의한 편향을 보였고, 모든 모델이 영어 프롬프트에도 히스패닉 학생에게는 자주 스페인어로 답했다. 한국어 사회편향 벤치마크 KoBBQ 로 평가한 모든 모델(GPT-3.5·GPT-4, Claude, CLOVA X 등)도 양의 편향 점수를 보였고 특히 맥락이 모호할 때 사회적 편향에 부합하는 답을 선호했다(J. Jin et al., 2024).

한국 원본 시험에서 문항을 모은 KMMLU 평가에서 가장 높은 점수는 GPT-4 의 59.95%로 원 시험의 합격선(80%)에 크게 못 미쳤다. 한국 특화 문항에서는 정확도가 더 떨어졌고, 모든 모델이 틀린 한국 특화 문항 73 개 중 31 개는 2017 년 시행 후 자주 개정된 한 법률에 관한 것이었다(Son et al., 2025). KoBBQ 저자들도 기계번역만으로는 모델의 편향을 측정하기에 불충분하며 수작업 번역과 문화적 적응이 중요하다고 결론지었다(J. Jin et al., 2024). 세 연구 모두 2023~2024 년 모델을 평가했으므로 최신 제품으로 일반화할 수 없다.

한국어 설명의 품질과 국내 제도·법령 질문의 정확성은 영어권 평가로 대신할 수 없으므로, 국내 도입 시 한국어 과업으로 따로 점검하는 것이 바람직하다.

사. 모드 전환의 한계

대규모 온라인 강좌에서 일반 LLM 모드와 RAG 기반 참고자료 모드 중 하나를 고르는 이중 모드 챗봇을 분석한 연구에서는 매 상호작용 시작마다 모드 설명 메시지가 나왔는데도 참여자의 41.3%가 모드 선택 기능을 몰랐거나 무관심했다(Han, 2026). Ⅱ 장 1 절 마의 CodeAid 강좌 배치 연구에서는 ChatGPT 사용 금지 정책에도 응답자 200 명 중 23%는 CodeAid 만, 38%는 CodeAid 와 ChatGPT 를 함께, 19%는 ChatGPT 만 사용했다(Kazemitabaar et al., 2024).

두 사례는 학습 모드 자체를 다루지 않았지만, 학습용 기능을 선택지로만 두면 쓰이지 않거나 일반 도구와 병용될 수 있음을 시사한다. 금지 정책 아래에서도 병용이 많았던 만큼 강제 설정만으로는 개인 계정의 일반 모드로 옮겨 가는 것을 막기 어려울 수 있어, 기본값 설정은 수업 목표 연계, 과정 제출·사용 공개 같은 과제 설계와 함께 운영하는 편이 적절해 보인다(Ⅳ 장 2 절 나).

아. 비용·종속·거버넌스 논쟁

현지 보도에 따르면 OpenAI 와 체제 단위 계약을 맺은 CSU 에서 자발적 AI 교육 이수율은 2026 년 4 월 기준 학생 0.7%, 교수 16%에 그쳤다. CSU 설문에서 학생의 67%는 교수가 AI 활용법을 가르치지 않는다고 느꼈고, 학생 대표는 계약 체결 과정에서 협의도 사전 통보도 받지 못했다고 밝혔다(Corzo, 2026). CSU 설문 보고서에서는 교수 응답자의 56%가 AI 의 긍정적 영향을, 52%가 부정적 영향을 보고했으며 두 문항은 별개로 측정되어 함께 동의할 수 있었다(California State University, 2026).

계약 비용을 두고도 논쟁이 이어졌다. CSU 담당자는 OpenAI 를 가장 비용이 적게 드는 선택지로 골랐다고 밝혔지만, 샌프란시스코주립대 일부 교수는 OpenAI 와의 계약 종료를 요구하는 청원을 시작했다(Corzo, 2026). CSU 는 예산 삭감 국면에서 논란이 된 OpenAI 계약을 갱신해 기관 우선순위 논쟁을 다시 불렀다(Rix, 2026).

CSU 사례에 비추어 볼 때 기관 도입은 비용 정당화, 교육 지원, 구성원 협의를 갖추지 못하면 정당성 논쟁에 부딪힐 가능성이 있다. 구성원 선호가 나뉜 상황의 단일 계약은 특정 사업자 종속을 낳을 수도 있다(Ⅱ 장 2 절 다). 다만 비용·거버넌스 논쟁의 근거는 사실상 OpenAI 와 계약한 CSU 한 사례의 언론 보도와 당사자 설문에 기대고 있어 일반화에는 제약이 따른다. 논쟁 사례가 OpenAI 계약에 몰린 것은 계약 조건과 논쟁을 함께 다룬 독립 보도가 이 사례에 집중되었기 때문으로 해석되며, 다른 사업자의 기관 계약에 대해서는 독립 평가 자료를 찾지 못했다.

[표 Ⅱ-4] 주요 쟁점 정리

쟁점	확인된 사실(요지)	근거의 성격	확실성(연구진 판단)	대학의 대응 수단
근거의 질	효과크기 이질성·출판 편향, 감사(Weidlich 외)에서 세 기준 충족 21%, 메타분석 1 편 철회	메타분석·감사·철회 공지	높음	통제 설계·지연 측정 확인, 철회 점검
환각·인용 오류	조작 참고문헌 GPT-3.5 55%·GPT-4 18%, 완전 허위 ChatGPT 1~22%·Gemini 0~26%	생성물 분석·독립 평가	높음(반복 확인)	출처 확인 절차, 자료 기반 응답
인지적 외주화·학습 깊이	비판적 사고와 부적 상관, 논증 품질·45 일 후 유지 저하, 얕은 지식 보고	상관 연구와 일부 실험	중간(인과는 일부 실험)	정답 보류·설명 중심 설계
학업 진실성	AI 생성 문장 직접 삽입 영국 학부생 12%(2026 년 조사), 대학 단계 표절 예방 교육 경험 42.7%	설문	중간	사용 공개, 과정 중심 평가
형평성	기관 유형·소득·소재지에 따른 사용 차이, 비용 외 요인 공존	횡단 설문	중간	기관 라이선스와 사용 지원
프라이버시	인적 검토 데이터의 학습 활용, 개인 계정 조건의 기관판·사업자별 차이	정부 점검·안내서·제공사 문서	높음	기관 계약, 입력 금지 안내
편향·한국어	집단별 콘텐츠 편향, 한국어 편향·한국 지식 공백	벤치마크(2023~2024 년 모델)	중간(최신 모델에서는 확인되지 않음)	한국어 과업 별도 점검
모드 전환	모드 미인지·무관심 41.3%, 금지 정책에도 일반 도구 병용	현장 연구	중간	기본값 설정과 과제 설계 병행
비용·종속·거버넌스	교육 이수 학생 0.7%, 학생 대표 협의 부재, 예산 삭감 국면의 계약 갱신, 종속 가능성	언론 보도·당사자 설문(OpenAI 계약 CSU 사례)·연구진 해석	중간(사례 기반)	단계적 도입, 구성원 협의, 갱신 조건 검토
규제	생성형 AI 고지·표시 의무(AI 기본법 제 31 조, EU AI Act 제 50 조), EU 학습성과 평가 AI 고위험 분류	법령	높음	고지·표시 이행 확인

자료: Ⅰ 장 1 절, Ⅱ 장 1~4 절, Ⅳ 장 1 절 가에서 인용한 연구·정책 자료를 바탕으로 연구진 정리

주: 확실성은 연구 설계와 반복 확인 여부에 따른 연구진 판단이다.

소결 Ⅱ장 4 절의 여덟 쟁점([표 Ⅱ-4]) 가운데 실험으로 인과가 확인된 위험은 일부여서 위험의 크기는 근거의 성격과 함께 읽어야 한다. 대부분은 모델 성능만으로 해소되기 어려워 출처 확인, 정답 보류형 설계와 과제 설계의 병행, 사용 공개와 과정 중심 평가, 기관 계약과 사용 지원, 비용 관리와 구성원 협의 같은 운영 설계로 관리할 성격으로 보이며, 이는 핵심 질문 Q3 후반부의 답으로 Ⅳ장에 이어진다.

Ⅲ. 대표 솔루션 심층 분석

이 장은 Ⅰ장 3절에서 고른 두 대표 솔루션을 프로필([표 Ⅲ-1]·[표 Ⅲ-2]), 평가척도 적용 결과, 비교 종합의 순서로 살펴본다. 판정은 제품의 보증이나 배척이 아닌 도입 검토 의견이며, 자료 기준일까지 공개된 문서로만 판정했다(Ⅰ장 4절).

1. 솔루션 프로필

ChatGPT 학습 모드는 Ⅱ장 1절에서 본 소크라테스식 질문·힌트·자기성찰 프롬프트에 더해 수준 진단 질문과 이전 대화 기억으로 개인화하고 퀴즈·개방형 질문으로 이해를 점검한다(OpenAI, 2025). 출시 때는 로그인 이용자에게 제공되었고 ChatGPT Edu 제공은 몇 주 뒤로 예고되었다(Ⅰ장 3절).

기관용 상품인 ChatGPT Edu(Ⅱ장 2절 나)는 출시 당시 GPT-4o를 기반 모델로 하고 데이터 분석 같은 고급 도구를 포함했다(OpenAI, 2024). ChatGPT Edu는 관리자에게 ChatGPT Enterprise와 같은 관리 기능을 제공하고, ChatGPT Edu 데이터는 OpenAI의 학생 데이터 개인정보 보호 계약에 따라 처리된다(OpenAI, 2026c). 2026년 6월 17일 공지에 따르면 Edu 워크스페이스의 데이터 내보내기는 기본값이 '꺼짐'으로 설정되어 있고 이 설정은 구성원 전체에 적용된다(OpenAI, 2026b).

[표 Ⅲ-1] ChatGPT 학습 모드·ChatGPT Edu 프로필

항목	내용
학습 기능·출시	학습 모드(study mode), 2025. 7. 29.
주요 상호작용	소크라테스식 질문·힌트·자기성찰 프롬프트, 수준 진단·기억 기반 개인화, 퀴즈·개방형 질문
이용 범위	출시 때 로그인 이용자 대상, ChatGPT Edu 제공은 출시 몇 주 뒤로 예고
제공사 현황	기관용 상품 ChatGPT Edu(대학 대상, 2024. 5. 30. 발표, 출시 당시 GPT-4o 기반, 가격 등 제공 조건 확인되지 않음)
데이터 처리(기관판)	기본값 모델 학습 미사용, 삭제 대화 30일 내 제거, 학생 데이터 개인정보 보호 계약
관리자 기능	Enterprise와 같은 관리 기능(관리 콘솔·그룹 권한·GPT 관리), 감사 로그(Compliance API), 데이터 내보내기 기본 꺼짐(2026. 6. 17.~), 학습 모드 지정·강제 설정 확인되지 않음
LMS 연동·확장	LTI 지원 확인되지 않음, 과목·과제용 맞춤형 GPT 워크스페이스 공유, 쿡북의 맞춤형 GPT Canvas API 연동 안내, Canvas 안 LLM 활용 과제는 Instructure 기능
이번 호 판정	공식 문서·정부 자료·독립 연구로 문서 판정 항목 11개 판정(실사용 진단 결과는 싣지 않음)

자료: OpenAI 공식 문서와 본문 Ⅱ장 1절·Ⅲ장 2절의 인용 자료를 바탕으로 연구진 정리

Gemini 가이드 학습은 Ⅱ장 1절에서 본 탐색적·개방형 질문과 단계별 설명에 더해 이미지·다이어그램·동영상·대화형 퀴즈 등 멀티모달 응답으로 답보다 과정에 초점을 둔다. 교육자는 가이드 학습 전용 링크를 Google Classroom에 게시하거나 학생과 공유할 수 있다(Heymans, 2025).

Google의 2025년 6월 발표에 따르면 기관용 상품인 Gemini for Education은 프리미엄 AI 모델 접근을 기본 제공하며, 소비자 무료판보다 높은 한도도 예고되었다. 같은 발표에서 Google은 Gemini가 18세 미만 사용자에게 더 엄격한 콘텐츠 정책을 적용한다고 밝혔다(Kirtikar, 2025).

[표 Ⅲ-2] Gemini 가이드 학습·Gemini for Education 프로필

항목	내용
학습 기능·출시	가이드 학습(Guided Learning), 2025. 8. 6.
주요 상호작용	탐색적·개방형 질문, 단계별 분해, 이미지·다이어그램·대화형 퀴즈 등 멀티모달 응답
이용 범위	개인 계정(요금제별 범위 확인되지 않음), 기관판 제공 범위 확인되지 않음
제공사 현황	기관용 상품 Gemini for Education(2025. 6. 30. 발표, Workspace for Education 요금제에 무료 포함)
데이터 처리(기관판)	엔터프라이즈급 데이터 보호, 교육 사용자 대화의 AI 모델 개선 미사용
관리자 기능	관리 콘솔(Gemini 앱 접근 관리)·Vault 대화 검색·사용 보고서, 가이드 학습 강제 설정 확인되지 않음
LMS 연동·확장	Gemini LTI(Canvas·Schoology 2024. 12.~, Moodle 2026. 5.~), Classroom 전용 링크 게시, 수업 자료 기반 Gem·노트북 배정(2025. 6. 발표)
이번 호 판정	공식 문서·정부 자료·독립 연구로 문서 판정 항목 11 개 판정(실사용 진단 결과는 싣지 않음)

자료: Google 공식 문서와 본문 Ⅱ장 1 절·Ⅲ 장 2 절의 인용 자료를 바탕으로 연구진 정리

두 프로필 표의 데이터 조건은 기관판 기준이며, 개인 계정은 같은 제공사라도 조건이 다르다(Ⅱ 장 4 절 마). 개인정보위 안내서는 개인용 무료 라이선스로 서비스형 LLM 을 쓰면 입력 데이터가 서비스 제공자의 LLM 학습 등에 활용될 우려가 있다고 본다(개인정보보호위원회, 2025).

