CONDERI

CONDERI ACADEMIC INSIGHT | 학술·연구 인사이트 | 대외 공개판

접근은 효과가 아니다:
AI 성과를 만드는 것은 배치와 상호작용의 밀도다

Access Is Not Effect:
AI Performance Comes From Placement and Interaction Density

본 리포트의 독자 및 활용 가이드

이 리포트는 2026년 8월부터 11월까지 공개된 정보시스템·경영·교육 분야 실증 연구 8편을 창업가·경영진·인사·교육 담당자가 바로 쓸 수 있는 언어로 정리한 학술 인사이트입니다. 실험 표본과 계수, 효과크기를 원문 그대로 인용했고, 자기보고와 행동 측정을 구분했습니다. 이번 호의 결론은 하나입니다 — **AI를 쓰게 하는 것과 AI와 일하게 하는 것은 다른 문제이며, 성과는 모델 성능이 아니라 정보 설계·조직 배치·참여 설계의 세 층에서 결정됩니다.**

문서 정보 / Document Information

문서등록번호 Document No.	CONDERI-AR-2026-11-15
발행일자 Published	2026. 11. 15. (일) Sunday
권호 Volume	VOL. 2026-11-15
분석범위 Coverage	2026. 08 ~ 2026. 11 공개 논문·실험 8편
배포구분 Distribution	Public | 일반공개
발행기관 Publisher	CONDERI AI산업전략연구실
조사·집필 Research	AI산업전략연구실

0.08 SD
AI 튜터 2년 실험의 학년당 효과
전면 참여 시 0.14 SD

−17%
AI를 '직원'으로 부를 때
관리자의 오류 탐지 감소

−35%
AI 결과정보만 제공할 때
개선 과업 위임 확률 하락

23%
관리자 1,261명 중
조직도에 AI 에이전트를 올린 기업

한 문장으로: AI의 효과는 모델이 아니라 설계에서 나온다 — 8편의 실증이 가리키는 세 층

정보 설계(기술) · 조직 배치(경영) · 참여 설계(교육) — 같은 방향의 증거를 한자리에 모았다

8편
정보시스템(ISR) 2편, 조직 실험 1편, 교육 RCT 4편, HCI 실험 1편

1,261명
AI 에이전트를 조직도에 올린 기업의 관리자 표본(가장 큰 조직 실험)

94,385건
AI 튜터 초안 메시지. 전문가가 94.7%를 수정 없이 승인, 오류 1,000건당 3건 미만

+0.33
고참여 미세 RCT의 효과크기(Hedges g). 표준화 검사에서는 0.02로 수렴

공통 결론 1 — 정보는 쌍으로 줘야 위임이 개선된다

ISR에 10월 28일 게재된 실험은 AI의 사전 확신도만, 또는 사후 결과만 제시하면 인간-AI 결합 성과가 나아지지 않거나 오히려 나빠진다는 것을 보였습니다. 결과 정보만 주면 "성과를 개선할 과업"을 위임할 확률이 35% 떨어졌습니다(β=−0.423, p<0.05). 두 정보를 함께 줬을 때만 위임 빈도와 위임 효율이 동시에 올라갔습니다(상호작용 β=+0.589/0.565, p<0.05). **설명 없이 결과만 보여주는 대시보드는 위임을 줄입니다.**

