CONDERI

CONDERI ACADEMIC INSIGHT | 학술·연구 인사이트 | 대외 공개판

위임의 구조: AI가 성과를 만드는 경로는 하나가 아니라 세 개였다

The Architecture of Delegation: Three Routes from AI Use to Performance, and the One Most Firms Miss

본 호의 문제의식

생성형 AI의 생산성 효과는 이미 여러 번 증명됐습니다. 그런데 조직 성과는 따라오지 않습니다. 이번 호는 그 간극을 '도입 여부'가 아니라 **경로**의 문제로 읽습니다. 이번 주 공개된 여섯 편의 실증 연구는 각각 다른 표본과 방법으로 같은 결론에 닿습니다 — AI는 **의사결정의 질**, **인지 노동의 분업**, **검증 절차**를 거칠 때만 성과가 됩니다. 도구를 준 조직과 구조를 바꾼 조직의 차이가 여기서 갈립니다.

문서 정보 / Document Information

문서등록번호 Document No.	CONDERI-AR-2026-11-13
발행일자 Published	2026. 11. 13. (금) Friday
권호 Volume	VOL. 2026-11-13
분석기간 Coverage	2026년 10월 ~ 11월 공개·게재 논문
배포구분 Distribution	Public | 일반공개
발행기관 Publisher	CONDERI AI산업전략연구실
조사·집필 Research	AI산업전략연구실

0.833
생성형 AI '주입'이 조직 효과성에 미치는 총효과 (n=615)

Q^2=0.524
알고리즘 관리·인간-AI 협업 모델의 업무성과 예측력 (n=128)

g=1.01
AI 맞춤학습의 학업성과 효과
단, 소표본 보정 시 구간이 0을 포함

+0.23 SD
AI 튜터 무작위실험 효과
무제한 접근이 제한 접근보다 +0.21 SD

한 문장으로: AI는 성과를 직접 만들지 않는다. 의사결정과 검증을 통과할 때만 만든다

이번 주 여섯 편의 실증 연구가 공통으로 가리킨 것 — 도입이 아니라 경로의 문제

완전매개
생성형 AI 협업은 '의사결정의 질'을 통해서만 조직 성과에 닿는다 (n=400)

두 갈래
텍스트형은 정보처리, 멀티모달형은 아이디어 창출에 더 강하다

이중 경로
알고리즘 관리와 인간-AI 협업은 서로 다른 매개로 성과에 이른다

효과 0.23
AI 튜터 RCT의 학업성과 향상(표준편차 단위). 단 조건에 따라 갈린다

1. 경영: '쓰는 것'과 '쓰게 하는 것'은 다른 변수다

국내 직장인 400명을 대상으로 한 연구는 생성형 AI 협업이 업무 효율성을 직접 높이지 않는다는 것을 보였습니다. 디지털 리터러시와 성과기대는 AI 협업을 유의하게 높였지만 노력기대는 그렇지 않았고, AI 협업은 **의사결정의 질을 완전매개**로 해서만 조직 성과에 닿았습니다. 같은 주에 발표된 싱가포르 연구(615명)도 방향이 같습니다. 조직 효과성에 가장 큰 총효과를 준 것은 개별 도구 사용이 아니라 **업무 프로세스에 AI가 '주입(infusion)'된 정도**(0.833)였습니다.

2. 교육: 효과는 크지만, 확실하지 않다

AI 맞춤학습의 통합 효과는 g=1.01로 컸습니다. 그런데 연구가 4편, 표본이 420명뿐이고 이질성이 매우 높아(I^2=89%) 소표본 보정을 적용하면 신뢰구간이 −0.10~2.11로 0을 포함합니다. 근거의 확실성은 '매우 낮음'입니다. 반면 334명 무작위실험은 AI 튜터가 +0.23 SD의 성과 향상을 냈고, **제한 없이 쓰게 한 집단이 제한한 집단보다 0.21 SD 더 높았습니다**. 효과의 크기보다 조건의 설계가 결과를 가릅니다.

기술
알고리즘 관리와 인간-AI 협업은 각각 다른 경로로 '근접 AI 활용'(코드·정보탐색)을 높였습니다. 하나의 지표로 관리하면 두 효과가 상쇄됩니다.

과학
891명 연구에서 AI 친숙도는 비판적 사고를 직접 높이지 않았습니다. 인지 노동의 분업과 결과물 재작업을 통해서만 유의했습니다(완전매개).

현장
같은 주 국내 에듀테크 데이터는 다른 얼굴을 보여줬습니다. 635억 원 플랫폼의 하루 방문율은 1.63%. 효과 이전에 사용이 없습니다.

문장 1
AI 활용은 성과를 직접 만들지 않는다. 의사결정의 질과 인지 노동의 분업을 거칠 때만 성과가 된다.

문장 2
조직 효과성을 결정하는 것은 도구 구매가 아니라 업무 프로세스에 AI가 내장된 정도다.

문장 3
교육에서 효과크기보다 중요한 것은 접근 규칙이다. 통제는 사용 리듬을 깨고, 개방은 점진적 통합을 만든다.

