CONDERI

ACADEMIC & PAPER RESEARCH INSIGHT | 대외 공개판

효과는 입증됐고 설계는 아직 비어 있다

Effects Proven, Design Missing — What Four 2026 Studies Say About AI in Practice

이 리포트가 다루는 것

이번 주 학술 지형은 한 문장으로 요약됩니다. 인공지능의 학습 효과는 이미 통계적으로 입증됐지만, 그 효과를 만드는 설계 원리는 아직 정리되지 않았습니다. 한국교육정보미디어학회지 32권 2호의 3수준 메타분석은 실험·준실험 54편, 효과크기 169개에서 평균 Hedges' g = 1.053을 보고했습니다. 그러나 효과는 상호작용 유형과 학습자 준비도에 따라 크게 달라졌습니다. 같은 기간 APJIS 36권 2호는 산업과 학계의 주제가 구조적으로 어긋나 있다는 점을 LDA·임베딩 분석으로 보였고, 한국경영정보학회 추계학술대회는 '에이전틱 AI에서 피지컬 AI로'를 주제로 논문 120여 편을 다뤘습니다.

g = 1.053
생성형 AI·학습자 상호작용
메타분석 평균 효과크기

82.3%
지식그래프 AI튜터군
지식 잔존율(대조군 63.8%)

120여 편
한국경영정보학회
2026 추계 발표 논문

54편
메타분석 대상 연구
효과크기 169개

Executive Summary / 요약

Four studies, one shared conclusion

오늘의 결론 — 네 편의 연구가 같은 곳을 가리킨다

첫째, 효과는 존재합니다. 생성형 AI와의 상호작용은 학습성과 전반에 큰 효과를 보였습니다(Hedges' g = 1.053, 54편·169개 효과크기). 둘째, 효과의 크기는 기술이 아니라 설계가 결정합니다. 정교화 피드백 중심 상호작용과 대화형 상호작용에서 효과가 컸고, 디지털 리터러시·과제 구조·개입 기간이 모두 유의한 조절변인이었습니다. 셋째, 조직 차원에서는 학계와 산업의 시간 감각이 어긋나 있습니다. APJIS 36권 2호는 이를 "구조적 시간 비동기성(structural temporal asynchrony)"으로 개념화했습니다. 넷째, 시스템 차원에서는 거버넌스가 정확도를 만듭니다. 데이터 거버넌스와 다중 에이전트 오케스트레이션을 결합한 연구는 거버넌스 적용 시 분석 정확도가 개선된다는 실험 결과를 보고했습니다.

[기술] 에이전틱에서 피지컬로

2026 추계학술대회 주제는 "Agentic AI에서 Physical AI로". AI가 스스로 과업을 수행하고 물리 환경에서 작동하는 단계로 논의 축이 이동했습니다.

[경영] 산학 격차의 정체

산업 뉴스와 IS 논문을 같은 언어 모델로 비교하면 주제 클러스터가 따로 형성됩니다. 문제는 관심의 차이가 아니라 지식 생산 방식의 시차입니다.

[과학] 거버넌스가 정확도를 만든다

멀티 에이전트 오케스트레이션 단독보다 데이터 거버넌스 프레임워크를 결합했을 때 비전문가의 분석 접근성이 높아졌습니다.

[교육] 유의하지만 조건부

동기 효과는 중간 수준, 몰입 효과는 불확실했습니다. 성과 유형별로 효과가 갈린다는 점이 도입 설계의 출발점입니다.