공통 결론 2 — 배치가 감독을 바꾼다

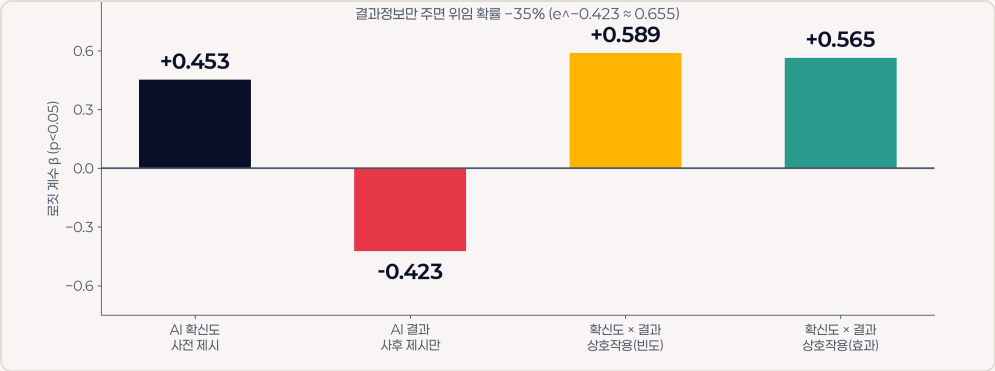
AI를 '도구'가 아니라 '직원'으로 조직도에 올린 기업의 관리자는 같은 문서에서 오류를 17% 덜 찾아냈고, 추가 검토 요청은 22%p 늘렸고, 책임은 자기에게서 9%p 줄이고 AI에 8%p 더 돌렸습니다. 성과 차이가 아니라 **책임 배분의 이동**입니다.

이번 호 결론 AI 성능이 아니라 배치·정보·참여의 설계가 성과를 정한다. 이번 호 8편의 실증은 세 층 모두에서 같은 방향을 가리킨다

구분	연구	표본·설계	핵심 계수
기술	ISR (2026-10-28) 위임 정보 설계	실험실 실험, 2×2 정보 조건	β=+0.589 상호작용, 결과만 −35%
기술	arXiv (2026-10-15) 위임 흔적 관찰	n=56, 영상 QA 392회	신뢰→위임 OR 14.97
경영	Wiles·Hsu·Bedard·Kropp (2026-09)	관리자 1,261명 무작위 실험	F1 −5.3%p, 재현율 −17%
경영	ISR (2026-04) 비용과 통제 선호	프롬프트 180만 건	유료화가 통제 선호를 높임
과학	ISR (2026-06) 아이디어 구체화	혁신 과제 실험	이미지 지원이 노력 감소
교육	NBER w35620 (2026-08) 칸미고	18개 중학교, 2년 군집 RCT	학기당 1.3 백분위, 0.06~0.08 SD
교육	Google·Eedi (2026-10-29) 제미나이 튜터	영국 10개교, n=1,179, 10주	표준화 검사 +0.02 SD
교육	Medly 미세 RCT (2026-09) GCSE	4주, n=644 완료	Hedges g = 0.33

I. 기술: AI에게 일을 맡기는 능력은 정보 설계의 함수다

ISR 2026-10-28 — 확신도만, 결과만 각각 제시하면 효과가 없거나 해롭고, 함께 줄 때만 위임이 개선된다



자료: Taudien, Fügener, Gupta & Ketter, Information Systems Research (2026-10-28). 로짓 계수, p<0.05

문제의식: 사람은 위임을 못한다

인간-AI 협업의 성과는 "누가 무엇을 하는가"의 배분에서 나옵니다. 그런데 선행연구는 인간이 위임을 너무 적게 하거나, AI가 더 잘하는 과업을 대신 맡기는 실수를 반복한다고 보고해 왔습니다. 이 연구는 그 원인을 성향이 아니라 **정보 부족**으로 재정의합니다. 실험은 동일한 과업에서 AI의 사전 확신도(정답 확률 추정)와 사후 결과(정답 여부)를 각각, 또는 함께 제시하며 위임 빈도와 위임 효율을 측정했습니다.