이번 호에 담은 여섯 편: 표본·방법·핵심 수치

국내 학술지 3편과 국제 저널·학회 3편을 분야별로 배치했다

분야	논문	표본·방법	핵심 발견
기술	알고리즘 관리와 인간-AI 협업이 AI 활용과 성과에 미치는 영향 Information Systems Review 28(1), 2026	n=128, PLS-SEM, PLSpredict	두 요인 모두 근접 AI 활용을 높이지만 매개 경로가 다름. 모델의 업무성과 예측력 Q^2=0.524
경영	생성형 AI 활용이 비즈니스 분석역량과 조직성과에 미치는 영향 한국빅데이터학회지 11(1), 2026	생성형 AI 사용 경험 근로자, 구조방정식	텍스트형은 정보처리·과부하 완화, 멀티모달형은 아이디어 창출에 강함. 둘 다 분석가용시간을 늘려 조직성과로 이어짐
경영	생성형 AI의 주입이 조직 민첩성과 효과성을 높이는가 Information, Communication & Society, 2026	n=615, TOE 프레임워크, 11개 가설	11개 가설 전부 지지. 조직 효과성 총효과 1위는 AI 주입(0.833), 2위는 시장 민첩성(0.468)
과학	생성형 AI와의 거래기억: 비판적 사고는 언제 작동하는가 arXiv 2610.38910, 2026	n=891 (미국 645·튀르키예 246)	AI 친숙도·협업 자신감은 비판적 사고를 직접 예측하지 못함. 분업·조정과 재작업을 통한 완전매개
교육	AI 기반 맞춤학습과 학업성과: 통제연구 체계적 문헌고찰·메타분석 Frontiers in Education, 2026	4편, 420명, PRISMA 2020	통합효과 g=1.01(0.33~1.68). 소표본 보정 시 −0.10~2.11. 근거 확실성 '매우 낮음'
교육	AI 튜터는 읽기 노력을 구축하는가: 무작위실험 IZA Discussion Paper 18338	n=334, 3개 조건 무작위배정	AI 튜터 접근은 +0.23 SD. 무제한 접근이 제한 접근보다 +0.21 SD 높음. 기초지식 낮고 자기조절 높은 학생에 효과 집중

선정 기준

① 2026년 10~11월 공개·게재 ② 표본과 방법이 공개돼 재현 가능 ③ 도구 도입이 아니라 **조직·학습 구조**를 다룸 ④ 수치(계수·효과크기·신뢰구간)가 본문에 명시됨. 초록만 공개된 연구는 효과크기만 인용하고 신뢰구간을 추정하지 않았습니다.

읽는 순서

기술 → 경영 → 과학 → 교육 순으로 배치했습니다. 각 장은 ① 문제의식 ② 실증 설계와 수치 ③ 경영진을 위한 시사점 ④ 적용 프레임워크의 네 단으로 구성했습니다. 마지막 두 장은 여섯 편을 하나의 실행 체계로 묶습니다.

논문	이 논문이 반박하는 통념	내일 바로 할 수 있는 일
알고리즘 관리·협업	"AI 사용률을 올리면 성과가 온다"	사용률 지표를 협업형·관리형 두 축으로 분리해 보고
활용 유형	"좋은 모델 하나면 된다"	정리형 과업과 생성형 과업에 서로 다른 도구를 배치
AI 주입	"경쟁 압력이 있으면 도입된다"	핵심 프로세스 하나를 골라 AI 산출물 입력 비율을 측정
거래기억	"많이 쓸수록 비판적 사고가 는다"	산출물 제출 시 출처·재작업 내역 1줄 첨부를 규칙화
맞춤학습 메타	"효과크기가 크니 확대 도입하자"	도입 전에 통제집단을 두는 소규모 검증 설계
AI 튜터 RCT	"먼저 혼자 풀게 해야 학습된다"	접근 제한을 단계적 개방으로 바꾸고 사용 로그를 확인

In One Sentence: AI Does Not Create Performance — It Passes Through Decisions and Verification

Five empirical studies published this week converge on the same architecture

Full mediation
GenAI collaboration reaches performance only through decision quality (n=400)

Two paths
Text-based use improves information processing; multimodal use improves idea generation

Dual route
Algorithmic management and human-AI collaboration operate through different mechanisms

0.23 SD
Effect of AI tutoring in a randomized experiment, conditional on access design

Management: usage is not the same variable as infusion

A Korean survey of 400 employees finds that generative AI collaboration does not raise team efficiency directly. Digital literacy and performance expectancy predicted collaboration, effort expectancy did not, and collaboration reached organisational outcomes **only through decision quality, as a full mediator**. A Singapore study of 615 respondents published in the same period points the same way: the largest total effect on organisational effectiveness came not from individual tool use but from the **infusion of GenAI into work processes (0.833)**, ahead of market agility (0.468), workforce agility (0.259) and operational agility (0.181).

Education: large effects, weak certainty

A meta-analysis of AI-based personalised learning reports a pooled Hedges g of 1.01 (95% CI 0.33–1.68) across four controlled studies and 420 students. With the Hartung-Knapp small-sample adjustment the interval widens to −0.10 to 2.11, heterogeneity is substantial (I^2=89%) and the certainty of evidence is very low. A separate randomised experiment with 334 students is more informative about design: AI tutor access raised performance by 0.23 SD, and **unrestricted access outperformed restricted access by 0.21 SD**, contradicting the assumption that early reliance undermines learning.

Technology

Algorithmic management and human-AI collaboration both increase proximal AI use — coding and general information tasks — but through different mediation paths. Managing both with a single metric hides one of them.

Science

Among 891 students in the US and Türkiye, familiarity with GenAI did not predict critical thinking directly. It mattered only through dividing cognitive labour with the tool and actively reworking its output.

Field reality

In the same week, Korean school data showed a 63.5 billion won AI platform with 1.63% daily visitation. The binding constraint is not the effect size but whether anyone opens the tool at all.

How to Read This Issue: Four Sections, One Framework

Each section follows problem, evidence, implication, and an operating rule

I. Technology — two mechanisms, not one

A PLS-SEM study of 128 employees separates algorithmic management from human-AI collaboration as second-order constructs across context, agency, interaction and outcome. Both raise proximal use; neither substitutes for the other. Predictive relevance for work performance is substantial (Q^2=0.524).

II. Management — the type of use decides the outcome

Text-based and multimodal use travel different routes: information processing and cognitive-load relief on one side, idea generation on the other. Both increase analyst available time, which in turn builds business analytics capability and organisational performance.

III. Science — division of labour is the mechanism

Transactive memory theory applied to human-AI partnerships explains why familiarity alone does nothing. Specialisation and coordination, plus reflective transformation of output, carry the effect. Interview studies add the mechanism's failure mode: uncritical acceptance that removes the human compensatory behaviour.

