CONDERI

CONDERI ACADEMIC INSIGHT | 학술·연구 인사이트 | 대외 공개판

위임의 값: 에이전트는 보고서를 잘 쓰지만, 책임을 쓸 줄은 아직 모른다

The Price of Delegation: Agents Draft Well, But Accountability Is Still Human Work

본 리포트의 독자 및 활용 가이드

이 리포트는 2026년 하반기 국내외 주요 학술지에 공개된 논문 가운데, AI 에이전트의 조직 도입과 인간-AI 협업의 성과·책임 구조를 실증한 7편을 골라 기술·경영·과학·교육의 네 섹션으로 정리했습니다. 각 논문은 문제의식 → 실증 설계와 수치 → 실전 시사점 순으로 다뤘고, 원문에 없는 해석은 콘데리 분석으로 구분했습니다. 초록 요약이 아니라 **의사결정에 쓸 수 있는 형태**로 재구성하는 것이 이 리포트의 목적입니다.

문서 정보 / Document Information

문서등록번호 Document No.	CONDERI-AR-2026-11-12
발행일자 Published	2026. 11. 12. (목) Thursday
권호 Volume	VOL. 2026-11-12
분석대상 Coverage	국내외 학술지 2026년 하반기 공개 논문 7편
배포구분 Distribution	Public | 일반공개
발행기관 Publisher	CONDERI AI산업전략연구실
조사·집필 Research	AI산업전략연구실

40% → 100%
멀티 에이전트의 출처 명시 충족률
단일 에이전트는 40%에 머물렀다

0.530
에이전틱 AI → 정보비대칭 감소
350개사 설문, p<.001

52편
K-12 AI 인과연구 (1,680편 중)
미국 교실 연구는 10편뿐

48시간
AI 파트너와 협업한 기억 효과 지속
인간 파트너와 유의한 차이 없음

한 문장으로: 에이전트는 초안을 쓰고, 책임은 사람에게 남는다

이번 주 논문 7편이 공통으로 가리킨 것은 성능이 아니라 위임의 구조 — 누가 승인하고 누가 검증하는가

≤30%
금융투자보고서의 '이용자 판단 고지' 충족률. 단일·멀티 에이전트 모두 30% 이하였다

0.037
AI 통합이 협업·유연성을 거쳐 성과로 가는 총간접효과. 단계마다 효과가 줄어든다

94%
4년간 줄어든 청년 일자리 28만5,000개 중 AI 노출 업종 비중 (한국은행)

19%
K-12 AI 인과연구 52편 중 실제 미국 교실에서 수행된 연구 비중 (10편)

기술: 멀티 에이전트는 출처를 살리고 확신을 죽였다

박호연·김경재(2026)는 동일한 언어 모델과 검색 환경을 통제한 뒤 단일 에이전트와 역할 분담형 멀티 에이전트를 비교했습니다. 출처 명시는 40%에서 100%로, 위험 고지는 90%에서 100%로 올랐지만 과도한 확정 표현 억제는 100%에서 80%로 떨어졌습니다. 이용자 판단 고지는 두 조건 모두 30% 이하였습니다. **구조를 나누면 검증 가능성은 올라가지만 책임의 언어는 자동으로 생기지 않는다**는 뜻입니다.

경영: 대리인 비용은 줄고, 거버넌스가 그 폭을 정한다

임원호(2026)는 350개사 설문으로 에이전틱 AI 도입이 정보비대칭 감소(0.530)와 감시효율성 향상(0.494)을 거쳐 대리인 비용을 낮춘다고 보고했습니다. AI 거버넌스의 조절효과는 0.113으로 작지만 유의했습니다. 유현주·정명석(2026)은 231명 자료에서 협업(0.406)과 프로세스 유연성(0.375)이 민첩성으로 이어지되 성과까지의 총간접효과는 0.037에 그친다고 밝혔습니다.

과학: 기억은 파트너가 누구든 협업에서 만들어진다

Lei·Liu(2026)는 AI로 표시된 파트너와의 협업이 이후 개인 회상 수행을 높였고, 그 효과가 48시간 뒤에도 유지됐으며 인간 파트너 조건과 유의한 차이가 없었다고 보고했습니다. 주관적 기억 부담은 오히려 늘지 않았습니다. Science(2026)의 '첫 1000일' 연구는 아동별 언어 습득이 입력의 밀도와 아간 재생(기억 공고화)에 함께 의존한다는 것을 8명의 종단 기록으로 보여줬습니다.

교육: 만족도와 성취는 분리된다

스탠퍼드 교육 AI 허브(2026-11)는 1,680편을 선별해 인과연구 52편을 분석했습니다. 개념 질문에는 긍정적, 과제 대필에는 부정적 효과가 나타났고, 학습 만족도 상승이 시험 성적으로 이어지지 않는 사례가 반복됐습니다. 미국 K-12 교실 연구는 10편(19%)에 불과했습니다. 현장에서는 교사 54.6%가 주 1회 이상 AI를 쓰지만 정책을 안다고 답한 교사는 17.3%였습니다.

이번 주 문헌의 지형: 네 섹션이 한 질문으로 모인다

기술 섹션의 두 논문은 에이전트 구조가 산출물의 검증 가능성과 기억을 어떻게 바꿈지 다뤘고, 경영 섹션의 네 편은 그 구조가 조직의 비용과 민첩성으로 전환되는 경로를 계수로 제시했습니다. 과학 섹션은 인간이 기계와 함께 기억하고 배우는 방식의 조건을, 교육 섹션은 그 조건이 교실에서 충족되지 않을 때 무엇이 무너지는지를 보여줬습니다. 네 섹션을 관통하는 질문은 같습니다 — **위임한 일의 결과를 누가, 어떤 기록으로, 언제 확인하는가.** 이 질문에 답이 있는 조직은 같은 모델로도 더 많은 성과를 남기고, 답이 없는 조직은 만족도만 얻습니다.