2. 평가척도 적용 결과

이번 호는 자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구로 문서 판정(DOC) 항목만 판정했다(Ⅰ 장 4 절). 실사용 진단이 있어야 판정할 수 있는 AUTO·HYBRID 14 개 항목은 미측정으로 두었다. 사용성 축의 일곱 항목(A1~A7)과 콘텐츠 품질·정확성(B2), 피드백·적응(B3), 교수설계·인지부하(B5), 상호작용 사용성·접근성(B7), 타당성·신뢰성과 환각 관리(C1), 공정성·편향 관리(C3), 안전·견고성·보안(C6)이 여기에 해당한다. 주관적 만족도(A8)는 문서 판정 항목이지만 고등교육 대상 공개 만족도 조사를 찾지 못해 미측정이다. 그 결과 두 제품 모두 같은 11 개 항목(필수 13 개 중 5 개)을 판정했고, 축 A 는 측정 항목이 없어 점수를 내지 않았다.

가. ChatGPT 학습 모드의 문서 판정

학습목표 정렬(B1)과 동기·참여(B4)는 제공사의 설계 서술에 기대어 부분충족으로 판정했다. OpenAI 는 학습 모드의 시스템 지시가 능동적 참여 유도, 인지부하 관리, 메타인지·자기성찰 개발, 호기심 촉진, 지지적 피드백을 반영한다고 밝혔다(OpenAI, 2025). 공개 자료로 평정한 국내 연구(K-12 맥락)에서도 ChatGPT 학습 모드는 모호한 요청에 학년·시험 범위·목표를 되묻고 소크라테스식 문답으로 사고 확장을 유도해 탐색 질문과 질문 구체화 변인에서 모두 2 점을 받았다(차덕원·홍후조, 2026). 그러나 강좌 학습목표와의 정렬은 과목용 맞춤형 GPT 공유 같은 별도 수단에 의존하고, 이 평정도 동기·몰입이 아닌 질문 방식을 본 결과다.

교수 통합·교수자 지원(B6)은 부분충족, 표준 준수·상호운용(B8)은 미충족이다. ChatGPT Edu 에서는 특정 과목·과제용 맞춤형 GPT 를 만들어 대학 워크스페이스에서 공유할 수 있고, 관리자는 그룹 권한과 GPT 를 관리한다(OpenAI, 2026a). 그러나 교수자가 강좌 단위로 학습 모드를 배정하는 경로와 교수자용 수업 설계 지원 자료는 문서에서 확인되지 않았다. 표준 연동에서는 공식 쿡북이 맞춤형 GPT

에 Custom Action 을 붙여 Canvas REST API 로 Canvas 와 연결하는 방법을 안내했을 뿐이다(Schule & Casson, 2024). 이 API 연동과 Canvas 안의 LLM 활용 과제(Ⅱ 장 1 절 라)는 LTI 같은 교육 표준 연동이 아니어서 B8 을 미충족으로 두었다.

효과성 증거 등급(B9)과 효과크기 정량(B10)은 미충족이다. 제공사와 독립적으로 수행된 고등교육 대상 실험·준실험 연구나 메타분석으로 학습 모드의 학습 효과를 검증한 근거가 없었고, 위 국내 평정은 학습 효과를 측정하지 않았다(Ⅱ 장 3 절 라). ChatGPT 일반 사용 메타분석의 효과크기는 제품 단위 근거가 아니어서 반영하지 않았다.

투명성·설명가능성(C2)은 부분충족이다. OpenAI 는 학습 모드가 맞춤형 시스템 지시로 작동한다고 공개하면서 이 방식이 대화마다 일관되지 않은 행동과 실수를 낳을 수 있다고 인정했다(OpenAI, 2025). 그러나 위 국내 평정에서 ChatGPT 학습 모드는 공식 데이터셋 없이 사용자 입력과 대화 맥락에 의존해 공식 데이터 기반과 출처 제시 변인에서 0 점을 받았다(차덕원·홍후조, 2026).

프라이버시·데이터 거버넌스(C4)도 부분충족이다. 기관판 문서에서 모델 학습 미사용, 관리자의 보관 기간 설정과 삭제 대화의 30 일 내 제거, 학생 데이터 개인정보 보호 계약, 데이터 내보내기 기본 꺼짐을 확인했다(Ⅲ 장 1 절). 개인정보위 안내서는 기업용 이용약관과 데이터 처리 부속서(DPA)로 입력데이터 소유권·재이용 제한·안전조치·데이터 파기를 명확히 규정하도록 권장한다(개인정보보호위원회, 2025). 이에 해당하는 근거와 국외 이전·수집 최소화·한국 개인정보 보호법 정합은 기관판 문서에서 확인되지 않았다. 한국 이용자 부록은 국외 이전을 거부하면 OpenAI 서비스를 더 이상 이용하지 못할 수 있다고 밝히는 개인·소비자 서비스용 문서여서 기관판 보호 장치로 보지 않았다(OpenAI, 2026e).

인간 통제·교수자 역할(C5)은 부분충족이다. Edu 관리자는 ChatGPT Enterprise 와 같은 관리 기능을 쓰고 Enterprise Compliance API 로 대화와 GPT 의 감사 로그에 접근할 수 있다(OpenAI, 2026c). 워크스페이스에서 과목용 맞춤형 GPT 를 공유할 수도 있지만(B6), 교수자나 관리자가 학습 모드를 지정하거나 강제하는 설정은 문서에서 확인되지 않았다. 책무성·거버넌스·정책 정합(C7)은 감사 로그와 한국 이용자 대상 국내대리인 고지를 확인했으나, 인공지능기본법 제 31 조의 사전 고지·결과물 표시 대응이 문서에서 확인되지 않아 부분충족으로 두었다(계도기간 중이어서 법적 이행 여부는 판단하지 않음, Ⅳ 장 1 절 가).

AI 리터러시·연령 적합성(C8)도 부분충족이다. 이용약관은 서비스 이용에 동의할 수 있는 최소 연령을 만 13 세 이상 또는 이용자 국가가 정한 연령 이상으로 두고, 만 18 세 미만 이용자는 부모 또는 법정대리인의 허가를 받아야 한다고 정한다(OpenAI, 2026d). 한국 이용자에게는 OpenAI 개인정보 처리방침의 아동 연령 기준이 13 세에서 14 세로 조정된다(OpenAI, 2026e). 학습 모드와 Edu 안의 AI 리터러시 지원 자료는 제품 문서에서 확인되지 않았다.

나. Gemini 가이드 학습의 문서 판정

학습목표 정렬(B1)과 동기·참여(B4)도 부분충족이다. Google 은 가이드 학습이 탐색적·개방형 질문으로 참여를 유도하고 문제를 단계별로 나누어 설명을 학습자에 맞추며, 이미지·다이어그램·동영상·대화형 퀴즈 등 멀티모달 응답으로 답보다 과정에 초점을 둔다고 밝혔다(Heymans, 2025). 이는 제공사의 설계 서술이어서 동기·참여를 독립적으로 확인한 근거로 보지 않았다. 강좌 학습목표와의 정렬은 교수자가 수업 자료 기반 Gem·노트북을 배정하는 등 별도 설정에 의존한다(Ⅱ 장 1 절 라).

교수 통합·교수자 지원(B6)과 표준 준수·상호운용(B8)은 충족이다. 교수자는 가이드 학습 전용 링크를 Google Classroom 에 게시하거나 학생과 공유할 수 있다(Heymans, 2025). LTI 는 학습 도구를 기관의 학습 환경에 연결하는 기술 표준이며 제품이 아니다(1EdTech Consortium, 2026). Gemini LTI 는 2024

년 12 월 Canvas·Schoology 와 연동되어 전면 배포되었고 Google Workspace Education Fundamentals·Standard·Plus 에서 제공된다(Google, 2024). 2026 년 5 월에는 Moodle 로 넓어졌으나 1 EdTech 인증 여부와 LTI 버전은 공지에서 확인되지 않았다(Ⅱ 장 1 절 라).

효과성 증거 등급(B9)과 효과크기 정량(B10)은 미충족이다. 제공사(Google)가 협력 기관과 수행한 시에라리온 중학교 수학 실험(자체 보고·K-12, Ⅱ 장 3 절 라)은 동료심사 전이고 고등교육 대상이 아니어서 점수에 반영하지 않았고, 제공사와 독립적으로 수행된 고등교육 대상 효과 근거는 확인되지 않았다.

투명성·설명가능성(C2)은 부분충족이다. Google 은 LearnLM 의 역량을 Gemini 에 넣어 가이드 학습(Guided Learning)을 출시했다고 밝혀 작동 방식을 공개했다(Heymans, 2025). 그러나 가이드 학습 응답의 한계 고지와 출처 제시 방식은 문서에서 확인되지 않았다.

프라이버시·데이터 거버넌스(C4)도 부분충족이다. Gemini for Education 은 엔터프라이즈급 데이터 보호와 관리자 통제를 갖춘 Workspace 핵심 서비스이고, Google 은 모든 교육 사용자의 대화 데이터를 AI 모델 개선에 쓰지 않는다고 밝혔다(Ⅱ 장 1 절 다). 그러나 보관·파기 기간, 수집 최소화·목적 제한, 국외 이전과 한국 개인정보 보호법 정합은 기관판 문서에서 확인되지 않았고, 데이터 처리 부속서(DPA)에 해당하는 근거도 찾지 못했다. 개인정보위 안내서는 LLM 서버가 해외에 있다면 개인정보 국외이전에 해당하는지 검토할 필요가 있다고 본다(개인정보보호위원회, 2025). Google 의 한국 거주자 대상 추가 정보는 수집한 개인정보가 국외의 Google 데이터 센터로 전송되어 보관된다고 밝히지만, Google 서비스 이용에 적용되는 개인정보처리방침의 추가 정보여서 기관판의 한국법 정합 근거로 보지 않았다(Google, 2026b).

인간 통제·교수자 역할(C5)은 부분충족이다. 교수자는 가이드 학습 링크를 Classroom 에 게시하고 관리자는 Gemini 앱 접근 관리·Vault 대화 검색·사용 보고서를 쓸 수 있지만(Ⅱ 장 1 절 다), 가이드 학습을 강제하는 관리자 설정은 문서에서 확인되지 않았다. 책무성·거버넌스·정책 정합(C7)은 ChatGPT 와 같은 기준으로 관리 수단과 국내대리인 고지를 확인했으나 인공지능기본법 제 31 조 대응이 확인되지 않아 부분충족이다.

AI 리터러시·연령 적합성(C8)도 부분충족이다. Gemini 는 18 세 미만 사용자에게 더 엄격한 콘텐츠 정책을 적용한다(Ⅲ 장 1 절). 한국에서 자신의 Google 계정을 직접 관리할 수 있는 최소 연령은 14 세이고, 이 연령 요건은 Google Workspace for Education 계정에는 적용되지 않을 수 있다(Google, Retrieved 2026). 대학생(성인) 학습자를 위한 AI 리터러시 지원의 구성은 확인되지 않았다.

다. 점수와 판정

문서 판정 항목은 두 제품 모두 공식 문서·정부 자료·독립 연구로만 판정했고, 효과 근거(B9·B10)에는 제공사와 독립적으로 수행된 고등교육 연구만 반영했다. 축 점수는 소수 첫째 자리까지 반올림하고 종합 점수는 반올림한 축 점수로 계산했으며, 결과는 [표 Ⅲ-3]과 같다.

[표 Ⅲ-3] 평가 결과 요약

구분	ChatGPT 학습 모드·ChatGPT Edu	Gemini 가이드 학습·Gemini for Education
판정 근거	자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구(실사용 진단 결과는 싣지 않음)	자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구(실사용 진단 결과는 싣지 않음)
측정 범위	11/26 항목(A 0·B 6·C 5), 필수 5/13	11/26 항목(A 0·B 6·C 5), 필수 5/13

구분	ChatGPT 학습 모드·ChatGPT Edu	Gemini 가이드 학습·Gemini for Education
축 A 사용성(가중치 30)	미산출(측정 항목 없음)	미산출(측정 항목 없음)
축 B 교육적 특성(가중치 35)	26.9	42.3
축 C AI 신뢰성·윤리·거버넌스(가중치 35)	50.0	50.0
종합 점수(축 B·C 참고값)	38.5	46.2
공통 11 개 항목 기준	38.5(종합과 같음)	46.2(종합과 같음)
게이트 항목	C1 미측정·C2 ◐·C4 ◐	C1 미측정·C2 ◐·C4 ◐
도입 검토 의견	판정 유보	판정 유보
핵심 보완점	미측정 항목의 실사용 진단, LTI 등 표준 연동·학습 모드 지정 수단, 응답 출처 제시, 기관판 데이터 처리 부속서·국외 이전 조건	미측정 항목의 실사용 진단, 가이드 학습 강제 설정, 응답 한계 고지·출처 제시, 기관판 보관 기간·국외 이전 조건

자료: 공식 문서·정부 자료·독립 연구에 근거한 문서 판정을 바탕으로 연구진 정리

주: ◐는 부분충족이며 미측정 항목은 점수 계산에서 뺐다. 종합 점수는 측정 항목이 있는 축 B·C 만으로 계산한 참고값이다. 항목별 판정과 근거는 부록 2 에 실었다.