IV. Education — access design beats tool quality

Meta-analytic effects are large but uncertain; the randomised evidence shows the design of access matters more than the model. Unrestricted, gradual integration outperformed gatekeeping that produced intensive prompting bursts disruptive to learning flow.

Section	Question	Evidence	Operating rule for the reader
Technology	Which management lever moves AI use?	n=128, two second-order constructs, Q^2=0.524	Measure collaboration and algorithmic management separately
Management	Why does usage not become performance?	n=400 full mediation; n=615 infusion effect 0.833	Attach every AI deployment to a named decision it improves
Science	What makes the partnership cognitive?	n=891 full mediation; 14 and 45 interviews	Require rework of outputs, not just review of them
Education	How should access be designed?	g=1.01 (uncertain); +0.23 SD randomised	Remove gatekeeping, add structure to the workflow

I. 기술: 알고리즘 관리와 인간-AI 협업은 같은 것이 아니다

Information Systems Review 28(1) · n=128 · PLS-SEM · 업무성과 예측력 Q^2=0.524

문제의식

AI가 도구에서 '자율적 에이전트'로 이동하면 조직 안에서 AI의 역할이 둘로 갈라집니다. 하나는 **알고리즘 관리(AIgM)** — AI가 업무 배분·평가·모니터링의 주체가 되는 것이고, 다른 하나는 **인간-AI 협업(HAC)** — 사람이 AI와 함께 과업을 수행하는 것입니다. 기존 연구는 'AI 채택'이라는 하나의 변수로 이 둘을 뭉갰습니다. 이 연구는 둘을 별개의 2차 요인으로 분해하고, 사용 패턴과 성과에 이르는 사회-기술적 경로를 묻습니다.

설계

Benbya 등(2021)의 프레임워크를 따라 HAC와 AIgM을 맥락·주체성·상호작용·결과의 4개 차원으로 개념화하고, Mohamed 등(2024)의 6가지 AI 사용 목적을 순차 경로 모델에 통합했습니다. 128명을 대상으로 PLS-SEM을 적용했고, PLSpredict로 예측력을 검증했습니다. 종속변수는 코드 작성·일반 정보탐색 같은 근접 활용, 튜링테스트·단순과업 같은 관계적 역학, 디버깅 같은 운영적 실행, 최종적으로 업무성과입니다.

발견

첫째, 둘 다 효과가 있다. HAC와 AIgM은 모두 근접 AI 활용을 유의하게 높였습니다. 즉 AI를 '관리자'로 쓰든 '동료'로 쓰든 사용량은 늘어납니다.

둘째, 경로가 다르다. 두 요인은 서로 다른 매개를 통해 작동했습니다. 같은 사용량 증가가 서로 다른 역량으로 축적된다는 뜻입니다. 조직이 'AI 사용률' 하나로 두 효과를 관리하면 한쪽의 기여가 다른 쪽에 가려집니다.

셋째, 예측력이 확인됐다. 모델은 업무성과에 대해 Q^2=0.524의 예측력을 보였습니다. PLS-SEM에서 0.5를 넘는 Q^2는 실무 예측에 쓸 수 있는 수준입니다.

경영진을 위한 시사점

AI 활용 지표를 하나만 두면 안 됩니다. 최소 두 축이 필요합니다 — ① 직원이 AI와 함께 일한 시간·과업 수(HAC) ② AI가 업무 배분·평가에 개입한 비중(AIgM). 두 축은 서로 다른 교육과 거버넌스를 요구합니다. HAC는 프롬프트·검증 훈련이, AIgM은 평가 기준의 투명성과 이의제기 절차가 핵심입니다.

적용: 두 축 진단 5문항

① 지난달 AI가 배정한 업무 비중을 숫자로 말할 수 있는가
② AI 평가 결과에 직원이 이의를 제기한 사례와 처리 절차가 있는가
③ AI와 협업한 과업의 '완료 정의'가 문서화돼 있는가
④ 디버깅·검증에 쓰인 시간을 따로 측정하는가
⑤ 두 축의 개선을 서로 다른 KPI로 관리하는가

차원	인간-AI 협업 (HAC)	알고리즘 관리 (AIgM)
맥락	사람이 과업을 소유하고 AI가 수행을 보조	AI가 과업 배분·모니터링의 주체로 배치
주체성	판단의 주도권이 사람에게 남음	평가·우선순위 결정에 AI의 규칙이 개입
상호작용	반복적 대화, 초안·수정의 순환	지표·알림·스코어카드 중심의 일방향 피드백
결과	근접 활용 증가, 업무 부담의 재분배	근접 활용 증가, 평가의 투명성 문제 부상
필요한 거버넌스	프롬프트·검증 훈련, 완료 정의 문서화	평가 기준 공개, 이의제기·재심 절차

II. 경영(1): 텍스트형 AI와 멀티모달 AI는 다른 능력을 키운다

한국빅데이터학회지 11(1) · 생성형 AI 활용 유형 → 비즈니스 분석역량 → 조직성과

문제의식

데이터가 넘치는 조직의 병목은 정보의 부족이 아니라 **과부하**입니다. 생성형 AI는 요약·재구성·자동화로 인지적 한계를 보완할 수 있지만, "AI를 쓴다"는 하나의 변수로 성과를 설명하려는 시도는 대부분 실패했습니다. 이 연구는 활용을 **텍스트 기반**과 **멀티모달 기반**으로 나누고, 그 효과가 어떤 역량을 거쳐 조직성과로 가는지 경로를 분해했습니다.

매개변수는 네 개입니다 — 아이디어 창출역량, 정보처리역량, 인지된 데이터 과부하 감소, 분석가용시간. 생성형 AI 사용 경험이 있는 근로자를 대상으로 설문하고 구조방정식으로 검증했습니다.