한 문장 결론 이번 주 문헌이 공통으로 말하는 것은 성능의 크기가 아니라 **책임의 위치**다 — 에이전트에게 일을 넘길수록 승인·로그·설명가능성의 설계가 성과의 변수가 되고, 그 설계를 가진 조직만 위임의 이익을 남긴다.

Executive Summary: Delegation Without Accountability

Seven studies converge on one point: the value of agents is set by the approval and verification layer, not the model

Technology: role division improves traceability, not accountability

Park and Kim (2026) held the model and retrieval environment constant and compared a single agent with a role-specialised multi-agent system on financial investment reports. Source attribution rose from 40% to 100% and risk disclosure from 90% to 100%, but suppression of overconfident phrasing fell from 100% to 80%. User-judgement notices stayed at or below 30% in both conditions. Splitting the workflow raises verifiability; it does not generate the language of responsibility.

Management: agency costs fall, governance sets the size

Lim (2026), surveying 350 firms, finds agentic AI adoption reduces organisational agency costs through lower information asymmetry (0.530) and better monitoring efficiency (0.494), with AI governance moderating the effect (0.113, p=.013). Yoo and Chung (2026), using data on 231 managers, report that AI integration raises collaboration (0.406) and process flexibility (0.375) and that agility predicts operational performance (0.190), yet the total indirect effect is only 0.037.

Science: memory is built in collaboration, whatever the partner

Lei and Liu (2026) find that collaborating with an AI-labelled partner produces post-collaborative memory benefits that persist after 48 hours and do not differ significantly from a human partner, while subjective memory burden does not rise. In Science, Raviv and colleagues (2026) trained child-specific models on dense home recordings from eight children over their first 1,000 days and show that learning depends on both input density and nightly replay of recent speech.

Education: satisfaction and achievement decouple

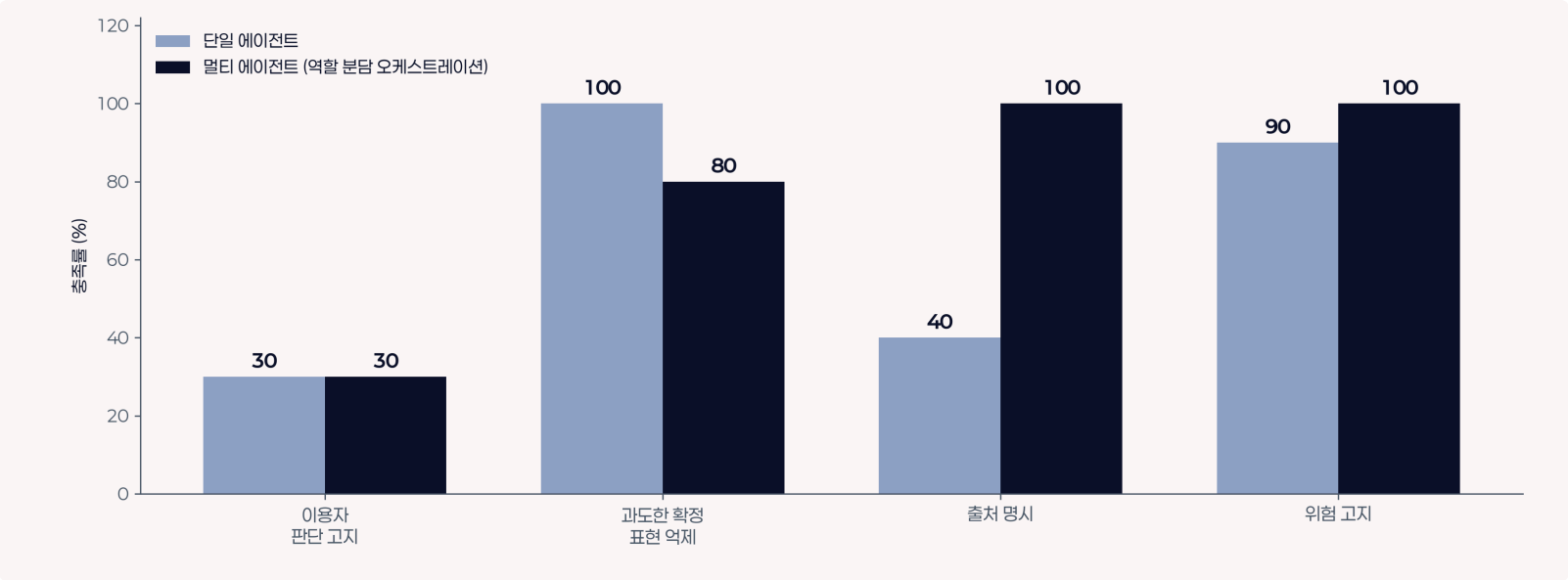
Stanford's AI Hub for Education (November 2026) screened 1,680 papers and analysed 52 causal studies. AI helps when it scaffolds thinking and harms when it substitutes for it. Gains in confidence and wellbeing did not translate into test scores, and only 10 studies (19%) were run in US K-12 classrooms. Among Korean teachers, 54.6% use AI weekly, 94.5% of them generic tools, and only 17.3% say they know the relevant policy.

52편	10편	0.113	15.1%	48h
K-12 AI 인과연구 (1,680편 선별)	미국 교실 현장 연구	AI 거버넌스 조절효과	청년 확장실업률	AI 협업 기억 효과 지속

KEY TAKEAWAY The evidence does not say agents are good or bad. It says the return on delegation is set by the layer that approves, records and explains — and that layer is still human work.

I. 기술(Tech): 멀티 에이전트의 두 얼굴

역할을 나누면 출처는 살아나고 확신은 죽는다 — 통제 실험이 드러낸 구조의 힘과 한계



[그림 1] RAI 체크리스트 충족률 비교. 동일 모델·동일 검색 환경 통제. 출처: 박호연·김경재(2026), 지능정보연구 32(3)

설계: 무엇을 나누면 좋아지는가

출처 명시 40%→100%, 위험 고지 90%→100%. 역할 분담은 '검증 가능한 산출물'을 만드는 데 강했다. 반면 처리시간은 늘고 과도한 확정 표현 억제는 100%→80%로 후퇴했다. 검증은 좋아지지만 단정은 늘어나는 역설이다.

