대한내과학회지: 제100권 제5호 2025 eISSN 2289-0769

Perspectives https://doi.org/10.3904/kjm.2025.100.5.197

거대 언어 모델 시대 연구자의 비판적 사고 보존 전략: 효율성 뒤에 숨겨진 인지 부하의 역설

인제대학교 의과대학 [1]약리학교실, [2]약물유전체연구센터, [3]심혈관대사질환센터

안상진[1,2,3]

Preserving Critical Thinking in the Age of Large Language Models: The Paradox of Cognitive Load and Efficiency

Sangzin Ahn[1,2,3]

[1]Department of Pharmacology, [2]Pharmacogenomics Research Center, [3]Cardiovascular and Metabolic Diseases Medical Research Center, Inje University College of Medicine, Busan, Korea

Rapid advancements in large language models (LLMs) have fundamentally transformed research practices across academic disciplines, with considerable adoption rates among researchers. While empirical studies have demonstrated the substantial positive effects of LLMs on learning outcomes and cognitive performance, these technological advances present a paradoxical challenge in maintaining critical thinking capabilities. LLMs offer unprecedented efficiency in research tasks in literature reviews, analysis, and writing, by significantly reducing task completion time while improving output quality. However, this efficiency stems largely from cognitive offloading, which is the delegation of mental processes to external systems, raising concerns about the potential weakening of human analytical abilities. Cognitive load theory provides a framework for distinguishing between the beneficial reduction of unnecessary cognitive burden and problematic offloading of essential cognitive processes required for deep understanding and critical analysis. Experimental evidence suggests that while LLM users experience reduced cognitive load across multiple dimensions, their critical reasoning performance may suffer compared with traditional search methods. The fundamental challenge lies in balancing the efficiency gains of LLM integration while preserving rigorous analytical thinking. Medical researchers must develop strategic approaches that leverage LLM capabilities while maintaining active engagement with primary sources and complex reasoning tasks. Success lies in recognizing that traditional research methods may represent essential investments in preserving critical thinking skills, suggesting LLM integration involves selective application rather than wholesale cognitive offloading. **(Korean J Med 2025;100:197-200)**

Key Words: Large language models; Thinking; Learning; Artificial intelligence; Workload

서 론

2022년 11월 ChatGPT 공개 이후 약 2년 반 동안 거대 언어 모델(large language model, LLM)은 빠르게 발전하며 연구 현장에 혁명적인 변화를 가져왔다. 2023년 말 의학 분야 상위 15개 학술지 교신저자 236명을 대상으로 한 설문조사에 따르면 LLM 사용법 교육 경험이 1% 미만임에도 불구하고 24%가 논문 작성에 LLM을 활용하고 있다고 답하였으며 79%는 앞으

Received: 2025. 6. 10 **Revised:** 2025. 6. 25 **Accepted:** 2025. 6. 27

Correspondence to:
Sangzin Ahn, M.D., Ph.D.
Department of Pharmacology, Inje University College of Medicine, 75 Bokji-ro, Busanjin-gu, Busan 47392, Korea
Tel: +82-51-890-5909, Fax: +82-893-1232, E-mail: sangzinahn@inje.ac.kr

로 연구에 LLM이 지대한 역할을 할 것이라고 예상하였다[1]. LLM의 교육적 효과에 대하여 다양한 맥락에서 관찰한 실증적 연구들이 다수 발표되었고 2025년 5월에는 51개의 정량적 교육 분야 연구에 대한 메타 분석이 발표되기도 하였다[2]. 이 분석에 따르면 LLM은 학습 성과를 향상시키는 데 통계적으로 유의미하고 큰 효과(g = 0.867)를 보였으며 고차원적 사고를 향상시키는 데에도 유의미한 중간 크기의 효과(g = 0.487)가 관찰되었다.