결과의 구조

텍스트 기반 활용은 정보처리역량 향상과 데이터 과부하 완화에 유의했습니다. 멀티모달 활용은 아이디어 창출역량에 더 큰 영향을 미쳤습니다. 두 유형 모두 분석가용시간을 늘렸고, 이 요인들이 비즈니스 분석역량을 거쳐 조직성과에 긍정적으로 작용했습니다.

해석: '무엇을 시켰는가'가 역량을 결정한다

이 결과는 AI 도입 담론을 두 가지로 갈라놓습니다.

첫째, 도구의 종류가 아니라 과업의 종류가 문제입니다. 텍스트형은 '정리'에, 멀티모달형은 '생성'에 강합니다. 같은 예산으로 두 도구를 사도, 조직이 어떤 과업에 배치하는지에 따라 축적되는 역량이 달라집니다.

둘째, 진짜 산출물은 '시간'입니다. 두 경로가 공통으로 거친 것은 분석가용시간의 증가였습니다. AI가 만든 시간을 다시 회의로 채우면 성과는 생기지 않습니다. 시간을 어디에 재투자했는지가 성과의 실제 변수입니다.

활용 유형	주로 강화하는 역량	적합한 과업	측정 지표 제안
텍스트 기반	정보처리역량, 인지된 과부하 감소	계약서 검토, 회의록 정리, 규정 비교, 고객 문의 분류	문서 1건당 처리 시간, 재작업률, 오류 발견 건수
멀티모달 기반	아이디어 창출역량	시각 자료 제작, 제품 콘셉트, 캠페인 시안, 교육 콘텐츠	시안 채택률, 첫 초안까지 걸린 시간, 채택 후 수정 횟수
공통 경로	분석가용시간 → 비즈니스 분석역량	의사결정 준비, 데이터 해석, 대안 비교	절감 시간의 재투자 내역, 의사결정 리드타임

한국 기업에 주는 의미

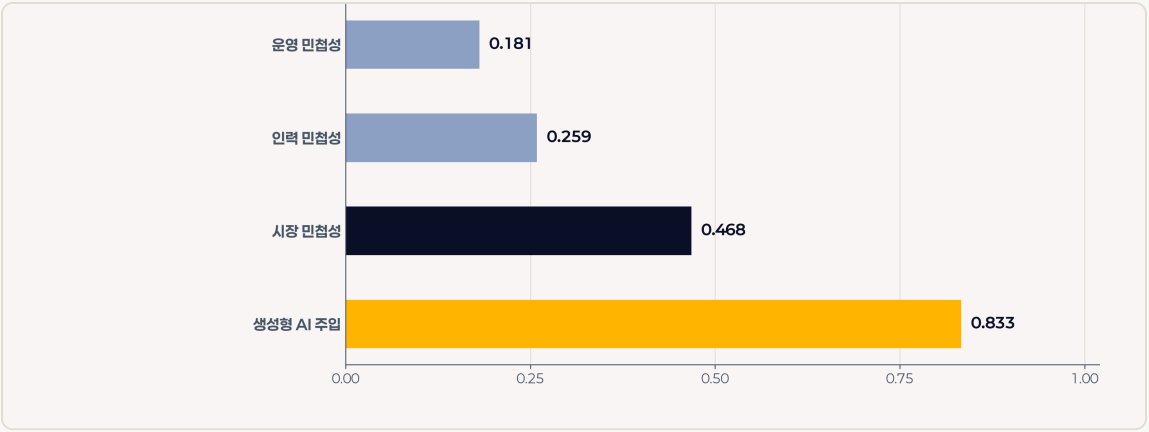
문서가 많은 조직(제조·금융·공공)은 텍스트형에서, 콘텐츠 조직(미디어·교육·브랜드)은 멀티모달에서 먼저 성과가 나타날 가능성이 큽니다. 같은 AI 예산이라도 배치가 다르면 축적되는 역량이 달라집니다.

주의할 해석

이 연구는 설문 기반 횡단 연구입니다. AI 활용이 역량을 키웠는지, 역량이 강한 조직이 AI를 더 잘 활용했는지는 구분되지 않습니다. 경로의 방향을 확정하려면 도입 전후를 비교하는 설계가 필요합니다.

II. 경영(2): 조직 효과성을 움직인 것은 도구가 아니라 '주입'이었다

Information, Communication & Society 2026 · n=615 · TOE 프레임워크 · 11개 가설 전부 지지



조직 효과성에 대한 총효과 계수. 자료: Koh & Yuen (2026)

설계

싱가포르 패널 11만5,000명 규모에서 성층화 무작위 표집으로 615명의 유효 응답을 확보했습니다. 생성형 AI를 전혀 쓰지 않는 조직의 응답자는 사전 질문으로 제외했습니다. 측정은 7점 리커트, 분석은 구조방정식입니다. 기술-조직-환경(TOE) 프레임워크로 선행요인을 정리했고, **11개 가설이 모두 지지**됐습니다.

총효과의 순위

1위 **생성형 AI 주입 0.833** — 도구를 '쓴다'가 아니라 업무 프로세스에 내장됐다는 정도입니다. 2위 시장 민첩성 0.468, 3위 인력 민첩성 0.259, 4위 운영 민첩성 0.181. 반대로 기술에 대한 신뢰(0.144)와 경쟁 압력(0.120)의 효과는 약했습니다. **외부 압력이나 기술 신뢰만으로는 조직이 움직이지 않는다**는 뜻입니다.

주입(infusion)을 어떻게 측정할 것인가

주입은 '사용 빈도'가 아니라 업무 흐름에의 내장 정도입니다. 실무 지표로 옮기면 세 가지가 됩니다 — ① 핵심 프로세스 중 AI 산출물이 입력되는 비율 ② AI 없이 같은 업무를 수행할 수 없는 정도(대체 불가성) ③ 예외 상황에서 AI를 우회하는 절차의 문서화 여부. 모두 '쓰는가'가 아니라 '얼마나 박혀 있는가'를 묻습니다.