한계: 책임의 언어는 생기지 않는다

이용자 판단 고지는 양 조건 모두 30% 이하였다. 구조를 나눠도 "이 보고서는 판단의 근거가 아니다"라는 문장은 스스로 나오지 않는다. 저자들은 HITL 검토와 도메인 특화 RAI 체크리스트의 결합을 제안한다.

실무 번역

에이전트 파이프라인의 마지막 노드는 '검증 에이전트'가 아니라 '판단 보류 문장'이어야 한다. 산출물 템플릿에 판단 고지·가정·한계를 고정 필드로 넣으면 이 결함은 구조적으로 막을 수 있다.

I-2. 기술: 조직 기억을 가진 에이전트

채봉석·박은혜(2026)의 3계층 구조 — 감지(Sensing) → 포착(Seizing) → 전환(Transforming)

문제의식: 예측은 모델이 아니라 기억이 한다

외식 수요예측은 프로모션·날씨·지역 이벤트가 겹치는 영역입니다. 단일 모델은 과거 패턴을 맞히지만 '왜 지난번 예측이 틀렸는지'를 축적하지 못합니다. 저자들은 이 간극을 동적역량 이론으로 재구성해, 수요예측을 **조직 학습과 기억의 문제**로 다시 정의했습니다. 6개 연속 기간의 워크플로 실증에서 감지→포착→전환의 3계층 에이전트 구조가 예측 정확도뿐 아니라 예측의 근거를 남기는 것을 보여줬습니다.

구조: 세 계층이 하는 일

감지는 외부 신호(날씨·이벤트·경쟁 프로모션)를 수집하고, **포착**은 그 신호를 기존 지식과 대조해 기회를 판별하며, **전환**은 실행 후 결과를 기억에 다시 씁니다. 핵심은 세 번째 단계입니다. 결과를 기록하지 않는 에이전트는 매번 같은 실수를 반복합니다. 이 구조는 '에이전트 오케스트레이션'을 기능 목록이 아니라 **기억의 순환**으로 설계해야 한다는 것을 시사합니다.

연결: 데이터 맥락이 성과를 가른다

박기준·조미영(2026)은 LLM 분석 에이전트에서 컬럼 의미·업무 용어·지표 산식 같은 데이터 맥락이 과업 유형별로 다르게 작동함을 통제 실험으로 보였습니다. 단순 조회는 기본 스키마만으로 천장효과, 지표 계산은 확장 메타데이터, 업무 해석은 시맨틱 레이어가 핵심이었습니다. 사전 매핑으로 맥락을 골라 주는 전략은 전체 제공보다 성과가 낮았고, 사후 필요맥락 포함률은 34.9%에 그쳤습니다.

시사점: '기억'과 '맥락'은 제품 사양이다

두 연구를 겹치면 실무 지침이 나옵니다. ① 결과·오류·수정 이력을 저장하는 기억 계층을 제품에 내장하라. ② 과업 유형을 먼저 분류하고 그에 맞는 맥락만 제공하라. ③ 맥락을 사전에 좁히지 말고 실행 중 필요에 따라 주입하라. **에이전트의 경쟁력은 모델이 아니라 축적 구조**에서 나옵니다.

조직 단위 거버넌스

Takahashi 등(2026)은 일본 산업재별의 AI 프로젝트 5개를 비교해, 에이전트를 **재사용 가능한 자산 포트폴리오(System of Agents)**로 관리할 때만 가치가 축적된다고 보고했습니다. 조건은 자산화, 표준화된 로그, 사람 개입 루틴, 그리고 재사용을 허용하는 의사결정 권한입니다. 프로젝트가 늘어도 학습이 남지 않으면 비용만 늘어납니다.

