

생성형 AI가 사용자 인지역량에 미치는 영향과 정책적 시사점*

임 일**

요약

대규모 언어 모델(Large Language Model, LLM)을 비롯한 생성형 AI의 급격한 확산은 지식 노동과 학습 환경에 전례 없는 생산성 향상을 제공하지만, 이와 동시에 인지적 외주화(cognitive offloading)를 유도하여 인간 고유의 비판적 사고, 독립적 추론, 장기 기억 능력을 저해할 위험을 안고 있다. 본 논문은 정보시스템(Information Systems: IS), 심리학, 인간-컴퓨터 상호작용(Human-Computer Interaction: HCI), 신경과학 및 교육학 분야의 최신 문헌을 종합하여 LLM이 인간의 인지역량에 미치는 영향에 대하여 이론적, 실증적 연구 결과를 조사하고 통합적인 재해석을 제시하였다. 종합하면, 습관적인 AI 의존은 (1) '인지 부채(cognitive debt)'의 누적, (2) 메타인지적 게으름(metacognitive laziness)과 시간 단축의 착각(speedup illusion), (3) '주권의 함정(sovereignty trap)'과 신념의 외주화(belief offloading), (4) 스키마 형성 붕괴에 따른 '속 빈 마음(hollowed mind)' 상태를 초래할 수 있는 것으로 나타났다. 본 연구에서는 이러한 인지적 쇠퇴 완화를 위한 방안으로 AI 설계 측면과 AI를 활용한 교육 측면에서 몇 가지 방안을 제시한다.

주제어 : 대규모 언어 모델(LLM), 인지적 외주화, 인지 부채, 주권의 함정, 속 빈 마음, 메타인지적 게으름

The Effects of Generative AI on User's Cognitive Capabilities and Its Policy Implications*

Il Im**

Abstract

The rapid proliferation of generative AI, including large language models (LLMs), has delivered unprecedented productivity gains in knowledge work and learning environments. At the same time, however, it carries the risk of inducing cognitive offloading that may undermine humans' inherent capacities for critical thinking, independent reasoning, and long-term memory formation. This paper synthesizes recent literature from information systems (IS), psychology, human-computer interaction (HCI), neuroscience, and education to provide a multifaceted theoretical and empirical analysis of the effects of LLMs on human cognitive abilities. Taken together, the reviewed evidence indicates that habitual reliance on AI can give rise to (1) the accumulation of 'cognitive debt,' (2) metacognitive laziness and a 'speedup illusion,' (3) a 'sovereignty trap' and belief offloading, and (4) a 'hollowed mind' resulting from the breakdown of schema formation. To mitigate such cognitive decline, this study suggests several approaches from two perspectives: the design of AI systems and the pedagogical use of AI in education.

Keywords: Large Language Model(LLM), cognitive offloading, cognitive debt, sovereignty trap, hollowed mind, metacognitive laziness

접수: 2026. 7. 30; 최종수정: 2026. 9. 7; 게재확정: 2026. 9. 7.

* 본 연구는 연세대학교 경영대학 연구경쟁력 강화사업(2025-22-0216)과 연세대학교 AI혁신연구원 연구비 (2025-22-0464) 지원을 받았습니다.

** 연세대학교 경영학과 교수 (il.im@yonsei.ac.kr, <https://orcid.org/0000-0001-7536-3170>)

I. 서론

ChatGPT를 비롯한 대규모 언어 모델(Large Language Model: LLM)은 학습, 글쓰기, 정보 탐색, 의사결정 지원 등 인간의 고차원적 인지 과업 전반에 빠르게 통합되고 있다. 이러한 기술적 전환은 단기적인 생산성을 극대화하고 즉각적인 피드백을 제공하는 등 지식집약적인 업무와 학습에 긍정적인 생산 혁신을 가져왔다(Martínez-Peláez et al., 2025). 그러나 프라이버시에 대한 우려를 가져오고 있으며(장창기, 2025), 동시에 효율성에 지나치게 의존한 나머지 인간 스스로 수행해야 할 본질적인 사고 과정마저 기계에 위임하는 인지적 외주화(cognitive offloading) 양상이 심화되고 있다(Gerlich, 2025a; Heersmink, 2024).

이러한 현상은 정보 검색 엔진과 인터넷이 인간의 기억 전략에 미친 영향을 규명한 기존 연구(Sparrow et al., 2011)의 연장선에 있다. 그러나 LLM은 단순한 정보 저장을 넘어 스스로 논증을 구성하고 판단을 내리는 등 고차원적 인지 과정을 대신 수행한다는 점에서 인터넷으로 인한 영향과는 질적으로 다른 심각한 결과를 가져올 수도 있다. AI의 무분별한 사용이 계속될 경우, 지식 습득의 기반이 되는 인지 구조가 잠식되어 장기적으로 비판적 사고와 독립적 추론 능력이 위축될 수 있다(Klein & Klein, 2025). 또한, 그 결과 인간의 학습능력 자체가 퇴보할 수도 있다. 최근에는 이러한 현상을 설명하기 위해 ‘인지 부채(cognitive debt)’(Kosmyrna et al., 2025), ‘주권의 함정(sovereignty trap)’과 ‘속 빈 마음(hollowed mind)’(Klein & Klein, 2025), ‘메타 인지적 게으름(metacognitive laziness)’(Choudhuri et al., 2026), ‘시간 단축의 착각(speedup illusion)’(Yu et al., 2026), ‘신념 외주화(belief offloading)’(Guingrich et al., 2026) 등 다수의 신조어가 학계에서 제안되고 있으며, 이는 문제의식이 특정 분야에 국한되지 않고 IS, 심리학, 인간-컴퓨터 상호작용(Human-Computer Interaction: HCI), 신경과학, 교육학 전반으로 급속히 확산되고 있음을 보여준다.

따라서 기술의 편의성과 인간의 인지적 성장 사이의 균형을 유지하는 것은 현시대의 가장 시급한 과제이다. 본 리뷰는 IS, 심리학, HCI, 교육학 및 신경과학 분야의 핵심 학술지, 학회에 발표된 최신 실증 연구와 이론적 논의, 그리고 아직 심사를 거치지 않은 예비 논문(preprint) 및 블로그와 같은 실무 자료를 통합적으로 검토한다. 이를 통해 생성형 AI가 인간의 인지역량에 미치는 구체적인 부작용을 규명하고 사용자의 인지적 자율성과 주권을 보호하기 위한 교육적, 시스템 설계적 개선 방안을 제시하고자 한다.

본 연구의 문헌조사를 위하여 IS 분야의 MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, 심리학·HCI 분야의 Computers in Human Behavior, Information & Management, Decision Support Systems, 주요 학술대회(ACM CHI, ICIS)를 대상으로 ‘large language model,’ ‘generative AI,’ ‘ChatGPT,’ ‘cognitive offloading,’ ‘critical thinking’을 검색어로 하여 2020년 이후 발행된 문헌을 검색하였다. Crossref API를 통한 데이터베이스 검색으로 상기 7개 IS·심리학 저널에서 총 93편의 문헌을 식별하였으며, 이에 더해 학회 프로시딩, 프리프린트(arXiv), 국내 저널 및 기타 관련 학술지에 대해서는 인용망 추적(citation chaining) 및 수기 검색(hand search)을 병행하였다. 식별된 문헌은 제목 및 초록을 기준으로 LLM 또는 생성형 AI가 인간의 인지역량(비판적 사고, 기억, 메타인지 등)에 미치는 영향을 직접적으로 다루는지 여부를 기준으로 1차 스크리닝하였고, 이후 전문 검토를 통해 IS, 심리학, HCI, 교육학 분야의 실증 및 이론 연구로 적격성을 재확인하였다. 또한, 검색결과에는 없지만 연구자 판단으로 적절하다고 생각되는 논문을 4편 추가하였으며 학술연구논문은 아니지만 중요하다고 생각되는 블로그도 2편을 추가하여 최종적으로 국내 문헌 4편, 국외 문헌 33편, 총 37편이 본 문헌연구에 포함되었다.

