JESK J Ergon Soc Korea 2026; 45(4): 265-275
https://dx.doi.org/10.5143/JESK.2026.45.4.265
https://jesk.or.kr eISSN:2093-8462

Cognitive Offloading and Deskilling in Creative Collaboration with Generative AI

Jieun Kwon

University of Minnesota, Human Factors and Ergonomics, United States, 55455

생성형 AI를 활용한 창조적 글쓰기 협업에서의 인지적 의존도와 탈숙련도

권 지 은

미네소타 대학교 인간공학과

Corresponding Author
Jieun Kwon
University of Minnesota, Human Factors and Ergonomics, United States, 55455
Email : jieunkwon14@gmail.com

Received : March 09, 2026
Revised : July 09, 2026
Accepted : July 16, 2026

Copyright@2026 by The Human Factors & Ergonomics Society of Korea. All right reserved.

Objective: This study aims to investigate how using generative AI such as ChatGPT influences users' cognitive dependency that results in cognitive offloading and deskilling.

Background: Generative AI, specifically LLM (Large Language Model)'s use had been rapidly growing in the creative domains such as writing. Some studies suggest that using LLM, or mainly, ChatGPT can support idea generation and expedite learnings, while others warn that users may experience deskilling and weakening on critical thinking. Most of the discussions on this topic have approached from the educational perspective focusing primarily on the ethical implications or the quality of outcome. On the other hand, relatively little attention has been paid to the process of using LLM and how human cognitions are affected.

Method: A total of 64 creative essays were collected from 32 University students. These students either used or not used ChatGPT to write the essays. Three writing experts (who identifies their job as a "writer" and has more than 5 years of writing experience) rated the essays on the number of assertions, the number of opposing views, the number of detailed cases introduced, concept quality, sense of agency, and structure diversity. For analysis, the study saw the difference between ChatGPT using group and no support group using paired-sample t-test.

Results: The results showed statistically significant differences between AI vs non-AI assisted groups. Participants who used ChatGPT during writing had used more counterclaims (p<0.001), while participants who wrote without ChatGPT demonstrated significant higher levels of authorial agency (p<0.001). Some marginal trends were found in the logical complexity, showing higher number of case use and complexity for the non-AI assisted group.

Conclusion: This study suggests that while generative AI can support humans, it may also potentially reduce and deskill individual cognitive abilities when articulating their arguments. In conclusion, this study suggests the need for generative AI systems from human factors perspective, so that the human-AI interaction helps support productivity while maintaining users' cognitive engagement and human agency.

Application: The results of this study may contribute to the development of AI interfaces and human-AI interaction design (AX), that support productivity while

reducing deskilling and cognitive offloading.

Keywords: AI-human interaction, Cognitive offloading, Deskilling, Creative writing

1. Introduction

2022년, 자연어를 기반으로 한 인공지능이 인간의 삶에 빠르게 스며들었다. ChatGPT가 대중에 소개된 지 1년만에 이미 의학 분야 학술지에서는 24%의 저자가 거대언어모델(LLM: Large Language Model)을 사용하고 있다고 답했고(Salvagno et al., 2024), 국제 공모전에도 AI와 함께 한 협업 창작물들이 폭발적으로 증가하고있다(Chun, 2025). 하지만, 빠르게 많은 이들의 일상을 변화시킨 만큼 그 활용에 대한 논란도 크다. 논쟁의 중심에는 AI의 활용 범위가 있다. 이는 본래 인간이 하던 다양한 일들을 인공지능이 "어디까지 하는 것이 좋을까"라는 질문에 대한 것이다. 특히, 그림이나 글쓰기와 같이 인간의 고유 영역이라 여겼던 창조적 영역에서, 인공지능이 꽤나 완성도 있는 퀄리티로 창조물을 내놓으면서 이 논쟁은 더 확산되고 있다(Kim, 2024). LLM이 텍스트로 이루어진 만큼, 이러한 논쟁은 특히 글쓰기 부분에서 더 집중적으로 연구되고 있다.

기존 글쓰기 교육과 관련된 몇몇 연구들에 따르면, 학습자는 일반적으로 스스로 써보고 실패하고 개선하는 시행착오를 겪으면서 글쓰기 능력을 향상해가는데, ChatGPT를 활용한 학습자들은 이와 관련하여 자료수집이나 검토와 같은 기초 능력을 모두 인공지능에게 맡기고, 사고력과 논리력과 같은 고차원적인 구성만 하는 방향으로 움직인다(Joo, 2023). 그 결과, 관련 연구자들은 글쓰기 교육에서의 AI 활용이 결국 학습의 과정 생략으로 이어질까 우려한다(Na, 2024). 더 나아가, 단기적으로 효율과 능률을 향상시키더라도 장기적으로는 선별되고 해석된 방대한 양의 정제된 정보를 받아들임으로써, 결과적으로 사용자의 비판적 사고 능력을 떨어뜨린다는 우려의 목소리도 커지고 있다(Ahn, 2025). 반대로, 그렇다고 해서 모든 글쓰기 학습 과정에서 AI 활용을 배제하는 것은 시대적으로 요구되는 디지털 리터러시 역량을 저하시키는 모순을 불러일으킨다. 이 때문에 Kim (2024)의 논의를 보면, 오히려 대학 글쓰기 과정에서 적극적으로 생성형 AI를 사용할 수 있도록 새로운 교육 과정을 모색해야 한다고 주장한다. 구체적으로는, ChatGPT를 초안에 활용함으로써 필자들이 가장 어려워하는 '아이디어 생성' 단계를 원활하게 수행하고, 이후에는 스스로 아이디어를 발전시키고 확장시키면서 결과적으로 성공적인 쓰기를 이룰 수 있도록 지원하는 것이다(Kim, 2024). 실제로 독일에서 학술적 글쓰기 과제에서 생성형 AI를 활용하게 한 결과, 학생들의 성취도가 높아지는 결과를 보였다(Meyer et al., 2024).