민첩성의 서열이 말하는 것

시장 민첩성이 인력·운영 민첩성보다 두 배 가까이 큰 총효과를 냈습니다. 고객과 시장에 닿는 프로세스에 AI를 넣는 것이 내부 운영 효율화보다 성과에 직결된다는 의미입니다. 내부 자동화만으로 시작한 조직은 효과가 늦게 나타날 수 있습니다.

한계

횡단 설문이므로 인과 방향을 확정할 수 없습니다. 성과가 좋은 조직이 AI를 더 깊이 주입했을 가능성도 배제할 수 없습니다. 또 유료 도구와 무료 도구 사용의 차이를 통제하지 않았습니다.

주입을 높이는 첫째 방법: 프로세스 재설계
기존 절차에 AI를 '추가'하는 것이 아니라, AI가 중간 단계를 대신하도록 절차를 다시 그립니다. 결재 단계를 줄이고 판단 지점만 남기는 방식입니다.

둘째: 고객 접점에 먼저
시장 민첩성의 총효과(0.468)가 내부 운영(0.181)보다 컸습니다. 내부 자동화만 반복하면 성과가 늦게 나타납니다.

셋째: 학습을 구조에 심기
조직 학습이 총효과 상위에 올랐습니다. AI 사용 기록을 개인 평가가 아니라 프로세스 개선의 입력으로 쓰는 설계가 필요합니다.

III. 과학: 비판적 사고는 AI 친숙도가 아니라 '분업 구조'에서 나온다

TMS 이론의 인간-AI 확장 · n=891(미국·튀르키예) · 인터뷰 연구 2편의 보완

이론: 거래기억체계(TMS)를 AI에 옮기면

TMS는 사람들 사이에서 '누가 무엇을 아는지'를 공유하는 체계입니다. 이를 인간-AI 파트너십에 적용하면 세 축이 나옵니다 — 전문성의 인식, 신뢰성 판단, 조정. 연구는 891명(미국 645명·튀르키예 246명)의 대학생을 대상으로 구조모형을 검증했습니다.

결과는 명확합니다. AI에 대한 자기평가(사고 능력)와 파트너십 신뢰는 실무 경험량을 늘렸지만, **비판적 사고를 직접 예측하지는 못했습니다**. 경험은 두 개의 하위 행동 — AI와 인지 노동을 나누고 조정하는 것, 산출물을 적극적으로 재작업하는 것 — 을 통해서만 유의했습니다. 두 국가 모두에서 구조가 대체로 동일했습니다(완전매개).

질적 연구가 보여준 실패 모드

14명의 지식노동자 인터뷰는 TMS가 인간-AI에서 **비대칭**으로 작동함을 보였습니다. 사람은 AI의 역량을 학습하지만, AI는 사람의 전문성을 일시적으로만 인식합니다. 신뢰는 양가적이어서, 검증되지 않은 정보에는 추가 검증 절차를 붙입니다. 조정은 위계적입니다 — "AI와 나는 상사-직원 같다", "AI는 늘 내가 준 것에만 반응한다".

다른 연구(45명 인터뷰)는 인지 오프로딩의 위험을 짚습니다. 오프로딩이 늘면 무비판적 수용이 늘고, 그만큼 사람의 **보상적 행동(검산·재구성)**이 줄어듭니다. 환각 위험이 모델이 아니라 사용 구조에서 증폭되는 이유입니다.

TMS 축	인간-AI에서 관찰된 것	조직이 설계할 수 있는 것
전문성 인식	사람은 AI의 강·약점을 학습하지만 AI는 사람을 기억하지 못함	AI가 참조할 '맥락 문서'를 상시 유지. 매 세션 재설명 비용을 구조로 흡수
신뢰성	검증된 맥락은 신뢰, AI가 새로 만든 정보는 불신 — 이중 기준	출처 요구를 기본값으로. 인용·수치가 필요한 산출물에는 근거 표기 의무화
조정	위계적 지시 구조. AI는 주도권이 없어 사람이 병목이 됨	역할 분담을 문서화(초안 AI·판단 사람·검증 제3자)하고, 병목 지점을 측정

행동 ① 분업
무엇을 AI가 하고 무엇을 사람이 하는지 사전에 나눕니다. 나누지 않으면 사람은 '검토자'로 전락하고, 검토는 형식이 됩니다. 분업의 단위는 과업이 아니라 **판단**이어야 합니다.

행동 ② 조정
AI가 참조할 맥락을 세션마다 다시 주지 않도록 문서·데이터로 고정합니다. 조정 비용이 사람에게 남아 있는 한 AI는 '빠른 검색창' 이상이 되지 못합니다.

행동 ③ 재작업
산출물을 '검토'하는 것이 아니라 다시 쓰는 것입니다. 재작업이 비판적 사고의 유일한 경로였습니다. 검토만 요구하는 규칙은 오프로딩을 그대로 둡니다.

역량 평가에 주는 시사점

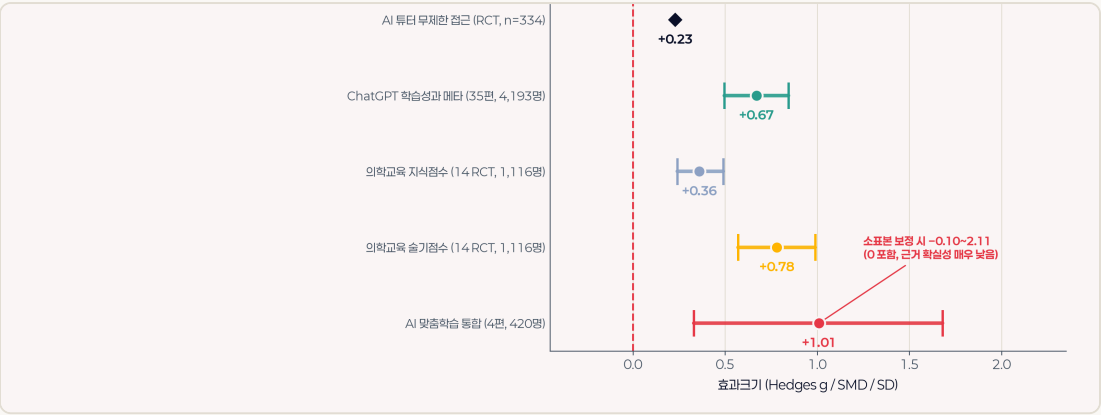
'AI 활용 능력'을 프롬프트 기술로 평가하면 이 연구가 설명한 경로를 놓칩니다. 평가해야 할 것은 ① 과업을 어떻게 나눴는가 ② AI가 만든 정보를 어떤 기준으로 검증했는가 ③ 산출물을 어느 정도 재작업했는가의 세 가지입니다.