II. 이론적 배경: 인지적 외주화와 확장된 인지

1. 거래적 기억과 구글 효과

인간이 정보 처리의 부담을 외부 도구나 타인에게 위임하는 인지 작용은 거래적 기억(transactive memory)과 인지적 외주화(cognitive offloading) 개념을 기반으로 설명된다. Sparrow et al.(2011)는 정보가 디지털 도구에 저장되어 있어 언제든지 접근할 수 있다고 믿을 때 사람들이 정보 자체보다 '정보를 어디서 찾을 수 있는지'의 경로를 더 잘 기억하는 이른바 구글 효과(Google effect)를 발견하였다. 이는 디지털 도구가 기억의 내용물이 아니라 정보를 처리하는 인지적 전략 자체를 바꾼다는 점을 시사하며, 이후 LLM의 인지적 영향을 다루는 다수의 연구가 이 이론적 틀을 계승하고 있다.

정동훈(2026)은 에이전틱 AI를 “인간의 부족한 기억력, 계산력, 탐색 능력을 보완하는 외주화된 인지(offloaded cognition) 시스템”으로 명시적으로 규정하고, 인간은 AI 산출물을 감독하고 가치 판단을 내리는 ‘지휘자’ 역할을 맡는다고 주장한다. 이 또한 AI에 의한 인지의 외주화의 일종이라고 할 수 있다.

2. 확장된 인지 이론과 LLM

최신 문헌들은 생성형 AI의 파급력이 이러한 기존의 검색 수단을 넘어선다고 경고한다. Heersmink(2024)는 LLM이 사용자의 사고 및 글쓰기 과정 자체에 역동적으로 개입하는 주체라는 점에서 이를 확장된 인지(extended cognition) 이론의 관점에서 재검토해야 한다고 강조한다. 그는 LLM이 사용자의 사고를 임시적으로 지탱해주는 비계¹⁾(scaffolding) 도구로 쓰이는가, 아

니면 사고 과정을 완전히 대신하는 대체(replacement) 도구로 쓰이는가에 따라 인지 발달의 궤적이 근본적으로 달라진다고 본다. 적절한 개입 없이 대체재로만 활용된다면 장기적으로 고차원 지적 역량의 영구적인 손실을 피할 수 없다는 것이 그의 핵심 주장이다.

이러한 이론적 프레임은 뒤에서 살펴볼 Klein and Klein(2025)의 ‘속 빈 마음(hollowed mind)’ 개념 과도 맞닿아 있다. 이들은 확장된 인지 이론을 정교화하여, 절차적, 검색적 정보의 외주화(예: 계산기, 검색 엔진 사용)와 LLM과 같은 통합적 추론(integrative reasoning)의 외주화는 질적으로 다른 문제이며, 후자의 경우 도구에 대한 의존이 곧 사고 능력 자체의 상실로 이어질 수 있다고 주장한다.

III. LLM이 인간 인지역량에 미치는 영향

1. 신경인지적 관점: ‘인지 부채(Cognitive Debt)’의 축적

인간이 기억과 추론, 글쓰기와 같은 핵심 인지적 활동을 LLM에 전적으로 외주화할 때 두뇌의 신경 가소성과 인지적 관여도는 저하되는 것으로 보고된다(Li, 2025). 이러한 현상의 생리학적 증거는 뇌파(EEG²⁾ 측정을 통해 실증되었다. Kosmyrna et al.(2025)이 MIT 미디어랩에서 수행한 4개월간의 종단적 실험(N=54)에서, 참가자들은 LLM 사용 집단, 검색 엔진 사용 집단, 도구 미사용 집단으로 나뉘어 에세이 작성 과제를 반복 수행하였다. 그 결과 LLM 사용 집단은 도구 미사용 집단에 비해 뇌의 신경 연결성(neural connectivity) 수준이 유의미하게 낮게 나타났으며, 자신이 생성한 글에 대한 소유감(ownership)과 기억 회상(recall) 능력이 현저히 저하되었다. 나아가 처음 LLM을 사용하던 집단이 이후 도구 없이 스스로 글을 써야 하는 상황(LLM-

1) 비계(scaffold)의 원래 의미는 공사장에서 인부들이 높은 현장에 접근을 돕기 위해 임시로 설치하는 발판 등의 구조물이다. 교육학에서는 학생의 학습을 돕기 위해서 보조적으로 사용하는 교재, 교육방식 등의 지원을 지칭하는 용어로 사용한다. 비계의 특성상 임시적으로 제공하고 점진적으로 없애는 것을 전제로 한다.

2) Electroencephalography (EEG): 전극을 통해 뇌의 전기적 활동(뇌파)을 기록하는 전기생리학적 측정방법을 말한다.

to-Brain 조건)으로 복귀했을 때도 신경 연결성의 활성 저하가 한동안 지속되는 경향이 관찰되었다. 연구자들은 이를 인지 능력의 누적 손실을 뜻하는 ‘인지 부채(cognitive debt)’로 개념화하였다. 다만 이 연구는 아직 동료 심사를 거치지 않은 예비 논문(preprint) 단계이며, 표본 크기(N=54)와 인과 해석의 적절성에 대한 학계의 비판적 의견이 존재한다. 따라서 해당 결과는 AI 수용이 신경생리학적 저하로 직결될 수 있음을 시사하는 강력하기는 하지만 아직은 잠정적인 결과로 이해하는 것이 적절하다. 또한, Li (2025)은 LLM을 ‘인지적 보철물(cognitive prostheses)’로 규정하며, 기억력, 비판적 사고력, 창의성, 동기 측면에서 장기적인 과의존의 위험을 종합적으로 논의하였다.

Jiang et al.(2025)은 EEG 신호를 활용해 LLM과의 상호작용이 주의(attention), 인지 부하(cognitive load), 의사결정 과정에 미치는 영향을 정량적으로 포착하는 프레임워크(Interaction-Aware Language Transformer 및 Interaction-Optimized Reasoning Strategy)를 제안하였다. 이 연구는 잘 설계된 LLM 인터페이스와의 상호작용이 정보처리를 효율화하여 인지 부하를 낮추고 주의 집중도를 높일 수 있음을 보여 주는데, 이는 일견 긍정적으로 해석될 수 있으나, 동시에 인지 부하의 감소가 정보에 대한 능동적 처리(deep processing)의 감소를 동반할 위험이 있다는 점에서 Kosmyna et al.(2025)의 인지 부채 논의와 함께 해석될 필요가 있다. 즉, 인지 부하 저하 자체는 중립적인 현상이지만, 그것이 ‘편안함’ 때문인지 ‘사고의 생략’ 때문인지를 구분하는 것이 향후 연구의 과제로 남는다.

2. 메타인지적 게으름과 비판적 사고의 저하

학습자가 AI를 일상적으로 사용할 때, 문제 해결과 사고의 과정을 덜어내면서 비판적 사고 역량이 저하되는 현상이 실증적으로 관찰되고 있다(Choudhuri et al., 2026). 북미 5개 대학의 과학, 기술, 공학, 수학(STEM) 전공 학생 299명을 대상으로 한 설문 연구

에 따르면, 일상적으로 생성형 AI를 신뢰하고 사용하는 학생일수록 성찰(reflection), 이해에 대한 욕구(need for understanding), 비판적 사고(critical thinking) 수준이 유의미하게 낮아지는 ‘메타인지적 게으름’을 보였다. 역설적으로 기술 친화적 동기(technophilic motivations)와 컴퓨터 자기효능감이 높은 학생일수록, 그리고 AI나 학업 경험이 많은 학생일수록 이러한 경향이 더 심한 것으로 나타났다(Choudhuri et al., 2026). 저자들은 이것이 습관적인 생성형 AI 사용이 점진적으로 지적 습관을 약화시켜 과의존과 사용 확대로 이어지는 ‘인지 부채의 악순환(cognitive debt cycle)’을 형성하기 때문으로 추측한다. 다만 이 연구는 2026년 초 공개된 예비 논문으로, 정식 학술지 게재 이전 단계임을 밝혀 둔다.