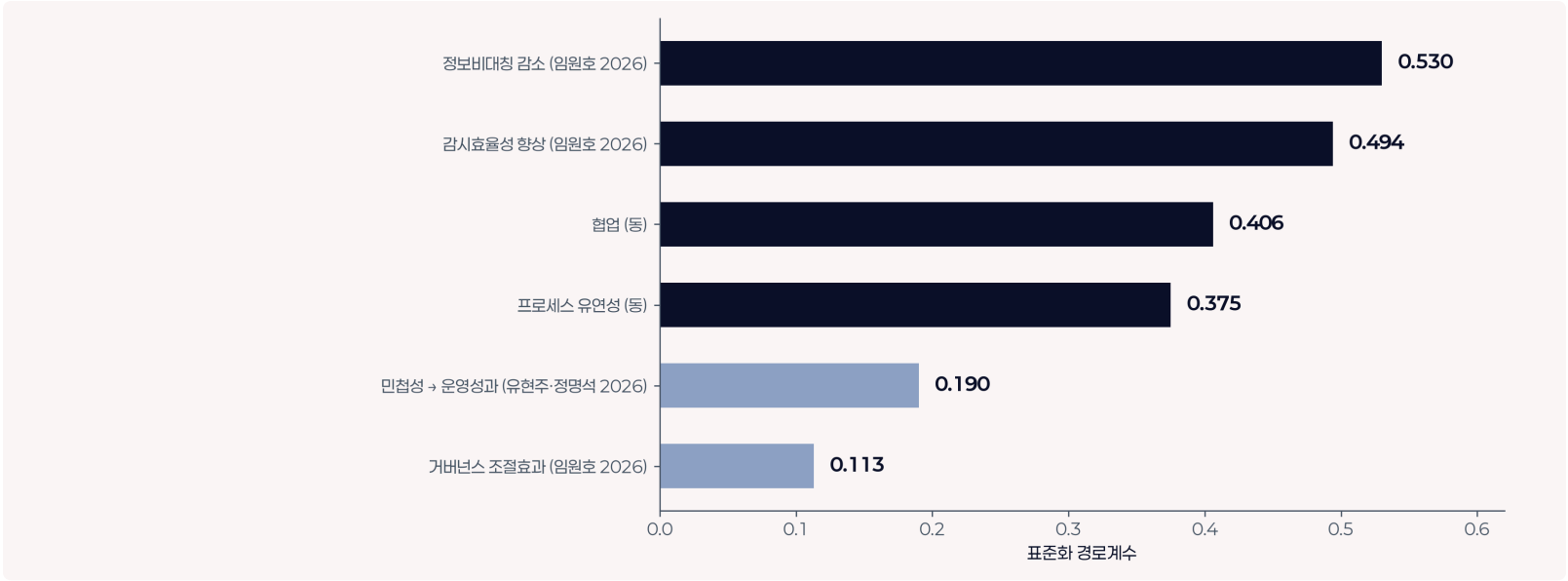
구현 체크리스트: 기억하는 에이전트를 만드는 다섯 단계

① **과업 분류** — 조회·계산·해석을 먼저 나눠라. 맥락 요구가 다르므로 같은 프롬프트로 묶으면 계산 과업에서 정확도가 떨어진다(박기준·조미영 2026). ② **기억 스키마** — 결과값만이 아니라 판단 근거·가정·틀린 이유를 함께 저장하라. 저장 단위가 곧 재사용 단위다. ③ **재생 루프** — 세션 종료 후 요약을 다시 입력에 넣는 절차를 자동화하라. Science(2026)의 야간 재생과 같은 기능이다. ④ **전환 기록** — 실행 결과를 기억에 재기록하는 단계를 생략하지 마라(채봉석·박은혜 2026). ⑤ **자산화** — 에이전트를 프로젝트 산출물이 아니라 재사용 자산으로 등록하고, 표준 로그와 사람 개입 지점을 문서화하라(Takahashi 등 2026).

KEY TAKEAWAY 모델을 바꾸지 않고 얻을 수 있는 개선이 구조에 남아 있다 — 과업 분류, 기억 스키마, 재생 루프, 전환 기록, 자산화의 다섯 단계다.

II. 경영(Management): 대리인 비용과 민첩성의 두 경로

에이전틱 AI는 정보비대칭을 줄이고, AI 통합은 협업을 바꾼다 — 그러나 성과로 가는 마지막 구간이 얇다



[그림 2] 국내 실증 연구의 표준화 경로계수. 출처: 임원호(2026) 산업진흥연구 11(3) / 유현주·정명석(2026) 한국컴퓨터정보학회논문지 31(8)

경로 ① 대리인 비용 (350개사)

정보비대칭 0.530, 감시효율 0.494로 둘 다 p<.001. 두 매개가 부분매개로 작동했고, AI 거버넌스 조절효과는 0.113(p=.013). 로그 추적·승인 체계·설명가능성·책임 소재가 갖춰질수록 같은 도입이 더 큰 비용 절감을 낳는다.

경로 ② 조직 민첩성 (231명)

AI 통합 → 협업 0.406, 프로세스 유연성 0.375. 민첩성 → 운영성과 0.190(p=.003). 두 연쇄 경로는 유의하지만 총간접효과는 0.037로 작다. 민첩성은 도입의 자동 산물이 아니라 협업 패턴의 재편 결과다.

두 연구의 교차점

효과는 '도입'이 아니라 '재편'에서 나온다. 거버넌스(승인·로그)와 협업 구조(프로세스 유연성)가 같은 방향을 가리킬 때 성과가 남는다. 도구만 지급하면 만족도만 오른다.

II-2. 경영: 알고리즘 관리와 인간-AI 협업은 다르게 작동한다

강현정(2026)은 두 역할을 2차 요인으로 분해해 서로 다른 매개 경로를 확인했다

설계: 4개 차원 × 6개 활용 목적

Benbya 등(2021)의 프레임워크를 따라 알고리즘 관리(AIgM)와 인간-AI 협업(HAC)을 맥락·주체성· 상호작용·결과의 4개 차원 2차 요인으로 개념화하고, 6개 AI 사용 목적을 순차 경로 모델에 넣었습니다. 정보시스템연구(ISR) 28(1)에 실린 이 연구는 128명 자료에 PLS-SEM을 적용했고, 모델의 예측력은 Q^2=0.524로 업무 성과를 상당히 설명했습니다.

발견: 같은 AI, 다른 경로

HAC와 AIgM은 모두 근접 AI 활용(코드 작성, 일반 정보 탐색)을 유의하게 높였지만 매개 경로가 달랐습니다. 협업은 관계적·평가적 역학(튜링 테스트, 단순 과업)을 거쳐, 관리는 운영적 실행 (디버깅)을 거쳐 성과로 이어졌습니다. **"AI를 관리자로 쓸 것인가 동료로 쓸 것인가"에 따라 조직이 얻는 역량이 다르다**는 실증입니다.

실무 적용 ①: 역할을 명시하라

같은 도구도 '평가자'로 쓰면 판단 역량이, '실행자'로 쓰면 디버깅 역량이 자랍니다. 팀 온보딩에서 AI의 역할을 문서로 정하고, 역할별로 기대 역량을 다르게 측정해야 합니다.

실무 적용 ②: 균형을 측정하라

관리와 협업 중 하나만 강하면 성과가 특정 과업에 갇힙니다. 분기마다 두 축의 사용 비중과 결과를 함께 기록하면, 어느 쪽이 병목인지 데이터로 판단할 수 있습니다.

주의: 자기보고의 한계

두 연구 모두 설문 기반 자기보고입니다. 효과 크기는 '인식된 성과'이며, 실제 매출·생산성과의 간극은 별도 검증이 필요합니다. 도입 효과를 주장할 때는 로그·산출물 지표를 함께 제시하십시오.