두 제품 모두 필수 13 개 항목 가운데 5 개(B1·B9·C2·C4·C5)만 측정되어 경계값 적용에 필요한 75%(10 개)에 못 미치므로 도입 검토 의견은 '판정 유보'다(Ⅰ 장 4 절). 게이트 항목 가운데 타당성·신뢰성과 환각 관리(C1)도 측정하지 못했다. 종합 점수는 축 B·C 만으로 계산한 참고값이어서 제품 순위로 읽을 수 없다.

공통 11 개 항목 기준 점수 차 7.7 점은 판정이 갈린 교수 통합·교수자 지원(B6)과 표준 준수·상호운용(B8)에서 나오고, 나머지 9 개 항목은 판정이 같다. 두 항목은 문서상 LTI 연동과 강좌 단위 배정 경로가 확인된 Gemini 만 충족이고 ChatGPT 는 각각 부분충족과 미충족이어서, 학습 모드 자체의 품질 차이로 보기는 어렵다. 판정 유보를 풀려면 필수 미측정 항목 8 개(A1·A4·A6·A7·B2·B3·C1·C3) 가운데 5 개 이상을 실사용 진단으로 측정해야 한다(Ⅳ 장 2 절 나).

3. 비교 종합

공개 문서로 본 두 학습 모드는 정답을 바로 주기보다 질문·힌트·단계별 설명으로 이해를 안내한다는 설계 목표가 같고, 구현 방식은 시스템 지시(OpenAI)와 학습용 모델 조정(Google)으로 갈린다(Ⅱ 장 1 절 가). 정답 보류가 실제로 지켜지는지, 응답이 정확하고 출처를 밝히는지는 실사용 진단으로 확인할 항목이어서 이번 호는 두 제품의 응답을 비교하지 않았다. OpenAI 는 시스템 지시 방식이 대화마다 일관되지 않은 행동과 실수를 낳을 수 있다고 인정했다(OpenAI, 2025).

Pastucha 외(2026)의 모델 수준 비교(이비인후과 지침 주제)에서는 웹 검색을 켠 두 버전(ChatGPT-4.1 51%, Gemini 2.5 Pro 41%)이 나머지 4 개 버전보다 참고문헌 정확도가 높았다. 완전 허위 참고문헌은 웹 검색을 켜면 ChatGPT-4.1 이 18~22%에서 1~3%로, Gemini 2.5 Pro 가 17~21%에서 0~3%로 줄었다. 학습 모드를 평가한 연구는 아니다.

모델 수준의 이 결과는 웹 검색 같은 근거 확인 기능이 출처 정확도를 좌우할 수 있음을 보여 주지만, 학습 모드 응답의 출처 제시 방식은 두 제품 모두 문서로 확인되지 않았다. 공개 문서상 두 제품이 갈리는 지점은 표준 연동과 강좌 단위 배정이다. Gemini 는 Gemini LTI 와 가이드 학습 전용 링크로 강좌에 연결되지만, ChatGPT Edu 는 과목용 맞춤형 GPT 를 워크스페이스에서 공유할 수 있을 뿐 LTI 지원이 확인되지 않고 Canvas 와의 연동도 쿡북의 API 연동 안내와 Canvas 의 LLM 활용 과제에 머문다 .

기관판 데이터 조건은 두 제품 모두 모델 학습 미사용을 밝히지만 데이터 처리 부속서와 국외 이전 조건은 기관판 문서에서 확인되지 않았고, 학습 모드를 강제하는 관리자 설정도 두 제품 모두 확인되지 않았다. 상대 제품 문서에서 확인되지 않았을 뿐 기능이 없다는 뜻은 아니므로, 대학은 우열보다 [표 Ⅲ-4] 마지막 열의 같은 질문으로 두 제품을 확인할 필요가 있다.

[표 Ⅲ-4] 대표 솔루션 비교

비교 차원	ChatGPT 학습 모드·Edu	Gemini 가이드 학습·for Education	대학 도입 시 확인할 점
학습 지원 설계 (제공사 서술)	시스템 지시로 소크라테스식 질문·힌트·자기성찰 프롬프트, 수준 진단·퀴즈(일관성 한계 인정)	LearnLM 조정 모델로 탐색적·개방형 질문, 단계별 분해, 멀티모달 응답	정답 보류 실사용 측정, 과제 설계
투명성·출처	작동 방식·한계 공개, 독립 평정에서 출처 제시 0 점(K-12 맥락)	작동 방식(LearnLM) 공개, 한계 고지·출처 제시 방식 미확인	출처 확인 절차, 웹 검색·자료 기반 응답 설정
데이터 조건(기관판)	모델 학습 미사용, 보관 기간 설정·삭제 대화 30 일 내 제거, 데이터 내보내기 기본 꺼짐	교육 사용자 대화 AI 모델 개선 미사용, 엔터프라이즈급 데이터 보호	DPA·보관 주기·국외 이전의 계약 명시
데이터 조건(개인 계정)	기본 학습 활용(opt-out 가능, 안내서 2025. 8. 3. 기준)	AI 와의 상호작용을 모델 개발·학습에 사용(한국 거주자 대상 고지)	개인 계정 입력 주의
관리·연동	Enterprise 와 같은 관리 기능·감사 로그, 과목용 맞춤형 GPT 공유, LTI 미확인(쿡북의 Canvas API 연동, Canvas LLM 활용 과제), 학습 모드 강제 설정 미확인	관리 콘솔·Vault·사용 보고서, Gemini LTI(Canvas·Schoology·Moodle) , 가이드 학습 전용 링크 (Classroom), 강제 설정 미확인	강제 설정·표준 연동 가능 여부
효과 근거	대학 대상 실험 미확인, 공개 자료 기반 독립 기능 평정(K-12 맥락, 효과 측정 아님)	대학 대상 실험·독립 평가 미확인, 제공사 K-12 RCT(자체 보고·동료심사 전, 점수 미반영)	자체 효과 측정 계획
평가 결과	판정 유보(참고 38.5 점, B6 ◐·B8 ○)	판정 유보(참고 46.2 점, B6·B8 ●)	미측정 항목의 실사용 진단

자료: 본문 Ⅱ장 1·3·4 절과 Ⅲ장의 인용 자료를 바탕으로 연구진 정리

주: '미확인'은 이 보고서가 근거로 삼은 공식 문서에서 해당 내용을 확인하지 못했다는 뜻이며, 기능이 없다는 뜻은 아니다.

Ⅳ. 시사점 및 종합

1. 국내 적용 시사점

법제, 데이터 보호, 대학 정책, 기관 도입 모델의 시사점은 [표 Ⅳ-1]에 수준별로 요약했다.

[표 Ⅳ-1] 수준별 시사점

수준	확인된 사실(요지)	시사점	점검 질문
국가	AI 기본법 고지·표시(계도기간 1 년 이상), 개인정보위 안내서(DPA·국외 이전 검토), 학생 개인정보 지침 권고	대학 자체 플랫폼의 의무 주체 검토	대학 제공 AI 서비스의 고지·표시 이행 주체는 누구인가
기관	도입 모델별 사용량 관리 방식, 두 모델 공통의 국외 이전·학습 제외 조건	계약 조건·운영 체계·학내 지침을 함께 설계	학습 제외·보관·국외 이전이 계약에 명시되었는가
교수자	정답 보류 설계의 근거 혼재, 모드 선택 기능 미인지, 총장 설문의 과정 중심 평가 우선	학습 모드 설정과 과제·평가 설계의 연계	과제가 학습 모드 사용을 전제하는가, 사용 공개 방식을 안내했는가
학습자	인용 오류와 검증 부담, 개인 계정의 데이터 처리, 개인정보 입력 제한	산출물 검증과 비판적 수용 교육	출처를 확인하고 개인정보를 입력하지 않았는가

자료: 본문 Ⅱ·Ⅲ·Ⅳ 장 인용 자료를 바탕으로 연구진 정리

가. 법제: AI 기본법의 고지·표시 의무와 의무 주체

인공지능기본법(AI 기본법) 제 31 조 제 1 항은 생성형 인공지능을 이용한 제품·서비스를 제공하려는 인공지능사업자에게 그 사실을 이용자에게 사전에 고지하도록 규정한다. 같은 조 제 2 항은 결과물이 생성형 인공지능으로 생성되었다는 사실을 표시하도록 하며, 이 법은 2026 년 1 월 22 일부터 시행되었고 인공지능사업자는 법인·단체·개인과 국가기관등을 포함한다(인공지능 발전과 신뢰 기반 조성 등에 관한 기본법, 2026). 과기정통부 가이드라인은 다른 개발사의 파운데이션 모델 API 를 활용해 이용자에게 서비스를 제공하는 경우 투명성 의무가 그 서비스를 제공하는 사업자에게 있다고 본다(과학기술정보통신부, 2026a).

과기정통부 설명자료에 따르면 무료 또는 비영리 목적이라도 생성형 AI 제품·서비스를 제공하는 인공지능사업자는 투명성 의무 대상이며, 정부는 1 년 이상의 계도기간 동안 사실조사와 과태료 부과를 유예한다(과학기술정보통신부, 2026b).

EU AI Act 제 50 조도 AI 와의 상호작용 고지와 생성형 AI 출력물의 기계 판독 가능한 표시를 요구한다. 부속서 III 은 수준을 가리지 않고 교육기관에서 학습성과를 평가하는 AI 시스템을 고위험 영역으로 열거한다(Regulation (EU) 2024/1689, 2024). 국내 AI 기본법은 유아교육·초등교육·중등교육에서의 학생 평가를 고영향 인공지능의 활용 영역으로 규정한다(인공지능 발전과 신뢰 기반 조성 등에 관한 기본법, 2026).

따라서 대학이 외부 모델 API 로 자체 플랫폼을 제공할 때 고지·표시 의무의 주체가 될 수 있는지, 도입 모델별로 제공사·공급사·대학 가운데 누가 이행하는지를 계약과 화면 구성에서 확인해야 한다(법적

판단은 아니다). 고등교육 평가의 고영향 해당 여부와 EU 고위험 의무의 조항별 적용 시점도 하위 법령과 공식 해석을 지켜봐야 한다.

나. 데이터 보호: 기관 라이선스·DPA·국외 이전

개인정보위 안내서는 서비스형 LLM 기반 AI 서비스를 제공하는 기업·기관이 이용자 데이터의 보관·재이용 여부와 국외 이전 여부를 사전에 검토할 필요가 있다고 본다. 개인용 무료 라이선스는 입력 데이터가 LLM 학습 등에 활용될 우려가 있다고 보고, 기업용 이용약관과 데이터 처리 부속서(DPA)로 소유권·재이용 제한·안전조치·파기를 명확히 규정하도록 권장한다(개인정보보호위원회, 2025). 국가인공지능전략위원회(2026)의 「대한민국 인공지능행동계획」은 교육부가 개인정보위와 협의해 고등교육법에 따른 학교를 포함한 각급 학교의 교육 AI 개발·활용 시 학생 개인정보 보호 지침을 2026년 4분기까지 마련하도록 권고했다.

FactChat(마인드로직, HAI-GPT의 기반)의 개인정보처리방침(2026년 5월 28일 시행)은 OpenAI·Anthropic·Google·Meta·X.AI·Perplexity(미국 소재)를 국외 이전 받는 자로 명시한다. 이전 항목은 질의와 대화 컨텍스트이고 보유 및 이용 기간은 30일이다. FactChat은 글로벌 LLM API를 마인드로직 명의로 호출해 이전 받는 자가 개인이나 소속 기관을 식별할 수 없다고 밝힌다(마인드로직, 2026).

안내서 사례는 의료기관을 전제로 해 대학 적용은 유추이지만(Ⅱ장 1절 다), 어느 도입 모델이든 해외 모델을 쓰는 한 국외 이전 검토는 피하기 어렵고 차이는 계약 상대와 문서 체계에 있다고 볼 수 있다. 계약 단계에서는 학습 제외·DPA·보관 주기·국외 이전 고지를 점검하는 것이 바람직하다(부록 4).

다. 대학 정책: 지침 마련과 제도화 수준

Ⅰ장 1절 가의 행동계획은 교육부가 대학 AI 활용 윤리 가이드라인을 개발·배포하도록 권고했고, 개별 대학도 자체 지침을 마련하고 있었다. 서울대는 2026년 1월 책임 있는 AI 활용을 위한 '서울대학교 AI 가이드라인'을 제정했다(서울대학교, 2026).

2026년 1~2월에 실시된 한국대학교육협의회 대학 총장 설문(140개교 응답)에서 생성형 AI 활용 정책을 채택 중인 대학은 40.0%(56개교)였고, '제정 검토 중'인 대학도 40.0%(56개교)였다. 2024년 설문의 채택·적용 대학 30개교와 비교하면 제도화 논의가 확산하는 추세로 해석되었다(한국대학교육협의회, 2026). 학생 쪽에서는 대학에 명확한 AI 지침·정책이 있다는 응답이 32.1%에 그쳤다(Ⅱ장 2절 마).

설문 수치로 미루어 보면 2026년 상반기 국내 대학의 생성형 AI 정책 제도화는 진행 단계에 있었던 것으로 읽힌다. 40개 대학 정책 분석에서 드러난 데이터 프라이버시·공평한 접근의 공백(Ⅱ장 4절 마)은 기관 도입과 맞닿으므로 도입과 학내 지침·평가 방식을 병행해 설계하는 편이 바람직하며, 국내 대학 지침의 체계적 분석은 후속 과제다.