한 줄 효과는 이미 증명됐다. 남은 질문은 "어떻게 설계했는가"다 — 네 연구의 공통 결론

I. 기술 (Tech) — 논의 축이 소프트웨어에서 물리 환경으로

KMIS 2026 추계학술대회 · 11월 12~13일 · 연세대

[기술] 에이전틱 AI에서 피지컬 AI로 — 2026 추계학술대회가 옮긴 논의 축

한국경영정보학회(회장 김범수)와 연세대 바른ICT연구소, ISACA Korea가 11월 12~13일 연세대 신촌캠퍼스에서 연 2026 추계학술대회의 주제는 "Agentic AI에서 Physical AI로: 지능형 정보시스템과 미래 사회 전환"이었습니다. 논문 120여 편이 AI 전환과 업무 혁신, 생성형 AI와 조직 변화, 인간-AI 협업을 주제로 발표됐습니다. 기조강연은 국가인공지능전략위원회 지원단장 김경만이 "국가 AI전략과 지속가능한 AI 생태계 조성을 위한 정부의 정책 방향"을 주제로 맡았습니다.

AI 코드 어시스턴트의 공급망 보안

바른ICT연구소 특별세션은 코드 생성 도구가 개발 파이프라인에 들어올 때 생기는 의존성·패키지 공급망 위험을 다뤘습니다. 생성 코드의 정확성보다 출처 검증이 먼저라는 문제 설정입니다.

AI 안경과 주변인 프라이버시

웨어러블 기기가 상시 촬영·인식 기능을 갖게 되면서 착용자가 아닌 주변인의 프라이버시 보호가 쟁점으로 부상했습니다. 피지컬 AI의 첫 규제 지점은 성능이 아니라 촬영 동의입니다.

AI 신뢰성 검증

ISACA Korea 세션은 AI 시스템의 신뢰성을 어떻게 검증할 것인가를 다뤘습니다. 감사·통제 관점에서 에이전트의 행위 기록과 권한 위임 체계가 핵심 논점이었습니다.

실무 시사점

에이전틱 AI 도입을 검토하는 조직은 세 가지를 먼저 문서화해야 합니다. 첫째, 에이전트가 어떤 권한을 위임받는지(권한 위임표), 둘째, 생성 코드·문서의 출처를 어떻게 검증하는지(공급망 검증 절차), 셋째, 물리 환경과 접점이 있는 기기의 촬영·인식 동의 절차입니다. 세 항목 모두 성능 지표가 아니라 통제 지표입니다.

II. 경영 (Management) — 산학 격차는 관심이 아니라 구조다

APJIS 36(2) 2026.06 · pp.210-237

[경영] 산업과 학계는 왜 같은 현상을 다르게 읽는가

Asia Pacific Journal of Information Systems 36권 2호(2026.06, pp.210-237)에 실린 김재희·최현정·김희웅(연세대)의 "Bridging the AI gap between Information Systems Research and Industry Practice"는 이 질문을 방법론으로 다룹니다. 연구진은 계산적으로 지원된 체계적 문헌고찰(computationally assisted SLR)을 설계해 잠재 디리클레 할당(LDA)과 임베딩 기반 의미 유사도 분석을 결합했습니다. 산업 뉴스와 IS 학술 논문의 담론 패턴을 같은 좌표에 놓고 비교한 것입니다.

발견	내용	실무 해석
구조적 시간 비동기성	산업과 학계의 지식 생산 방식이 달라 시차가 구조적으로 발생	격차는 관심 부족이 아니라 생산 주기의 차이
동일 현상의 다른 해석	공유된 주제 영역에서도 해석 렌즈가 다름	같은 용어로 대화해도 결론이 다를 수 있음
영역별 고유 클러스터	한쪽에만 형성되는 주제군이 존재	산업만 다루는 의제가 연구 공백
연구 아젠다 제안	Padmanabhan 등(2022) 프레임 확장	중장기 전략 수립에 학술 근거 연결

경영진을 위한 세 가지 독법

첫째, 학술 논문의 결론이 늦게 도착하는 것은 게으름이 아니라 검증 절차의 비용입니다. 둘째, 산업 담론만으로 의사결정을 하면 검증되지 않은 개념을 표준처럼 쓰게 됩니다. 셋째, 반대로 학술 근거만 요구하면 시장 타이밍을 놓칩니다. 이 연구의 실천적 제안은 "산업 이슈를 연구 언어로 번역해 사내 아젠다에 올리는 절차"를 만드는 것입니다.

