CONDERI

ACADEMIC & RESEARCH INSIGHT | 대외 공개판

에이전트를 더 붙인다고 더 좋아지지 않는다

More Agents Do Not Make Better Answers — Seven Measured Limits from a 120-Paper Conference

이 리포트가 다루는 것

11월 12~13일 열린 한국경영정보학회 추계학술대회는 '에이전틱 AI에서 피지컬 AI로'를 주제로 논문 120여 편을 발표했습니다. 이 리포트는 그중 기술·경영·과학·교육 네 영역에서 실무 적용 가치가 높은 연구를 골라, 하나의 질문으로 묶었습니다. **에이전트를 여러 개 붙이면 정말 더 좋은 답이 나오는가.** 답은 조건부입니다. 역할 분담은 출처 명시율을 40%에서 100%로 올렸지만 과도한 확정 표현 억제율은 100%에서 80%로 떨어졌고, 이용자 판단 고지 충족률은 두 구조 모두 30% 이하였습니다. 논문 7편이 남긴 측정값으로 에이전트 팀을 설계하는 다섯 단계를 제시합니다.

40% → 100%
역할 분담 시
출처 명시율

100% → 80%
과도한 확정 표현
억제율 하락

≤30%
이용자 판단 고지
충족률(두 구조 공통)

6%
EBIT 5% 이상
AI 기여 기업

I–II 기술
역할 분담의
임계 조건

III 경영
쓰는 89%,
버는 37%

IV–V 과학·교육
정체성 표류와
적절한 의존

VI 프레임워크
다섯 단계
설계안

i. 요약 — 수가 아니라 검증 계층

Not the number of agents, but the layer that verifies them

120여 편	32(3)	4개 영역	7편	5단계
KMIS 2026 추계 발표 논문	지능정보연구 2026년 가을호	기술·경영 과학·교육	이 리포트가 심층 분석	에이전트 팀 설계 프레임

한 문장

이번 학회의 공통 결론은 **"에이전트 수"가 아니라 "검증 계층"**이다. 역할을 나누면 추적성은 좋아지지만 비용과 확신 왜곡이 함께 늘고, 판단을 이용자에게 되돌려주는 고지 장치는 자동으로 생기지 않는다.

방법론적 특징

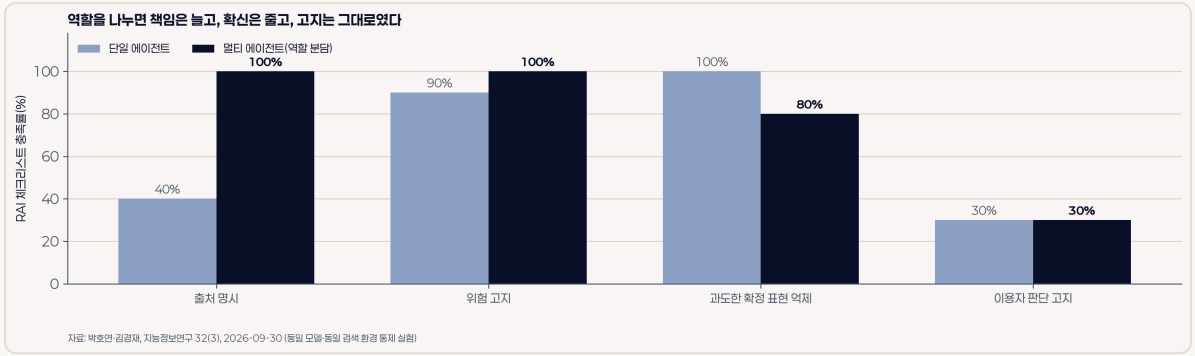
지능정보연구 32(3)의 비교 연구는 동일 언어모델·동일 검색 환경을 통제한 실험 설계로 싱글 에이전트와 역할 분담형 멀티 에이전트를 처리시간·응답길이·실행 안정성·RAI 체크리스트 네 축으로 평가했습니다. 구조만 바꾼 순수 효과를 분리했다는 점에서 실무 적용 가치가 큽니다.

독자를 위한 안내

이 리포트는 공개된 학술지 논문과 학술대회 발표 목록·초록을 근거로 작성했습니다. 실증 수치는 각 논문의 보고값을 그대로 인용했고, 산업 통계(맥킨지 2026 설문)는 논문 외 보조 근거로 구분해 표시했습니다. 해석과 프레임워크는 콘데리 AI산업전략연구실의 것입니다.