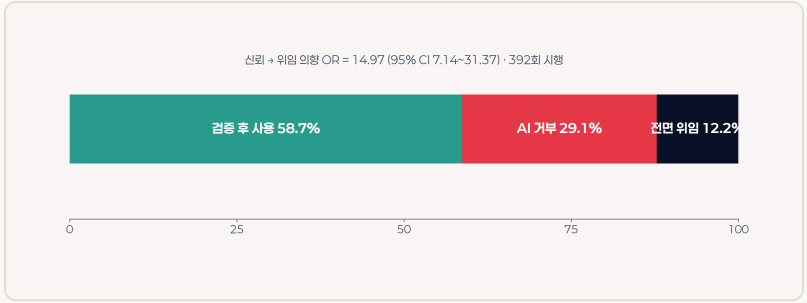
발견: 신호는 함께 줄 때만 작동한다

사전 확신도는 위임 확률을 유의하게 올렸습니다(β=+0.453). 반면 사후 결과만 제시하면 위임을 통해 성과를 개선할 수 있는 과업의 위임 확률이 35% 떨어졌습니다(β=−0.423). 두 신호를 함께 주면 상호작용이 유의하게 플러스였습니다(β=+0.589, 0.565). 저자들은 확신도가 기대를 보정하고, 결과 정보가 그 기대를 확인·반박하면서 AI에 대한 정신모형이 정확해진다고 해석합니다.

설계 결정	흔한 구현	실험 결과가 말하는 것	권고
신뢰도 표시	"정확도 92%" 단일 수치	과업 인스턴스 수준의 확신도가 필요	과업별 확신도 + 판단 근거 한 줄
결과 피드백	정오답만 표시	단독으로는 위임을 줄인다	기대값과 실제 결과를 함께 비교 표시
책임 표시	"AI가 작성" 문구	책임 소재가 흐려지면 검토 강도가 낮아진다	최종 승인자와 검토 범위 명시
위임 단위	문서 단위 통째 위임	과업 분해가 위임 효율을 높인다	문장·표·수치 검증을 분리 위임

II. 과학: 신뢰는 위임의 충분조건이 아니다 — 사람은 확인하면서 맡긴다

신뢰→위임 의향 OR 14.97, 그러나 정책 선택은 검증 58.7%·거부 29.1%·전면 위임 12.2%



자료: Chiatti, Matteucci & Schiaffonati, arXiv 2610.21948 (2026-10-15). n=56, 영상 질의응답 392회

의도된 위임과 실행된 위임은 다르다

참가자들은 응답을 본 뒤 신뢰도를 평가하고, 다음 유사 과업에서 세 가지 정책 중 하나를 고릅니다 — 거부, 검증 후 사용, 전면 위임. 신뢰는 위임 의향을 강하게 예측했습니다(OR=14.97, 95% CI 7.14~31.37). 그러나 신뢰와 위임은 같은 변수가 아니었습니다. 신뢰가 높은 구간에서도 검증 정책이 가장 많이 선택됐고, 과업 순서가 뒤로 갈수록 전면 위임이 서서히 늘었습니다(OR=1.22, 1.02~1.45). **사람은 신뢰해도 확인하려 하고, 확인하는 경험이 쌓이면 조금씩 맡깁니다.**

설계 함의

이 결과는 "신뢰도 조사"로는 실제 사용을 예측할 수 없다는 뜻입니다. 조직이 알아야 하는 것은 선언된 신뢰가 아니라 **실제로 어떤 단계에서 무엇을 확인했는가**입니다. 위임 흔적(delegation trace)을 남기는 로그 설계가 감독 체계의 최소 단위가 됩니다.

같은 달의 보조 증거: 비용이 붙으면 사용자는 더 통제한다

ISR에 4월 게재된 연구는 생성형 AI 프롬프트 180만 건을 분석해, 과금이 시작되면 사용자가 싸구려 응답을 받아들이는 대신 프롬프트를 정교하게 다듬어 통제권을 확보하는 방향으로 행동이 바뀐다는 것을 보였습니다. 무료 크레딧이 사라지는 순간 사용자의 사용 방식이 '검색'에서 '설계'로 이동한다는 뜻이며, 이는 기업용 AI 도입에서 사용량 과금이 오히려 품질을 끌어올릴 수 있는 경로를 시사합니다.