라. 기관 도입 모델의 선택: 단일 벤더형과 멀티 LLM 통합형

서울대 제안요청서는 SSO 통합인증과 관리자의 프롬프트 데이터 조회·추출, 사용자가 정하는 대화 보관 주기와 자동 삭제, 입력·생성 데이터의 상용 모델 학습 배제를 요구했다(서울대학교 정보화본부, 2026). 순천향대 규격서는 통합로그인(SSO) 연동을 사업 내역에 넣고, RAG 원문·임베딩·메타데이터의 클라우드 저장과 사용자 데이터 격리·보호를 요구했다(순천향대학교, 2025).

사용량 관리 방식은 두 모델에서 다르게 나타난다([표 IV-2]). 단일 벤더형인 CSU는 전 캠퍼스에 교육용 ChatGPT 무제한 접근을 제공했다(II장 2절 다). 멀티 LLM 통합형에서는 순천향대 규격서가 그룹·개별 사용자별 월간 사용 한도를 설정해 정해진 예산 안에서 운영하는 관리자 기능을 요구했다(순천향대학교, 2025). 본교 HAI-GPT도 구성원별 월 크레딧을 둔다(II장 2절 라).

[표 IV-2] 두 도입 모델 비교

구분	단일 벤더형(서울대 ChatGPT Edu, CSU)	멀티 LLM 통합형(순천향대·본교)
대표 사례	서울대(2026. 6.~), CSU(체제 단위)	순천향대 조달 공고, 본교 HAI-GPT(이해관계)
대상 인원	서울대 약 3만 9,000명, CSU 학생 47만여 명·교직원 6만 3,000명	순천향대 최대 1만 계정, 본교 공개 자료 없음
기간	서울대 확인되지 않음, CSU 18개월 뒤 3년 갱신	순천향대 4개월, 본교 1년(학내 보도)
예산(공고 기준, 계약 금액 아님)	서울대 3,288,986,000원, CSU 공고 예산 확인되지 않음(계약 지불액은 [표 II-2])	순천향대 60,000,000원, 본교 공개 자료 없음
데이터 보호 조건	상용 모델 학습 배제, 보관 주기 설정·자동 삭제(서울대 요구), 국외 이전(기관판 조건 확인되지 않음)	사용자 데이터 격리·보호(순천향대 요구), 공급사 경유 국외 이전
관리·통제 수단	SSO, 관리자 프롬프트 조회·추출(서울대 요구)	SSO 연동, 그룹·개별 월간 사용 한도 설정(순천향대 요구)
사용량 관리	서울대 확인되지 않음, CSU 무제한 접근	순천향대 사용자별 월간 한도 설정 요구, 본교 재학생 월 5,000 크레딧·교직원 월 2만 크레딧
주요 위험(연구진 해석)	비용 정당화 논쟁(CSU), 특정 제공사 종속	공급사 경유 국외 이전·보관, 중개 공급사 의존, 고유 학습 기능 제공 여부

자료: 본문 II장 2·4절과 IV장 1·2절의 인용 자료를 바탕으로 연구진 정리

주: 멀티 LLM 통합형 사례 두 곳은 모두 FactChat(마인드로직) 기반이다(연구진 확인).

두 모델의 조달 요구사항은 학습 제외·보관·데이터 보호와 통합 인증에서 비슷하지만, 사용량 관리 방식, 데이터 경로와 계약 구조, 제공사 고유 학습 기능의 이용 가능성은 다르다고 볼 수 있다. 이 비교의 멀티 LLM 통합형 사례는 모두 본교 HAI-GPT와 같은 공급사의 플랫폼이다(예산 비교의 한계는 IV장 3절 Q2).

2. 본교·연구회 활용 제언

가. 본교 적용 사례: HAI-GPT(관계 공개)

이 보고서는 HAI-GPT를 운영하는 본교 AI에듀테크센터의 연구과제 산출물이며, 연구진에는 멀티 LLM 도입을 추진한 본교 교무 부서와 센터 소속 구성원이 포함된다. HAI-GPT는 마인드로직의 FactChat을 기반으로 한다. 이러한 이해관계를 고려해 HAI-GPT는 평가척도로 진단하지 않고 공개된 운영 사실만 기술한다.

본교는 LLM 27종을 하나의 플랫폼에서 골라 쓰는 멀티 LLM 구조로 고비용 추론 모델을 대학 계정으로 제공한다(2026년 2월 발표)(한림대학교, 2026). HAI-GPT는 2026년 3월 3일 운영을 시작했고

재학생에게 월 5,000 크레딧, 교직원에게 월 2 만 크레딧이 무료로 제공되며 1 년 계약이 완료된 상태다(한림학보 보도)(강지오, 2026).

본교는 AI 사용 과정에서 생성되는 데이터를 대학이 직접 소유·관리하는 구조를 구축했고 이를 교육·연구·행정에 환류할 계획이다(한림대학교, 2026). 한편 FactChat 처리방침(2026 년 5 월 28 일 시행)은 도입 기관명이 빈칸인 기관 공용 서식이다(마인드로직, 2026). 국외 이전 조건은 IV장 1 절 나와 같다.

개인정보위 안내서는 서비스형 LLM 기반 AI 서비스를 제공하는 기업·기관이 이용자 데이터의 보관·재이용(AI 학습 포함) 여부와 국외이전 여부를 사전에 검토할 필요가 있다고 본다(개인정보보호위원회, 2025). 대학 계정으로 제공되는 이 플랫폼에 대해서도 실제 정책값, 국외 이전 데이터의 보관 주체와 기간, 대학 소유·관리의 계약상 범위를 확인해 구성원에게 알릴 필요가 있다.

나. 제언

첫째, 교무 부서와 교수학습개발센터는 도입한 플랫폼이 학습 모드(정답 대신 질문·힌트·이해 점검으로 안내하는 기능)와 그 강제·배정 수단(Ⅲ장 3 절)을 제공하는지 다음 학기 전에 확인하고, 제공하지 않으면 강좌용 지시문이나 별도 도구로 보완하는 방안을 검토할 수 있다. Han(2026)은 학습자 자율성을 키우고 챗봇 기능을 수업 목표에 맞추는 의도적 설계가 필요하다고 결론지었다. 다만 정답 보류형 설계의 효과 근거는 의사결정 과제 실험과 K-12 연구가 중심이고 대학 실험의 결과는 일관되지 않으며 강제 설정만으로는 개인 계정으로 옮겨 가는 것을 막기 어려울 수 있어(Ⅱ장 3 절 다·4 절 나·사), 과정 제출·사용 공개 같은 과제 설계를 병행하는 편이 적절해 보인다.

둘째, 교양교육 부서와 연구회는 교양 AI 교육과 학습 도구 운영을 연결할 수 있다. 대학 AI 기본교육과정 지원사업 기준에서 AI 기초 교양 교과는 다양한 전공 학생이 수강 가능한 난이도로 구성하고 AI 윤리, 비판적 사고, 기초 도구 실습을 포함해야 한다(교육부·한국대학교육협의회, 2026). 본교의 AI 기초 교과에서도 학습 모드 사용과 산출물 검증을 연구회의 교양 AI 교육 프레임워크(연구회 내부 설계안, 미발표)와 연결할 수 있다.

셋째, AI 에듀테크센터와 개인정보 보호 부서는 고지·표시와 데이터 보호 항목을 점검표(부록 4)로 정리해 HAI-GPT 정책값과 외부 서비스 이용 안내에 적용할 수 있다. 사전 고지는 이용약관·서비스 가입 절차·계약서나 화면 표시로 생성형 AI 활용을 명시하는 방식으로 할 수 있다(과학기술정보통신부, 2026a). 개인정보위는 인적 검토 사실의 명확한 고지와 입력 데이터 삭제 기능의 접근성 개선을 권고했다(개인정보보호위원회, 2024).

넷째, 교무 부서와 AI 에듀테크센터는 학내 효과 측정과 운영 자료 활용을 함께 설계할 수 있다. 행동계획은 학교 현장에 도입되는 교육 AI 서비스의 안전성·윤리성·학습효과에 대한 공적 검증 체계가 필요하다고 보았다(국가인공지능전략위원회, 2026). 첫째 제언의 확인 결과 학습 모드(또는 강좌용 지시문)를 기본값으로 둘 수 있으면, 학내 측정은 교양 교과 여러 분반에서 기본값과 자율 선택을 비교하고 몇 주 뒤 AI 없이 지연 검사를 하는 방식이 될 수 있다. 이용 데이터 환류(IV 장 2 절 가)는 대학 소유·관리의 계약상 범위를 확인한 뒤 비식별 집계 지표부터 쓰고, 수집 범위·가명 처리·이용 고지를 학생 개인정보 보호 지침(IV 장 1 절 나)에 맞출 수 있다.

연구회는 이번 호에서 미측정으로 남은 AUTO·HYBRID 항목을 측정하는 실사용 진단을 후속 과업으로 삼는다. 두 학습 모드에 같은 표준 과업(정답 직접 요구, 오개념·허위 전제, 출처 요청, 오답 입력과 오류 복구 등)을 적용하고, 기관 계정과 개인 맞춤을 끈 조건에서 반복 측정해 판정 유보의 해소 여부를 확인할 수 있다(Ⅲ장 2 절 다). 진단 절차는 전공 교과 'AI 와 에듀테크'와 'AI 활용 소프트웨어 설계와 구현'의 실습 과제로 연결할 수 있으며, 결과는 연구회 정기회의에서 공유한다.

3. 종합

Q1(학습 모드의 기술적 차이와 교수 설계 원리)에 대해, 학습 모드는 일반 대화와 같은 모델에 교수용 시스템 지시를 얹거나(OpenAI) 교수법 지시를 따르도록 조정한 모델을 쓰며(Google), Anthropic·Microsoft 의 배치형은 구현 방식이 공개되지 않았다. 설계 원리는 제공사 문서에서 표방되지만, 제품 단위 독립 평가는 ChatGPT 학습 모드를 공개 자료로 평정한 국내 연구(K-12 맥락) 한 건이었다. 이 연구에서 학습 모드는 상호작용·탐색 영역은 만점이었으나 정보 신뢰성·근거와 학교 맥락·교사 연계 점수는 낮았다. 정답 보류도 구현 방식만으로는 보장되지 않는다. OpenAI 는 시스템 지시 방식이 대화마다 일관되지 않은 행동을 낳을 수 있다고 인정했고, 독립 평가(2024 년 모델 수준)에서도 정답 비노출 비율이 모델에 따라 크게 달랐다. Ⅲ 장에서는 문서로 판정할 수 있는 11 개 항목만 판정해 두 제품 모두 '판정 유보'(참고 점수 ChatGPT 38.5 점, Gemini 46.2 점)였으며, 점수 차는 문서상 LTI 연동·강좌 배정에서 나와 학습 모드의 품질 차이를 뜻하지 않는다.

Q2(두 기관 배포 모델의 차이)에 대해, 두 모델 모두 해외 모델을 쓰는 한 국외 이전과 학습 제외·보관 조건을 계약으로 확인해야 한다([표 Ⅳ-2]). 비용은 공고 예산의 대상·기간·범위가 달라 공개 자료로 단가를 비교할 수 없다. 차이는 데이터가 한 제공사의 기관판 환경에서 처리되는지, 공급사 플랫폼을 거쳐 여러 제공사로 전달되는지에 있고, 사용량 관리 방식도 달랐다(CSU 무제한 접근, 순천향대 규격서·본교 HAI-GPT 의 월간 한도). 단일 벤더형은 제공사 고유의 학습 모드·관리 기능을 쓸 수 있으나 특정 제공사 종속과 비용 정당화 논쟁에 부딪힐 수 있다. 멀티 LLM 통합형은 모델 선택 폭이 넓으나 중개 공급사 의존과 공급사의 처리방침 변경, 고유 학습 기능의 제공 여부에 영향을 받을 수 있다.

Q3(학습 효과의 근거와 위험 관리)에 대해, 대학생 메타분석의 큰 효과크기는 이질성·출판 편향·설계 문제로 과대 또는 과소 추정되었을 수 있어 신중하게 해석해야 한다. 정답 보류 설계가 제한 없는 사용의 학습 저해를 줄인 직접 근거는 K-12 현장실험이고, 대학 실험에서는 무제한 사용도 지식 향상에서 뒤지지 않았으며 설계된 튜터의 우위도 일관되지 않았다. 과의존 위험은 일부 실험(45 일 후 유지 저하, 논증 품질 저하)이 시사하나 상관 연구가 많고, 상용 학습 모드의 대학생 효과는 아직 검증되지 않았다. 학습 모드를 강제하는 관리자 설정은 두 제품 모두 공개 문서에서 확인되지 않았고, 모드를 선택지로만 두면 쓰이지 않거나 일반 도구와 병용될 수 있다(Ⅱ 장 4 절 사). 따라서 위험 관리는 과정 중심 평가·사용 공개·출처 검증 교육과 학내 효과 측정을 병행하는 방향이 될 가능성이 크다.

범용 LLM 의 학습 모드는 교수 설계 원리를 표방하지만 정답 보류는 보장되지 않고 대학 현장의 학습 효과는 아직 검증되지 않았으며, 공개 문서만으로는 두 대표 솔루션의 도입 검토 의견을 내기 어려워 판정을 유보했다. 기관 도입은 어느 모델이든 비용·데이터 조건과 교육 지원을 관리할 과제를 안는다. 이번 호는 대학이 학습 모드의 탑재 여부보다 계약상 데이터 조건, 과제·평가 설계, 학내 효과 측정을 갖추며 효과를 확인해 가야 한다고 결론짓는다.