연구	표본·방법	핵심 계수	조직에 주는 의미
임원호 (2026)	350개사 설문, 구조방정식	정보비대칭 0.530 / 감시효율 0.494 / 거버넌스 0.113	거버넌스가 갖춰질수록 같은 도입이 더 큰 비용 절감을 낳는다
유현주·정명석 (2026)	231명 관리자, PLS-SEM 5,000회 부트스트랩	협업 0.406 / 유연성 0.375 / 성과 0.190 / 총간접 0.037	민첩성은 자동 산물이 아니다. 협업 패턴 재편이 필요하다
강현정 (2026)	128명, PLS-SEM·PLSpredict	예측력 Q^2=0.524	관리와 협업은 다른 경로로 성과에 도달한다
박기준·조미영 (2026)	통신 도메인 합성데이터, 통제 실험	사후 필요맥락 포함률 34.9%	맥락은 사전 선별보다 실행 중 주입이 효과적이다

KEY TAKEAWAY 네 편의 연구가 같은 곳을 가리킨다 — 효과는 도입이 아니라 재편에서 나온다. 거버넌스와 협업 구조가 함께 움직일 때만 성과가 남는다.

III. 과학(Science): 기억은 어디에서 만들어지는가

Science의 '첫 1000일' 언어 습득 모델과 AI 파트너 협업기억 연구가 같은 곳을 가리킨다 — 입력의 밀도와 재생

Science (2026-11-05): 아동별 모델, 8명의 1000일

Raviv 등은 'First 1000 Days' 프로젝트의 초밀도 가정 녹음(아동 1인당 수천 시간)으로 8명 각각에 대해 별도의 언어 학습 모델을 만들었습니다. 사전 지식 없이 그 아이가 실제로 들은 발화만을 시간 순서대로 학습시켰고, 수면의 기억 공고화 역할을 모사해 하루 끝에 최근 입력을 재생했습니다. 결과는 두 가지였습니다. 첫째, **입력의 밀도**가 결정적이었습니다. 일부 녹음만 준 모델은 훨씬 적은 단어를 학습했습니다. 둘째, **야간 재생**이 그에 못지않게 중요했습니다. 재생이 거의 없는 모델은 입력이 풍부해도 단어 학습에 실패했습니다.

Current Psychology (2026-11-09): AI 파트너와의 협업기억

Lei·Liu는 범주화된 단어목록으로 세 조건(AI로 표시된 파트너, 인간 파트너, 개인)을 비교했습니다. 협업 후 개인 회상 검사를 즉시와 48시간 뒤에 실시한 결과, AI 파트너 조건에서도 협업 후 이득이 유의했고 48시간 뒤까지 유지됐습니다. 인간 파트너 조건과의 유의한 차이는 없었습니다. 또한 협업은 주관적 기억 부담을 증가시키지 않았습니다. **'AI라는 표시'가 인지적 파트너로서의 효능을 훼손하지 않는다**는 것이 이 연구의 핵심입니다.

Nature Machine Intelligence (2026): 뇌 신호로 추론을 유도하다

Xiao 등은 fMRI로 측정한 인간의 연역추론 뇌 활성 패턴과 10개 오픈소스 LLM의 내부 표상을 비교하고, 뇌 신호로 모델 표상을 유도했을 때 연역추론 성능이 개선되며 기존 언어 기반 지도학습과 상보적임을 보였습니다. 인간과 모델의 표상이 부분적으로 정렬된다는 증거입니다.

해석: 설계 원칙 세 가지

① 학습·업무 시스템에 '재생(replay)'을 설계하라 — 회고·복기 없는 자동화는 기억을 남기지 않는다. ② 입력의 밀도가 결과를 정한다 — 데이터 양보다 **해당 개인·조직에 실제로 도달한 정보**가 중요하다. ③ 파트너가 AI라는 사실은 학습 효과를 떨어뜨리지 않는다 — 인터페이스의 문제이지 인지의 문제가 아니다.

제품 함의

AI 튜터의 성능은 모델 크기보다 '그 학생의 실제 입력'과 '복기 세션'에서 나온다. 개인화의 최소 단위는 학급이 아니라 학생이며, 세션 종료 후 요약·재생을 제품 루프에 넣는 것이 학습 효과의 실질적 레버다.

인지과학이 제품에 던지는 세 가지 질문

첫째, 우리 제품은 사용자의 어떤 기억을 대신하고 있는가. 대신하는 기억이 많을수록 사용자는 편해지지만 학습은 줄어든다. AI 튜터가 '정답'을 기억하고 학생이 '과정'을 기억하도록 설계해야 한다. **둘째, 사용자가 무엇을 다시 꺼내게 하는가.** 요약·오답 노트·복기 알림은 기능이 아니라 학습 루프의 필수 부품이다. **셋째, 파트너가 AI라는 사실이 성과를 떨어뜨리는가.** Lei·Liu(2026)의 결과는 '아니오'다. 문제는 정체성이 아니라 인터페이스다 — 언제 말을 걸고, 어떤 형태로 확인을 요구하며, 실패했을 때 어떻게 설명하는지가 신뢰를 만든다.

KEY TAKEAWAY 재생 없는 학습은 없다. 입력의 밀도와 최근 경험을 다시 꺼내는 절차가 함께 있을 때만 기억이 남는다 — 아동의 언어도, 조직의 예측도, AI와의 협업도.