I. 기술 — 역할을 나누면 무엇이 좋아지고 무엇이 나빠지나

What role division buys and what it costs



지료: 박호연·김경재, 지능정보연구 32(3), 2026-09-30 (doi 10.13088/jiis.2026.32.3.039)

좋아진 것 — 추적성

출처 명시 40%→100%, 위험 고지 90%→100%. 역할을 나누면 각 에이전트가 자기 산출물의 근거를 남기게 되고, 그 기록이 보고서에 그대로 실립니다. 규제 산업에서 이 차이는 큽니다.

나빠진 것 — 확신의 절제

과도한 확정 표현 억제는 100%→80%로 하락했습니다. 검토자 역할이 붙어도 "~이다"라고 단정하는 문장을 스스로 걸러내는 능력은 오히려 줄었습니다. 처리시간도 늘었습니다.

실무 해석 — 역할 분담은 공짜가 아니다

이 실험의 조건은 동일 모델·동일 검색 환경입니다. 즉 성능 차이는 오직 구조에서 나왔습니다. 실무에서 멀티 에이전트를 도입할 때 기대할 수 있는 것은 정답률 상승이 아니라 **근거의 흔적**입니다. 따라서 도입 지표도 정확도가 아니라 "출처가 붙은 문장의 비율"과 "고지 누락 건수"로 잡아야 합니다. 그리고 확신 표현 억제는 프롬프트가 아니라 **검토 단계의 규칙**(금지어 목록, 확신도 표기)으로 다시 넣어야 합니다. 구조가 대신해 주지 않기 때문입니다.

II. 기술 — 복구는 구조가 아니라 전환부에서 결정된다

Recovery is decided at the hand-off, not by the architecture

발표	연구 질문	핵심 발견
D1-3 · KAIST	When Does Role Division Help? — 역할 분담은 언제 도움이 되는가	모델 규모가 작을 때와 클 때 역할 분담의 효과가 다르게 나타남. 공공 문서 검토 과제에서 역할 분담의 이득은 모델 규모와 과제 구조에 조건부
D1-2 · 경희대	단계별 통제 개입을 활용한 LLM 에이전트 오류 복구 병목 식별	복구 실패는 한 지점이 아니라 단계 전환부에 집중. 어느 단계에 개입해야 복구율이 오르는지 병목을 식별하는 방법을 제시
P-16 · 가천대	사용자 수정 행동 기반 AI 에이전트 개인화: 단일·다중 에이전트 비교	선호 추론의 정확도와 사용자 부담은 상충. 다중 구조가 항상 개인화에 유리하지 않음
F1-4 · 방송대	Do Marketing Tricks Work on Machines? — 설득 전략에 대한 LLM 취약성 감사	LLM은 인간을 겨냥한 설득 전략에 측정 가능한 수준으로 흔들림. 에이전트가 대신 읽는 세상에서는 마케팅 문구가 곧 조작 벡터

세 연구가 함께 말하는 것

역할 분담은 **임계 조건**이 있습니다. 모델이 작으면 역할을 나눠 이득이 나지만, 모델이 커지면 한 번에 처리하는 편이 낫거나 비용만 늘어납니다(KAIST). 복구 능력도 구조가 아니라 **전환부**에서 결정되므로, 개입 지점을 설계하지 않으면 에이전트를 늘려도 실패는 같은 자리에서 반복됩니다(경희대). 개인화는 사용자 부담과의 교환관계이므로, 다중 구조를 도입할 때는 "정확도 몇 %p"와 "사용자가 몇 번 더 고쳐야 하는가"를 함께 재야 합니다(가천대).