위와 같은 논쟁은 아직 LLM이 대중적으로 사용된 지 얼마 안된 지금 시점에 해야 하는 매우 필수적이고 중요한 논의라고 생각된다. 그러나 대부분의 연구가 교육학적 관점에서 다루어지고 있으며, 논의의 내용이 아직 구체적인 LLM 활용의 과정보다는 과연 LLM을 활용해도 되는지에 대한 윤리적, 결과적 접근에 제한되고 있는 것도 사실이다. 다시 말해, 대부분의 LLM와 글쓰기 관련 연구는 LLM이라는 도구를 사용하여 어떠한 결과가 초래되었는지에 대한 연구이지만, 그 과정에서 LLM이라는 "도구로 인해" 인간의 인지 과정에는 어떠한 변화가 있었는지에 대해서는 이야기하지 않는다는 것이다. 따라서 본 연구에서는 생성형 AI, 특히 LLM 을 사용하는 과정에서, 실제로 인간에게 어떠한 변화가 일어나는지 구체적으로 알아보고자 한다. 본 연구는 특히 다음과 같은 두 가지 주요 측면을 포함하고 있다. 첫째, ChatGPT를 교육학적 관점에서 학습의 도구로 적합한지를 평가한 기존 연구들과 달리 처음부터 인지적 외주화의 수단으로 보고, 그 외주화에서 일어나는 구체적인 인간 인지의 변화를 인간공학적 관점에서 보고자 한다. 두 번째로, 인간-시스템간 상호작용에서 일어난 인지의 변화가 결과적으로 태스크 숙련도에 어떤 영향을 미치는지 보고자 한다. 이는 LLM에 의존한 사용자의 결과를 그 산출물의 질로 평가해온 기존 연구들과는 달리, LLM 사용 과정에서 사용자의 숙련도가 저하되는지 알아봄으로써 결과물이 아닌 과정에 집중하고자 하는 것이다. 이 연구는 LLM의 활용을 보다 구체적인 과정으로 들여다봄으로써, 인간-AI 상호작용의 심리학적, 인간공학적 측면을 분석하고 최종적으로는 인간의 주체성과 숙련을 유지할 수 있는 AI 협업 원칙을 제시하는데 기여한다.

2. Theoretical Background

2.1 Cognitive offloading

인지적 외주화는 인지 부하 이론(Cognitive Load Theory - CLT)에서 다루어지는 현상으로, 인간이 어떤 태스크를 실행할 때 필요한 인

지적 노력을 낮추기 위해 외부의 도구를 활용하는 것을 이야기한다(Young et al., 2014; Risko and Gilbert, 2016). 인지 부하 이론에 따르면 인지 부하, 또는 정신적인 노력은 맥락에 따라 크게 내재적 부하(task complexity), 외재적 부하(non-essential processing), 그리고 학습 부하(schema construction) 세가지로 나누어진다(Young et al., 2014). 내재적 부하는 태스크 자체의 복잡성에서 오는 인지적 노력이고, 외재적 부하는 잘못된 설명 방식이나 중복 정보 등 불필요한 정보 제거 과정에서 발생한다. 세 번째인 학습 부하는 어떤 정보를 스스로의 지식으로 만드는데 발생하는 노력으로, 예를 들어 어떤 주장을 형성하거나, 개념을 재정의 할 때, 또는 구체적인 사례를 생각하는 등의 과정에서 발생한다.

인지적 외주화는 이중에서 보통 내재적 부하와 외재적 부하를 완화시키는 역할을 할 수 있는데, 세 번째 학습 부하에서는 오히려 사용자의 학습 능력을 감소시킬 수 있다는 연구 결과가 보고된 바 있다(Risko et al., 2014; Moritz, 2019; Eskritt and Ma, 2014; Sparrow et al., 2011). 특히, 연구에 따르면 사용자가 인지적 외주화를 하는 경우 단순한 암기 능력도 감소하였지만 더 나아가 목표 태스크를 실행하기 위해 주어진 정보를 어떻게 활용할지 어려워하는 모습을 보였다(Eskritt and Ma, 2014). 예를 들어, 최근 Jin et al. (2024)가 고령층을 대상으로 한 "모바일 기기를 활용한 인지적 외주화" 연구를 보면, 외부 장치 기억 저장소로 활용할 수 있도록 한 그룹은 그렇지 않은 그룹보다 정보를 기억하거나 추론하는 능력이 현저히 떨어졌다. 또한 Ferguson et al. (2015)의 연구에서는, 상식 퀴즈를 풀 때 인터넷 접속 권한을 가진 그룹이 그렇지 않은 그룹보다 "모르겠다"라고 답할 확률이 높았고, 자신이 답한 답에 대해서도 "내가 답을 알고 있다는 느낌"이 더 낮게 나왔다. 이는, 인지적 외주화가 높을 수록 학습 능력도 저하되며, 학습이 되지 않고 있다는 느낌의 메타인지까지 이루어진다는 것을 보여주는 것이다.