III-2. 과학: 세 편의 연구를 하나의 표로 읽기

대상·설계·핵심 수치·실무 함의를 같은 형식으로 정리했다

연구	설계	핵심 발견	실무 함의
Raviv 등 (2026) **Science 394(6824)**	아동 8명, 1000일 가정 녹음, 아동별 증분 학습 모델	입력 밀도와 어간 재생이 모두 필요. 두 요건이 빠지면 단어 학습이 붕괴. 3개 내부 파라미터가 학습 속도 차이를 설명	개인화 데이터의 도달 밀도와 복기 루프를 제품 사양으로 명문화. 학급 단위 평균은 개인차를 지운다
Lei·Liu (2026) **Current Psychology**	단어목록 협업 회상, 3조건, 즉시·48시간 지연 검사	AI 파트너 조건에서 협업 후 이득 유의, 48시간 유지, 인간 파트너와 차이 없음. 주관적 부담은 증가하지 않음	AI를 '동료'로 배치한 학습·업무 설계가 인지적으로 유효. 브랜딩보다 인터페이스 설계가 중요
Xiao 등 (2026) **Nature Machine Intelligence**	fMRI 연역추론 데이터, LLM 10종 표상 비교 및 유도	인간 뇌와 모델 표상의 부분적 정렬. 뇌 신호 유도가 추론 성능을 개선하고 언어 기반 지도와 상보적	인지과학적 신호를 모델 평가·개선 루프에 활용할 여지. 규제·설명가능성 논의의 근거
채봉석·박은혜 (2026) **KMIS 춘계**	외식 수요예측, 6개 연속 기간 워크플로 실증	감지→포착→전환 3계층 구조가 예측과 근거를 함께 남김. 조직 학습·기억이 성과의 매개	에이전트 결과를 기억에 재기록하는 단계를 파이프라인에 필수로 포함

공통 명제: 재생 없는 학습은 없다

아동의 언어 습득, AI와의 협업 기억, 조직의 수요예측 — 세 층위에서 반복되는 조건은 '최근 경험을 다시 꺼내는 절차'입니다. 자동화가 이 절차를 없애면 효율은 올라가지만 학습은 멈춥니다.

측정 제안: 세 가지 로그

① 입력 도달률(사용자별 실제 사용 데이터 비중) ② 복기 실행률(세션 후 요약 열람·재검토 비율) ③ 수정 이력 재사용률(과거 오류·수정이 다음 과업에 반영된 비율). 세 지표가 오르지 않는 도입은 만족도만 올린다.

연구의 한계

아동 8명은 표본이 작고, 협업기억 연구는 단기 실험실 과제입니다. 생태적 타당성은 후속 연구의 몫입니다. 본 리포트는 수치를 일반화하지 않고 설계 원칙의 근거로만 사용했습니다.

왜 층위가 다른 연구를 함께 읽는가

아동의 언어 습득(개인), AI와의 협업기억(쌍), 조직의 수요예측(조직)은 분석 단위가 다르지만 같은 메커니즘을 공유합니다. 개인 수준에서 확인된 '재생'과 '입력 밀도'가 조직 수준에서 '전환 기록'과 '데이터 도달률'로 번역되고, 제품 수준에서는 '세션 후 요약'과 '개인화 데이터'로 구현됩니다. 한 층위에서 확인된 원리가 다른 층위에서 반복되는지 확인하는 것이 실무적으로 유용한 이유는, **제품 기능을 인지 원리에 대응시켜 설명할 수 있기 때문**입니다. 투자자·구매자에게 "왜 이 기능이 필요한가"를 모델 성능이 아니라 학습 원리로 답할 수 있게 됩니다.

KEY TAKEAWAY 개인·쌍·조직의 층위에서 반복되는 원리는 같다 — 도달한 정보의 밀도, 그리고 그것을 다시 꺼내는 절차. 제품 사양서에 이 두 줄을 넣는 것이 개인화의 출발점이다.

IV. 교육(EduTech): 52편의 인과연구가 말한 것

스탠퍼드 교육 AI 허브의 2026년 11월 검토 — AI는 만능키도 재앙도 아니고, 쓰임이 결과를 가른다

리뷰의 구조: 1,680편 → 52편 → 10편

연구팀은 전 세계 관련 논문 1,680여 편 가운데 인과관계를 엄격히 검증한 실증 연구 52편을 정밀 분석했습니다. 3월 첫 검토(20편)보다 32편이 늘었습니다. 그러나 이 가운데 실제 미국 K-12 교실에서 학생을 대상으로 한 연구는 10편(19%)에 불과했습니다. 나머지는 성인·대학생이거나 통제된 단기 실험실 환경이었습니다. 노스웨스턴대 엘리자베스 팁튼 교수의 지적처럼, 효과도 부작용도 충분히 모르는 상태에서 전면 도입이 진행되고 있습니다.

발견 ① 보조와 대필은 정반대다

개념 이해를 위한 질문, 맞춤형 연습문제 생성처럼 **사고를 지원하는 사용**은 개념 이해도와 학습 지속성을 높였습니다. 반면 과제·에세이를 통째로 맡기는 **외주화**는 인지적 노력 감소와 학습 결손으로 이어졌습니다. 수잔나 로브 교수는 "종이가 좋은지 나쁜지 묻는 것과 같다"며 "교실 내 명확한 가이드라인이 없으면 오히려 학습을 빼앗길 수 있다"고 경고했습니다.

발견 ② 만족도와 성적은 분리된다

많은 학생이 AI 사용 후 자신감·웰빙·만족도가 높아졌다고 답했지만, 그것이 시험 성적이나 장기 개념 습득으로 이어지지 않았습니다. '공부가 잘되고 있다'는 느낌이 실제 성취와 어긋날 수 있다는 점이 이번 리뷰의 가장 실무적인 경고입니다.

발견 ③ 교사용 AI는 경력에 따라 다르다

수업 계획·맞춤 피드백에서 AI는 업무 부담을 줄이고 피드백 질을 높였습니다. 그러나 교육 경험이 적은 신규 교사가 과도하게 의존하면 교재 연구로 축적해야 할 전문성이 약화되고 학생 몰입도가 떨어졌습니다. 경력 단계별로 다른 가이드가 필요합니다.

한국 현장: 인식의 격차

KEDI 조사(교사 271명)에서 54.6%가 주 1회 이상 AI를 사용하고 94.5%가 범용 생성형 AI를 씁니다. 개인정보·학습데이터 문제를 겪은 교사는 40.6%, 관련 법·정책을 안다고 답한 교사는 17.3%였습니다. 사용은 보편화됐지만 규범은 아직 도착하지 않았습니다.