이러한 발견은 지식 노동자를 대상으로 한 대규모 조사에서도 확인된다. Lee et al.(2025)은 319명의 지식 노동자를 대상으로 설문을 실시하고, 생성형 AI 활용 과정에서 실제로 936건의 사례를 수집, 분석하였다. 그 결과 과업에 대한 자기 확신(task-specific self-confidence)이 높을수록 비판적 사고가 더 많이 발휘되는 반면, 생성형 AI에 대한 확신(confidence in GenAI)이 높을수록 비판적 사고에 들이는 노력이 감소하는 것으로 나타났다. 이는 AI 산출물에 대한 신뢰가 곧 사용자의 검증 동기를 대체하는 기제로 작동함을 시사하며, 앞서 논의한 ‘인지적 외주화’가 개인의 인지 성향이 아니라 AI에 대한 상대적 신뢰 수준에 의해 조절될 수 있음을 보여준다.

이와 유사한 맥락에서 Gerlich(2025a)는 설문조사 연구(N=666)에서 AI 도구 사용 빈도와 비판적 사고력 사이의 유의한 역의 상관관계를 확인하였으며, 이 관계가 인지적 외주화에 의해 매개됨을 입증하였다. 즉, AI 도구를 많이 사용할수록 비판적 사고력이 약화되는데, ‘인지적 외주화’ 경향이 클수록 이 관계가 더 강해진다는 것이다. 특히 이러한 효과는 AI 산출물에 대한 의존도가 높고 독립적 평가 능력이 상대적으로 낮은 젊은 연령층에서 더 뚜렷하게 나타났다. 이 연구는 상관관계 분

석이라는 방법론적 한계가 존재하기는 하지만, LLM을 포함한 AI 도구에 대한 습관적 의존이 정보를 스스로 평가, 정당화하는 인식적 주체성을 약화시킬 수 있다는 실증적 근거를 제공한다.

3. 시간 단축의 착각(Speedup Illusion)

AI와의 상호작용에서 사용자는 종종 인지적 노력이 줄어든 것을 작업 시간이 단축된 것으로 착각한다(Yu et al., 2026). 이 연구에서는 대규모 (N=1,237) 행동연구를 통해서 사용자가 직접 과제를 수행할 때와 AI의 도움을 받을 때의 실제 작업 완료 시간에는 유의미한 차이가 없었음에도 불구하고, 사용자는 AI가 작업을 훨씬 더 빠르게 해주고 주관적 노력(subjective effort)이 실제로 더 덜 들었다고 응답하였다. 저자들은 이를 ‘시간 단축의 착각(speedup illusion)’으로 명명하였다. 이는 사용자가 시간의 절감과 노력의 절감을 혼동하며, AI의 매끄러운 답변을 수용함으로써 인지적 참여를 스스로 포기하는 결과를 낳는다는 점에서 위에서 논의한 메타인지적 게으름과 비슷한 맥락에 있다. 다만 이 연구도 2026년 발표된 예비 논문 단계로, 후속 검증이 필요하다.

4. ‘주권의 함정(Sovereignty Trap)’과 신념 외주화

IS 및 HCI 관점에서는 AI 산출물에 대한 신뢰와 의존 양상이 인간의 비판적 사고를 무디게 하는 과정을 추적해 왔다. Klein and Klein(2025)은 AI가 제공하는 답변에 손쉽게(frictionless) 접근할 수 있는 환경이 사용자로 하여금 심층 학습에 필수적인 인지 과정을 체계적으로 우회하게 만드는 ‘속 빈 마음(hollowed mind)’ 상태를 유발한다고 주장한다. 이들은 이러한 심리적 기제를 ‘주권의 함정(sovereignty trap)’으로 명명하는데, 이는 AI가 보여주는 권위 있는 유능함이 사용자로 하여금 정보 접근을 실제 능력으로 착각하게 하여 스스로의 지적 판단을 AI에 양도하도록 유혹하는 현상을 가리킨다. 저자들은 절차적, 검색적 정보의 외주화와 LLM 사

용과 같은 통합적 추론의 외주화를 구분해야 한다고 제안하며, 전자는 기존의 계산기, 검색엔진 사용과 유사하게 정당화될 수 있지만, 후자의 경우 도구 의존이 곧 능력의 공동화로 이어질 위험이 크다고 강조한다.

나아가, 사용자가 자신의 신념과 가치관 형성 과정에서 AI에 양도하는 현상도 보고되고 있다. Guingrich et al.(2026)은 ‘신념 외주화(belief offloading)’라는 개념을 제안한다. 이는 신념을 형성하고 유지하는 인지 과정 자체가 AI 시스템에 위임되어, 그 결과가 사용자의 행동과 신념 체계 전반에 영향을 미치는 현상을 말한다. 저자들은 철학, 심리학, 컴퓨터 과학 문헌을 종합하여 신념 외주화가 발생하는 경계 조건에 대한 기술적(descriptive) 분류 체계를 제시하고, 그 규범적 함의를 논의한다.

이러한 우려는 실제 조직 맥락에서도 확인된다. Blanding(2025)은 전문성을 갖춘 실무자조차 AI의 그럴듯하지만 틀린 답변에 설득되어 자신의 판단을 유보하는 사례들을 소개하며, 조직 내 의사결정에서 AI에 대한 ‘과도한 위임(deferring to AI)’이 초래할 수 있는 위험을 경고한다. 이 자료는 동료 심사를 거친 학술 논문이 아니라 실무 지향적 칼럼(practitioner commentary)이라는 점에서 인용 시 유의가 필요하나, 학계의 실증 연구 결과가 실제 조직 현장에서도 재현되고 있음을 시사하는 정황적 근거를 제시한다.

IV. LLM이 의사결정과 창의성에 미치는 영향

1. 알고리즘 신뢰와 의사결정에서의 인지적 관여

IS 분야에서는 LLM을 포함한 알고리즘적 의사결정 보조 도구에 대한 인간의 신뢰와 의존 양상이 오랜 연구 주제였다. Jussupow et al.(2024)은 인간이 알고리즘의 조언을 인간의 조언보다 더 신뢰하는 ‘알고리즘 선호(algorithm appreciation)’와 반대로 더 불신하는 ‘알고리즘 회피(algorithm aversion)’ 현상이 공존하는 이유를 개인 특성, 과업 특성, 알고리즘 특성이라는 세 층

위로 통합하여 설명한다. 이 프레임워크는 LLM 기반 의사결정 지원 도구에 대한 인간의 인지적 관여 수준이 고정적이지 않고 맥락에 따라 체계적으로 달라짐을 보여준다.

Klingbeil et al.(2024)은 실험을 통해, 조언이 단순히 AI에 의해 생성되었다는 사실을 아는 것만으로도 사람들이 해당 조언에 과도하게 의존하는 경향을 보인다는 것을 입증하였다. 참가자들은 AI의 조언이 맥락 정보나 자신의 판단과 상충하는 경우에도 이를 따르는 경향을 보였으며, 이러한 과의존은 실제로 비효율적인 의사결정 결과로 이어졌다. 이는 LLM이 제공하는 답변의 정확성과 무관하게, 'AI가 생성했다'는 출처 표시 자체가 사용자의 비판적 검토 과정을 단축시킬 수 있음을 시사한다.