다음 호 예고

Vol.02 적응형 학습 시스템(2026 년 7 월분) — 이번 호에서 경계로 남긴 학습 경로·난이도 조정 문제를 이어받아, 범용 LLM 의 진단·개인화 기능과 전용 적응형 학습 엔진의 구조·고등교육 효과 근거를 비교한다

부록

부록 1. 참고 자료 목록

가. 국내 자료

Google. (2026b, 5 월 13 일). 한국 거주자를 위한 개인정보 관련 추가 정보. https://policies.google.com/privacy/additional?hl=ko&gl=KR

OpenAI. (2026d, 1 월 1 일). 이용약관. https://openai.com/ko-KR/policies/terms-of-use/

OpenAI. (2026e, 3 월 27 일). OpenAI 대한민국 이용자를 위한 개인정보 처리방침 부록. https://openai.com/ko-KR/policies/kr-privacy-policy-addendum/

강지오. (2026, 3 월 9 일). [기획] 다양한 AI 도구 한곳에...'HAI-GPT' 오픈. 한림학보. https://news.hallym.ac.kr/news/articleView.html?idxno=20722

개인정보보호위원회. (2024, 3 월 28 일). 개인정보위, 주요 인공지능(AI) 서비스 사전 실태점검 결과 발표: LLM 사업자에 대해 학습 과정 및 서비스 관련 개선권고 [보도자료]. https://pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10027

개인정보보호위원회. (2025). 생성형 인공지능(AI) 개발·활용을 위한 개인정보 처리 안내서. https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=11410

과학기술정보통신부. (2026a). 인공지능 투명성 확보 가이드라인. https://www.msit.go.kr/bbs/view.do?sCode=user&mId=102&mPid=100&bbsSeqNo=81&nttSeqNo=3148988

과학기술정보통신부. (2026b, 1 월 28 일). 인공지능 기본법 시행 관련 추가 설명 [보도참고자료]. https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&bbsSeqNo=94&nttSeqNo=3186809

과학기술정보통신부, 한국지능정보사회진흥원. (2026, 3 월 31 일). "국민 44.5% 생성형 인공지능 경험, 전년 대비 11.2%P 증가" 과기정통부,「2025 인터넷 이용 실태조사」결과 발표 [보도자료]. https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=208&mId=307&nttSeqNo=3187098

교육부. (2025, 11 월 10 일). 모두를 위한 인공지능(AI) 인재양성 방안 발표 [보도자료 및 별첨「AI for All: 모두를 위한 AI 인재양성방안」]. https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=104462

교육부, 한국대학교육협의회. (2026, 2 월 23 일). 2026 년 대학 인공지능(AI) 기본교육과정 개발 지원사업 선정 공고 [보도자료 및 별첨「2026 년 대학 AI 기본교육과정 개발 지원 사업 추진계획」]. https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&boardSeq=105401&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=020402&opType=N

국가인공지능전략위원회. (2026). 대한민국 인공지능 행동계획: 인공지능 기본계획(2026~2028). https://www.aikorea.go.kr/web/board/brdDetail.do?menu_cd=000011&num=328

김지하, 김정아, 김은영, 백승주, 손윤희, 조옥경. (2025). 데이터 기반 대학 교육 혁신 모니터링 연구(Ⅲ): 대학의 교수·학습 실태 및 수준 분석 (기술보고 TR 2025-02). 한국교육개발원. https://www.kedi.re.kr/khome/main/research/selectPubForm.do?plNum0=16267

마인드로직. (2026, 5 월 28 일). 개인정보처리방침(FactChat). https://factchat-public.s3.ap-northeast-2.amazonaws.com/public/pdf/ko_privacy_policy.pdf

박돈하, 이효숙. (2026). 생성형 인공지능 시대 대학의 표절예방 교육실태와 과제. 빌둥 Bildung, 7(1), 93–127. https://doi.org/10.37998/le.2026.7.1.93

백다은, 손완상, 송지훈, 유명현. (2024). 생성형 AI(Generative AI) 활용 학습 효과 메타분석. 교육정보미디어연구, 30(4), 1261–1285. https://doi.org/10.15833/kafeiam.30.4.1261

산업통상자원부, 소프트웨어정책연구소. (2025). 2024 년 이러닝산업 실태조사. https://spri.kr/posts/view/23880?code=stat_sw_e-learning_reports

서울대학교. (2026, 6 월 2 일). 서울대, 'AI Native Campus' 전환 본격화: 전 구성원 대상 ChatGPT Edu 도입 [보도자료]. https://www.snu.ac.kr/snunow/press?md=v&bbsidx=170942

서울대학교 정보화본부. (2026). 구성원 선호 반영 교육·연구용 생성형 AI 단일 서비스 도입 제안요청서(수정) (나라장터 입찰공고 R26BK01475290-000). 조달청 나라장터. https://www.g2b.go.kr/

순천향대학교. (2025, 10 월 21 일). 순천향대학교 AX/DX 고도화를 위한 통합 LLM 플랫폼 도입 [입찰공고 교육혁신원 2025-02 및 첨부「통합 LLM 플랫폼 입찰 제안 안내 및 물품 규격」]. https://home.sch.ac.kr/sch/06/030000.jsp?mode=view&article_no=2025102113054344601 7

윤혜준, 정은진, 김혜정. (2025). AI 시대의 학습에 대한 대학생의 인식과 불안 (KRIVET Issue Brief 제 307 호). 한국직업능력연구원. https://www.krivet.re.kr/kor/sub.do?menuSn=14&pstNo=PB0000000598

이윤빈. (2024). 대학생 필자의 글쓰기에서 챗 GPT 답변의 활용 유형 및 양상 연구. 교양교육연구, 18(4), 59–79. https://doi.org/10.46392/kjge.2024.18.4.59

인공지능 발전과 신뢰 기반 조성 등에 관한 기본법, 법률 제 21311 호 (2026). https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=282791&chrClsCd=010202&urlMode=lsInfoP&efYd=20260721&ancYnChk=0

차덕원, 홍후조. (2026). 고등학교 진로·진학 지원 AI·디지털 플랫폼의 기능적 특성 비교와 개선 방향 탐색. 한국교육학연구, 32(2), 401–429. https://doi.org/10.29318/ker.32.2.15

한국대학교육협의회. (2026, 3 월 6 일). 2026 KCUE 대학 총장 설문(I). https://www.kcue.or.kr/notice/sub06.php?at=view&idx=658532

한림대학교. (2026, 2 월 9 일). 한림대학교, 멀티 LLM 도입으로 캠퍼스 내 AI 접근성 확대 [보도자료]. https://www.hallym.ac.kr/bbs/pr/876/382246/artclView.do

나. 국외 자료

1EdTech Consortium. (2026). Learning Tools Interoperability (LTI). https://www.1edtech.org/standards/lti

Anthropic. (2025a, April 2). Introducing Claude for Education. https://www.anthropic.com/news/introducing-claude-for-education

Anthropic. (2025b, July 9). Advancing Claude for Education. https://www.anthropic.com/news/advancing-claude-for-education

Anthropic. (2025c, August 28). Updates to consumer terms and privacy policy. https://www.anthropic.com/news/updates-to-our-consumer-terms

Baek, C., Tate, T., & Warschauer, M. (2024). "ChatGPT seems too good to be true": College students' use and perceptions of generative AI. Computers and Education: Artificial Intelligence, 7, Article 100294. https://doi.org/10.1016/j.caeai.2024.100294

Barcaui, A. (2025). ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention. Social Sciences & Humanities Open, 12, Article 102287. https://doi.org/10.1016/j.ssaho.2025.102287

Bassner, P., Lenk-Ostendorf, B., Beinstingel, R., Wasner, T., & Krusche, S. (2026). Less stress, better scores, same learning: The dissociation of performance and learning in AI-supported programming education. Computers and Education: Artificial Intelligence, 10, Article 100537. https://doi.org/10.1016/j.caeai.2025.100537

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

California State University. (2026). Ahead of the curve: What the nation's largest public university system is learning about AI. https://www.calstate.edu/impact-of-the-csu/technology/ai-empowered-csu/Documents/csu-ahead-of-the-curve.pdf

Chen, S., & Cheung, A. C. K. (2025). Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis. Educational Research Review, 49, Article 100737. https://doi.org/10.1016/j.edurev.2025.100737

Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., Zhong, A., Hu, X., Liang, J., Yu, P. S., & Wen, Q. (2025). LLM agents for education: Advances and applications. In Findings of the Association for Computational Linguistics: EMNLP 2025 (pp. 13782–13810). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-emnlp.743

Corzo, A. (2026, May 1). Cal State struck a deal with OpenAI. Some students and faculty refuse to use it. CalMatters. https://calmatters.org/education/2026/05/california-state-university-open-ai-chatgpt-contract/

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). 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

Gajos, K. Z., & Mamykina, L. (2022). Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In Proceedings of the 27th International Conference on Intelligent User Interfaces (pp. 794–806). Association for Computing Machinery. https://doi.org/10.1145/3490099.3511138

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

Google. (Retrieved 2026). Age requirements on Google Accounts. Google Account Help. https://support.google.com/accounts/answer/1350409?hl=en

Google. (2024, December 9). Introducing Gemini Learning Tools Interoperability (LTI™). Google Workspace Updates. https://workspaceupdates.googleblog.com/2024/12/Gemini-Learning-Tools-Interoperability.html

Google. (2026a, May 12). Google Workspace Assignments LTI™ and Gemini LTI™ are now available for Moodle. Google Workspace Updates. https://workspaceupdates.googleblog.com/2026/05/workspace-assignments-lti-and-gemini-lti-for-moodle.html

Han, S. (2026). Learner autonomy and mode selection in dual-mode chatbot: Students' interaction patterns and learning outcomes in an online course. Computers & Education, 240, Article 105464. https://doi.org/10.1016/j.compedu.2025.105464

Heymans, M. (2025, August 6). Guided Learning in Gemini: From answers to understanding. Google. https://blog.google/products-and-platforms/products/education/guided-learning/

Instructure. (2025, July 23). Instructure and OpenAI announce global partnership to embed AI learning experiences within Canvas [Press release]. https://www.instructure.com/press-release/instructure-and-openai-announce-global-partnership-embed-ai-learning-experiences

Jin, J., Kim, J., Lee, N., Yoo, H., Oh, A., & Lee, H. (2024). KoBBQ: Korean bias benchmark for question answering. Transactions of the Association for Computational Linguistics, 12, 507–524. https://doi.org/10.1162/tacl_a_00661

Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence, 8, Article 100348. https://doi.org/10.1016/j.caeai.2024.100348

Kazemitabaar, M., Ye, R., Wang, X., Henley, A. Z., Denny, P., Craig, M., & Grossman, T. (2024). CodeAid: Evaluating a classroom deployment of an LLM-based programming assistant that balances student and educator needs. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1–20). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642773

Kirtikar, A. (2025, June 30). New Gemini tools for students and educators. Google. https://blog.google/products-and-platforms/products/education/gemini-iste-2025/

LearnLM Team. (2024). LearnLM: Improving Gemini for learning (arXiv:2412.16429). arXiv. https://doi.org/10.48550/arXiv.2412.16429

LearnLM Team & Fab AI. (2026). Teaching with Gemini: Measuring the impact of Guided Learning on student mathematics progress in Sierra Leone [Technical report]. Google. https://storage.googleapis.com/deepmind-media/LearnLM/learnLM_sierraleone_may26.pdf

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

Lee, H.-Y., Chen, P.-H., Huang, Y.-M., & Wu, T.-T. (2026). From immediate answers to structured inquiry: A Socratic method-enhanced ChatGPT system for blended learning discussions. Journal of Educational Computing Research, 64(5), 1254–1293. https://doi.org/10.1177/07356331261429306

Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. International Journal of Educational Technology in Higher Education, 21(1), Article 16. https://doi.org/10.1186/s41239-024-00447-4

Liu, R., Zenke, C., Liu, C., Holmes, A., Thornton, P., & Malan, D. J. (2024). Teaching CS50 with AI: Leveraging generative artificial intelligence in computer science education. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (pp. 750–756). Association for Computing Machinery. https://doi.org/10.1145/3626252.3630938

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. Journal of Empirical Legal Studies, 22(2), 216–242. https://doi.org/10.1111/jels.12413

MarketsandMarkets. (2024, November 27). AI in education market worth $5.82 billion by 2030 [Press release]. https://www.marketsandmarkets.com/PressReleases/ai-in-education.asp

Maurya, K. K., Srivatsa, K. V. A., Petukhova, K., & Kochmar, E. (2025). Unifying AI tutor evaluation: An evaluation taxonomy for pedagogical ability assessment of LLM-powered AI tutors. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 1234–1251). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.57

Melumad, S., & Yun, J. H. (2025). Experimental evidence of the effects of large language models versus web search on depth of learning. PNAS Nexus, 4(10), Article pgaf316. https://doi.org/10.1093/pnasnexus/pgaf316

Microsoft Education Team. (2026, May 13). Study and Learn: AI built for your student. Microsoft Education Blog. https://www.microsoft.com/en-us/education/blog/2026/05/study-and-learn-ai-built-for-your-student/

Mo, F., Huang, J., Yang, Y., Özen, Z., Maeda, Y., & Olenchak, F. R. (2026). Undergraduate students' learning outcomes with ChatGPT: A meta-analytic study. Computers and Education: Artificial Intelligence, 10, Article 100536. https://doi.org/10.1016/j.caeai.2025.100536

Mordor Intelligence. (2025, November 21). AI in education market size & share analysis: Growth trends and forecast (2025–2030). https://www.mordorintelligence.com/industry-reports/ai-in-education-market

OpenAI. (2024, May 30). Introducing ChatGPT Edu. https://openai.com/index/introducing-chatgpt-edu/

OpenAI. (2025, July 29). Introducing study mode. https://openai.com/index/chatgpt-study-mode/