III. 과학 (Science) — 자율성의 상한은 모델이 아니라 데이터 계약

데이터 거버넌스 × 멀티 에이전트 오케스트레이션

[과학] 거버넌스가 붙은 멀티 에이전트가 더 정확했다

최승혁·송민(연세대)의 연구는 데이터 거버넌스 프레임워크와 다중 생성형 AI 에이전트 오케스트레이션 프레임워크를 결합한 빅데이터 자동 분석 시스템을 제안합니다. 문제의식은 명확합니다. 데이터 탐색·처리·추출·시각화의 전 과정이 전문 분석가의 영역에 머물러 있어 비전문가가 접근하기 어렵다는 점입니다. 자연어만으로 분석을 수행하는 자율 분석 어시스턴트를 제안하되, 거버넌스 계층을 함께 두어 데이터 품질 개선과 분석 자동화를 동시에 달성하려 했습니다.

실험 설계

비교 조건은 거버넌스 프레임워크 적용 여부입니다. ① 거버넌스 없이 에이전트 오케스트레이션만 사용한 조건의 정확도, ② 거버넌스를 적용한 조건의 정확도, ③ 업무 과제 수행 시간을 측정했습니다.

시사점 1 — 정확도는 계층에서 나온다

에이전트 수를 늘리는 것보다 데이터 품질 규칙을 먼저 세우는 편이 정확도에 유리했습니다. 자율성의 상한은 모델 성능이 아니라 데이터 계약입니다.

시사점 2 — 비전문가 접근성

자연어 인터페이스는 분석가 부재를 대체하는 것이 아니라 분석 요청의 병목을 푸는 도구입니다. 사내 도입 시 요청 큐의 길이가 지표가 됩니다.

인간-AI 협업 연구로의 연결

이 결과는 인간-AI 협업 문헌의 교류기억체계(transactive memory) 논의와 맞닿습니다. 에이전트가 무엇을 기억하고 누가 그것을 검증하는지가 정해지지 않으면, 자동화는 속도만 높이고 신뢰는 낮춥니다. 실무적으로는 에이전트별 기억 범위와 검증 책임자를 표로 관리하는 것이 출발점입니다.

IV. 교육 (EduTech) — 효과는 조건부, 설계는 구체적

2026년 상반기 실증 연구 세 편

연구	설계	핵심 수치	시사점
GAI-학습자 상호작용 메타분석 (한국교육정보미디어학회지 32(2), 2026.04)	실험·준실험 54편, 효과크기 169개, 3수준 메타분석·RVE	Hedges' g = 1.053 정교화 피드백·대화형 상호작용에서 효과 최대	효과는 상호작용 설계의 질에 좌우. 몰입 효과는 불확실
지식그래프 결합 AI 튜터 (BMC Medical Education, 2026.05)	무작위 대조 3군, 간호학 1학년 362명 (완료 301명), 16주	모듈 점수 82.5 vs 77.6 vs 74.1 (η^2=0.16) 지식 잔존 82.3% vs 63.8%	구조화된 지식 위에 튜터를 얹을 때 효과가 커진다
AI 디지털 티처 (Smart Learning Environments, 2026.05)	학기 단위 무작위 대조, 84명, 인문학 전공	분석 에세이 83.5 vs 71.3 외재적 인지부하 3.2 vs 5.4 (p<.05)	인지부하 감소가 성과 개선의 경로. 정서 궤적도 함께 개선

[교육] 세 연구를 겹쳐 읽으면 설계 원칙이 남는다

첫째, 정교화 피드백입니다. 정답을 알려주는 피드백보다 학습자의 추론을 확장하는 피드백이 효과가 컸습니다. 둘째, 구조입니다. 지식그래프처럼 개념 간 관계가 명시된 환경에서 튜터의 효과가 커졌습니다(지식그래프 이해도 η^2=0.41). 셋째, 부하 관리입니다. AI가 설명을 대신할 때 외재적 인지부하가 낮아지고(3.2 vs 5.4) 성과가 올랐습니다. 넷째, 조건부 효과입니다. 학습동기는 중간 수준, 학습몰입은 불확실했으므로 "몰입이 좋아진다"는 홍보 문구는 근거가 약합니다.