Stadler et al.(2024)의 실험 연구는 이러한 경향을 학습 맥락에서 정량적으로 뒷받침한다. 이들은 대학생 91명을 ChatGPT-3.5 사용 집단과 구글 검색 사용 집단으로 무작위 배정하여 과학적 탐구 과제(자료 조사 및 추천안 정당화)를 수행하게 하였다. 그 결과 ChatGPT 사용 집단은 인지 부하에서 유의하게 낮은 수준을 보고 하였으나, 최종 추천안에 대한 근거 제시와 추론의 질은 오히려 구글 검색 집단보다 낮았다. 저자들은 이를 '인지적 안락함의 대가(cognitive ease at a cost)'로 명명하며, 인지 부하 감소가 반드시 학습이나 추론의 질적 향상으로 이어지지는 않는다는 점을 실증적으로 보여주었다. 이 결과는 위에서 논의한 메타인지적 게임 및 시간 단축의 착각 개념을 서로 다른 실험 패러다임에서 교차 검증한 결과로 볼 수 있다. 김수정(2026)은 생성형 AI와의 지속적 상호작용이 초래하는 이와 같은 '인지적 경계의 붕괴'를 AI의 '아부(sycophancy)' 성향이 이용자의 자기 서사를 과도하게 강화해 현실 인식과 사회적 판단 능력을 약화시키는 과정으로 설명하기도 한다.

한편 Li et al.(2025)은 'LLM 의존 척도(LLM dependence scale)'를 개발, 검증하였다. 이들은 탐색적, 확인적 요인분석을 통해 의존을 도구적 의존(instrumental dependency, 의사결정과 인지적 과

업에서의 기능적 의존)과 관계적 의존(relationship dependency, LLM을 사회적으로 의미 있는 존재로 인식하는 경향)이라는 두 요인으로 구분하였다. 이 척도는 '보조(assistance)'에서 '의존(reliance)'으로의 이행 과정을 계량적으로 포착할 수 있는 도구를 제공함으로써, 향후 LLM 사용이 인지역량에 미치는 종단적 영향을 측정하는 데 기여하고 있다.

Alavi et al.(2024)은 지식경영(knowledge management) 프레임워크를 생성형 AI 맥락으로 확장하여 생성형 AI가 지식의 창출, 저장, 전달, 적용이라는 지식경영 과정 전반에 미치는 영향을 논의한다. 이들은 생성형 AI가 정보처리력을 향상시키고 개인 및 조직 학습을 촉진할 잠재력을 지니는 동시에, AI 편향(bias)과 인간 간 사회화 감소, 특히 경험이 적은 주니어 지식노동자의 주변화(marginalization)라는 조직적 위험을 수반한다고 지적한다. 이는 위에서 논의한 인지적 외주화의 세대, 숙련도별 비대칭적 영향이 조직 수준에서도 재현될 수 있음을 시사한다.

한편 Schulz and Knierim(2024)은 웨어러블 EEG 장비를 활용하여 생성형 AI 보조 도구가 실제로 지식 노동의 인지 부하를 감소시키는지를 실증적으로 검증하였다. 예비 연구 결과, 머신러닝 모델링 과정을 수행하는 동안 AI 어시스턴트를 사용한 참가자들의 인지 부하는 기존의 인터넷 검색을 사용한 경우와 비교하여 전반적으로 감소하지 않았으며, 오히려 특정 조건에서는 인지 부하가 증가하는 것으로 나타났다. 이 결과는 위에서 논의한 인지 부하 감소 및 시간 단축의 착각에 대한 통념에 중요한 반대 증거를 제공하며, 생성형 AI의 인지적 효과가 과업의 성격과 도구의 설계에 따라 상당히 이질적일 수 있음을 시사한다. 다만 이 연구는 학술대회 발표 단계의 예비 연구(pilot study)로, 표본 규모가 작아 향후 대규모 후속 검증이 필요하다.

2. 공동 창작, 의사결정, 창의성

HCI 분야에서는 LLM과의 상호작용이 사용자의 사

고 내용 자체에 미치는 영향에 주목해 왔으며, Jakesch et al.(2023)은 CHI 2023에서 발표한 대규모 온라인 실험(N=1,506)을 통해, 특정 견해를 더 자주 생성하도록 설계된 LLM 기반 글쓰기 보조 도구를 사용한 참가자들이 실제로 해당 견해 쪽으로 태도가 이동하고, 작성한 글에도 그 편향이 반영됨을 확인하였다. 참가자들은 이러한 영향을 스스로 인식하지 못한 채 도구의 견해를 자신의 것으로 흡수하는 경향을 보였다. 이 연구는 LLM이 단순히 표현을 돕는 도구를 넘어, 사용자의 신념과 태도 형성 과정, 즉 인식적 자율성(epistemic autonomy)에까지 관여할 수 있음을 실증적으로 보여준 대표적 사례로 평가되며, 위에서 논의한 '신념 외주화' 개념의 실증적 연구로 볼 수 있다.

1) 글쓰기 지원과 비계(Scaffolding) 설계

Dhillon et al.(2024)은 현장 실험(N=131)을 통해, 논증적 글쓰기 과제에서 AI 비계 수준(무지원, 문장 단위 제안인 저수준 비계, 문단 단위 제안인 고수준 비계)과 글쓰기 품질, 생산성 사이에 U자형 관계가 존재함을 발견하였다. 저수준 비계는 오히려 품질과 생산성을 저해하였으나, 고수준 비계는 유의한 향상을 가져왔으며 특히 평소 글쓰기에 익숙하지 않거나 기술 친숙도가 낮은 사용자에게 더 큰 이득을 주었다. 이는 비계의 '양'보다 '설계 방식'이 인지적 결과를 좌우함을 보여주는 중요한 증거로, 아래에서 논의할 구조화된 프롬프팅 설계 원칙에 실증적 근거를 제시한다.

한편 Göldi et al.(2024)은 무작위 실험(N=120)을 통해, 동료 평가 글쓰기 과제에서 아이디어 생성과 평가를 돕는 지능형 지원 도구를 제공했을 때 학생들이 도구가 미리 제시한 아이디어나 평가를 그대로 수용하기보다 스스로 선택적으로 활용하는 경향을 보였으며, 지원이 제공되었음에도 글쓰기 자체에는 오히려 더 적극적으로 참여했음을 확인하였다. 이는 위에서 논의한 인지적 외주화 우려에 대한 중요한 반례로서, 도구가 완성된 답을 즉시 제공하기보다 사용자의 관련 인지 과정(예: 아이디어 발상, 평가적 판단)을 '유발'하는 방식으로 설

계될 경우 오히려 인지적 참여가 증가할 수 있음을 시사한다. 이러한 결과는 인지적 영향이 도구 자체가 아니라 설계 방식에 의해 좌우된다는 점을 뒷받침한다.

교육 현장의 실제 경험을 다룬 연구도 존재한다. Park and Ahn(2024)은 질적 연구에서, 한 학기 동안 ChatGPT를 실제로 활용한 대학생 30명과 교수, AI 전문가 등 5명을 대상으로 포커스 그룹 워크숍과 참여적 디자인 세션을 진행하였다. 학생들은 ChatGPT의 대화형 인터페이스와 즉각적 피드백을 긍정적으로 평가하였으나, 동시에 복잡한 개념을 지나치게 단순화하여 받아들이는 경향, 그리고 스스로 검증하지 않고 답을 그대로 수용하는 행태에 대한 우려도 함께 보고되었다. 이 연구는 실제 고등교육 현장에서 나타나는 기회와 위험이 공존함을 실증적으로 보여준다.