그러나 이러한 기술 혁신은 의학 연구자들에게 근본적인 고민거리를 제시한다. LLM이 문헌 종합, 데이터 분석, 문서 작성 등 업무에 전례 없는 편리와 효율을 제공하지만 이러한 효율성은 인간의 인지적 과정을 외부 시스템에 위임하는 인지적 오프로딩에 상당 부분 기반하기 때문에 오히려 인간의 인지적 추론 능력이 약화될 수 있다는 역설적인 우려가 제기되고 있다. 어렵게 공부한 것이 오래 기억에 남는 것처럼 LLM을 통해 업무를 쉽게 처리할수록 근원적인 비판적 사고와 문제 해결 능력이 약화될 수 있다는 것이다. 본 논문에서는 LLM의 이러한 이중적 특성을 심층적으로 검토하고 의학 연구자로서 비판적 사고를 보존하기 위한 전략을 모색하고자 한다.

본 론

글쓰기 업무의 효율성 증진 효과

LLM은 글쓰기 관련 영역에서 탁월한 효율성을 보여준다. 459명의 독일 고등학생을 대상으로 한 무작위 대조군 실험에서 학술적 글쓰기 과제에 LLM의 피드백을 받은 학생들이 성취도, 동기 부여, 긍정적 감정 측면에서 유의미한 개선을 보였다[3]. 다양한 분야의 전문가 453명을 대상으로 한 실험에서도 LLM을 사용하였을 때 글쓰기 형태의 업무 시간이 40% 감소하고 품질은 18% 증가하는 결과가 나타났다[4]. 자문 전문가 758명을 대상으로 한 무작위 대조군 실험에서도 업무 처리 속도가 25% 증가하고 결과물의 품질이 40% 증가하는 유사한 결과가 관찰되었다[5]. 이처럼 글쓰기와 관련된 업무에서 LLM의 생산성 및 효율성 증진 효과는 일관되고 설득력 있게 나타나고 있다.

인지 부하의 역설

LLM이 업무를 효율화하고 특히 인지 부하를 줄여주는 명백한 이점에도 불구하고 지나친 의존에 대한 걱정이 공존하며 이는 근거 없는 우려가 아니다. 91명의 독일 대학생을 대상으로 자료를 찾아 과학적 논증을 구성하는 과제를 부여하였을 때 LLM을 활용한 집단은 내재적, 외재적, 본유적(본질적) 인지 부하가 모두 유의미하게 낮다고 보고하였지만 검색 엔진을 활용하여 자료를 찾은 학생들보다 논증 타당도가 유의미하게 낮은 평가를 받았다[6]. 매개 분석을 통해 세 가지 인지 부하 중 본유적 인지 부하(germane cognitive load)가 추론의 품질에 영향을 미친다는 연관성이 밝혀졌다.

인지 부하 이론에 따르면 내재적 인지 부하는 학습 내용 자체의 난이도와 복잡성으로 인해 발생하며 학습 내용을 잘 분류하고 단계적으로 소화함으로써 관리해야 할 대상이다. 반면 외재적 인지 부하는 학습 내용이 학습자에게 전달되는 과정에서 발생하는 불필요한 부하로 최소화해야 할 대상으로 간주된다. LLM을 적절히 활용하면 이 두 가지 인지 부하를 줄여 작업과 학습의 효율을 높일 수 있다. 그러나 학습 내용을 기존의 지식과 통합해 스키마를 구성하는 데 들어가는 노력인 본유적 인지 부하까지 인지적 오프로딩 형태로 LLM에게 맡겨버리면 이해의 깊이가 얕아지고 비판적 사고 함양이 저해될 수 있다.

인지 부하 조절을 통한 학습 효과 증진 사례

LLM이 학습자의 본유적 인지 부하를 감소시켜 깊이 있는 학습을 저해할 수 있다는 우려와 달리 외재적 인지 부하를 줄여 학습을 촉진하는 긍정적 결과도 있다. Huang 등[7]이 187명의 치과대학 학생을 대상으로 진행한 무작위 대조군 연구가 좋은 사례이다. 이 연구에서 실험군은 실기 교육 영상을 시청하며 LLM을 보조 학습 도구로 활용하였고 대조군은 영상 자료만으로 학습하였다. 실험군 학생들은 이해하기 어렵거나 불명확한 부분을 즉시 챗봇에 질문하여 해답을 얻음으로써 외재적 인지 부하를 줄였고 그 결과 시험에서 유의미하게 높은 점수를 기록했다. 이는 LLM이 외재적 인지 부하를 줄여준 덕분에 학생들이 학습 내용 자체에 더 많은 정신적 자원(본유적 인지 부하)을

투입하여 학습 효과를 높인 것으로 해석할 수 있다.