III. 경영: AI를 조직도에 올리면, 관리자는 덜 읽고 더 넘긴다

관리자 1,261명 실험 — AI를 '직원'으로 부르는 순간 오류 탐지 −17%, 추가 검토 요청 +22%p, 책임은 AI로 이동



자료: Wiles(Boston University)·Hsu·Bedard·Kropp(BCG), "Putting AI on the Org Chart" (2026-09-19). 관리자 1,261명

실험 설계: 같은 문서, 다른 이름표

관리자 1,261명에게 오류가 심어린 동일한 문서를 주고 검토하게 했습니다. 무작위로 문서를 "AI 도구가 만든 초안", "AI 직원이 만든 초안", "인간 직원이 만든 초안"으로 제시했습니다. 전체 평균 효과는 작았습니다. 그러나 **이미 조직도에 AI 에이전트를 올려 둔 기업의 관리자들**에서는 뚜렷한 패턴이 나타났습니다 — 저자들이 부르는 '뜨거운 감자(hot potato)' 효과입니다.

수치로 본 감독의 이동

AI 직원 프레이밍은 이 집단에서 검토 성과(F1)를 5.3%p 낮췄고, 재현율은 5.8%p (baseline 0.41 대비 약 17%) 떨어뜨렸습니다. 추가 검토 요청은 22%p 늘었고(0.50→0.72), 책임 배분은 관리자 자신 −9%p, AI 시스템 +8%p로 이동했습니다. 같은 문서를 인간 직원의 것이라고 했을 때는 이런 감소가 없었습니다.

조직 설계 AI를 조직도에 올리는 일은 기술 도입이 아니라 권한과 책임의 부여다. 이름표를 바꾸면 검토 강도가 바뀌고, 책임의 위치도 바뀐다

IV. 경영: 책임은 사라지지 않고 옮겨간다 — 옮겨간 곳을 설계해야 한다

같은 문서, 다른 이름표 실험의 전체 수치와 반대 해석. 도입과 감독 설계는 같은 결정이어야 한다

측정 항목	AI 도구로 제시	AI 직원으로 제시	변화	해석
검토 성과 F1	0.41	0.357	−5.3%p	정확도와 재현율의 조화평균 하락
재현율(오류 발견)	0.41	0.352	−5.8%p(−17%)	심어둔 오류를 놓치는 비율 증가
정밀도	—	—	소폭 하락	잘못 지적하는 비율도 함께 늘어남
추가 검토 요청	0.50	0.72	+22%p(+44%)	비용이 있는 상황에서도 요청 증가
책임 배분(관리자)	—	—	−9%p	자기 책임으로 인식하는 비중 감소
책임 배분(AI)	—	—	+8%p	시스템으로 책임 전가
인간 직원 대조군	—	—	효과 없음	단순한 위임 회피가 아님

왜 이 결과가 조직 설계 문제인가

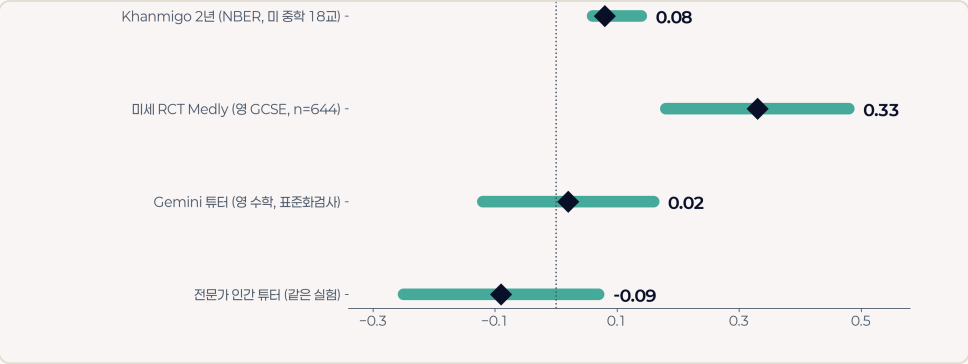
AI 에이전트를 조직도에 올리는 일은 기술 도입이 아니라 **권한과 책임의 부여**입니다. 직함을 주는 순간 동료는 그 산출물을 "검토가 끝난 업무"로 취급하고, 관리자는 "내가 최종 승인자"라는 인식에서 멀어집니다. 저자들은 AI를 조직 어디에 두는가 자체가 거버넌스 결정이라고 정리합니다. 실무적으로는 ①AI 산출물에 대한 최종 승인자를 직무기술서에 명시 ②검토 범위(수치·인용·정책)를 체크리스트화 ③검토 시간을 업무 시간으로 계상하는 세 가지가 필요합니다.