현업 적용 — 무엇을 먼저 재나

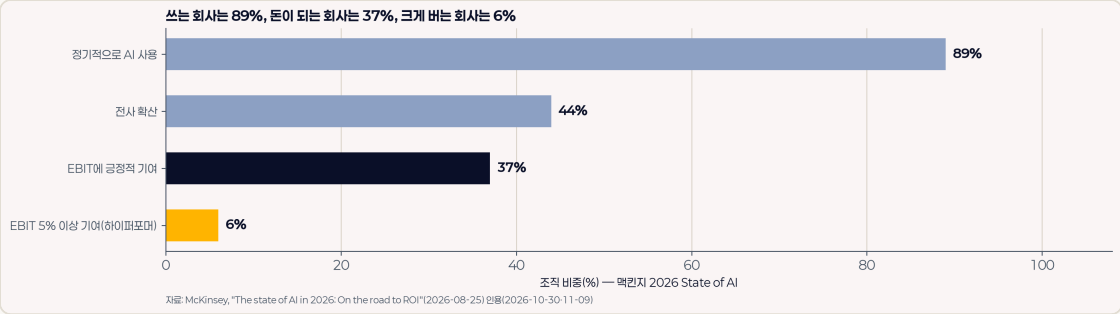
에이전트 도입 심사를 할 때 첫 장표에 정확도를 올리지 마십시오. 대신 세 개의 숫자를 올리십시오. 출처가 붙은 문장 비율, 인계 지점에서 걸러진 오류 건수, 사람이 최종 확인한 비율입니다. 이 세 숫자가 오르지 않으면 에이전트 수를 늘려도 결과는 같습니다.

주의 — 발표 초록 기준

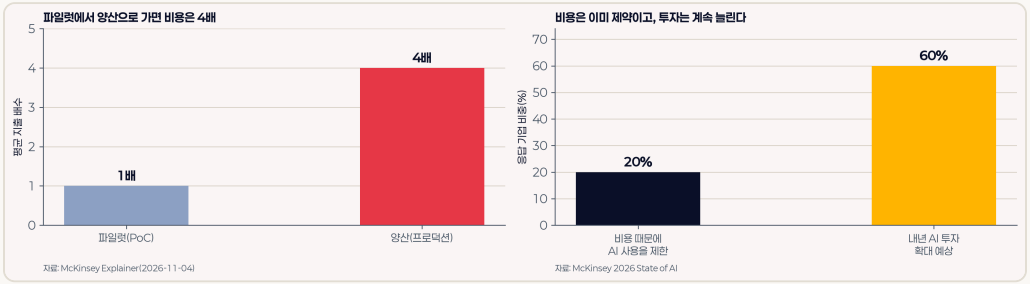
학술대회 발표 논문은 정식 게재 전 단계라 초록·발표 자료를 근거로 했습니다. 수치가 보고되지 않은 연구는 방향과 설계만 인용했고, 게재본이 나오면 표본과 효과 크기를 다시 확인해야 합니다.

III. 경영 — 쓰는 회사 89%, 돈이 되는 회사 37%

Adoption is broad, value is narrow, and cost decides the middle



자료: McKinsey, "The state of AI in 2026: On the road to ROI"(2026-08-25) 인용(2026-10-30, 11-09)



자료: McKinsey Explainer(2026-11-04), McKinsey 2026 State of AI

논문과 산업 통계가 만나는 지점

학술대회가 측정한 것은 구조의 효과이고, 산업 설문이 측정한 것은 조직의 결과입니다. 두 데이터가 같은 곳을 가리킵니다. 도입률(89%)과 성과(37%) 사이의 52%p 격차는 모델 성능이 아니라 **프로세스 재설계**의 부재에서 옵니다. 파일럿에서 양산으로 넘어갈 때 지출이 평균 4배가 되고, 다섯 곳 중 한 곳은 비용 때문에 사용을 줄입니다. 즉 "에이전트를 더 붙이는" 결정은 곧 비용 결정이며, 가치가 확인된 좁은 영역에 집중하는 편이 통계적으로 유리합니다.

IV. 경영·과학 — 에이전트가 새는 세 곳

Three places where agents leak: secrets, context, identity

비밀정보 — 코드가 새는 경로

P-05(연세대) AI 코딩 에이전트가 만든 PR의 하드코딩 비밀정보 노출을 같은 저장소의 인간 PR과 비교했습니다. 병합 이후에도 노출이 지속되는지까지 추적한 점이 핵심입니다. 에이전트는 편의를 위해 키를 코드에 넣고, 리뷰는 그 줄을 지나칩니다. 대응은 탐지 도구가 아니라 **비밀정보를 코드 밖으로 강제하는 파이프라인**입니다.

프라이버시 — 검색이 정확할수록

P-02(연세대) RAG의 검색 성능과 프라이버시 위험의 교환관계를 분석했습니다. 검색이 정확해질수록 더 많은 맥락이 프롬프트에 실리고, 그 맥락이 다시 저장·전송됩니다. P-04는 금융 RAG 벡터 저장소의 임베딩 역전 위험을, P-07은 잘못된 사용자 주장에 대한 프롬프트 방어의 구성요소별 효과를 다뤘습니다.