2) 의사결정과 적정 신뢰(Appropriate Reliance)

AI 지원 의사결정 맥락에서 기존 연구들은 신뢰 보장의 중요성을 강조한다. Ma et al.(2024)은 세 편의 연속 연구를 통해, 인간의 자기확신(self-confidence)을 보정하는 메커니즘이 AI에 대한 합리적 신뢰를 촉진하고 인간-AI 팀의 의사결정 성과를 향상시킴을 입증하였다. 이는 위에서 논의한 IS 분야의 적정 신뢰 연구(Klingbeil et al., 2024)와 상호보완적으로, 사용자의 메타인지적 자기평가 능력이 AI 의존의 적절성을 좌우하는 핵심 변수임을 보여준다.

Kim et al.(2025)은 LLM이 제공하는 설명(explanation)과 출처(source)가 사용자의 적정 신뢰에 미치는 영향을 정교하게 분석하였다. 설명은 정답과 오답 모두에 대한 의존을 동시에 높이는 반면, 출처 제공은 정답에 대해서는 적정 신뢰를 높이면서도 오답에 대한 과의존은 감소시키는 것으로 나타났다. 흥미롭게도 설명 내에 내적 불일치가 존재할 경우 오히려 오답에 대한 과의존이 줄어들었는데, 이는 완벽하게 매끄러운 설명이 사용자의 비판적 검토를 무력화할 수 있음을 시사한다.

실제 검색 행동에 대한 대규모 실험도 이루어졌다.

Spatharioti et al. (2025)은 LLM 기반 검색을 사용한 참가자들이 기존 검색엔진 사용자보다 더 빠르고 적은 수의 질의로 과제를 완료하며 더 높은 만족도를 보고하였음을 확인하였다. 그러나 LLM이 제공한 정보가 틀린 경우에도 사용자들은 이를 그대로 수용하여 과의존하는 경향을 보였으며, 이는 색상 강조 등 오류 인지를 돕는 인터페이스 개입을 통해서만 완화될 수 있었다.

3) 창의성에 대한 이중 효과

LLM이 인간의 창의적 사고에 미치는 영향은 양가적인 것으로 나타난다. Kumar et al.(2025)은 대규모 사전등록 실험(N=1,100, 두 개의 병렬 실험)을 통해, 확산적 사고와 수렴적 사고라는 창의성의 두 핵심 요소에 대한 LLM 지원의 효과를 검증하였다. 그 결과 LLM의 도움을 받는 동안에는 창의적 수행이 단기적으로 향상되었으나, 이후 LLM 없이 동일한 과제를 수행할 때에는 오히려 독립적 창의 수행 능력이 저해되는 것으로 나타났다. 이는 위에서 논의한 Kosmyna et al.(2025)의 ‘인지 부채’ 개념이 창의성 영역에서도 재현됨을 보여주는 증거이다.

시각적 아이디어 발상 과제를 다룬 Wadinambiarachchi et al.(2024)의 실험(N=60)은 이러한 우려를 뒷받침한다. 이들은 참가자를 AI 이미지 생성기(Midjourney) 지원 조건과 비지원 조건으로 무작위 배정하였는데, AI 지원을 받은 참가자들은 초기 예시에 대한 고착(fixation)이 더 강하게 나타났으며, 산출한 아이디어의 수, 다양성, 독창성 모두에서 기준 집단보다 낮은 성과를 보였다. 질적 분석에 따르면 이러한 효과는 참가자가 AI에 프롬프트를 구성하는 방식과 AI의 제안에 대응하는 전략에 따라 상당히 조절되는 것으로 나타나, 창의성에 대한 LLM의 영향 역시 위에서 논의한 것처럼 설계와 사용 방식에 따라 달라질 수 있음을 시사한다.

이상의 결과는 앞서 논의한 Kosmyna et al.(2025)의 소유감 저하 결과 및 Klein and Klein(2025)의 속 빈 마음 개념과 함께, LLM과의 공동 작업이 산출물의 질적 향상과는 별개로 사용자의 주체적 사고 과정

을 잠식할 수 있다는 점을 보여준다. 그러나 Göldi et al.(2024)의 연구가 보여주듯, 이러한 잠식 효과는 보편적이거나 불가피한 것이 아니라 도구의 설계 방식—비계의 수준, 설명과 출처의 제시 방식, 오류 가시성—에 따라 상당히 조절될 수 있다. 요컨대 위에서 리뷰한 신경과학, 교육, 심리학, IS, HCI 등의 서로 다른 학문분야의 연구가 모두 유사한 결론, 즉 ‘검토와 정당화 과정을 생략한 AI 의존은 인지적 자율성을 잠식하지만, 이는 설계적 개입을 통해 완화될 수 있다’는 결론으로 수렴하고 있다는 점이 핵심 시사점이다.

앞의 문헌연구를 종합하면 몇 가지 공통된 패턴이 드러난다. 첫째, IS, 심리학, HCI, 신경과학, 교육학 모두 LLM의 영향은 사용자가 ‘전략적 활용’과 ‘무비판적 의존’ 중 어느 쪽에 가까우가에 따라 상반된 결과를 낳는다는 점이다(Gerlich, 2025a, 2025b; Jussupow et al., 2024; Klein & Klein, 2025). 둘째, 다수의 연구가 인지적 영향의 핵심 기제로 ‘검토와 정당화 과정의 생략’을 지목한다는 점에서, LLM에 대한 과의존은 단순 능력 저하가 아니라 인지적 노력을 기울이는 습관 자체의 변화로 이해될 필요가 있다. 셋째, 본 리뷰가 통합한 개념들, 즉 인지 부채, 주권의 함정, 속 빈 마음, 메타 인지적 게으름, 시간 단축의 착각, 신념 외주화는 서로 독립적인 현상이라기보다 ‘검토 생략→신뢰 과잉→스키마 미형성→자율성 상실’로 이어지는 하나의 연쇄적 과정을 서로 다른 각도에서 포착한 것으로 재해석할 수 있다.

이상의 문헌연구결과를 종합하면 <표 1>과 같다.

V. 인지 저하 극복 및 인지적 주권 회복을 위한 개선 방안

이러한 부작용에 대처하기 위해, 개입 전략을 제시한다. 인지 저하를 극복하기 위해서는 AI 시스템 자체의 설계를 바꿔야 할 필요도 있고, 또한 AI를 사용하는 사용자, 특히 인지역량의 형성을 주도하는 교육과정의 변화가 필요하다고 생각된다. 따라서, 개입 전략은 AI 시