반대 해석도 기록해 둔다

이 실험은 문서 검토라는 단일 과업, 미국 관리자 표본, 자기보고 책임 배분이라는 한계를 가집니다. 전체 표본의 평균 효과가 작았다는 점은 "AI 직원" 개념이 조직 안에서 제도화된 곳에서만 효과가 나타난다는 것을 뜻합니다. 따라서 이 결과는 AI 에이전트를 도입하지 말라는 결론이 아니라, **도입과 동시에 감독 설계를 함께 넣으라**는 증거로 읽어야 합니다.

V. 교육: 2년짜리 대규모 실험의 효과가 작았던 이유는 모델이 아니라 참여였다

칸미고 2년 RCT 0.06~0.08 SD, GCSE 미세 RCT 0.33 — 같은 AI 튜터가 다른 결과를 낸 이유



자료: NBER w35620(2026-08), arXiv 2609.14789(2026-09), Google·Eedi LearnLM RCT(2026-10-29). 점 = 점추정, 선 = 신뢰구간

가장 큰 실험의 가장 작은 효과

테네시주 중학교 18곳에서 2년간 진행된 군집 무작위 배정 실험에서, 칸 아카데미의 AI 튜터 칸미고를 쓰도록 배정된 학생의 수학 성취는 학기당 전국 백분위 1.3, 학년 기준 0.06~0.08 표준편차 올랐습니다. 전면 참여 기준으로 환산하면 0.14 SD입니다. 이는 AI 없는 칸 아카데미 연습과 비슷한 수준입니다. 저자들이 찾은 병목은 성능이 아니라 **참여**였습니다 — 96%가 한 번은 시도했지만 중위 학생은 연습일의 3분의 1에만 메시지를 보냈고, 오답이 난 세션의 17%에서만 질문했습니다.

참여가 높은 실험은 효과가 크다

영국 GCSE 과학 4주 미세 RCT(완료 644명)에서는 Hedges g=0.33(0.18~0.48)이 나왔고, 물리 0.31·화학 0.32·생물 0.52로 과목별로도 일관됐습니다. 다만 탈락률 30.7%, 성취도 측정이 교육과정 정렬 문항이라는 한계가 있어 저자들도 예비적 결론으로 다룹니다.