V. 적용 — 12주 실행 프레임워크

From evidence to a documentable design

① 측정
가입률이 아니라 주간 활성 활용률을 성과 지표로 삼습니다. 효과크기는 상호작용 유형에 따라 갈리므로, 사용 시간이 아니라 피드백 유형별 사용 비중을 기록합니다.

② 설계
지식 구조를 먼저 만듭니다. 개념 관계가 명시된 지식그래프나 커리큘럼 맵 위에 AI를 얹는 순서가 반대보다 효과적입니다.

③ 통제
에이전트 권한 위임표와 데이터 계약을 문서화합니다. 거버넌스 계층이 없는 자동화는 정확도가 아니라 오류 전파 속도를 높입니다.

단계	활동	산출물
1주	상호작용 유형 정의(정교화 피드백·대화형·단순 응답)	피드백 유형 분류표
2~4주	지식 구조 구축 및 파일럿 2개 반 운영	지식그래프 초안, 파일럿 로그
5~8주	거버넌스 계층 적용 및 A/B 비교	정확도·시간 비교 리포트
9~12주	인지부하·정서 지표 포함 재측정	분기 개선안

적용 효과는 설계에서 나오고, 설계는 측정 가능한 형태로 문서화될 때 자산이 된다

English Summary / 영문 요약

Four 2026 studies on AI effects, gaps and governance

Meta-analysis of learner interactions with generative AI

A three-level meta-analysis in the Korean Association for Educational Information and Media journal (volume 32, issue 2, April 2026) pooled 54 experimental and quasi-experimental studies and 169 effect sizes using robust variance estimation. The overall effect of learner interactions with generative AI was large, with a Hedges g of 1.053. Elaborative-feedback and dialogic interactions produced the largest effects. Digital literacy, task structure and intervention duration were significant moderators. Effects were large for self-efficacy, self-regulated learning and achievement, medium for motivation, and uncertain for engagement.

Two randomised trials on AI tutoring

A three-arm randomised controlled trial with 362 first-year nursing students (301 completers) over 16 weeks in BMC Medical Education compared a generative AI tutor integrated with a knowledge graph, blended teaching and traditional teaching. Module scores were 82.5, 77.6 and 74.1 (F=28.64, eta squared 0.16), knowledge retention 82.3%, 71.5% and 63.8%, and case-based inference accuracy 91.2%, 78.5% and 67.3%. A separate semester-long trial in Smart Learning Environments with 84 literature undergraduates found analytical essay scores of 83.5 against 71.3 and lower extraneous cognitive load, 3.2 against 5.4.

Industry-academia gap and agent governance

Asia Pacific Journal of Information Systems 36(2) published a computationally assisted systematic literature review combining latent Dirichlet allocation with embedding-based semantic similarity. It conceptualised the gap between industry practice and information systems research as structural temporal asynchrony and proposed an expanded research agenda. Separately, a multi-agent orchestration study presented at the Korea Society of Management Information Systems reported that combining a data governance framework with agent orchestration improved analysis accuracy for non-expert users. The society held its 2026 autumn conference on November 12 and 13 under the theme from agentic AI to physical AI, with around 120 papers presented.

Summary **Effects are established; design and governance determine their size**

발행 CONDERI AI산업전략연구실 | 문서등록번호 CONDERI-AR-2026-11-19 | 문의 contact@conderi.com