정체성 — 여러 사용자가 한 에이전트를 쓸 때

E2-5(연세대 바른ICT연구소) 다중 사용자 에이전트의 정체성 표류(identity drift)를 다룹니다. 사용자마다 다른 기대와 말투가 누적되면서 에이전트의 역할 정체성이 이동하는 현상입니다. 기업용 에이전트에서 이는 곧 **브랜드 목소리의 표류**입니다.

신뢰 — 적절한 의존의 설계

P-19(KAIST·베이징대) 설명과 사용자의 인지 방식이 맞을 때 적절한 의존(appropriate reliance)이 일어나는 조건을 AI 건강 조언 맥락에서 검증했습니다. 설명을 많이 붙이는 것이 아니라 **사용자가 검증할 수 있는 형태**로 붙여야 신뢰가 과신도 과소신도 되지 않습니다.

네트워크 관점 — 왜 이 누수가 동시에 생기나

네 연구는 서로 다른 현상을 다루지만 구조는 같습니다. 에이전트는 **맥락을 가진 채로 다음 단계로 넘어가고**, 그 맥락에는 사용자가 입력한 것과 시스템이 주입한 것이 함께 들어 있습니다. 비밀정보(코드), 개인정보(검색 맥락), 역할 정체성(대화 이력)은 모두 이 "넘어가는 맥락"에 실립니다. 따라서 통제 지점도 하나로 모입니다 — 인계 시점에 무엇을 넘길지 화이트리스트로 정하고, 넘긴 뒤에는 로그를 남기는 것입니다. 이 한 가지 설계가 세 누수를 동시에 줄입니다.

누수 에이전트 팀의 세 가지 누수 — 비밀정보는 코드로, 맥락은 프롬프트로, 정체성은 대화 이력으로 샌다

V. 교육 — 가르치는 능력보다 의심하는 능력

Teaching learners to doubt, not just to ask

발표	연구	교육 현장에 주는 것
D2-4 · 경북대	영어회화 교육업체의 AX 전환을 위한 AI Tutor 생성 솔루션 TutorBridge	기존 강사·교재 자산을 유지한 채 튜터를 얹는 전환 설계. 학원·어학원 같은 중소 교육사업자의 실행 가능한 경로를 제시
F1-5 · 경희대	LLM 금융자문에서 정보 프레이밍과 고객 투자성향이 위험자산 추천 비중에 미치는 영향	같은 정보라도 제시 방식에 따라 추천 비중이 달라짐. 금융·진로 상담형 AI의 프레이밍 통제가 필요하다는 근거
F1-7 · KAIST·연세대	AI 생성 검색 결과의 위험 고지 구체성이 정보 검증 및 회피 행동에 미치는 영향	고지가 구체적일수록 이용자의 검증 행동이 늘어남. 교육용 AI에서 고지 문구의 설계가 학습 태도를 바꿈
P-19 · KAIST	Explanation–User Fit and Appropriate Reliance on AI Health Advice	설명의 양보다 적합도. 학생·학부모 대상 AI 조언에서 신뢰 캘리브레이션의 설계 원칙

교육 영역의 결론 — 튜터보다 고지

교육 분야 발표에서 반복되는 것은 "AI가 얼마나 잘 가르치는가"가 아니라 **"학습자가 AI를 얼마나 적절히 의심하는가"**입니다. 고지의 구체성이 검증 행동을 늘리고(F1-7), 설명의 적합도가 적절한 의존을 만듭니다(P-19). 프레이밍이 추천을 흔든다는 결과(F1-5)는 진로·금융 상담형 서비스에 직접적인 규제·설계 함의를 줍니다. 공교육 AI 도입이 확대되는 국면에서 학교가 먼저 정해야 할 것은 콘텐츠가 아니라 **고지 문구와 검증 절차**입니다.

학교·학원이 지금 할 수 있는 것

첫째, AI가 만든 답에 "확인 필요" 표시를 붙이는 규칙을 정합니다(F1-7). 둘째, 튜터에게 정답을 주게 하지 않고 **질문을 만들게** 합니다. 셋째, 상담형 AI에는 프레이밍 금지 목록(공포·희소성·권위 호소)을 둡니다(F1-5).