교육·훈련 설계

같은 주의 국내 데이터(11개 교육청)는 AI 문해 교육의 초점이 '도구 사용법'에 머물러 있음을 보여줍니다. 실증 결과는 초점을 **분업·검증·재작업**으로 옮기라고 말합니다.

IV. 교육(1): 효과크기는 크지만, 근거는 아직 얇다

AI 맞춤학습 메타분석 g=1.01(0.33~1.68) · I²=89% · 근거 확실성 매우 낮음



AI 학습 개입의 효과크기와 신뢰구간. 자료: 각 논문 공개 수치

왜 1.01을 그대로 믿으면 안 되는가

PRISMA 2020 절차로 84건을 검토해 조건을 만족한 연구는 4편, 표본은 420명이었습니다. 통합 효과는 g=1.01(95% CI 0.33~1.68)로 '매우 큰' 수준입니다. 그러나 **Hartung-Knapp 소표본 보정**을 적용하면 구간이 −0.10~2.11로 넓어져 0을 포함합니다. 이질성 I²=89%, 예측구간 −0.42~2.43. 4편 중 2편은 비뚤림 위험이 '심각'이었고, 근거 확실성은 GRADE 기준 **매우 낮음**입니다.

연구가 모두 컴퓨팅·공학 과정에서 이뤄졌다는 점도 한계입니다. 인문·사회 과목이나 초·중등으로 일반화할 근거는 아직 없습니다.

더 큰 표본의 메타분석

ChatGPT 학습성과 메타분석(35편·4,193명)은 g=0.670(0.495~0.844)을 보고했습니다. 조절변수 중 과목, 실험 기간, 수업 방식이 유의했고, 교육 단계와 지식 유형은 유의하지 않았습니다. 의학교육 14개 무작위실험(1,116명)에서는 지식 SMD 0.36, 술기 0.78, 만족도 0.97이었습니다. **지식보다 술기에서 효과가 크다**는 점이 일관됩니다.

효과크기를 읽는 규칙

① 통합값보다 신뢰구간과 이질성을 먼저 본다 ② 표본 수 4편 이하면 '예비적 증거'로만 인용한다 ③ 효과크기는 개입의 질이 아니라 대조군의 상태에 의존한다 — 대조군이 전통 강의일 때 효과가 커진다 ④ 같은 결과라도 지식·술기·만족도를 분리해 읽는다.

현장 데이터와의 간극

실험실의 g=1.01과 교실의 1.63%는 모순이 아닙니다. 전자는 **쓰게 했을 때의 효과**, 후자는 **쓰지 않을 때의 현실**입니다. 도입률이 아니라 사용률이 성과의 상한을 결정합니다.

이 장의 한 줄 효과크기 1.01은 '기대값'이 아니라 '가능성의 상한'이다. 확실성을 올리는 방법은 표본을 늘리는 것이고, 유용성을 올리는 방법은 대조군을 설계하는 것이다.

IV. 교육(2): 접근을 열어주는 것이 통제하는 것보다 나았다

IZA 무작위실험 n=334 · 무제한 접근 +0.23 SD · 제한 접근 대비 +0.21 SD

설계

대학생 334명을 인센티브가 걸린 시험에 대비하도록 세 조건에 무작위 배정했습니다. ① 교재만 제공 ② 먼저 독립적으로 읽어야 AI 튜터를 쓸 수 있는 제한 접근 ③ 학습 기간 내내 제한 없는 접근. 결과는 예상과 달랐습니다. AI 튜터 접근은 대조군보다 **0.23 표준편차** 높은 성과를 냈고, **무제한 접근이 제한 접근보다 0.21 표준편차 더 높았습니다**. 조기 의존이 학습을 해친다는 우려와 반대의 결과입니다.

행동 분석이 설명한 이유

무제한 접근 집단은 AI 사용이 **점진적으로 통합**됐습니다. 반면 제한 집단은 해제 시점에 프롬프트를 몰아서 쓰는 **집중 폭발**이 나타나 학습 흐름이 끊겼습니다. 같은 도구, 같은 시간이라도 접근 규칙이 사용 리듬을 바꾸고, 사용 리듬이 학습을 바꿉니다. 효과는 이질적이었습니다 — 기초지식이 낮고 자기조절이 강한 학생에게 효과가 집중됐습니다.

한국의 현장: 같은 주의 두 데이터

① 사용률. 11개 시도교육청의 인공지능 맞춤형 교수학습플랫폼은 가입자 127만1,120명, 예산 635억 원, 하루 평균 방문율 1.63%(7~9월)입니다. 9월만 보면 2.25%입니다.

② 규제. 경기도교육청은 초등의 범용 생성형 AI 단독 이용을 제한하고, 고등은 출처·이용 내역 공개를 전제로 허용하는 기준안을 제시했습니다. 내년 1월 5일 최종안을 발표합니다.

두 데이터를 겹쳐 읽으면 질문이 바뀝니다. "AI를 허용할 것인가"가 아니라 **"어떤 접근 규칙이 점진적 통합을 만드는가"**입니다. 실험 결과는 규제의 방향이 아니라 설계의 방향을 지지합니다.