OpenAI. (2026a). ChatGPT Edu: Bring AI to campus at scale. ChatGPT. https://chatgpt.com/business/education/

OpenAI. (2026b). ChatGPT Enterprise and Edu release notes. OpenAI Help Center. https://help.openai.com/en/articles/10128477-chatgpt-enterprise-and-edu-release-notes

OpenAI. (2026c, January 8). Enterprise privacy at OpenAI. https://openai.com/enterprise-privacy/

Pastucha, M., Skarżyński, H., Kochanek, K., & Jędrzejczak, W. W. (2026). Reference accuracy in large language model chatbots: A metric for inherent misinformation? Medical Science Monitor, 32, Article e950916. https://doi.org/10.12659/msm.950916

Paul, M., & Dhami, D. (2025, September 8). 6 ways to use NotebookLM to master any subject. Google. https://blog.google/innovation-and-ai/models-and-research/google-labs/notebooklm-student-features/

Pichai, S. (2025, August 6). Bringing the best of AI to college students for free. Google. https://blog.google/products-and-platforms/products/gemini/google-ai-pro-students-learning/

Quarnstrom, D. (2025, October 15). Designing Microsoft 365 Copilot to empower educators, students, and staff. Microsoft Education Blog. https://www.microsoft.com/en-us/education/blog/2025/10/designing-microsoft-365-copilot-to-empower-educators-students-and-staff/

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., Abdulla, A. A., Akopyan, A., Aldana Segura, M. W., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydın, F., Badran, A., Balbontín-Alvarado, R., Ben Saad, H., . . . Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. PLOS ONE, 20(2), Article e0315011. https://doi.org/10.1371/journal.pone.0315011

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives

2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). (2024). Official Journal of the European Union, L, 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

Rix, K. (2026, May 27). CSU renews controversial OpenAI deal despite faculty criticism, budget cuts [Republished from EdSource]. Local News Matters. https://localnewsmatters.org/2026/05/27/csu-renews-openai-chatgpt-contract/

Schule, K., & Casson, J. (2024, September 17). GPT Actions library - Canvas Learning Management System. OpenAI Cookbook. https://developers.openai.com/cookbook/examples/chatgpt/gpt_actions_library/gpt_action_canvas

Shi, Y., Yu, K., Dong, Y., & Chen, F. (2026). Large language models in education: A systematic review of empirical applications, benefits, and challenges. Computers and Education: Artificial Intelligence, 10, Article 100529. https://doi.org/10.1016/j.caeai.2025.100529

Sok, S., & Shin, H. W. (2025). Do interactions with ChatGPT influence L2 learners' oral speaking ability, summarization ability, and perceptions of generative AI tasks? TESOL Quarterly, 59(S1). https://doi.org/10.1002/tesq.70001

Son, G., Lee, H., Kim, S., Kim, S., Muennighoff, N., Choi, T., Park, C., Yoo, K. M., & Biderman, S. (2025). KMMLU: Measuring massive multitask language understanding in Korean. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 4076–4104). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.naacl-long.206

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, Article 108386. https://doi.org/10.1016/j.chb.2024.108386

Stephenson, R., & Armstrong, C. (2026). Student generative AI survey 2026 (HEPI Report 199). Higher Education Policy Institute. https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/

Swacha, J., & Gracel, M. (2025). Retrieval-augmented generation (RAG) chatbots for education: A survey of applications. Applied Sciences, 15(8), Article 4234. https://doi.org/10.3390/app15084234

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports, 13(1), Article 14045. https://doi.org/10.1038/s41598-023-41032-5

Wang, J., & Fan, W. (2026). Retraction note: The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. Humanities and Social Sciences Communications, 13(1), Article 528. https://doi.org/10.1057/s41599-026-07310-z

Weidlich, J., Gašević, D., Drachsler, H., & Kirschner, P. (2025). ChatGPT in education: An effect in search of a cause. Journal of Computer Assisted Learning, 41(5), Article e70105. https://doi.org/10.1111/jcal.70105

Weissburg, I., Anand, S., Levy, S., & Jeong, H. (2025). LLMs are biased teachers: Evaluating LLM bias in personalized education. In Findings of the Association for Computational Linguistics: NAACL 2025 (pp. 5665–5713). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-naacl.314

부록 2. 평가척도 채점표

두 대표 솔루션의 항목별 판정과 근거는 다음과 같다. 이번 호는 자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구로 문서 판정(DOC) 항목만 판정했고, 실사용 진단 결과는 싣지 않았다(Ⅰ장 4절).

대상 솔루션 ChatGPT 학습 모드·ChatGPT Edu (OpenAI) **평가자** 연구진(문서 판정, 채점 초안 작성에 생성형 AI 보조) **접근 방식** 자료 기준일까지 공개된 제공사 문서·정책·연구로 판정했다(실사용 진단 결과는 이번 호에 싣지 않음).

ID	항목	측정	가중	판정	근거·출처
A. 사용성 (가중치 30%)					
A1	과업 완료 성과	AUTO	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A2	효율성·반응성	AUTO	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A3	시스템 상태 가시성	HYBRID	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A4	오류 예방·복구	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A5	일관성·표준 준수	HYBRID	보조	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A6	접근성 — 자동 검출 영역	AUTO	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A7	접근성 — 수동 검증 영역	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A8	주관적 만족도(SUS 등)	DOC	권장	미측정	고등교육 대상 SUS 등 공개 만족도 조사를 찾지 못해 공개 문서로 판정할 근거가 없었고, 소규모 사용자 평정도 실시하지 않아 판정하지 않았다.
B. 교육적 특성 (가중치 35%)					
B1	학습목표 정렬	DOC	필수	◐	제공사는 학습 모드의 시스템 지시가 능동적 참여·인지부하 관리·메타인지와 자기성찰·호기심 촉진·지지적 피드백을 반영하고, 수준 진단 질문과 퀴즈·개방형 질문으로 이해를 점검한다고 밝힌다. 그러나 강좌 학습목표와의 정렬은 사용자가 자료를 올려 답변 근거를 제한하거나 특정 과목용 맞춤형 GPT 를 만들어 워크스페이스에 공유하는 등 별도 수단에 의존하므로 부분충족으로 판정했다. (출처: OpenAI, 2025; 차덕원·홍후조, 2026; OpenAI, 2026a)
B2	콘텐츠 품질·정확성	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).

ID	항목	측정	가중	판정	근거·출처
B3	피드백·적응	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B4	동기·참여	DOC	권장	◐	공식 발표문은 정답 대신 소크라테스식 질문·힌트·자기성찰 프롬프트로 안내하고 퀴즈·개방형 질문과 맞춤 피드백, 호기심 촉진을 반영한 참여 설계를 밝힌다. 그러나 이는 제공사의 설계 서술이며 동기·몰입을 독립적으로 확인한 근거는 없다. 공개 자료 기반 독립 기능 평정(고등학교 진로·진학 맥락)의 탐색 질문·질문 구체화 만점도 동기·몰입이 아닌 질문 방식을 본 결과여서 부분충족으로 판정했다. (출처: OpenAI, 2025; 차덕원·홍후조, 2026)
B5	교수설계·인지부하	HYBRID	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B6	교수 통합·교수자 지원	DOC	권장	◐	ChatGPT Edu 에서는 특정 과목·과제용 맞춤형 GPT 를 만들어 대학 워크스페이스에 공유할 수 있고, 관리자는 그룹 권한과 GPT 를 관리한다. 그러나 교수자가 강좌 단위로 학습 모드를 배정하는 경로와 교수자용 수업 설계·차별화 지원 자료는 등재 문서에서 확인되지 않았다. Instructure 와 OpenAI 가 발표한 Canvas 의 LLM 활용 과제는 교수자가 Canvas 안에서 AI 상호작용과 학습 목표를 정하는 Canvas 기능이어서 판정을 바꾸지 않았다. (출처: OpenAI, 2026a; Instructure, 2025)
B7	상호작용 사용성·접근성(교육 맥락)	HYBRID	보조	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B8	표준 준수·상호운용	DOC	보조	○	OpenAI 공식 문서에서 LTI 등 교육 표준 연동은 확인되지 않았다. 공식 쿡북(2024 년 9 월)은 맞춤형 GPT 의 Custom Action 으로 Canvas REST API 에 연결하는 방법을 안내할 뿐 LTI 등 교육 표준 연동은 다루지 않는다. Instructure 와 OpenAI 의 Canvas 협력(2025 년 7 월)은 Canvas 안의 기능으로 ChatGPT Edu 자체의 LTI 연동이 아니다. 비표준 API 연동은 LTI 같은 교육 표준(학습 도구를 기관 학습 환경에 연결하는 기술 표준)의 준수로 보지 않았다. (출처: Schule & Casson, 2024; Instructure, 2025; 1EdTech Consortium, 2026)

ID	항목	측정	가중	판정	근거·출처
B9	효과성 증거 등급	DOC	필수	○	제공사와 독립적으로 수행된 고등교육 대상 실험·준실험 연구, 메타분석, 공인 인증으로 학습 모드의 학습 효과를 검증한 근거가 확인되지 않아 독립 효과 근거 없음(증거 등급 최하 단계)으로 판정했다. 학습 모드를 다룬 독립 연구인 공개 자료 기반 기능 평정(고등학교 진로·진학 맥락)은 학습 효과를 측정하지 않았고, ChatGPT 일반 사용 메타분석은 제품 단위 근거가 아니어서 제외했다. (출처: 차덕원·홍후조, 2026)
B10	효과크기 정량	DOC	권장	○	제공사와 독립적으로 수행된 고등교육 대상 연구에서 학습 모드에 특정된 학습 성과 효과크기는 보고되지 않았다. 제공사 발표문에도 학습 효과 연구 결과는 없고, ChatGPT 일반 사용 메타분석의 효과크기는 학습 모드에 적용하지 않았다. (출처: OpenAI, 2025)
C. AI 신뢰성·윤리·거버넌스 (가중치 35%)					
C1	타당성·신뢰성과 환각 관리	HYBRID	필수*	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음). 참고: 독립 연구(모델 수준 평가)에서 웹 검색 없이 ChatGPT-4.1이 낸 참고문헌의 18~22%가 완전 허위였고 웹 검색을 켜면 1~3%였으나 학습 모드 대상 측정은 아니다. (출처: Pastucha et al., 2026)
C2	투명성·설명가능성	DOC	필수*	◐	OpenAI는 발표문에서 학습 모드가 맞춤형 시스템 지시로 작동한다고 공개하고, 이 방식이 대화마다 일관되지 않은 행동과 실수를 낳을 수 있다고 인정했다. 그러나 공개 자료 기반 독립 평정에서 학습 모드는 출처 제시 변인 0점을 받았고, 산출 근거를 이해 가능한 형태로 제시하는 장치는 문서로 확인되지 않았다. (출처: OpenAI, 2025; 차덕원·홍후조, 2026)
C3	공정성·편향 관리	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음). 참고: 독립 연구는 GPT-4o·Gemini 1.5 Pro 등이 학생의 인구학적 특성에 맞춰 교육 콘텐츠를 생성·선택하는 방식에서 유의한 편향을 보고했으나 학습 모드 대상 측정은 아니다. (출처: Weissburg et al., 2025)

ID	항목	측정	가중	판정	근거·출처
C4	프라이버시·데이터 거버넌스	DOC	필수*	◐	기관판(ChatGPT Edu) 문서로 데이터와 대화의 OpenAI 모델 학습 미사용, 관리자의 보관 기간 설정과 삭제 대화의 30 일 내 제거(법적 보존 의무가 있는 경우 제외)를 확인했다. 학생 데이터 개인정보 보호 계약에 따른 처리와 워크스페이스 데이터 내보내기 기본 꺼짐도 확인했다. 그러나 국외 이전·수집 최소화·한국 개인정보보호법 정합은 기관판 문서로 확인되지 않았고, 개인정보위 안내서가 권고하는 데이터 처리 부속서(DPA)와 국외 이전 검토에 해당하는 기관판 근거도 찾지 못했다. 미국 서버 보관·국외 이전 거부권·국내대리인·아동 기준 14 세를 밝힌 한국 이용자 부록은 OpenAI 개인정보 처리방침에 덧붙는 문서로 기관 계약 조건을 담지 않는다. 이 부록은 국외 이전을 거부하면 서비스를 이용하지 못할 수 있다고도 밝혀 기관판의 보호 장치로 보지 않았다. 개인정보위 안내서의 비교표(2025 년 8 월 기준)에서 개인용은 기본적으로 학습에 활용(opt-out 가능)되므로 이 판정은 기관판 도입을 전제로 한다. (출처: OpenAI, 2024; OpenAI, 2026c; OpenAI, 2026b; 개인정보보호위원회, 2025; OpenAI, 2026e)
C5	인간 통제·교수자 역할	DOC	필수	◐	관리자는 Enterprise Compliance API 로 대화와 GPT 의 감사 로그에 접근할 수 있고, 워크스페이스에서는 특정 과목용 맞춤형 GPT 를 만들어 공유할 수 있다. 그러나 교수자나 관리자가 학습 모드를 지정하거나 강제하는 설정은 등재 문서에서 확인되지 않았고, OpenAI 는 학습 모드가 대화마다 일관되지 않은 행동과 실수를 보일 수 있다고 인정했다. 평가·배치 같은 고부담 결정 기능은 없어 해당 요건은 적용하지 않았다. (출처: OpenAI, 2026c; OpenAI, 2026a; OpenAI, 2025)
C6	안전·견고성·보안	HYBRID	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음). 참고: 문서로 확인한 저장·전송 구간 암호화(AES-256, TLS 1.2+)는 판정에 쓰지 않았다. (출처: OpenAI, 2026a)