위와 같은 연구들을 기반으로, 본 연구에서는 글쓰기라는 작업을 실행할 때 ChatGPT를 활용하여 인지적 외주화를 할 경우, 사용자들의 학습 부하 능력이 저하될 것을 예상한다. 여기에서 글쓰기와 관련된 학습 부하는 다음과 같이 측정하고자 한다:

주장의 개수: 사용자가 어떠한 작업을 실행할 때 인지적 외주로 인해 학습 부하에서의 처리 능력을 감소시킨다면, 자기만의 주장을 형성하는 것이 어려울 것으로 본다. Petty and Cacioppo (1986) Elaboration Likelihood Model에 따르면 정교화(elaboration)은 정보에 대한 인지적 처리가 많고 깊다는 것을 의미한다. 즉, 어떠한 주장을 할 때 더 많은 근거와 설명이 따를 경우 이것은 필자가 그 주장에 대한 정보가 많고 또 깊이 있게 들여야 봤다는 것이다. 때문에, 인지적 외주화의 종속변수로 측정을 주장의 개수로 보고자 한다.

논증 복잡성과 반론 제시 여부: 학습 부하에서의 처리 능력이 저하되면, 복잡한 논리를 구성하기 어려워진다(Risko and Gilbert, 2016). Flower and Hayes (1981)의 Cognitive process theory of writing에 따르면, 글쓰기에 숙련도가 높은 사람은 논증 구조를 더 복합적으로 설계하고 반복적인 재구성이 가능하다고 한다. 즉, 논증 구조가 복잡할 수록 그 내용을 잘 알고, 그 주장을 여러 방향으로 고민해 보았다는 것이다. 또한, 어떠한 주장을 제시할 때 그 주장에 대한 반론이나 예상되는 반박에 대해 언급하는 것 역시 그 주장에 대한 인지적 처리가 많았음을 암시한다(Kuhn, 1991; Toulmin, 2003). 이에 근거하여, 필자가 사용한 논증 단계의 복잡성과 반론 제시 여부를 보고자 한다.

구체적 사례 수: 마지막으로, 구체적 사례 사용은 정보 처리에 매우 중요한 부분이다. 정보 처리가 깊을 수록 더 많은 의미를 알 수 있고, 주장에 대한 맥락적인 확장이 가능해진다(Craik and Lockhart, 1972). 인지적 외주화를 활용하여 학습 부하에서의 처리 능력이 저하되면, 어떠한 개념을 정확하게 사례에 적용하기 어려울 것으로 보고, 글에서의 구체적인 사례 수를 보고자 한다.

2.2 Deskilling

탈숙련화란 기술의 상실이나 기술의 감퇴, 전문성의 퇴화라고도 불리며 인간이 이전에 습득한 전문적인 기술이 자동화나 AI 등의 사용과 의존을 통해 서서히 감소되거나 상실되는 현상을 의미한다(El Tarhouny and Farghaly, 2026). 이는 산업화의 일환으로, 반복적이고 표준화된 업무로 인해 인간의 전문성이 퇴화하는 현상을 일컫던 단어에 그 뿌리가 있으며 최근에는 AI로 인한 업무 대체나 자동화로 인해 인간의 스킬이 저하되는 현상에서 더 많이 사용되고 있다(Natali et al., 2025). 여기에서 감소되는 인간의 '스킬'에는 기술적인 것 뿐만 아니라 인간이 장기간 다양한 환경에서 같은 업무를 하며 쌓인 경험에 의한 '노하우'도 해당되며(Ferdman, 2025), 회계 업무와 같은 전문 분야도 포함된다(Rinta-Kahila et al., 2023).

물론, 여기에도 논쟁은 있다. 몇몇 연구들에 따르면 AI로 인한 탈숙련화는 전문성의 저하를 뜻하는 것이 아니라, 초보자와 전문가의 평등화라고 주장한다. 즉, AI를 활용함으로써 자신이 본래 잘하던 것을 못하게 되는 것이 아니라, 그저 기존에 초보자였던 사람들이 AI를 통해 전문가처럼 잘하게 됨으로써 전문가가 시장에서의 경쟁력을 잃는 것이라는 말이다(Campero et al., 2022; Noy and Zhang, 2023; Crowston and Bolici, 2025). 여기에서 탈숙련화(deskilling)는 사실상 초보자의 숙련화(upskilling)을 의미한다. 더 나아가 일부 연구에서는 AI를 통한 자동화가 전문가가 더더욱 높은 성과를 내는 데에만 도움이 되고, 초보자에겐 큰 도움이 안되었던 사례도 있다. 예를 들어, Jia et al. (2024) 연구에서는 40명의 전화 상담사가 3,144명의 고객을 대상으로 카드에 가입하도록 유도하였는데, AI를 통해 '가입할 확률이 높은 고객'을 추천 받을 경우 숙련된 상담사들은 추천 받기 전보다 더 높은 성과를 냈고, 초보 상담사들은 큰 차이를 보이지 못했다.