기업 교육에도 같은 원칙

직무 교육에 AI를 붙일 때 성과 지표를 "수료율"에서 "AI 답을 검증한 비율"로 바꾸면, 교육의 목적이 도구 사용에서 판단 훈련으로 이동합니다. 이는 맥킨지가 지적한 프로세스 재설계와 같은 방향입니다.

VI. 프레임워크 — 에이전트 팀을 설계하는 다섯 단계

Five design decisions before adding another agent

단계	설계 결정	근거가 된 측정값
1. 과제 분해	역할 분담이 이득인지 먼저 계산 — 모델 규모·과제 구조·비용 3축	KAIST D1-3, 지능정보연구 32(3)
2. 전환부 설계	에이전트 사이 인계 지점에 검증 게이트 1개 — 여기서 복구율이 결정	경희대 D1-2
3. 근거 강제	출처 없는 문장은 보고서에 실리지 않게 스키마로 차단(40%→100% 사례)	지능정보연구 32(3)
4. 확신 통제	금지어·확신도 표기를 프롬프트가 아니라 규칙으로(100%→80% 하락 방지)	지능정보연구 32(3)
5. 사람 고지	이용자 판단 고지 ≤30% — 자동화로 해결되지 않는 유일한 항목	지능정보연구 32(3), P-19

도입 전에 답해야 할 세 질문

첫째, 이 과제에서 틀린 답의 비용이 얼마인가. 비용이 크면 5단계(사람 고지)를 생략할 수 없습니다. 둘째, 우리 데이터에 비밀정보가 섞여 있는가. 섞여 있다면 에이전트보다 **시크릿 관리**가 먼저입니다. 셋째, 이 에이전트의 목소리는 누구의 것인가. 다중 사용자 환경이라면 정체성 표류를 막는 시스템 프롬프트 고정과 정기 감사가 필요합니다. 세 질문에 답하지 못한 채 에이전트 수를 늘리는 것은 비용을 늘리는 가장 확실한 방법입니다.

이번 주 문헌이 남긴 하나의 문장

지능정보연구 32(3)의 결론부는 정직합니다. "에이전트 구조의 자동화만으로 보고서의 책임성을 충분히 확보하는 데 한계가 있다." 이 문장은 AI 도입을 미루라는 뜻이 아니라, **책임을 설계해야 비로소 자동화가 이익이 된다**는 뜻입니다. 검증 게이트 한 개, 고지 문구 한 줄, 인계 로그 한 장이 모델 업그레이드보다 큰 차이를 만드는 국면입니다.

프레임워크 에이전트 팀의 성과는 모델 수가 아니라 검증 게이트의 위치로 결정된다

English Summary

For global readers and archive

English Summary

The Korean Society of Management Information Systems (KMIS) held its 2026 autumn conference on November 12 and 13 under the theme "From Agentic AI to Physical AI", presenting about 120 papers. Journal of Intelligence and Information Systems 32(3) published, on September 30, a controlled comparison of single-agent and multi-agent financial report generation using the same model and retrieval environment. The multi-agent structure raised source attribution from 40 to 100 percent and risk disclosure from 90 to 100 percent, but reduced the suppression of overconfident phrasing from 100 to 80 percent, and both structures satisfied the user-judgment notice item in fewer than 30 percent of cases. Processing time increased.

What the program adds

Role division helps only under conditions: a KAIST study asks when role division helps and finds the answer depends on model scale and task structure. A Kyung Hee study locates agent failure recovery bottlenecks at stage hand-offs rather than inside stages. A Yonsei study compares hard-coded secrets in AI-generated pull requests with human pull requests in the same repository and tracks exposure after merge. Other papers examine the utility-privacy trade-off in retrieval-augmented generation, embedding inversion in Korean financial vector stores, identity drift in multi-user agents, and appropriate reliance on AI health advice. A KAIST and Peking University study finds that explanation-user fit, not explanation volume, drives calibrated trust.

Management evidence

McKinsey 2026 State of AI: 89 percent of organisations use AI regularly, 44 percent are scaling it, 37 percent report positive EBIT impact and 6 percent attribute at least 5 percent of EBIT to AI. Spending nearly quadruples from pilot to production, and one in five organisations limits use because of operating cost. The practical implication is to concentrate agents where value is measured, place a verification gate at each hand-off, and keep human notice as a design requirement rather than an afterthought.

Summary **Traceability is what agents add; restraint is what they remove**