ID	항목	측정	가중	판정	근거·출처
C7	책무성·거버넌스·정책 정합	DOC	권장	◐	Enterprise Compliance API 를 통한 대화·GPT 감사 로그, 전용 워크스페이스와 관리 콘솔, 학생 데이터 개인정보 보호 계약, 한국 이용자 대상 고지의 국내대리인 지정(개인정보 보호법·정보통신망법)을 확인했다. 인공지능기본법 제 31 조의 사전 고지·결과물 표시 대응은 등재 문서에서 확인되지 않았고, 계도기간 중이어서 법적 이행 여부는 판단하지 않았다. (출처: OpenAI, 2026c; OpenAI, 2026a; OpenAI, 2026e; 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법, 2026; 과학기술정보통신부, 2026b)
C8	AI 리터러시·연령 적합성	DOC	보조	◐	이용약관은 서비스 이용에 동의할 수 있는 최소 연령을 만 13 세(국가 기준 우선)로 정하고 만 18 세 미만에게 부모 또는 법정대리인의 허가를 요구하며, 한국 이용자 부록은 아동 연령 기준을 14 세로 조정한다. 학습 모드와 Edu 안의 AI 리터러시 지원 자료는 제품 문서에서 확인되지 않았다. (출처: OpenAI, 2026d; OpenAI, 2026e)
축 점수 A - · B 26.9 · C 50.0 → 종합 38.5 | 고부담 게이트 해당 없음 | 도입 검토 의견 판정 유보					

26 개 항목 중 문서 판정(DOC) 11 개를 판정했고(필수 13 개 중 5 개), 실사용 진단이 필요한 AUTO·HYBRID 14 개 항목과 공개 근거를 찾지 못한 A8 은 미측정으로 두어 점수 계산에서 제외했다. 필수 항목 측정이 75%에 못 미쳐 도입 검토 의견은 판정 유보이고, 축 A 는 측정 항목이 없어 점수를 내지 않았으며 종합 점수는 축 B·C 만으로 계산한 참고값이다. Gemini 도 같은 11 개 항목을 판정해 공통 항목 기준 점수는 종합 점수와 같다. 두 제품 모두 자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구로만 판정했고, 효과 근거(B9·B10)는 제공사와 독립적으로 수행된 고등교육 연구만 반영했다. 축 점수는 소수 첫째 자리까지 반올림하고 종합 점수는 반올림한 축 점수로 계산했다.

대상 솔루션 Gemini 가이드 학습·Gemini for Education (Google) **평가자** 연구진(문서 판정, 채점 초안 작성에 생성형 AI 보조) **접근 방식** 자료 기준일까지 공개된 제공사 문서·정책·연구로 판정했다(실사용 진단 결과는 이번 호에 싣지 않음).

ID	항목	측정	가중	판정	근거·출처
A. 사용성 (가중치 30%)					
A1	과업 완료 성과	AUTO	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A2	효율성·반응성	AUTO	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A3	시스템 상태 가시성	HYBRID	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A4	오류 예방·복구	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A5	일관성·표준 준수	HYBRID	보조	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).

ID	항목	측정	가중	판정	근거·출처
A6	접근성 — 자동 검출 영역	AUTO	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A7	접근성 — 수동 검증 영역	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
A8	주관적 만족도(SUS 등)	DOC	권장	미측정	고등교육 대상 SUS 등 공개 만족도 조사를 찾지 못해 공개 문서로 판정할 근거가 없었고, 소규모 사용자 평정도 실시하지 않아 판정하지 않았다.
B. 교육적 특성 (가중치 35%)					
B1	학습목표 정렬	DOC	필수	◐	제공사는 가이드 학습의 목표를 정답보다 깊은 이해에 두고 탐색적·개방형 질문, 단계별 분해, 대화형 퀴즈로 활동을 구성한다고 밝힌다. 강좌 학습목표와의 정렬은 교수자가 수업 자료 기반 Gem·NotebookLM 노트북을 배정하는 등 별도 설정에 의존하므로 부분충족으로 판정했다. (출처: Heymans, 2025; Kirtikar, 2025)
B2	콘텐츠 품질·정확성	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B3	피드백·적응	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B4	동기·참여	DOC	권장	◐	공식 발표문은 탐색적·개방형 질문으로 참여를 유도하고 이미지·다이어그램·동영상·대화형 퀴즈를 포함한 멀티모달 응답으로 답보다 과정에 초점을 두는 설계를 밝힌다. 그러나 이는 제공사의 참여 설계 서술이며 동기·참여를 독립적으로 확인한 근거는 없어 부분충족으로 판정했다. (출처: Heymans, 2025)
B5	교수설계·인지부하	HYBRID	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B6	교수 통합·교수자 지원	DOC	권장	●	교수자는 가이드 학습 전용 링크를 Google Classroom 에 게시하거나 학생과 공유할 수 있다. Gemini LTI 는 2024 년 12 월 Canvas·Schoology 와 연동되고 2026 년 5 월 Moodle 로 넓어졌다. 제공사는 수업 자료 기반 Gem·NotebookLM 노트북을 Classroom 과 Gemini LTI(Schoology·Canvas)로 학생에게 배정하는 기능을 발표했다. Gemini in Classroom 은 교사의 수업 계획과 차별화를 돕는 기능을 함께 제공한다. (출처: Heymans, 2025; Kirtikar, 2025; Google, 2024; Google, 2026a)

ID	항목	측정	가중	판정	근거·출처
B7	상호작용 사용성·접근성(교육 맥락)	HYBRID	보조	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음).
B8	표준 준수·상호운용	DOC	보조	●	Gemini LTI 는 학습 도구를 기관 학습 환경에 연결하는 기술 표준인 LTI 방식으로 Canvas(Instructure)·Schoology(PowerSchool) 등 타사 LMS 에 통합되어 2024 년 12 월부터 배포되었다. 2026 년 5 월에는 Moodle 로 확장되어 Moodle 관리자는 LTI Dynamic Registration 규격으로 등록한다. 이용하려면 Workspace 관리자가 Google Workspace LTI 서비스와 Gemini 서비스를 켜야 하며, 1EdTech 인증 여부와 LTI 버전은 공지에서 확인되지 않았다. (출처: Google, 2024; Google, 2026a; 1EdTech Consortium, 2026)
B9	효과성 증거 등급	DOC	필수	○	제공사와 독립적으로 수행된 고등교육 대상 실험·준실험 연구, 메타분석, 공인 인증으로 확인된 가이드 학습의 효과 근거가 없어 독립 효과 근거 없음(증거 등급 최하 단계)으로 판정했다. 참고: 제공사(Google)가 협력 기관과 수행한 사전등록 무작위 통제 실험 보고서(시에라리온 중학교, 동료심사 전)는 점수에 반영하지 않았다. (출처: LearnLM Team & Fab AI, 2026)
B10	효과크기 정량	DOC	권장	○	제공사와 독립적으로 수행된 고등교육 대상 연구에서 가이드 학습의 학습 성과 효과크기는 보고되지 않았다. 참고: 제공사(Google)가 협력 기관과 수행한 시에라리온 중학교 수학 무작위 통제 실험(7·8 학년 1,763 명)은 의도처치 기준 +0.258 표준편차(95% 신뢰구간 0.027~0.488)를 보고했다. 이 보고서는 동료심사 전이고 고등교육 대상이 아니어서 점수에 반영하지 않았다. (출처: LearnLM Team & Fab AI, 2026)
C. AI 신뢰성·윤리·거버넌스 (가중치 35%)					
C1	타당성·신뢰성과 환각 관리	HYBRID	필수*	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음). 참고: 독립 연구(모델 수준 평가)에서 웹 검색 없이 Gemini 2.5 Pro 가 낸 참고문헌의 17~21%가 완전 허위였고 웹 검색을 켜면 0~3%였으나 가이드 학습 대상 측정은 아니다. (출처: Pastucha et al., 2026)
C2	투명성·설명가능성	DOC	필수*	◐	제공사는 가이드 학습이 학습용으로 미세조정한 LearnLM 모델군의 역량을 Gemini 에 넣은 기능이라고 작동 방식을 공개한다. 그러나 등재 문서에서 가이드 학습 응답의 한계 고지와 산출 근거(출처) 제시 방식은 확인되지 않아 부분충족으로 판정했다. (출처: Heymans, 2025)

ID	항목	측정	가중	판정	근거·출처
C3	공정성·편향 관리	HYBRID	필수	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음). 참고: 독립 연구는 GPT-4o·Gemini 1.5 Pro 등이 학생의 인구학적 특성에 맞춰 교육 콘텐츠를 생성·선택하는 방식에서 유의한 편향을 보고했다. 제공사 실험 보고서는 기초 수학 실력이 높은 학생일수록 이득이 컸다고 보고했다(형평성 쟁점). (출처: Weissburg et al., 2025; LearnLM Team & Fab AI, 2026)
C4	프라이버시·데이터 거버넌스	DOC	필수*	◐	기관판 발표문으로 Gemini for Education 이 엔터프라이즈급 데이터 보호와 관리자 통제를 갖춘 Workspace 핵심 서비스이고, 모든 교육 사용자의 대화 데이터를 AI 모델 개선에 쓰지 않는다는 점을 확인했다. 그러나 보관·파기 기간, 수집 최소화·목적 제한, 국외 이전과 한국 개인정보보호법 정합은 기관판 문서로 확인되지 않았고, 개인정보위 안내서가 권고하는 데이터 처리 부속서(DPA)에 해당하는 근거도 찾지 못했다. 국외 이전·국내대리인을 밝힌 Google 의 한국 거주자 대상 추가 정보는 Google 서비스 이용에 적용되는 개인정보처리방침의 추가 정보여서 기관판의 한국법 정합 근거로 보지 않았다. 이 추가 정보는 Gemini 앱 같은 AI 와의 상호작용을 모델 개발·학습에 쓴다고 밝히므로 이 판정은 기관판 도입을 전제로 한다. (출처: Kirtikar, 2025; 개인정보보호위원회, 2025; Google, 2026b)
C5	인간 통제·교수자 역할	DOC	필수	◐	교수자는 가이드 학습 링크를 Classroom 에 게시할 수 있고 수업 자료 기반 Gem·노트북을 Classroom·LTI 로 배정하는 기능도 발표되었으며, 관리자는 Gemini 앱 접근 권한·Vault 대화 검색·사용 보고서를 관리한다. 그러나 가이드 학습을 강제하는 관리자 설정은 등재 문서에서 확인되지 않았다. 평가·배치 같은 고부담 결정 기능은 없어 해당 요건은 적용하지 않았다. (출처: Heymans, 2025; Kirtikar, 2025)
C6	안전·견고성·보안	HYBRID	권장	미측정	실사용 진단이 필요한 항목이라 공개 문서만으로는 판정하지 않았다(실사용 진단 결과는 이번 호에 싣지 않음). 참고: 문서로 확인한 18 세 미만 사용자 대상 강화 콘텐츠 정책과 엔터프라이즈급 데이터 보호는 판정에 쓰지 않았다. (출처: Kirtikar, 2025)

ID	항목	측정	가중	판정	근거·출처
C7	책무성·거버넌스·정책 정합	DOC	권장	◐	관리자의 Gemini 앱 접근 관리, Vault 대화 검색, 사용 보고서 등 책무성 수단과 한국 거주자 대상 고지의 국내대리인 지정(개인정보 보호법·정보통신망법)을 확인했다. 인공지능기본법 제31 조의 사전 고지·결과물 표시 대응은 등재 문서에서 확인되지 않았고, 계도기간 중이어서 법적 이행 여부는 판단하지 않았다. (출처: Kirtikar, 2025; Google, 2026b; 인공지능 발전과 신뢰 기반 조성 등에 관한 기본법, 2026; 과학기술정보통신부, 2026b)
C8	AI 리터러시·연령 적합성	DOC	보조	◐	Gemini 는 18 세 미만에게 더 엄격한 콘텐츠 정책과 AI 리터러시 자료를 담은 청소년 온보딩을 제공한다. 개인 계정의 한국 최소 연령은 14 세이고 이 기준은 Workspace for Education 계정에는 적용되지 않을 수 있다. Gemini LTI 공지는 학생이 안전하고 생산적인 AI 활용을 위한 교육 자료를 이용할 수 있다고 밝히지만 그 내용은 공지에 없다. 대학생(성인) 학습자를 위한 AI 리터러시 지원의 구성은 확인되지 않았다. (출처: Kirtikar, 2025; Google, Retrieved 2026; Google, 2026a)
축 점수 A - · B 42.3 · C 50.0 → 종합 46.2 | 고부담 게이트 해당 없음 | 도입 검토 의견 판정 유보					

26 개 항목 중 문서 판정(DOC) 11 개를 판정했고(필수 13 개 중 5 개), 실사용 진단이 필요한 AUTO·HYBRID 14 개 항목과 공개 근거를 찾지 못한 A8 은 미측정으로 두어 점수 계산에서 제외했다. 필수 항목 측정이 75%에 못 미쳐 도입 검토 의견은 판정 유보이고, 축 A 는 측정 항목이 없어 점수를 내지 않았으며 종합 점수는 축 B·C 만으로 계산한 참고값이다. ChatGPT 도 같은 11 개 항목을 판정해 공통 항목 기준 점수는 종합 점수와 같다. 두 제품 모두 자료 기준일까지 공개된 공식 문서·정부 자료·독립 연구로만 판정했고, 효과 근거(B9·B10)는 제공사와 독립적으로 수행된 고등교육 연구만 반영했다. 축 점수는 소수 첫째 자리까지 반올림하고 종합 점수는 반올림한 축 점수로 계산했다.