이렇듯, AI로 인한 탈숙련화의 다양한 개념 중, 본 연구는 AI의 등장 전부터 차용되어온 인간공학적 관점에서의 탈숙련화를 차용한다. 이는 결국 자동화와 기술을 통한 루틴작업에서 사람은 연습 부족이 발생하고, 이것이 기술의 저하로 이어진다는 관점이다(Parasuraman and Riley, 1997; Onnasch et al., 2014). 또한 자동화는 사람의 작업 개입을 줄이는데, 이것이 결국 상황인식 능력을 저하시키고 결과적으로는 사람이 본래 가져왔던 기술을 손실하게 만든다는 관점이다(Endsley and Kiris, 1995). 즉, 인간공학적 관점에서의 탈숙련화는 자동화 시스템의 사용으로 인한 인간 개입 감소로, 판단이나 구조 설계 능력이 저하되는 것을 의미하며, 이를 글쓰기에 대입하면 탈숙련화는 다음과 같이 정의할 수 있다(Table 1 참고).

Table 1. Deskilling of writing function

	Deskilling definition	Measure
Sense of agency	Fails to present their own argument	The frequent use of the word "I think", "My thoughts are"
Concept	Fails to define and introduce new concepts and perspectives	The frequent use of the word "Generally..." "It is known"
Sentence diversity	Repeats same sentence structure	Sentence length, and use of diverse connecting vocabularies such as "therefore", "In addition", "However" etc.

여기에서 주체성(Sense of agency)은 자신의 주장을 펼칠 수 있는 능력으로, "A는 B이다"와 같은 명시적 주장과 "나는~ 생각한다" 등의 표현을 이야기한다. 그렇다면 주체성의 탈숙련도는 이러한 표현의 부재로 평가할 수 있다. 두 번째, 개념(concept)은 어떠한 새로운 개념의 정의 혹은 개념의 적용을 이야기하며, 개념의 탈숙련도는 "일반적으로..."와 같은 표현과 개념이 적용된 사례의 부재로 평가할 수 있다. 마지막으로, 복잡성(Sentence diversity)는 문장 구조의 복잡성을 이야기하며, 복잡성의 탈숙련도는 "반대로"와 같은 표현의 사용, 시각이 다양화와 문장 자체의 길이 등으로 평가가 가능하다. 이를 정리하면 Table 2와 같다.

Table 2. Writing functions' evaluation criteria defined

Writing functions	Evaluation criteria	Definition
Sense of agency	Statements	A production of arguments or statements
	Sense of agency	A formation and expression of self-thoughts
Concept	Concept	An introduction of new concept
	Cases	Introduction of example cases based on new concept
Sentence diversity	Diversity	Use of diverse sentence structures
	Complexity	A well-developed reasoning based on different reasonings
	Counterclaims	Generation of different perspectives on the same subject matter

3. Method

3.1 Data collection

본 연구에서는 서울소재의 대학에서 동일한 전공의 1~4 학년생을 대상으로 32명을 모집하여 2번의 과제를 실행하도록 하였다. 먼저, 32명을 그룹 1, 그룹 2로 나누고 1그룹은 "소셜 미디어는 청소년에게 해인가?"라는 주제로, 그리고 2그룹은 "ChatGPT는 대학생에게 유익한가?"라는 주제로 A4 용지 1~2쪽 이내 논술을 하도록 하였다. 그 과정에서 1그룹과 2그룹은 다시 두 개의 소그룹으로 나뉘어, 한 소그룹은 AI를 활용할 수 있었고 나머지 소그룹은 AI를 활용할 수 없었다(Figure 1 참고).

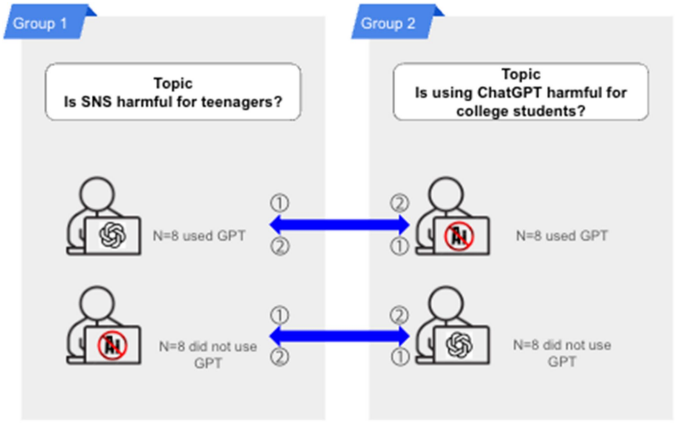

Figure 1. Participant groups and conditions

글쓰기 주제는 대학생들이 쉽게 의견을 제시할 수 있고 친숙하게 다룰 수 있는 주제를 선정하였다.

본 연구에서는 두 가지 글쓰기 주제를 제시하였는데, 하나는 "소셜 미디어는 청소년에게 해인가?" (SNS)라는 주제로, 그리고 다른 그룹은 "ChatGPT는 대학생에게 유익한가?"(GPT)였다. 이 주제는 다음과 같은 절차를 통해 선정되었다.

1. 대학생들이 쉽게 의견을 제시할 수 있고 친숙하게 다룰 수 있는 분야 4가지를 먼저 선정(환경, 경제, 교육, 미디어).
2. 각 분야에 대해 기존의 지식만으로도 찬반을 논할 수 있는 글쓰기 주제를 2개씩 개발(예: 대학 등록금은 낮아져야 하는가? 일회용품은 강하게 규제되어야 하는가? 등).

그 중 글쓰기 교육 전문가(박사학위자) 2인의 의견을 검토하여 대학생들이 현재 사용하는 기술과 관련된 주제 2개를 선정하였다. 주제 난이도를 맞추기 위해 질문의 형식을 일관되게 하였으며(A는 B에게 유익한가/해인가?) 지시문도 명료하게 제시하였다.