활용 유형	인과연구 결과	교실 적용 원칙
개념 학습·맞춤 연습	긍정 (이해도·지속성 향상)	질문 생성·힌트·연습문제를 사고 과정에 붙인다. 정답 제시는 마지막 단계로 제한한다
과제 대필·외주화	부정 (인지 노력 감소, 학습 결손)	제출물에 과정 기록(질문·수정 이력)을 요구하고, 초안 작성은 교실 안에서 시작한다
교사 수업 설계·피드백	긍정 (업무 부담 감소, 피드백 질 향상)	피드백 초안은 AI, 최종 판단은 교사. 채점 기준표를 사전에 고정한다
신규 교사의 과의존	부정 (전문성·몰입 저하)	경력 3년 미만 교사는 AI 사용 전 교재 연구 시간을 의무 배정한다
정서·만족도	성과와 분리 (자신감↑ ≠ 성적↑)	만족도 조사를 성과 지표로 쓰지 않는다. 성취도 연계 데이터를 별도로 수집한다

KEY TAKEAWAY AI는 만능키도 재앙도 아니다. 사고를 지원하면 성과가 오르고 대신하게 하면 학습이 빠진다 — 교실의 규칙은 기능이 아니라 용도에 적어야 한다.

IV-2. 교육: 학교가 요구하는 다섯 줄의 계약

앤트로픽·교원노조 협약과 1EdTech 표준이 사실상의 구매 사양이 되고 있다

요구 조항	내용 (앤트로픽 × AFT·UFT 구속력 협약)	제품·영업 대응
학습 금지	학생·교사 데이터를 모델 학습·미세조정·벤치마크·개선에 사용 금지. 비식별화·합성데이터 원료 사용도 금지, 가수집분 소급 적용	데이터 처리 계약서에 목적 제한 조항을 넣고 로그로 증명 체계를 갖춘다
삭제 180일	삭제 요청 시 운영 시스템에서 180일 내 영구 삭제, 백업도 동일 기한. EU·북미 내 저장	삭제 파이프라인과 보관 위치를 문서화. 국내 저장 요구에는 리전 선택 옵션으로 대응
통지 72시간	침해 확인 또는 의심 시 72시간 내 학교에 통지. SOC 2 Type II·ISO 27001/27701·ISO 42001 유지	인증 로드맵을 입찰 서류에 포함. 사고 대응 훈련 기록을 첨부
에이전트 기본 꺼짐	메시지 전송·과제 제출·구매 등 대신 행동하는 기능은 학생 계정에서 기본 비활성. 관리자 승인 후 활성화	권한 위임 UI를 '기본 거부' 원칙으로 재설계. 관리자 콘솔에 승인 로그 노출
동반자형 기능 금지	정서적 의존 유도, 학습에 필요한 시간보다 오래 붙잡기, 친구 흉내내기 금지. 위험 주제 대응 절차 필수	체류시간 최적화 지표를 교육 제품에서 분리. 위기 대응 프로토콜을 기획 단계에 포함

왜 계약이 먼저인가

스탠퍼드 리뷰가 보여준 것은 효과의 조건부성입니다. 조건이 확정되지 않은 상태에서 학교가 선택할 수 있는 유일한 확실한 기준은 데이터·안전 조항입니다. 기능 비교표보다 계약 조항이 구매 결정을 먼저 가릅니다.

표준의 이동

1EdTech는 '신뢰할 수 있는 이동형 학습 맥락(TPLC)' 6원칙으로 데이터 범위를 교육기관이 관리하는 구조를 논의 중입니다. 한국에는 교육과정·직무역량 데이터를 국제 표준과 연결할 것을 요청했습니다. 표준을 먼저 구현한 기업이 수출에서도 유리합니다.

검증 과제

리뷰가 지적한 가장 큰 공백은 '현장 근거'입니다. 국내 기업이 강점을 만들 수 있는 지점은 교실 단위 무작위 비교, 성취도 연계 데이터, 교사 개입 지점의 표준화입니다. 마케팅이 아니라 논문으로 증명하는 회사가 다음 입찰을 가져갑니다.

국내 사례	데이터 접점	계약 준비도 (콘데리 평가)	우선 보완 지점
서울 1,300개교 AI 챗봇	질의응답 로그, 관리자 대시보드	로그 보유·모니터링 충족, 삭제 기한 명시 미확인	보관 기간·삭제 절차의 계약 명문화
테크빌 마이클 AI	한글 문서·생활기록부 초안	교내 양식 호환 강점, 학습 활용 금지 조항 필요	모델 학습 미사용 증명·감사 로그
비주얼캠프 시선추적	읽기 행동·문해력 평가 30만 건	대규모 행동 데이터 보유, 국외 이전 고지 필요	저장 지역·재식별 방지 설계
충남 미주온·아이글	학습 지표 5종, 손글씨 답안	교육청 주도 거버넌스, 교사 최종 판단 구조	AI 1차 피드백의 설명가능성 문서화

※ 계약 준비도는 공개 보도와 서비스 설명을 근거로 한 콘데리 평가이며, 각 사의 실제 계약 조항을 확인한 것이 아닙니다.

V. 종합: 7편이 공통으로 말하는 네 가지

성능·구조·기억·책임 — 도입 효과를 결정하는 변수는 모델 밖에 있다

1. 구조가 성능의 절반이다

역할 분담형 멀티 에이전트는 출처 명시를 2.5배로 끌어올렸습니다(40%→100%). 3계층 오케스트레이션은 예측의 근거를 남겼습니다. 모델을 바꾸지 않고도 얻을 수 있는 개선이 구조에 남아 있습니다.

2. 기억이 없으면 학습이 없다

아동의 언어 습득(재생), 조직의 수요예측(전환), AI와의 협업기억(48시간 유지)이 같은 원리를 가리킵니다. 결과를 다시 꺼내는 절차가 제품과 조직에 내장돼야 효과가 축적됩니다.