주: 판정은 ●(충족)·◐(부분충족)·○(미충족)·미측정으로 표시하며, 미측정 항목은 점수 계산에서 제외했다. 측정 열의 AUTO 는 도구 실측, HYBRID 는 실측 후 검토자 확인, DOC 는 문서 판정을 뜻한다. 실사용 진단이 필요한 AUTO·HYBRID 항목은 공개 문서만으로 판정하지 않아 미측정이며, 축 A 는 측정 항목이 없어 점수를 내지 않았다. 가중 열에 별표가 붙은 항목은 도입 검토 의견의 게이트 항목(C1·C2·C4)이다. 필수 항목 측정이 75%에 못 미쳐 두 제품 모두 도입 검토 의견은 판정 유보이고, 종합 점수는 축 B·C 만으로 계산한 참고값이다. 요약 행의 고부담 게이트(C1·C3·C4·C5 가 모두 충족되어야 함)는 학습자 평가·배치처럼 고부담 결정에 쓰는 도구에만 적용하며, 두 제품은 그런 기능이 없어 해당하지 않는다.

부록 3. 용어 해설

본문에 나오는 주요 용어를 이 보고서의 쓰임에 맞추어 풀이한다.

용어	설명
대규모 언어 모델(LLM)	방대한 텍스트로 학습해 문장을 이해하고 생성하는 AI 모델. ChatGPT·Gemini·Claude 같은 대화형 서비스의 기반이다
학습 모드	정답을 곧바로 주는 대신 질문·힌트·이해 점검으로 학습 과정을 안내하도록 설계된 범용 LLM 의 대화 모드(이 보고서의 조작적 정의). OpenAI 는 study mode(ChatGPT 학습 모드), Google 은 Guided Learning(가이드 학습), Anthropic 은 Learning mode(Claude 학습 모드)로 부른다
시스템 지시(system instructions)	사용자 입력과 함께 모델에 전달되어 응답 방식을 정하는 사전 지시문. ChatGPT 학습 모드는 맞춤형 시스템 지시로 작동한다

용어	설명
미세조정(fine-tuning)	사전 학습된 모델을 특정 목적의 데이터로 추가 학습시키는 방법. Google 의 LearnLM 은 학습용으로 미세조정한 모델군이다
검색 증강 생성(RAG)	질문과 관련된 자료를 먼저 찾아 그 내용에 근거해 응답을 생성하는 방식. 환각을 줄이지만 없애지는 못한다
환각(hallucination)	사실이 아니거나 존재하지 않는 내용(예: 허위 참고문헌)을 그럴듯하게 생성하는 현상
정답 노출(answer leakage)	튜터 역할의 AI 가 학습자가 스스로 풀어야 할 정답을 직접 드러내는 것
가드레일	AI 가 정해진 범위를 벗어난 응답(예: 과제 정답 전체 제공)을 하지 않도록 두는 지시·점검 장치
학습 도구 상호운용성(LTI)	학습 도구를 도구별 로그인 없이 대학의 학습관리시스템(LMS)에 연결하는 1EdTech 컨소시엄의 기술 표준
기관 라이선스	대학 등 기관이 구성원을 위해 계약하는 이용권. 개인 구독과 데이터 처리·관리 조건이 다르다
단일 벤더형·멀티 LLM 통합형	대학이 한 제공사의 서비스를 직접 또는 국내 공급사를 거쳐 기관 단위로 계약하는 방식(단일 벤더형)과, 공급사의 플랫폼을 구독해 구성원이 여러 제공사의 모델을 골라 쓰게 하는 방식(멀티 LLM 통합형)
데이터 처리 부속서(DPA)	기관 이용계약에 붙여 입력 데이터의 소유권·재이용 제한·안전조치·파기 등을 정하는 문서
자기조절학습(SRL)	학습자가 목표를 세우고 학습 과정을 점검·조절하며 결과를 성찰하는 학습
인지적 외주화(cognitive offloading)	기억·판단 같은 인지 과정을 외부 도구에 맡기는 것. 비판적 사고와의 관계는 주로 상관 연구로 보고되었다(Ⅱ 장 4 절 나)
무작위 통제 실험(RCT)	참여자를 처치집단과 통제집단에 무작위로 배정해 효과를 추정하는 연구 설계
의도처치(ITT) 분석	무작위 배정된 집단을 실제 참여 정도와 관계없이 배정된 대로 비교하는 분석 방식
효과크기(g, g+, d)	집단 간 차이를 표준편차 단위로 나타낸 값. g 는 Hedges' g, d 는 Cohen's d 이며, 일부 메타분석은 통합한 평균 효과크기를 g+로 적는다
이질성(I^2)	메타분석에 포함된 연구들의 효과크기가 서로 다른 정도. I^2 는 전체 변동 가운데 연구 간 차이가 차지하는 비율로, 클수록 평균 효과를 일반화하기 어렵다
출판 편향	유의하거나 큰 효과를 보고한 연구가 더 많이 출판되어 메타분석의 평균 효과가 부풀려질 수 있는 현상. Egger 검정 등으로 징후를 점검한다
동료심사 전 자료(프리프린트)	학술지 동료심사를 거치기 전에 공개된 논문·보고서. 게재본과 내용이 다를 수 있다
미측정	판정 근거가 없어 판정하지 않은 평가척도 항목. 점수 계산의 분자와 분모에서 모두 뺀다. 이번 호는 실사용 진단이 필요한 AUTO·HYBRID 항목과 공개 근거를 찾지 못한 항목을 미측정으로 두었다(Ⅰ 장 4 절)
판정 유보	필수 항목 측정이 75%에 못 미치는 등 측정 범위가 부족해 도입 검토 의견의 경계값을 적용하지 않는 상태. 평가척도 v1 에 없던 범주로 이번 호에서 더했으며 두 대표 솔루션이 이에 해당한다(Ⅰ 장 4 절, Ⅲ 장 2 절 다)

부록 4. 기관 도입 점검표

본문에 흩어진 도입 점검 항목을 단계별로 모았다. 담당은 예시이며 각 대학의 조직에 맞게 바꾸어 쓸 수 있다.

단계	점검 항목	담당(예시)	관련 절
계약 전	학습 제외 범위, 보관·삭제 주기, 사람의 검토 여부를 계약 문서와 데이터 처리 부속서로 확인했는가	정보화·개인정보 보호	Ⅱ장 1절 다, Ⅳ장 1절 나
계약 전	국외 이전 여부와 이전받는 자·보관 기간, 공급사 경유 여부를 확인했는가	개인정보 보호	Ⅳ장 1절 나·라
계약 전	사용량 한도, 크레딧 과금 구조, 갱신 조건을 확인했는가	교무·정보화	Ⅱ장 2절 다·라, Ⅳ장 1절 라
계약 전	고지·표시 의무를 제공사·공급사·대학 가운데 누가 이행하는지 정했는가	정보화	Ⅳ장 1절 가
계약 전	LTI 지원과 발표된 연동의 실제 제공 여부를 확인했는가	정보화	Ⅱ장 1절 라
도입 설정	학습 모드 제공 여부와 기관 차원의 강제 설정·강좌 배정 수단을 확인했는가	교수학습개발센터·정보화	Ⅱ장 1절 가, Ⅲ장 3절
도입 설정	통합 인증(SSO) 연동을 확인하고 관리자 통제와 감사 로그의 접근 범위·열람 권한을 정했는가	정보화·개인정보 보호	Ⅱ장 1절 다, Ⅳ장 1절 라
도입 설정	자료 기반 응답과 출처 표시 설정을 확인했는가	교수학습개발센터	Ⅱ장 1절 나, Ⅲ장 3절
도입 설정	한국어 지원 수준과 한국어 설명·국내 제도 질문의 정확성을 한국어 과업으로 점검했는가	교수학습개발센터	Ⅰ장 3절, Ⅱ장 4절 바
수업 운영	과제가 학습 모드 사용을 전제하는지와 AI 사용 공개 방식을 안내했는가	교수학습개발센터	Ⅱ장 4절 다·사, Ⅳ장 2절 나
수업 운영	개인정보·비공개 자료 입력 금지와 외부 서비스의 약관·보관 정책 확인을 안내했는가	교수학습개발센터·개인정보 보호	Ⅱ장 4절 마, Ⅳ장 1절 나
수업 운영	출처 확인을 과제 절차의 한 단계로 넣었는가	교수학습개발센터	Ⅱ장 4절 가
점검·평가	이용률·한도 도달률·교육 이수율 같은 집계 지표를 개인 식별 없이 점검하는가	교무·정보화	Ⅱ장 4절 아, Ⅳ장 2절 나
점검·평가	통제 비교와 지연 측정을 포함한 학내 효과 측정을 설계했는가	교무·교수학습개발센터	Ⅱ장 3절 라, Ⅳ장 2절 나

PUBLICATION INFORMATION

발간 정보

항목	내용
보고서명	월간 에듀테크 솔루션 분석 보고서 Vol.01 LLM 기반 학습 도구
대상 월	2026 년 6 월분 (전 8 호 중 제 1 호)
발간일	2026. 6. 30.
자료 기준일	2026. 6. 30. — 제품의 기능·가격·정책 정보는 이 날짜 현재 기준임
발행처	한림대학교 AI 에듀테크 연구과제
과제	글로컬대학 30 한림대학교 AI 에듀테크센터 연구과제 「고등교육 AI 에듀테크의 통합적 발전 방안 연구 — 전공 운영 모델·연구 방향성 정립·기초교양 AI 리터러시 교육」 [연구 2] 월간 에듀테크 솔루션 분석
연구책임자	김슬기 (한림대학교 일송자유교양대학)
집필진	김슬기(연구책임자)·고영웅·김진환·김병정·김성우·김용수·서동기·육청민·이정근·정혜선·진서연·한민정·한수미·홍수현·황현석 (이상 한림대학교)
문의	skkim@hallym.ac.kr

인용 표기

김슬기·고영웅·김진환·김병정·김성우·김용수·서동기·육청민·이정근·정혜선·진서연·한민정·한수미·홍수현·황현석(2026). 월간 에듀테크 솔루션 분석 보고서 Vol.01: LLM 기반 학습 도구. 한림대학교 AI 에듀테크 연구과제.

이용 안내

○ 이 보고서의 저작권은 한림대학교에 있습니다. 출처를 밝히면 비영리 목적으로 인용·활용할 수 있습니다.

○ 이 보고서에서 다룬 제품·서비스는 분석 목적으로 선정한 것이며 특정 제품의 홍보나 보증을 뜻하지 않습니다.

○ 제품의 기능·가격·정책은 자료 기준일(2026. 6. 30.) 이후 바뀔 수 있습니다.

○ 이 보고서의 내용은 연구진의 연구 결과이며 한림대학교 또는 AI 에듀테크센터의 공식 견해와 다를 수 있습니다.

○ 이 보고서는 HAI-GPT 를 운영하는 한림대학교 AI 에듀테크센터의 연구과제 산출물로, HAI-GPT 는 평가하지 않고 사례로만 다루었습니다. 연구진에는 멀티 LLM 도입을 추진한 교무 부서와 센터 소속 구성원이 포함되며, 비교한 멀티 LLM 통합형 사례(순천향대)는 HAI-GPT 와 같은 공급사(마인드로직 FactChat)의 플랫폼입니다.

사사

본 과제(결과물)는 2026 년도 교육부 및 강원특별자치도의 재원으로 강원 RISE 센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE) 글로컬대학 30 의 결과입니다.(2026-RISE-10-009)

This research was supported by the Regional Innovation System & Education(RISE) Glocal University 30 Project program through the Gangwon RISE Center, funded by the Ministry of Education(MOE) and the Gangwon State(G.S.), Republic of Korea.(2026-RISE-10-009)

MONTHLY EDTECH SOLUTION ANALYSIS REPORT

월간 에듀테크 솔루션 분석 보고서 시리즈

호	대상 월	주제
Vol.01	2026 년 6 월분	**LLM 기반 학습 도구**
Vol.02	2026 년 7 월분	적응형 학습 시스템
Vol.03	2026 년 8 월분	AI 튜터링 플랫폼
Vol.04	2026 년 9 월분	학습분석 도구
Vol.05	2026 년 10 월분	AI 평가·피드백 도구
Vol.06	2026 년 11 월분	인간-AI 공동조절 학습 지원 도구
Vol.07	2026 년 12 월분	교수자용 생성형 AI 도구
Vol.08	2027 년 1 월분	AI 거버넌스·학업 진실성 솔루션
종합	2027 년 1 월	종합 동향 리포트

이번 호 소개

「월간 에듀테크 솔루션 분석 보고서」 Vol.01(2026 년 6 월분)은 LLM 기반 학습 도구를 기술 구조, 시장 동향, 교육적 적용, 한계·논쟁점의 4 분면으로 분석했다. 학습 모드는 정답 보류를 보장하지 않고 메타분석의 큰 효과는 신중히 읽어야 하며 상용 학습 모드의 대학 실험 근거는 확인되지 않았다. ChatGPT 학습 모드와 Gemini 가이드 학습은 기준일까지 공개된 문서로 판정해 둘 다 판정 유보였다. 기관 도입에는 AI 기본법과 개인정보 안내서에 맞춘 데이터 조건과 과제·평가 설계가 필요하다.

키워드 LLM 기반 학습 도구 · 학습 모드 · 고등교육 · 기관 도입 · 데이터 보호

한림대학교 AI 에듀테크 연구과제

글로컬대학 30 한림대학교 AI 에듀테크센터 연구과제「고등교육 AI 에듀테크의 통합적 발전 방안 연구」[연구 2] 산출물