예:
주제: "소셜 미디어는 청소년에게 득인가, 해인가? 이에 대한 장단점을 논하고, 본인의 의견을 제시하시오."

유의사항:
- 서론–본론–결론의 구조로 작성
- 득과 해를 균형 있게 제시할 것
- 분량: A4 1~2쪽 내외

첫 번째 과제가 끝난 후, 각 그룹은 다시 주제와 AI 활용 여부를 반대로 하여 다시 한번 과제를 실행하였다. 예를 들어, 첫 번째에서 SNS 주제로 AI를 활용할 수 있었다면, 그 다음 순서에서는 GPT 주제로 AI를 활용하지 않고 글을 작성하였다(Figure 1. 참고). 이러한 방법으로 순서효과(AI를 먼저 쓰는 경우 vs. AI 없이 먼저 작성해보는 경우)와 주제효과(SNS 주제 vs. GPT 주제)를 통제하고자 하였다. AI는 모든 학생들이 손쉽게 사용할 수 있는 ChatGPT로 통일하였으며, AI 활용 그룹에 있는 경우 각 학생은 ChatGPT에 로그인하지 않은 상태로 도움을 받을 수 있었다.

3.2 Data analysis

32명의 참여자가 2번씩 글을 써서, 총 64개의 글을 분석하였으며 3명의 글쓰기 전문가가 채점자로 참여하여 모든 글에 대해 평가하였다. 3명의 글쓰기 전문가는 '작가'를 본업으로 하여 글에 대한 수입이 있으며, 5년 이상의 경력을 가진 이들로 구성되었다. 각 전문가는 학생들의 ChatGPT 사용 여부를 모르는 상태에서 모든 글을 채점했으며, 채점 가이드라인은 다음과 같이 제공하였다(Table 3).

Table 3. Essay writing guidelines for the raters

	Coding guidelines and criteria
Statements	How many core arguments does the author present as 'their own'? (e.g., "I think...", "my thoughts on...")
Counterclaims	How many counterarguments appear in the text? (e.g., "but", "however", "on the other hand" etc)
Cases	How many detailed examples are provided in the text? (e.g., "for example", "as a case" etc)
Concept score	The key concept discussed in the text is clearly defined and explained (1-very low, 5-very high)
Sense of agency	The author expressed their own idea rather than focusing on general or commonly stated opinions (1=very low, 5=very high)
Diversity	The author used a variety of sentence structures (e.g., subordinate, contrastive, conditional…etc) (1=very low, 5=very high)
Complexity	The author demonstrated a complex reasoning rather than using a single reasoning (e.g., "A is B") (1=very low, 5=very high)

각 평가 기준에 대해서 평가자간 신뢰도를 계산하였고, 주장, 반론, 사례 등은 연속적인 점수가 아닌 개수에 대한 코딩임으로 ICC (Intraclass Correlation Coefficient) 보다는 Krippendorff Alpha를 활용하여 계산하였다. Krippendorff Alpha는 분산이 아닌 각 평가자간의 불일치 크기(difference)를 기준으로 계산하며, 본 연구와 같이 개수 기반의 코딩에 적합하다고 판단하였다(Krippendorff, 2018). 3명의 전문가 점수는 평균을 내어 활용하였고, 본 논문의 데이터 분석에는 마이크로소프트 엑셀과 SPSS을 활용하여 분석하였다.

4. Results

4.1 Interrater reliability

세 명의 평가자가 각각 글을 독립적으로 평가한 내용에 대해서, 상호-평가자간 신뢰도를 Krippendorff Alpha로 확인한 결과 Table 4과 같이 분석되었다. 총 7개의 항목에 대하여 주장수(Statement)를 제외한 6가지 항목에서 "높은 수준"의 신뢰도를 확인하였다.

Table 4. Krippendorff Alpha for three raters' scores

Krippendorff Alpha	Topic 1	Topic 2
Statements	0.6063	0.5651

Table 4. Krippendorff Alpha for three raters' scores (Continued)

Krippendorff Alpha	Topic 1	Topic 2
Counterclaims	0.8601	0.6879
Cases	0.7091	0.7336
Concept score	0.7857	0.8219
Sense of agency	0.7755	0.8302
Diversity	0.7589	0.8046
Complexity	0.7701	0.8621

*Inter-rater correlation is very high if $\alpha \geq 0.800$, medium if $0.667 \leq \alpha < 0.800$, low if $\alpha < 0.667$

반면, 주장수 항목에 대해서 알파 값이 0.606 수준으로 낮게 나타났으며, 이는 "주장"에 대한 평가가 평가자에 따라 다르게 해석되었기 때문으로 보인다. 또한, 글에서 "내 생각"으로 시작한 문장이 한 개 이더라도, 이어지는 다른 문장 또한 저자의 생각인지 등을 판단하는 데에서 평가자간 차이가 생겼을 것으로 예상한다. 그 결과, 본 연구의 분석에서는 신뢰도가 "낮음"에 해당되는 '주장수' 요인에 대해서는 조심스럽게 접근할 것을 권고하고, 추후 보완적 하기 위한 추가적인 연구가 필요할 것을 미리 명시한다.

4.2 AI use group vs No use group

먼저, 본 연구는 한 집단이 AI를 활용하여 글을 쓰고 AI를 비 활용하여 글을 쓰는 두 가지 라운드를 모두 경험하여 이루어진 데이터이므로 표본의 독립성 가정에 위배된다고 판단한다. 따라서, 통계분석에서는 32명의 참가자별로(AI-noAI) 라운드를 각자 매칭하여 Paired sample *t*-test를 진행하였다(Table 5).