접근 설계	관찰된 행동	학습 결과	교육기관·에듀테크에 주는 조치
무제한 통합	점진적 사용, 필요 시점에 짧게 반복	+0.23 SD (대조군 대비)	도구를 과제 흐름 안에 상시 배치. 로그인 장벽·차시 제한을 줄일 것
사전 독립학습 후 해제	해제 직후 집중 프롬프트 폭발	무제한 대비 −0.21 SD	'먼저 혼자' 규칙은 사용 리듬을 깨뜨릴 수 있음. 단계적 개방으로 설계
교사 감독 하 교육용 AI	수업 내 제한적 활용	정책 기준안(경기)	감독의 초점을 '차단'이 아니라 '재작업 지도'에 둘 것

접근 규칙 설계의 세 원칙

① **상시성.** 과제가 시작되는 순간부터 도구가 있고, 언제든 짧게 쓸 수 있어야 합니다. 차시 단위 개방은 '몰아쓰기'를 만듭니다.
② **점진성.** 초반에는 교사가 사용 장면을 보여주고, 후반에는 학생이 스스로 질문을 구성하게 합니다.
③ **기록성.** 무엇을 물었고 무엇을 고쳤는지 남깁니다. 기록 자체가 평가의 근거가 됩니다.

에듀테크 사업자에게

이 실험은 제품 기능이 아니라 **사용 흐름의 설계**가 효과를 만든다는 것을 보여줍니다. 로그인·승인·차시 제한은 학습 효과를 깎는 마찰입니다. 반대로 재작업 기록·출처 확인·질문 이력은 효과를 만드는 기능입니다. 계약 갱신 때 제시할 지표도 여기에 맞춰야 합니다.

V. 종합: 여섯 편이 가리키는 하나의 구조 — 위임 · 검증 · 책임

경로가 다른 연구들이 같은 세 개의 관문을 통과할 때만 성과가 발생했다

관문 1. 위임 — 무엇을 넘길 것인가

텍스트형·멀티모달형 연구는 위임의 **종류**가 역량을 결정한다고 말합니다. 알고리즘 관리 연구는 위임의 **주체**(AI가 관리자인가, 동료인가)가 별개의 경로를 만든다고 말합니다. 위임을 정의하지 않은 도입은 측정도 개선도 불가능합니다.

관문 2. 검증 — 누가 무엇을 확인하는가

TMS 연구의 '재작업', 오프로딩 연구의 '보상적 행동', 교육 연구의 '점진적 통합'은 모두 검증 절차의 다른 이름입니다. 검증이 없는 위임은 사용량만 늘립니다 — 615명 연구에서 기술 신뢰(0.144)가 약했던 이유입니다.

관문 3. 책임 — 결정의 소유자

400명 연구에서 AI 협업은 **의사결정의 질**을 통해서만 성과에 닿았습니다. 의사결정의 질은 '누가 그 결정을 소유하는가'와 분리되지 않습니다. AI가 만든 초안에 사람의 서명이 붙는 순간, 책임 구조가 성과 구조가 됩니다.

연구	성과로 이어진 경로	경로를 막는 요인	조직이 통제 가능한 변수
알고리즘 관리·협업	HAC·AIgM → 근접 활용 → 업무성과 (Q^2=0.524)	두 축을 하나의 지표로 관리	지표의 분리, 교육·거버넌스의 분리
활용 유형	정보처리·아이디어 → 분석가용시간 → 분석역량 → 성과	절감 시간의 회수(회의 재투입)	절감 시간의 재투자 규칙
AI 주입	주입 0.833 → 조직 효과성	기술 신뢰·경쟁 압력만으로 접근	핵심 프로세스 내장률
거래기억	분업·조정 + 재작업 → 비판적 사고	무비판적 수용, 검증 생략	재작업 의무, 출처 요구
AI 맞춤학습	개입 → 학업성과(g=1.01, 불확실)	소표본·비뚤림·과목 편중	효과 검증 설계(통제집단)
AI 튜터	무제한 접근 → 점진적 통합 → +0.23 SD	제한 접근이 만든 집중 폭발	접근 규칙과 사용 리듬

관문이 무너지는 장면 ①

사용률만 보고되는 조직. 대시보드에는 'AI 사용 시간' 하나가 있고, 협업형과 관리형이 구분되지 않습니다. 두 효과가 상쇄되고 개선점이 보이지 않습니다(n=128).

관문이 무너지는 장면 ②

절감 시간을 다시 회의로 채우는 조직. AI가 만든 시간이 의사결정의 질로 전환되지 않고, 분석가용시간은 늘었는데 성과는 그대로입니다(n=400).

관문이 무너지는 장면 ③

검토만 요구하고 재작업은 요구하지 않는 조직. 오프로딩이 늘고 보상적 행동이 줄어 오류가 늦게 발견됩니다(n=891, 45명 인터뷰).

세 관문의 순서 위임이 먼저 정의되지 않으면 검증은 형식이 되고, 검증이 없으면 책임은 사람에게 남지 않는다. 여섯 편의 연구는 이 순서를 한 번도 뒤집지 않았다.

VI. 실전 프레임워크: AI 위임 4단계 설계도

이번 주 여섯 편의 실증 결과를 하나의 도입 절차로 압축했다

① 과업 등급화

과업을 **정리형**(텍스트형 AI)과 **생성형**(멀티모달)으로 나누고, 각각에 측정 지표를 붙입니다. 등급이 없으면 효과도 없습니다.

② 위임 수준 결정

초안(AI) → 판단(사람) → 검증(제3자)의 3역 역할을 문서화합니다. AI를 '관리자'로 쓰는 경우에는 평가 기준 공개와 이의제기 절차가 선행 조건입니다.

③ 검증 지점 설치

산출물마다 '출처 요구'와 '재작업 의무'를 붙입니다. 검증이 없는 사용은 인지 오프로딩을 늘리고 보상적 행동을 줄입니다.