〈표 1〉 LLM의 사용과 인지역량의 관계에 대한 연구 요약

연구 카테고리	독립변수	종속변수(DV)	결과 요약	논문
신경인지적 지표 (뇌연결성·인지부하)	도구 유형(LLM/검색엔진/미사용, 4개월 반복); LLM 인터페이스 설계·상호작용 방식	신경 연결성(EEG), 소유감, 기억 회상; 주의, 인지 부하, 의사 결정 과정	LLM 사용 집단은 신경 연결성이 최저이고 소유감·회상이 저하되며, 도구 제거 후에도 저하가 지속('인지 부채'). EEG 기반 정량 프레임워크는 잘 설계된 인터페이스가 인지 부하를 낮추고 주의를 높일 수 있음을 보여주지만, 능동적(deep) 정보처리 감소 위험도 공존한다는 양가적 시사점을 제공	Kosmyna et al., 2025; Jiang et al., 2025
	도구 유형(ChatGPT-3.5 vs 구글 검색; AI 어시스턴트 vs 인터넷 검색)	인지 부하(외재적·내재적·본유적, EEG), 추론·근거제시의 질	두 연구 모두 AI 도구 사용이 기대와 달리 인지 부하를 유의하게 낮추지 못하거나(일부 조건에서 오히려 증가), 낮아진 경우에도 추론의 질은 오히려 저하됨('인지적 안락함의 대가')	Stadler et al., 2024; Schulz and Knierim, 2024
비판적 사고력·메타인지	AI 신뢰·사용빈도, 기술 친화적 동기/자기효능감, 과업 자기확신 vs AI 확산, AI 지원 여부	성찰·이해욕구·비판적 사고('메타인지적 게으름'), 비판적 사고 발휘 노력, 비판적 사고력, 지각된 시간/노력('시간 단축의 착각')	네 편의 연구가 수렴적으로, AI에 대한 신뢰·사용빈도가 높을수록 비판적 사고에 들이는 노력이 감소함을 보여줌. 특히 자기효능감이 높거나 AI가 실제로보다 빠르다고 착각하는 사용자일수록 검증 없이 수용하는 경향이 강해 '인지 부채의 악순환'을 형성	Choudhuri et al., 2026; Lee et al., 2025; Gerlich, 2025a; Yu et al., 2026
	LLM 의존 전반(개념적 논의)	기억력, 비판적 사고력, 창의성, 동기	LLM을 '인지적 보철물(cognitive prostheses)'로 규정하고, 장기 과의존의 다차원적 위험을 종합 논의	Li, 2025
신뢰·의사결정 (적정의존)	조언 출처 표지(AI/인간), 자기확신 보정 메커니즘, 설명·출처 제공 여부·불일치, 검색도구 유형(LLM기반/기존)	의존도·과의존, 의사 결정 효율성/성과, 적정 신뢰 수준, 결정 속도·정확도	네 편을 종합하면, 'AI가 생성했다'는 표지 자체는 맹목적과의존을 유발하지만, 설명·출처의 제시 방식·사용자의 자기확신 보정·인터페이스 상 오류 가시성(색상강조)에 따라과의존이 완화될 수 있음이 확인됨	Klingbeil et al., 2024; Ma et al., 2024; Kim et al., 2025; Spatharioti et al., 2025
	개인·과업·알고리즘 특성(알고리즘 수용); 생성형 AI의 지식경영 통합 정도	알고리즘 선호/회피(인지적 관여 수준); 지식 창출·저장·전달·적용, 조직적 위험	두 편 모두 개념적 통합 리뷰로서, AI에 대한 신뢰·관여 수준이 맥락에 따라 체계적으로 달라진다는 점과, 생성형 AI가 조직 학습을 촉진하는 동시에 주니어 인력 주변화 같은 위험을 수반한다는 점을 이론적으로 논의	Jussupow et al., 2024; Alavi et al., 2024
신념·인지적 자율성·정서적의존	AI의 손쉬움(frictionless) 정보 접근; 신념 형성 과정의 AI 위임; 생성형 AI와의 지속적 상호작용('관계화')	'속 빈 마름'·주권 상실(개념적); 신념 체계·행동('신념 외주화'); 인지적 경제·정서적 웰빙·소통역량	세 편의 이론 논문이 서로 다른 개념(주권의 함정, 신념 외주화, 관계화)으로 동일한 현상—AI에 대한 손쉬운 접근이 사용자의 독립적 판단·신념형성·현실인식 능력을 잠식하는 과정—을 각기 다른 각도에서 포착	Klein and Klein, 2025; Guingrich et al., 2026; 김수정, 2026

〈표 1〉 계속

연구 카테고리	독립변수	종속변수(DV)	결과 요약	논문
신념·인지적 자율성· 정서적의존	편향되도록 설계된 AI 글쓰기 도구; AI 답변의 유창함·그럴듯함	태도·의견 변화; 판단 유보·과도한 위임	두 연구 모두 위 이론적 개념들이 실제로 나타남을 보여줌—대규모 실험에서 사용자는 AI가 유도한 견해를 무의식적으로 흡수했고, 전문 실무자조차 그럴듯한 오답에 판단을 유보한 사례가 다수 보고됨	Jakesch et al., 2023; Blanding, 2025
	(척도 개발 연구)	도구적 의존 (instrumental) / 관계적 의존 (relationship)	‘보조 → 의존’ 이행 과정을 계량적으로 포착하는 2요인 구조 척도를 개발·검증	Li et al., 2025
창의성· 글쓰기 지원	AI 비계 수준(무지원/ 저수준 비계/고수준 비계); 지능형 지원 도구 제공 여부	글쓰기 품질, 생산성; 글쓰기 참여도 (engagement)	두 연구 모두 인지적 외주화가 보편적이지 않음을 보여주는 반례로, 지원의 ‘양’이 아니라 ‘설계 방식’ 에 따라 결과가 갈림—적절히 설계된 고수준 비계· 유발형 지원은 오히려 품질과 참여를 높임	Dhillon et al., 2024; Göldi et al., 2024
	LLM 지원 여부 (사용 중/사용 이후); AI 이미지 생성기 지원 여부	확산적·수렴적 사고 (창의성); 디자인 고착 (fixation), 아이디어 수·다양성·독창성	두 연구 모두 AI 지원이 창의적 산출물의 단기적 질은 높이지만, 초기 예시에 대한 고착을 강화하고 이후 독립적 창의 수행 능력을 저해함을 실증— ‘인지 부재’가 창의성 영역에서도 재현됨	Kumar et al., 2025; Wadinambiarachchi et al., 2024
	ChatGPT 실제 활용 경험(한 학기)	인식(기회·위험 지각)	대화형 인터페이스·즉각 피드백은 긍정 평가; 단순화·미검증 수용에 대한 우려도 동시 보고	Park and Ahn, 2024

시스템 개발의 관점과 AI를 사용하는 교육에서의 전략으로 나누어서 살펴본다.

1. AI 시스템 설계에 대한 시사점

1) 구조화된 프롬프팅과 비계(Scaffolding)의 점진적 제거

학습자가 AI에 최초 질문으로 완성된 출력값을 기대하는 단방향 의존성에서 벗어나도록 유도해야 한다. AI가 사용자로 하여금 해결 과정을 직접 도출하게 하거나 반론 제기 및 가설 검증을 제안하도록 설계하는 ‘구조화된 프롬프팅(structured prompting)’ 기법의 활용을 일상화할 필요가 있다. Gerlich(2025b)에 따르면, 이러한 통제된 방식의 대화식 활용은 인지적 외주화를

현저히 경감시키고, 사용자의 성찰적 몰입(reflective engagement)을 극대화하는 것으로 나타났다. 이는 동일 저자의 상관연구(Gerlich, 2025a)에서 제기된 우려에 대해 실험적으로 통제 가능한 해법을 제시했다는 점에서 의미가 있다.

또한, 교육 측면에서 AI는 항구적인 도구가 아닌 임시 지원 장치로서의 비계(scaffold)로 상정되어야 하며, 학습자의 주체적 이해도가 높아질수록 개입 빈도를 낮춰 독립적 사고를 하도록 디자인되어야 한다. Vendrell and Johnston(2026)은 고등교육 맥락에서 생성형 AI를 통합할 때 지켜야 할 설계 원칙들을 제시하며, 특히 비계의 수준을 학습자의 숙련도에 따라 동적으로 조정하는 ‘점진적 해체(fading)’ 원리를 강조한다.

2) 메타인지 프롬프트(Metacognitive Prompts)의 시스템적 개입

대학생들을 대상으로 한 실험에서 AI 검색 과정 중 학습자의 성찰을 유도하는 ‘메타인지 프롬프트’(이해 점검, 다양한 관점 고려, 통합 및 성찰 등)를 제공했을 때, 학생들은 더 다양한 주제를 탐색하고 심층적인 후속 질문(critical queries)을 던지는 등 비판적 사고가 촉진되었다. Singh et al. (2025)은 40명 규모의 사용자 연구를 통해 단순히 AI의 답을 수동적으로 수용하는 대신, AI의 한계와 편향성을 평가하고 누락된 관점을 파악하도록 돕는 시스템적 개입이 효과적임을 보였다. 이는 인터페이스 설계 수준에서 비판적 사고를 할 수 있도록 ‘기본값(default)’을 조정할 필요가 있을 수도 있음을 시사한다.