Table 5. Paired-sample *t*-test

Category	AI mean	No AI mean	Mean diff (AI-no AI)	SD of diff	*t*	*p*-value	Cohen's d
Statements	3.260	3.365	-0.104	1.032	-0.571	0.572	-0.101
Counterclaims	2.031	1.063	0.969	0.595	9.213	$p<.001^{***}$	1.629
Cases	1.542	1.781	-0.240	1.027	-1.319	0.197	-0.233
Concept score	2.760	2.844	-0.083	1.033	-0.456	0.652	-0.081
Sense of agency	2.854	3.635	-0.781	0.910	-4.855	$p<.001^{***}$	-0.858
Diversity	3.375	3.406	-0.031	0.652	-0.271	0.788	-0.048
Complexity	3.000	3.260	-0.260	0.862	-1.708	0.098	-0.302

M ± SD $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$

그 결과, 반론 수와 주체성에서 유의미한 차이를 보였다. 반론 수는 가설과는 반대로 AI를 활용하였을 때 더 높게 나왔고, 주체성은 가설과 동일하게 AI를 활용할 경우 더 낮아지는 것으로 나왔다. 특히, AI 활용 시 반론/ 대안점 관점의 제시가 거의 2배로 증가하는 것을 볼 수 있는데, 이는 인지적 외주화의 숨겨진 '생산적 이득'으로 볼 수 있다. 반대로 AI 사용시 사고/작업에 대한 주체성은 현저히 낮아지는 패턴을 볼 수 있는데, 이는 인지적 외주화의 '심리적' 비용으로 생각할 수 있다. 추가적으로, 논리 복잡성도 통계적으로는 비유의적 차이(p=0.098)를 보였는데 AI를 활용한 경우 복잡성이 더 낮아지는 추세를 보인다. 이는 "AI를 활용할 경우 논리를 설계하는 능력에서 탈숙련도를 보일 것이다"라는 가설과 같은 방향성을 보인 것이다.

4.3 Order effect and topic effect

추가적으로, A는 AI 활용을 먼저, B는 AI 활용을 나중에 하여 글을 쓴 만큼, 결과에 순서효과가 있는지 각 요소에 대하여도 1라운드와 2라운드 글의 점수를 짝지어 Paired *t*-test 분석으로 확인하였다. 그 결과, 문장 구조 다양성에서는 AI를 먼저 활용할 경우 AI를 활용한 글에서 문장 구조가 더 다양(p<0.05)하게 나왔다(Table 6).

Table 6. Order effect paired sample *t*-test

Variable	n	R1 mean	R1 SD	R2 mean	R2 SD	*t*	*p*-value
Statements	32	3.469	0.846	3.156	0.580	1.791	0.083
Counterclaims	32	1.500	0.713	1.594	0.553	-0.463	0.647
Cases	32	1.719	0.709	1.604	0.735	0.617	0.542
Concept score	32	2.875	0.692	2.729	0.658	0.803	0.428
Sense of agency	32	3.115	0.701	3.375	0.915	-1.250	0.221
Diversity	32	3.510	0.535	3.271	0.564	2.236	0.033*
Complexity	32	3.250	0.616	3.010	0.721	1.560	0.129

M ± SD *p<.05, **p<.01, ***p<.001

더 나아가, 주제효과에 대한 분석 결과 반론 수와 주체성에서 주제효과가 발견되었다(Table 7). 반론 수에서는 GPT 주제(M=1.73, SD=0.65)가 SNS 주제(M=1.31, SD=0.56) 보다 높게 나타났으며(p<0.05), 반대로 주체성에서는 SNS 주제(M=3.44, SD=0.75)가 GPT 주제(M=3.06, SD=0.88) 보다 높게 나타났다. 나머지 다섯 가지 변수에서는 유의미한 차이가 나타나지 않았다.

Table 7. Topic effect paired sample *t*-test

Dependent Var.	n	GPT mean	GPT SD	SNS mean	SNS SD	SD diff.	*t*	*p*-value
Statements	32	3.344	0.765	3.237	0.726	1.034	0.240	0.812
Counterclaims	32	1.733	0.646	1.312	0.557	1.075	2.073	0.048*
Cases	32	1.656	0.740	1.645	0.720	1.040	0.178	0.860
Concept score	32	2.800	0.720	2.785	0.647	1.025	-0.060	0.952
Sense of agency	32	3.056	0.880	3.441	0.752	1.163	-2.077	0.047*
Diversity	32	3.500	0.478	3.301	0.640	0.646	1.533	0.136
Complexity	32	3.200	0.730	3.086	0.655	0.856	0.434	0.668

M ± SD *p<.05, **p<.01, ***p<.001

5. Discussion and Conclusion

본 연구는 LLM, 특히 ChatGPT를 활용한 글쓰기 중 사용자에게 나타나는 인지적 외주화 현상과 탈숙련도를 인간공학적 관점에서 분석하였다. 이를 위해 총 32명의 대학생이 두 가지 주제에 대해 AI를 활용하거나 활용하지 않고 글을 쓰는 태스크를 수행하였고, 그렇게 만들어진 총 64개의 글을 세 명의 "글쓰기 전문가"가 평가하였다. 분석 결과, 가설과 일치하는 부분과 오히려 반대인 부분들이 발견되었다. 예를 들어, 자신의 생각을 썼는지 확인하는 '주체성'에서는 AI를 활용할 경우 주체성이 더 낮아지는 모습을 발견하였고 이는