연구자가 경계해야 할 사항

LLM의 이러한 잠재력에도 불구하고 의학과 같이 정확도가 중요한 분야에서는 적용에 신중을 기해야 한다. 가장 큰 우려는 LLM이 때때로 부정확하거나 불완전한 정보, 이른바 환각을 생성할 수 있다는 점이다. 이러한 가능성을 염두에 두고 인간은 비판적인 감시자 역할을 수행해야 한다. Stadler 등[6]은 검색 엔진 활용 시 사용자가 다양한 검색 결과에서 상충하는 정보를 분석하고 비판적으로 수용하는 과정을 통해 비판적 사고 훈련이 이루어진다고 보았는데 이는 LLM 사용 시 어떤 자세를 견지해야 할지에 대한 통찰을 준다. 최근 다양한 LLM 서비스가 제공하는 심층 리서치(deep research) 도구는 기존 문헌 검색과 달리 검색, 선별, 해석이 완료된 결과물을 제공한다. 이렇게 정제된 정보를 무비판적으로 수용하는 태도를 반복하면 본유적 인지 부하를 충분히 경험하지 못하게 되어 비판적 분석 능력이 저하될 위험이 있다. 환각을 가려내는 것뿐만 아니라 연구자 자신의 사고력을 유지하기 위해서도 LLM의 결과물을 맹신하지 않고 근거 자료를 폭넓게 탐색하며 원자료에 대한 해석이 적절한지 검토하는 과정을 생략해서는 안 된다.

결 론

LLM의 적용은 업무 효율과 학습 능률 향상이라는 분명한 이점을 제공하지만 동시에 비판적 사고의 약화를 야기할 수 있는 인지적 오프로딩의 위험성도 내포하고 있다. 이러한 양면성을 인식하고 어떤 종류의 인지 부하를 LLM에 외주화해도 안전한지 구분하여 LLM을 전략적으로 활용해야 한다. 연구 자료 수집 과정에서 본유적 인지 부하를 유발하는 활동을 직접 수행하는 것은 단기적으로 비효율적으로 보일 수 있으나 비판적 사고 능력을 보존하기 위한 가치 있는 장기 투자로 간주해야 한다. 때로는 조금 돌아가는 듯 보이는 방식이 오히려 더 효율적일 수 있다는 인식을 갖고 LLM을 통해 확보한 정신적 자원을 현명하게 배분하는 전략을 구사해야 할 것이다.

중심 단어: 거대 언어 모델; 사고; 학습; 인공지능; 작업 부하

CONFLICTS OF INTERESTS

No potential conflict of interest relevant to this article was reported.

FUNDING

This research was supported by the National Research Foundation of Korea (NRF), funded by the Korean government (MSIT) (RS-2025-02214129).

AUTHOR CONTRIBUTIONS

Sangzin Ahn is solely responsible for all aspects of this article.

ACKNOWLEDGEMENTS

None.

REFERENCES

1. Salvagno M, Cassai A, Zorzi S, et al. The state of artificial intelligence in medical research: a survey of corresponding authors from top medical journals. PLoS One 2024;19:e0309208.
2. Wang J, Fan W. The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. Humanit Soc Sci Commun 2025;12:621.
3. Meyer J, Jansen T, Schiller R, et al. Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. Comput Educ Artif Intell 2024;6:100199.
4. Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. Science 2023;381:187-192.
5. Dell'Acqua F, McFowland E III, Mollick ER, et al. Navigating the jagged technological frontier: field experimental evidence of the effects of ai on knowledge worker productivity and quality [Internet]. Rochester (US): SSRN, c2023 [cited 2025 May 10]. Available from: https://ssrn.com/abstract=4573321.
6. Stadler M, Bannert M, Sailer M. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Comput Human

Behav 2024;160:108386.
7. Huang S, Wen C, Bai X, et al. Exploring the application capability of ChatGPT as an instructor in skills education for dental medical students: randomized controlled trial. J Med Internet Res 2025;27:e68538.