교육 실증 AI 튜터의 성과는 모델의 정답률이 아니라 학생이 대화를 시작하는 비율에서 갈렸다. 접근권은 조건이고, 참여가 원인이다

VI. 교육: 감독이 붙은 AI 튜터는 정확했고, 그래도 시험 점수는 오르지 않았다

영국 10개교 1,179명 — 초안 9만 4천 건 중 94.7% 무수정 승인, 오류 0.3% 미만, 표준화 검사 차이는 0.02 SD

조건	설계	결과	해석
표본	영국 10개 공립 중등학교, 1,179명, 10주	4개 조건 무작위 배정	교실 규모·표준화 검사 동시 적용
조건 4개	표준 보충학습 / 전문가 인간 튜터 / AI 튜터 / 개인맥락 AI 튜터	—	동일 교육과정·동일 시간
감독 방식	AI 초안 94,385건을 전문가가 승인·재생성·수정	94.7% 수정 없이 승인	오류 1,000건당 3건 미만
근접 측정	플랫폼 내 이해도·중간 퀴즈	AI 튜터 ≈ 전문가 인간 튜터 > 표준 보충	가까운 전이에서는 효과 확인
표준화 검사	Renaissance Star Maths(외부 시행·채점)	AI +0.02 SD, 개인맥락 AI −0.01 SD	광범위 전이에서는 차이 미검출
인간 튜터	같은 기간 전문가 1:1	−0.09 SD(−0.25~0.07)	표준 보충보다 낮을 확률 87.1%
해석	심화 오답 해결에 시간을 써서 진도 범위가 좁아짐	여학생에서 효과가 더 우호적	시간·범위·평가 설계의 상호작용

이 실험이 반박한 두 가지 통념

첫째, “AI 튜터가 인간 튜터를 대체한다”는 주장은 근접 측정에서만 성립합니다. 표준화 검사에서는 네 조건의 차이가 사라졌고, 전문가 인간 튜터조차 표준 보충보다 낮은 점추정치를 보였습니다. 둘째, “AI는 틀린 답을 가르친다”는 우려는 이 표본에서 확인되지 않았습니다. 감독자가 94.7%를 수정 없이 승인했고 오류율은 1,000건당 3건 미만이었습니다. **문제는 정확도가 아니라 무엇을 얼마나 오래 다루게 하는가였습니다.**

교육 제품에 주는 설계 지침

①근접 지표와 표준화 지표를 함께 보고한다 ②오답 세션에서 대화를 강제하는 장치를 넣는다 ③심화와 진도 범위의 트레이드오프를 학부모·교사에게 숫자로 보여준다 ④사람 감독의 승인률·수정률을 로그로 남긴다.

VII. 교육: 답을 주는 튜터보다 구조를 주는 튜터가 전이를 만든다

간호학 362명 16주 3군 실험 — AI+지식그래프군 82.5점, 지식그래프 이해 η^2=0.41

측정	AI 강화군	혼합군	전통군	효과크기
모듈 시험	82.5 ± 7.3	77.6 ± 8.1	74.1 ± 8.5	F=28.64, η^2=0.16
최종 종합점수	84.2 ± 6.8	78.9 ± 7.5	75.2 ± 7.9	F=32.16, η^2=0.18
지식그래프 이해	4.54 ± 0.50	3.81 ± 0.70	2.96 ± 0.88	F=103.72, η^2=0.41
지식 잔존율	82.3%	—	—	F=48.52, η^2=0.25
사례 추론 정확도	91.2%	—	—	F=67.14, η^2=0.31

왜 이 실험이 중요한가

중국 창저우 의학전문대 간호학과 1학년 362명(완료 301명)을 16주간 세 집단으로 나눈 무작위 실험입니다. 생성형 AI 튜터에 **지식그래프**를 결합한 조건이 가장 높은 성과를 냈고, 효과크기가 가장 큰 항목은 점수가 아니라 지식그래프 이해(η^2=0.41)와 사례 추론(η^2=0.31)이었습니다. 즉 AI가 답을 대신하는 방식이 아니라, 개념 사이의 관계를 구조로 보여주는 방식이 학습 전이를 만들었습니다. 앞 장의 표준화 검사 무효과와 모순되지 않습니다 — 구조를 준 집단에서만 전이가 나타났습니다.

국내 적용 지점

한국에서 AI 디지털교과서가 교육자료로 지위가 바뀐 뒤 학교별 선택 시장이 열렸습니다. 이 실험은 그 시장에서 무엇이 팔리는지를 시사합니다. 문항 생성이 아니라 개념 구조·오답 경로·교사 확인 화면입니다. 검증과 기록이 구매 기준이 된다는 점은 시장 동향에서 관찰한 흐름과 같습니다.