가설(AI를 활용할 경우 인지적 외주화와 탈숙련이 일어난다)과 일치한다. 또한, 구체적인 사례수도 유의수준이 미미하긴 하지만, AI를 활용할 때보다 활용하지 않을 때 더 많은 사례를 서술한 것으로 확인되어 그 방향성을 확인하였다. 반대로, 반론 수에서는 AI를 활용할 경우 반론 수가 더 적어질 것으로 예상한 것과는 다르게, AI를 활용할 경우에 반론 수가 더 늘어나는 양상을 보였다. 이를 종합적으로 보면, AI 사용 시 특별히 어떤 종류의 문장(사례, 주장, 아이디어 등)이 많아지거나 적어지지 않지만 새로운 종류의 사고(반론)을 유도하는 모습을 볼 수 있다.

한편, 추가적으로 진행한 순서효과 분석에서는 AI를 먼저 활용한 그룹에서 문장 구조 다양성이 증가하는 것을 발견하였다. 이는 AI를 활용하는 것이 이후 택하는 글쓰기 전략에 영향을 주는 것임을 알 수 있다. 또한 주제효과 분석에서는 ChatGPT와 관련된 주제에 대해서 글을 쓸 때, 반론이 증가하고 반대로 SNS 주제로 글을 쓸 때에는 주체성이 높아지는 양상을 보였다. 이러한 결과에 대한 해석에는 주의가 필요하다. 본 연구의 실험 설계상, 이는 엄밀히 "주제만"의 효과는 아닌 것으로 보인다. 참가자들은 GPT와 SNS 주제에 대하여 어떤 경우 AI를 활용하여 글을 작성했고, 어떤 경우에는 AI 없이 작성했다. 즉, 주제효과가 조건효과와 혼합되어 나타나는 것이다. 그 결과 추가적으로 혼합효과에 대해 살펴본 결과, GPT를 활용하여 GPT에 대하여 서술할 때 주체성이 더 높아지는 것을 발견하였는데, 이는 사용자가 AI에 대해서 논의할 때는, AI에게 조금 더 의존하는 양상을 나타내는 것으로 보인다. 이 또한 자세한 원인을 알기 위해서는 추가적인 질적 연구가 필요할 것으로 보인다.

본 연구의 한계로는, 제한된 시간과 특정 주제 하에 글쓰기 태스크를 수행하도록 하여 현실과는 괴리가 있는 통제된 환경에서의 결과라는 점과 두 번째로는 사용자의 AI 사용 방식에 대한 구체적인 내용(어떤 대화를 나누었는지)에 대해서는 분석되지 않았다는 점이 존재한다. 먼저, 본 연구에서는 매우 단순한 특정 주제들에 대해 글을 쓰는 태스크를 수행하였으므로, 이러한 결과는 실제 우리 주변에서 ChatGPT를 활용하는 다양한 환경에 대해서 모두 대변할 수 없다. 또한, 태스크를 수행함에 있어서 구체적인 목표가 없었던 것(예: 질 높은 글을 써서 높은 점수를 받아야 한다 등)이 실제 상황과는 다를 수도 있다는 점을 한계로 인정하고 있다. 또한, 사용자가 ChatGPT와 대화하며 어느 정도의 탈숙련도가 일어났다는 분석 결과는 보여주지만, 어떠한 방식으로 탈숙련도가 발생했는지 그 과정에 대한 내용을 볼 수 없다는 것이 한계이다. 이는 추가적인 질적 연구를 통해, 사용자가 어떤 질문을 했고 또 ChatGPT는 어떤 답을 했는지 등을 자세하게 분석하여 보완할 수 있다. 이러한 상호작용 방식에 대한 연구를 통해 향후 AI 활용 전략이나 프롬프트 유형을 보다 세분화하여 제안할 수 있을 것이다.

본 연구는 위와 같은 한계점도 있지만, 또한 사회, 학술적으로도 기여한다. 먼저, 본 연구는 생성형 AI를 활용한 과제 수행에서, AI 활용이 결과물에만 영향을 주는 것이 아니라 사용자의 주체성을 약화시키고 탈숙련도를 발생시킬 수 있다는 가능성을 보여주었다. 이러한 결과는 단순히 생성형 AI가 도구로써 어떤 의미가 있는지에 대한 평가적 접근을 넘어서서, 인간과 AI의 상호작용에서 인간이 어떠한 영향을 받는지 알아본 것에 그 의미가 있다.

특히, 본 연구의 결과는 생성형 AI가 논점을 제시하는데 도움을 주거나 생각을 확장하는데 도움을 줄 수 있지만, 한편으로는 사용자가 자신의 논리를 형성하는 과정에서 일정 수준의 의존성을 유발 할 수 있음을 시사한다. 이는 추후 교육현장에서 학습자가 AI를 활용하면서도 자신의 관점과 논리를 스스로 형성하도록 지원하는 체계를 제안하며, 더 세분화된 인간-AI 협업 모델 설계에 대한 기초적인 자료로 활용 될 수 있다. 또한, 본 연구의 결과는 생성형 AI 활용을 단순히 인간의 생산성을 높이는 도구로 보지 않고, 사용자의 인지적 개입과 주체성을 유지할 수 있는 인터페이스의 설계가 필요함을 시사한다. 이러한 관점은 인간-AI 상호작용에서 인간공학적 요소를 반영한 시스템 설계가 필요함을 강조하며, 향후 생성형 AI 설계에서 인간의 인지적 역할과 숙련도를 유지할 수 있는 방향에 대한 중요한 기초자료가 될 수 있다.