3. 책임은 자동으로 생기지 않는다

이용자 판단 고지는 멀티 에이전트에서도 30% 이하였습니다. 거버넌스의 조절효과는 0.113으로 작지만 유의했습니다. 승인·로그·설명가능성은 비용이 아니라 성과의 조건입니다.

4. 효과는 마지막 구간에서 감쇠한다

AI 통합 → 협업(0.406) → 민첩성(0.245) → 운영성과(0.190)로 갈수록 계수가 줄고, 총간접효과는 0.037이었습니다. 도입 초기의 큰 효과가 성과로 그대로 이어지지 않는다는 뜻입니다. 감쇠를 줄이는 방법은 의사결정 권한의 재배치와 프로세스 재설계이지, 도구의 추가 도입이 아닙니다.

창업가를 위한 세 문장

첫째, 에이전트 제품의 차별점은 오케스트레이션과 기억 계층에서 만드십시오.
둘째, 교육·공공 시장은 계약 조항이 기능보다 먼저 심사됩니다.
셋째, 성과 주장에는 로그와 산출물 지표를 붙이십시오. 만족도는 성과가 아닙니다.

2027년을 준비하는 세 가지 결정

첫째, 도입 예산의 일부를 '기록 예산'으로 전환하라. 새 도구를 하나 더 사는 대신 승인 로그·산출물 검증·복기 세션에 예산을 배정하라. 임원호(2026)의 거버넌스 조절효과(0.113)는 작지만, 그것이 없으면 다른 효과도 성과로 남지 않는다.
둘째, 교육·공공 제품은 계약서를 먼저 써라. 학습 금지·180일 삭제·72시간 통지·에이전트 기본 꺼짐은 이제 사실상의 구매 사양이다. 제품 기획 단계에서 이 다섯 줄을 요구사항으로 고정하면 영업 사이클이 짧아진다.
셋째, 성과 주장의 단위를 바꿔라. 만족도·사용량이 아니라 검증 가능한 산출물 지표(오류율, 수정 재사용률, 재작업 시간)로 보고하라. 스탠퍼드 리뷰가 보여준 만족도와 성취의 분리는 교육 시장에만 국한된 현상이 아니다.

KEY TAKEAWAY 성능·구조·기억·책임 — 네 변수 가운데 모델 밖에 있는 것이 셋이다. 2027년의 격차는 어떤 모델을 쓰는지가 아니라, 그 셋을 누가 설계하는지에서 갈린다.

부록: 문헌 목록 및 유의사항

References and Notes

분석 문헌 (2026년 공개)

① 박호연·김경재, 「단일 에이전트와 멀티 에이전트 금융투자보고서 생성 비교 평가」, 지능정보연구 32(3), 2026. DOI 10.13088/jiis.2026.32.3.039
② 임원호, 「에이전틱 AI 도입이 조직 대리인 비용 감소에 미치는 영향」, 산업진흥연구 11(3), pp.271-279, 2026. DOI 10.21186/IPR.2026.11.3.271
③ 유현주·정명석, 「AI 통합이 조직 민첩성에 미치는 영향」, 한국컴퓨터정보학회논문지 31(8), pp.243-252, 2026. DOI 10.9708/jksci.2026.31.08.243
④ Hyunjeong Kang, 「The Impact of Algorithmic Management and Human-AI Collaboration in AI-Augmented Organizations」, Information Systems Review 28(1), 2026. DOI 10.14329/isr.2026.28.1.355
⑤ 박기준·조미영, 「기업 LLM 분석 에이전트의 데이터 맥락이 분석 성과에 미치는 영향」, 한국지능정보시스템학회 32(3), pp.449-473, 2026
⑥ 채봉석·박은혜, 「An Agentic AI System to Improve Organizational Learning and Memory」, 한국경영정보학회 춘계통합학술대회, pp.978-981, 2026
⑦ N. Takahashi, Y. Kohda, H. Gokon, 「Governing Agentic AI as Organizational Portfolios for Sustained Value」, Journal of Creating Value, 2026. DOI 10.1177/23949643261450309
⑧ H. Raviv et al., 「The first 1000 days: Modeling language learning from individual children's everyday speech」, Science 394(6824), 2026. DOI 10.1126/science.aeh9311
⑨ H. Lei, X. Liu, 「Collaboration with an AI-labelled partner elicits durable post-collaborative benefits」, Current Psychology, 2026. DOI 10.1007/s12144-026-09993-8
⑩ M. Xiao et al., 「Beyond representational alignment with brain-guided language models for robust reasoning」, Nature Machine Intelligence, 2026. DOI 10.1038/s42256-026-01278-w
⑪ Stanford AI Hub for Education, 「The Evidence Base on AI in K-12: A November 2026 Review」, 2026-11-10
⑫ 한국교육개발원(KEDI) 교사 AI 활용 조사, 경향신문 보도(2026-10-19), 교사 271명 응답

유의사항

인용한 계수는 각 논문이 보고한 표준화 경로계수·충족률이며, 서로 다른 표본과 설계에서 나온 값이므로 직접 비교와 합산은 제한적입니다. 설문 기반 연구의 효과는 자기보고 성과를 포함합니다. 본 리포트의 해석과 '실무 번역'은 콘데리 분석이며 원문의 결론이 아닙니다. 스탠퍼드 리뷰의 수치(1,680편·52편·10편)는 공개 보도와 보고서 요약을 근거로 했습니다.

동시 발행: 트랙 1 시장·산업 인사이트
이자의 시대 — 코어위브·TSMC·소프트뱅크와 청년 고용 (CONDERI-IR-2026-11-12)

동시 발행: 포커스 칼럼 6호
칩은 국경을 넘었고, 책임은 아직 국경 안에 있다

동시 발행: 글로벌 옴니버스 4호
AI 인프라의 회계 — 6대 매체 교차 분석 (CONDERI-GO-2026-11-12)