전이의 조건 근접 지표의 효과와 표준화 검사의 무효과는 모순이 아니다. 전이는 구조를 준 집단에서만 나타났고, 그 구조는 개념 사이의 관계였다

VIII. 통합 프레임워크: 정보·배치·참여·전이·기록의 다섯 층

8편의 실증을 하나의 설계표로 압축했다. 각 층마다 측정 지표와 실패 신호가 있다

층	설계 변수	실증 근거	측정 지표	실패 신호
1층 정보	AI 확신도 + 사후 결과 동시 제공, 과업 단위 확신도	ISR 위임 실험(β=+0.589)	위임 빈도·위임 후 성과 개선율	결과만 표시하는 대시보드, 정확도 단일 수치
2층 배치	AI 산출물의 최종 승인자 지정, 검토 범위 체크리스트	조직도 실험(F1 −5.3%p)	검토 시간, 오류 발견율, 책임 배분	"AI가 작성" 표기만 있는 워크플로
3층 참여	오답·실패 지점에서 상호작용 강제, 심화와 범위의 균형	칸미고 2년 RCT, GCSE 미세 RCT	세션당 대화 턴 수, 오답 세션 대화율	접근권만 부여, 사용 로그 없음
4층 전이	개념 구조·지식그래프·사례 추론 경로 제공	간호학 3군 실험(η^2=0.41)	표준화 검사, 사례 추론 정확도	근접 지표만 보고
5층 기록	위임 흔적 로그, 승인·수정 이력	arXiv 위임 흔적 연구(OR 14.97)	정책 선택 분포, 검증 비율	신뢰도 설문만 존재

통합 프레임워크 AI 도입의 성과는 모델 선택이 아니라 다섯 층의 설계 문서에서 결정된다. 이번 호 8편은 각 층에서 무엇을 측정해야 하는지를 알려준다

IX. 실무 적용: 이번 주에 점검할 일곱 개 항목

각 항목은 질문 한 줄과 최소 조치 한 줄로 구성했다. 모델을 바꾸지 않고도 실행 가능하다

점검 항목	현재 상태를 확인하는 질문	최소 조치
위임 단위	AI 산출물을 문서 전체로 승인하는가, 문장·수치 단위로 검증하는가	수치·인용 검증을 별도 단계로 분리
정보 표시	사용자가 AI의 확신도와 사후 결과를 함께 보는가	과업별 확신도와 정오답 피드백 동시 노출
책임 배분	AI 산출물의 최종 승인자가 이름으로 지정돼 있는가	직무기술서·워크플로에 승인자 명시
감독 시간	검토 시간을 업무 시간으로 계상하는가	검토 소요를 공수로 산정해 예산에 반영
참여 설계	교육·고객 지원에서 실패 지점에 대화를 강제하는가	오답·이탈 지점에서 인터랙션 트리거
전이 측정	근접 지표와 외부 표준 지표를 함께 보는가	도입 3개월·6개월 표준 지표 비교
기록	신뢰도 설문 외에 실제 위임 흔적이 남는가	승인·수정 이력 로그와 월간 리뷰

창업가에게

고객이 AI 기능을 사는 이유는 성능표가 아니라 **검증 가능성**입니다. 도입 제안서에 승인자·검토 범위·기록 항목을 함께 넣으면, 같은 모델로도 구매 결정이 빨라집니다.

교육 사업자에게

접근권 판매에서 참여 설계 판매로 이동해야 합니다. 오답 세션 대화율, 심화 학습 시간, 표준화 검사 전이를 계약 지표로 제시하는 것이 다음 구매 기준입니다.

English Summary / 국문 요약의 영문판

Access is not effect: evidence from eight studies on delegation, placement and engagement

Executive Summary (English)

Eight empirical studies published between August and November 2026 point in one direction: AI outcomes are determined less by model capability than by how information, organisational placement and interaction are designed. In a field experiment published in Information Systems Research on 28 October, showing users only the AI's ex-post outcome reduced the odds of delegating a task that would have improved performance by 35% (beta = -0.423, $p < 0.05$). Showing ex-ante certainty alone had no significant effect on delegation effectiveness. Only when both signals were presented together did delegation frequency and effectiveness rise (interaction beta = +0.589 and +0.565, $p < 0.05$).