References

Ahn, S., "Large language model usage guidelines in Korean medical journals: a survey using human-artificial intelligence collaboration." *Journal of Yeungnam Medical Science*, 42, 2025.

Campero, A., Vaccaro, M., Song, J., Wen, H., Almmatouq, A. and Malone, T.W., A test for evaluating performance in human-computer systems. *arXiv Preprint arXiV*, 2022.

Chun, H., "Study on Prompting in Generative AI: A Creative Writing Tool for the Digital Era." *Journal of Digital Contents Society*, 26(2), 275-285, 2025.

Craik, F.I.M. and Lockhart, R.S., Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684, 1972.

Crowston, K. and Bolici, F., Deskilling and upskilling with AI systems. *Information Research an International Electronic Journal*, 30 (iConf), 2025.

El Tarhouny, S. and Farghaly, A., Deskilling dilemma: brain over automation. *Frontiers in Medicine*, 13, 1765692, 2026.

Endsley, M.R. and Kiris, E.O., The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381-394, 1995.

Eskritt, M. and Ma, S., Intentional forgetting: Note-taking as a naturalistic example. *Memory & Cognition*, 42(2), 237-246, 2014.

Ferdman, A., AI Deskilling is a structural problem. *AI & SOCIETY*, 1-13, 2025.

Ferguson, A.M., McLean, D. and Risko, E.F., Answers at your fingertips: Access to the Internet influences willingness to answer questions. *Psychonomic Bulletin & Review*, 22(3), 880-887, 2015.

Flower, L. and Hayes, J.R., A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387, 1981.

Jia, N., Luo, X., Fang, Z. and Liao, C., When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, 67(1), 5-32, 2024.

Jin, Y., Li, D., Shi, J., Hao, P., Sun, F., Zhang, J. and Fang, B., Robot gpt: Robot manipulation learning from chatgpt. *IEEE Robotics and Automation Letters*, 9(3), 2543-2550, 2024.

Joo, M.J., "How to Teach Writing in the Age of Generative AI Reimagining Writing Instruction -Focusing on the analysis of university learners' perceptions of the utility of utilizing ChatGPT in the 'pre-writing stage'-." *DONAM OHMUNHAK*, 44, 71-103, 2023.

Kim, K.H., "A Way of University Writing Education in the Era of Generative AI -Focused on Use of Generative AI in Argumentative Writing-." *Urimalgeul: The Korean Language and Literature*, 101, 63-86, 2024.

Krippendorff, K., *Content analysis: An introduction to its methodology*. Sage publications, 2018.

Kuhn, D., *The skills of argument*. Cambridge University Press, 1991.

Meyer, J., Jansen, T., Schiller, R., Liebenow, L.W., Steinbach, M., Horbach, A. and Fleckenstein, J., Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 2024.

Moritz, J., *The influence of representation control on task performance and learning*, Eberhard Karls Universität Tübingen, Ph.D Dissertation, 2019.

Na, E.M., "In the Age of ChatGPT, Reflections on University Writing Education and Research on Educational Directions." *Hansung Literature Studies*, 51, 79-106, 2024.

Natali, C., Marconi, L., Dias Duran, L.D. and Cabitza, F., AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond. *Artificial Intelligence Review*, 58(11), 356, 2025.

Noy, S. and Zhang, W., Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192, 2023.

Onnasch, L., Wickens, C.D., Li, H. and Manzey, D., Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, 56(3), 476-488, 2014.

Parasuraman, R. and Riley, V., Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253, 1997.

Petty, R.E. and Cacioppo, J.T., The elaboration likelihood model of persuasion. In *Advances in Experimental Social Psychology* (Vol. 19, pp. 123-205). Academic Press, 1986.

Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W. and Ruissalo, J., The vicious circles of skill erosion: A case study of cognitive automation. *Journal of the Association for Information Systems*, 24(5), 1378-1412, 2023.

Risko, E.F. and Gilbert, S.J., Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676-688, 2016.

Risko, E.F., Medimorec, S., Chisholm, J. and Kingstone, A., Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science*, 38(3), 537-564, 2014.

Salvagno, M., Cassai, A.D., Zorzi, S., Zaccarelli, M., Pasetto, M., Sterchele, E.D., Chumachenko, D., Gerli, G.A., Azamfirei, R. and Taccone, F.S., The state of artificial intelligence in medical research: A survey of corresponding authors from top medical journals. *Plos One*, 19(8), 2024.

Sparrow, B., Liu, J. and Wegner, D.M., Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043), 776-778, 2011.

Toulmin, S.E., *The uses of argument*. Cambridge university press, 2003.

Young, J.Q., Van Merrienboer, J., Durning, S. and Ten Cate, O., Cognitive load theory: implications for medical education: AMEE Guide No. 86. *Medical Teacher*, 36(5), 371-384, 2014.

Author listings

Jieun Kwon: jieunkwon14@gmail.com
Highest degree: PhD, College of Design, Human Factors and Ergonomics Program, University of Minnesota Twin-Cities
Position title: Lecturer, Department of Global AI Convergence, Seoil University
Areas of interest: Artificial Intelligence Experience (AX), Design Thinking, Interaction Design