④ 책임 기록

의사결정 로그에 'AI 초안 여부·검증자·최종 승인자'를 남깁니다. 이 기록이 곧 다음 계약과 감사의 근거가 됩니다.

도입 체크리스트 (연구 근거 연결)

① 우리 조직의 AI 지표가 '사용률' 하나인가 → 두 축(HAC·AIgM)으로 분리 (n=128)
② AI가 만든 시간을 어디에 재투자했는지 기록하는가 → 분석가용시간의 재투자 (n=400)
③ 핵심 프로세스에 AI 산출물이 입력되는 비율을 아는가 → 주입 0.833 (n=615)
④ 산출물에 출처 표기와 재작업을 요구하는가 → 재작업이 비판적 사고의 경로 (n=891)
⑤ 교육·훈련의 효과를 통제집단으로 검증했는가 → g=1.01의 불확실성
⑥ 접근 규칙이 사용 리듬을 깨뜨리고 있지 않은가 → 무제한이 제한보다 +0.21 SD (n=334)

30일 실행 순서

1주. 과업 목록을 정리형·생성형으로 분류하고, 각 과업의 현재 소요 시간을 측정.
2주. 두 개 과업에만 AI를 배치하고 3역(초안·판단·검증)을 지정. 검증 체크리스트 작성.
3주. 절감 시간의 재투자 내역을 기록. 의사결정 로그에 AI 개입 여부 표기 시작.
4주. 같은 과업의 처리 시간·재작업률·의사결정 리드타임을 비교. 효과가 없으면 도구가 아니라 **경로**를 바꿉니다.

역할	초안 단계	판단 단계	검증 단계
AI	대안 생성, 자료 구조화	—	출처 후보 제시
실무자	과업 정의, 맥락 제공	대안 선택, 가정 명시	수치·인용 대조
관리자	과업 등급 지정	의사결정 승인	재작업 요구 여부 결정
조직	도구 배치, 접근 규칙	의사결정 로그 보관	검증 기준·감사 추적

이 프레임워크의 전제

여섯 편의 연구는 서로 다른 표본·방법으로 같은 경로를 관찰했습니다. 그러나 어느 연구도 '이 순서로 실행하면 성과가 난다'를 직접 검증하지는 않았습니다. 이 도식은 실증 결과를 실행 절차로 번역한 것이며, 적용 후 전후 비교로 확인해야 합니다.

실패를 빨리 아는 방법

30일 뒤 세 숫자를 봅니다 — ① 같은 과업의 처리 시간 ② 재작업률 ③ 의사결정 리드타임. 세 숫자가 모두 그대로라면 도구가 아니라 과업 선정이 틀린 것입니다. 도구를 바꾸기 전에 과업을 바꾸십시오.

출처·방법·한계

인용한 논문의 서지 정보와, 이 호에서 확정한 것과 확정하지 않은 것

서지 정보	표본·방법	이 호에서 인용한 값
강현정 (2026). The Impact of Algorithmic Management and Human-AI Collaboration in AI-Augmented Organizations. *Information Systems Review*, 28(1), 355-.	n=128, PLS-SEM, PLSpredict	HAC·AIgM의 이중 매개 경로, Q^2=0.524, 2차 요인 4개 차원
조영은·우종필·김중안 (2026). 생성형 AI의 활용이 비즈니스 분석역량과 조직성과에 미치는 영향. *한국빅데이터학회지*, 11(1), 133-156.	생성형 AI 사용 경험 근로자, SEM	텍스트·멀티모달의 차별 효과, 4개 매개변수, 분석가용시간 경로
Koh, L. Y., & Yuen, K. F. (2026). Infusion of generative artificial intelligence. *Information, Communication & Society*.	n=615, TOE, SEM	11개 가설 지지, 총효과 0.833 / 0.468 / 0.259 / 0.181
Akgun, M., & Toker, S. (2026). When Does GenAI Use Support Critical Thinking? arXiv:2610.38910.	n=891 (미국 645·튀르키예 246)	완전매개 구조, 분업·조정과 재작업의 경로
Govea Sotomayor, J., & Garcia Ortiz, J. S. (2026). AI-based personalized learning and academic performance. *Frontiers in Education*.	4편, 420명, PRISMA 2020	g=1.01(0.33~1.68), 보정 후 −0.10~2.11, I^2≈89%, GRADE 매우 낮음
Fischer, M., Rau, H. A., & Rilke, R. M. (2025). AI Tutoring Enhances Student Learning Without Crowding Out Reading Effort. IZA DP 18338.	n=334, 3조건 무작위	+0.23 SD, 무제한−제한 = +0.21 SD, 집중 폭발 현상
보조 인용: Wu et al. (2026) *Humanities and Social Sciences Communications* 35편·4,193명 g=0.670; Pillai et al. (2026) *Postgraduate Medical Journal* 14 RCT·1,116명; Meske et al. (2026) arXiv:2610.28343 14명 인터뷰; Xu et al. (2026) AMCIS 45명 인터뷰.	각 논문 참조	효과크기 비교, TMS 비대칭, 인지 오프로딩의 실패 모드

방법론적 한계

인용한 연구 대부분이 횡단 설문이므로 인과 방향은 확정되지 않습니다. 메타분석은 포함 연구 수가 적어 이질성 추정이 불안정합니다. 인터뷰 연구는 표본이 14명·45명으로 일반화에 한계가 있습니다. 효과크기는 서로 다른 척도(Hedges g, SMD, 표준편차 단위)를 함께 표기했으므로 절대값 비교는 주의해야 합니다.

확정한 것과 확정하지 않은 것

확정: 각 논문이 보고한 계수·효과크기·표본 수, 그리고 연구 설계.
미확정: 우리 조직에 적용했을 때의 효과 크기. 이 호의 프레임워크는 논문의 경로를 실행 절차로 옮긴 제안이며, 검증된 처방이 아닙니다. 적용 후에는 반드시 자체 통제집단이나 전후 비교로 확인하십시오.