Placement changes oversight

A randomized experiment with 1,261 managers (Wiles, Hsu, Bedard and Kropp, September 2026) found that 23% of organisations already list AI agents on organisational charts. Among managers at those firms, framing an identical document as the work of an "AI employee" rather than an "AI tool" reduced the F1 score of their review by 5.3 percentage points and recall by 5.8 points (about 17%), increased requests for additional review by 22 percentage points, and shifted perceived accountability away from the manager by 9 points and toward the AI system by 8 points. Human-employee framing produced no such effect.

Education: access is not engagement

A two-year cluster randomized trial in 18 Tennessee middle schools (NBER w35620) found that assignment to Khanmigo raised maths achievement by 1.3 national percentile ranks per term, or 0.06 to 0.08 standard deviations a year (0.14 with full participation) - similar to Khan Academy practice without AI. The binding constraint was engagement: 96% of students tried the tutor, but the median student messaged it on only a third of practice days. An expanded UK RCT with Gemini (1,179 students, 10 schools) found AI tutoring on par with expert human tutors on proximate measures, with 94.7% of 94,385 drafts approved unchanged and fewer than three errors per 1,000 messages, yet no detectable advantage on standardized Star Maths tests (+0.02 SD).

0.06~0.08 SD	**−17%**	**+0.33**	**58.7%**
Annual effect of a two-year AI tutoring trial	Error detection by managers when AI is framed as an employee	Effect size in a high-engagement four-week micro-RCT	Share of trials in which users chose verify-then-use over full delegation

One line **Stop asking what the model can do. Ask where it sits, what information it shows, and what makes people engage with it**

맺음: 성과는 모델이 아니라 설계 문서에 적힌다

정보·배치·참여 — 세 층의 설계가 바뀌지 않으면 더 좋은 모델도 같은 결과를 낸다

이번 호가 확인한 것

8편의 실증은 세 층에서 같은 방향을 가리킵니다. **정보를 쌍으로 주지 않으면 위임은 개선되지 않고, AI를 조직의 행위자로 부르면 감독은 줄고, 접근권만 주면 참여가 병목이 되고, 구조를 주면 전이가 일어납니다.** 특히 교육 분야의 두 대규모 실험은 "AI 튜터가 사람을 대체한다"는 서사와 "AI는 위험하다"는 우려를 동시에 반박했습니다. 정확도는 감독 아래 충분했고, 효과는 참여 설계에 달려 있었습니다.

이번 호가 확인하지 못한 것

조직도 실험은 문서 검토 단일 과업과 미국 관리자 표본이라는 한계를 가집니다. GCSE 미세 RCT는 탈락률 30.7%로 결과 해석에 주의가 필요하고, 저자들도 예비적 결론으로 다룹니다. 간호학 실험은 단일 기관·단일 과목입니다. 국내 조직·학교를 대상으로 한 동일 설계의 재현 연구는 아직 확인하지 못했습니다.

연구	표본	주요 수치	실무 번역
ISR 위임 정보 설계	실험실 2×2	상호작용 β=+0.589	확신도와 결과를 함께 표시
arXiv 위임 흔적	n=56·392회	검증 58.7%	로그로 감독을 측정
조직도 실험	관리자 1,261명	F1 −5.3%p	AI 산출물 승인자 지정
칸미고 2년 RCT	18개교·2년	0.06~0.08 SD	참여율을 KPI로
Gemini 튜터 RCT	10개교·1,179명	표준화 +0.02 SD	표준 지표 병행 측정
Medly 미세 RCT	n=644	g=0.33	오답 세션 대화 강제
간호학 3군 실험	n=301	지식그래프 η^2=0.41	개념 구조를 제품화

발행 CONDERI AI산업전략연구실 | 문의 contact@conderi.com | 인용된 계수는 원문 공개값이며 재현 연구에 따라 달라질 수 있습니다.