3) 재생적 경험과 인지적 주권(Cognitive Sovereignty)의 확립

종래의 기술 도입이 주로 생산성 및 속도 ‘최적화(optimization)’ 패러다임 하에 진행되었는데, AI에서는 이와 같은 패러다임은 인간 인지의 위축이라는 역풍을 낳고 있다. 향후 시스템 설계는 인공지능이 인간의 두뇌를 지적 퇴화로부터 지키고, 뇌의 신경 가소성과 역동적 유연성을 오히려 회복해 주는 ‘재생적 경험(regenerative experience, RX)’ 패러다임으로 진화해야 한다(Gee, 2025). Gee(2025)는 AI 시스템 설계 목표를 효율성 극대화에서 인간의 인지적, 건강적 재생으로 전환할 것을 촉구하며 이를 뒷받침하는 윤리적 거버넌스와 리더십의 필요성을 논의한다. 이 자료는 심사 이전의 예비 논문이라는 한계가 있으나, 본 논문이 검토한 여러 실증 연구들이 공통적으로 지향하는 설계 철학을 개념적으로 통합하고 있다는 점에서 논의의 가치를 지닌다.

궁극적으로 사용자는 AI 출력물의 매끄러운 수식어 이면을 끊임없이 의심하고 교차 검증을 실행하는 인지적 긴장감을 품어야 하고, 결과물의 최종 수용 및 책임 귀속은 오로지 인간 주체에 놓인다는 ‘인지적 주권

(cognitive sovereignty)’의 가치관을 함양해야 한다(Klein & Klein, 2025). 만일 이를 개개인이 하기 어렵다면 시스템의 설계나 제도에서도 이를 지원하도록 할 필요가 있을 것이다.

2. 일반 교육에서의 시사점

1) ‘선(先) 수동 학습(Manual First)’과 ‘AI-Free Zones’의 확립

두뇌 장기기억 시스템에 확실한 개념적 스키마를 형성하기 위해서는 일련의 인지적 노이즈와 마찰, 즉 ‘바람직한 어려움(desirable difficulties)’이 내재되어야 한다(최승락, 2025). 이를 달성하기 위해, 기본 개념 학습 단계에서는 생성형 AI 활용을 철저히 배제하고 개념을 체화하는 ‘선(先) 수동 학습’ 기간이 보장되어야 한다. 이는 위에서 논의한 Klein and Klein(2025)의 절차적 외주화와 통합적 추론 외주화의 구분과 직접적으로 연결되는 설계 원칙으로, 통합적 추론 능력이 아직 형성되지 않은 학습 초기 단계에서는 AI 개입을 원천적으로 차단해야 함을 의미한다.

아울러, 전체 업무 혹은 학습 과정에서 과업 기획 단계, 핵심 아이디어 브레인스토밍, 최종 논리 검증 단계 등 중추적 고등 인지 영역에서는 생성형 AI 사용을 차단하는 ‘AI 없는 구역(AI-Free Zones)’을 제도화함으로써 사용자의 인지적 자율성을 제도적으로 사수해야 한다. 이러한 제안은 국내 문헌에서도 확인되는데, 최승락(2025)은 생성형 AI 시대의 교육적 활용 방안을 다룬 논의에서 인지적 외주화의 위험을 완화하기 위한 교육 현장 차원의 실천적 지침을 제시한 바 있다. 다만 이 자료는 소속 기관에서 발간한 보고서 성격임을 고려할 필요가 있다.

2) 과정 중심 평가와 비교 판단(Comparative Judgment) 훈련

AI가 결과물의 유창함과 형식적 그럴싸함을 정교하게 흉내 낼 수 있는 환경에서는 기존의 결과물 중심 평

가 모델이 무력화된다. 따라서 사용자가 결과물을 얻기 위해 AI와 프롬프트 수준에서 주고받은 맥락적 상호작용 및 교정 과정을 평가하는 과정 중심 평가 체제로 전환해야 한다. Martínez-Peláez et al.(2025)은 LLM이 고등교육에서 능동적 학습과 맞춤형 피드백을 촉진할 잠재력을 지니는 동시에, 평가 방식이 결과물이 아닌 과정 중심으로 재설계될 때에만 비판적 사고와 문제 해결 역량 함양에 실질적으로 기여할 수 있다고 주장한다. 이 연구는 LLM과 인지능력에 대한 중요한 시사점을 제공하지만 LLM이 교육환경에서 사용되는 경우에 초점을 맞추고 있다는 점에서 본 논문과의 차이점이 존재한다.

특히 학습자에게 고의로 오류나 할루시네이션(환각), 논리적 약점이 섞인 AI 다중 출력 답변들을 제시한 후 직접 결함을 추출하고 서열을 책정하게 만드는 '비교 판단(comparative judgment)' 훈련은 인지적 나태성(lazy thinking)을 방지하고 메타인지 역량을 환기시키는데 효과적인 것으로 나타났다(House, 2025). 이는 채점 및 평가 플랫폼을 운영하는 실무 기관의 블로그 자료에서 제안된 방법론으로, 심사를 거친 학술적 근거는 충분하지 않으나, 앞에서 언급된 메타인지 프롬프팅 연구(Singh et al., 2025) 및 구조화된 프롬프팅 연구(Gerlich, 2025b)와 개념적으로 일치한다.

VI. 결론 및 향후 연구 방향

본 리뷰는 다학제적 통합 고찰을 통해 생성형 AI가 인간의 인지역량에 가하는 잠재적 위험 요인과 이를 상쇄하기 위한 설계 원칙들을 종합하였다. LLM은 올바른 제어 장치와 메커니즘 하에 활용될 경우 생산성 제고와 지적 협업을 돕는 '확장 도구'로 유용하지만, 무비판적이고 습관적인 외주화 수단으로 방치된다면 '인지 부채'의 가속과 '주권의 함정'을 이끌며 인지 생태계를 황폐화할 위험을 지닌다. 결국, AI는 인간 사유의 대체제가 될 수 없으며, 구조화된 프롬프팅, 메타인지 프롬프트, 선 수동 학습과 AI 없는 구역, 과정 중심 평가, 그리고 재생적 인텔리전스 패러다임과 같은 제도적 가이드

라인과 설계적 마찰이 뒷받침될 때 비로소 인간과 AI 간의 진정한 인지적 파트너십이 실현될 수 있을 것이다.

향후 연구는 다음과 같은 방향성을 가져야 할 것이다. 첫째, Li et al. (2025)이 수립한 'LLM 의존 척도'와 같은 검증된 측정 도구의 범용성을 넓혀, 실제 외주화 행위가 기억 및 추론 수준에 미치는 장기 종단적 영향의 인과적 고리를 밝혀야 한다. 둘째, 다양한 과업 맥락(정량적 분석, 감성적 글쓰기, 사회적 상호작용 등)에 따라 AI 도구의 수용 수준과 인지 저하 폭이 상이하게 일어나는지에 대한 광범위하고 심도 있는 연구가 이루어져야 한다. 셋째, Heersmink(2024) 및 Klein and Klein(2025)이 주창한 절차적 외주화와 통합적 추론 외주화의 구분, 그리고 최승락(2025) 등이 제안한 점진적 비계 구조와 'AI 없는 구역'의 실제적 유효성을 비교 행동 분석 및 뇌 활동 측정을 통해 체계적으로 검증해야 할 것이다. 넷째, 현재 가설적으로 제안되고 있는 인지 부채, 시간 단축의 착각, 신념 외주화 등의 핵심 개념들이 심사를 거친 대규모 반복 연구를 통해 재현되는지 확인하는 것이 학문적 신뢰성 확보를 위한 선결 과제이다.

■ 참고문헌

- 김수정 (2026). 생성형 AI의 일상화와 인간-인공지능 관계의 재구성: 인공지능의 '관계화'가 제기하는 새로운 정책 쟁점. <정보화정책>, 33권 2호, 3-28. <https://doi.org/10.22693/NIAIP.2026.33.2.003>
- 장창기 (2025). 인공지능 기반 서비스 이용에 있어서 개인의 정보평가 성향의 영향: 프라이버시 염려 및 보호 행동과의 조절된 매개효과 분석. <정보화정책>, 32권 3호, 21-42. <https://doi.org/10.22693/NIAIP.2025.32.3.021>
- 정동훈 (2026). 에이전틱 AI 시대의 인간-기계 상호작용 재구성에 관한 이론적 고찰. <정보화정책>, 33권 1호, 3-24. <https://doi.org/10.22693/NIAIP.2026.33.1.003>
- 최승락 (2025). <생성형 AI 시대, AI를 어떻게 활용할 것인가?: 인지적 외주화의 위험과 교육적 해법>. 한림대학교 인문사회융합연구원 제1회 AI 융합교육 콜로키움

- 발표자료. [비(非) 동료심사 콜로키움 발표문]
- Alavi, M., Leidner, D. E. & Mousavi, R. (2024). A knowledge management perspective of generative artificial intelligence. *Journal of the Association for Information Systems*, 25(1), 1-12. <https://doi.org/10.17705/1jais.00859>
- Blanding, M. (2025). *The dangers of deferring to AI: It seems so right even when it's wrong*. Harvard Business School Working Knowledge.
- Choudhuri, R., Sanchez, C. A., Burnett, M. & Sarma, A. (2026). Thinking less, trusting more: GenAI's impacts on students' cognitive habits (arXiv: 2601.22430) [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2601.22430>
- Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S. & Robert, L. P. (2024). *Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models*. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). ACM. <https://doi.org/10.1145/3613904.3642134>
- Gee, R. O. W. (2025). Stewarding regenerative experience (RX): From AI-driven efficiency to human cognitive and health regeneration [Preprint]. Preprints.org.
- Gerlich, M. (2025a). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Gerlich, M. (2025b). From offloading to engagement: An experimental study on structured prompting and critical reasoning with generative AI. *Data*, 10(11), 172. <https://doi.org/10.3390/data10110172>
- Göldi, A., Wambsganss, T., Neshaei, S. P. & Rietsche, R. (2024). *Intelligent support engages writers through relevant cognitive processes*. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). ACM. <https://doi.org/10.1145/3613904.3642549>
- Guinrich, R. E., Mehta, D. & Bhatt, U. (2026). Belief offloading in human-AI interaction (arXiv: 2602.08754) [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2602.08754>
- Heersmink, R. (2024). Use of large language models might affect our cognitive skills. *Nature Human Behaviour*, 8, 805-806. <https://doi.org/10.1038/s41562-024-01859-y>
- House, M. (2025). Cognitive offloading is real and new learners are most at risk: How RM Compare keeps the thinking human. RM Compare Blog.
- Jakesch, M., Bhat, A., Buschek, D., Zalmanson, L. & Naaman, M. (2023). *Co-writing with opinionated language models affects users' views*. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), Article 111. ACM. <https://doi.org/10.1145/3544548.3581196>
- Jiang, T., Wu, J. & Leung, S. C. H. (2025). The cognitive impacts of large language model interactions on problem solving and decision making using EEG analysis. *Frontiers in Computational Neuroscience*, 19, 1556483. <https://doi.org/10.3389/fncom.2025.1556483>
- Jussupow, E., Benbasat, I. & Heinzl, A. (2024). An integrative perspective on algorithm aversion and appreciation in decision-making. *MIS Quarterly*, 48(4), 1575-1590. <https://doi.org/10.25300/MISQ/2024/18512>
- Kim, S. S. Y., Vaughan, J. W., Liao, Q. V., Lombrozo, T. & Russakovsky, O. (2025). *Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies*. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). ACM. <https://doi.org/10.1145/3706598.3714020>
- Klein, C. R. & Klein, R. (2025). The extended hollowed mind: Why foundational knowledge is indispensable in the age of AI. *Frontiers in Artificial Intelligence*, 8, 1719019. <https://doi.org/10.3389/frai.2025.1719019>
- Klingbeil, A., Grützner, C. & Schreck, P. (2024). Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160, 108352. <https://doi.org/10.1016/j.chb.2024.108352>

- Kosmyrna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I. & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task (arXiv:2506.08872) [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2506.08872>
- Kumar, H., Vincentius, J., Jordan, E. & Anderson, A. (2025). *Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking*. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). ACM. <https://doi.org/10.1145/3706598.3714198>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. & Wilson, N. (2025). *The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers*. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). ACM.
- Li, B. (2025). Study on the long-term cognitive effects of excessive LLM use among students and general users. *Journal of Contemporary Educational Research*, 9(10), 203-214. <https://doi.org/10.26689/jcer.v9i10.12112>
- Li, Z., Zhang, Z., Wang, M. & Wu, Q. (2025). From assistance to reliance: Development and validation of the large language model dependence scale. *International Journal of Information Management*, 83, 102888. <https://doi.org/10.1016/j.ijinfomgt.2025.102888>
- Ma, S., Wang, X., Lei, Y., Shi, C., Yin, M. & Ma, X. (2024). "Are you really sure?" Understanding the effects of human self-confidence calibration in AI-assisted decision making. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). ACM. <https://doi.org/10.1145/3613904.3642671>
- Martínez-Peláez, R., Mena, L. J., Toral-Cruz, H., Ochoa-Brust, A., Gonzalez Potes, A., Flores, V., Ostos, R., Ramirez Pacheco, J. C., Felix, R. A. & Felix, V. G. (2025). Can large language models foster critical thinking, teamwork, and problem-solving skills in higher education?: A literature review. *Systems*, 13(11), 1013. <https://doi.org/10.3390/systems13111013>
- Park, H. & Ahn, D. (2024). *The promise and peril of ChatGPT in higher education: Opportunities, challenges, and design implications*. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). ACM. <https://doi.org/10.1145/3613904.3642785>
- Schulz, T. & Knierim, M. T. (2024). *Cognitive load dynamics in generative AI-assistance: A NeuroIS study*. Proceedings of the 45th International Conference on Information Systems (ICIS 2024). Association for Information Systems.
- Singh, A., Guan, Z. & Rieh, S. Y. (2025). *Enhancing critical thinking in generative AI search with metacognitive prompts*. Proceedings of the Association for Information Science and Technology, 62(1). <https://doi.org/10.1002/pra2.1287>
- Sparrow, B., Liu, J. & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776-778. <https://doi.org/10.1126/science.1207745>
- Spatharioti, S. E., Rothschild, D. M., Goldstein, D. G. & Hofman, J. M. (2025). *Effects of LLM-based search on decision making: Speed, accuracy, and overreliance*. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25). ACM. <https://doi.org/10.1145/3706598.3714082>
- Stadler, M., Bannert, M. & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160, 108386. <https://doi.org/10.1016/j.chb.2024.108386>
- Vendrell, M. & Johnston, S.-K. (2026). Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. *Computers and Education:*

Artificial Intelligence, 10, 100572. <https://doi.org/10.1016/j.caeai.2026.100572>

- Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q. & Velloso, E. (2024). *The effects of generative AI on design fixation and divergent thinking*. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). ACM. <https://doi.org/10.1145/3613904.3642919>
- Yu, S., Cheng, M., Jabbar, A., Sucholutsky, I., Collins, K. M., Jurafsky, D. & Hawkins, R. D. (2026). Cognitive offloading and the speedup illusion in human-AI interaction (arXiv:2605.23177) [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2605.23177>