Check for updates

컴퓨터교육학회 논문지 2026년 제29권 제7호
https://doi.org/10.32431/kace.2026.29.7.004

공학계열 대학생의 생성형 AI 고도 활용 유형에 따른 AI 리터러시 수준 분석: 프롬프트 작성 전략과 비판적 검증을 중심으로

AI Literacy Levels Across Profiles of Advanced Generative AI Use among Engineering Undergraduates: Focusing on Prompting Strategies and Critical Verification

이대영†
Daeyeong Lee†

요약

본 연구는 공학계열 대학생의 생성형 AI 고도활용을 프롬프트 작성 전략과 비판적 검증의 결합 양상으로 유형화하고, 유형별 AI 리터러시 수준 차이를 분석하는 데 목적이 있다. 분석 결과, 공학계열 대학생은 비판적 검증보다 프롬프트 작성 전략을 더 높은 수준(=2.745)으로 활용하고 있었다. 그리고 생성형 AI 고도활용에 대한 유형화 결과 전반적 중저활용형, 고프롬프트-검증결핍형, 저프롬프트-검증우세형, 전반적 고활용형의 네 개의 유형으로 도출되었다. 또한 이러한 유형에 따라 AI 리터러시 전체와 하위변인 수준 모두에서 유의한 차이가 나타났다. 이를 바탕으로 프롬프트를 능숙하게 작성하는 능력이 곧 산출물을 검증하는 능력으로까지 이어지지 않는다는 점을 규명하였다. 이는 생성형 AI 활용 교육에서 프롬프트 작성 능력뿐 아니라 산출물의 정확성, 근거, 출처를 비판적으로 검증하는 교육이 함께 이루어져야 한다는 점을 시사한다.

주제어 생성형 AI, AI 리터러시, 생성형 AI 고도활용, 프롬프트 작성 전략, 비판적 검증

ABSTRACT

This study aims to classify engineering undergraduates' advanced use of generative AI based on the combined patterns of prompting strategies and critical verification, and to analyze differences in AI literacy levels across the identified profiles. The results showed that engineering undergraduates used prompting strategies at a higher level than critical verification (= 2.745). The profile analysis of advanced generative AI use identified four profiles: low-to-moderate use, high-prompting/verification-deficit, low-prompting/verification-dominant, and high use. In addition, significant differences were found in both overall AI literacy and all of its subdomains across these profiles. Based on these findings, this study clarified that the ability to skillfully construct prompts does not necessarily extend to the ability to verify AI-generated outputs. This suggests that education on generative AI use should address not only prompting skills, but also the critical verification of the accuracy, evidence, and sources of AI-generated outputs.

Keywords Generative AI, AI literacy, Advanced generative AI use, Prompting strategies, Critical verification

†정회원 한국공학대학교 전공교육혁신센터 연구교수

논문투고 2026년 05월 29일
심사완료 2026년 06월 15일
게재확정 2026년 06월 30일
발행일자 2026년 07월 13일

1. 서론

생성형 AI(Generative Artificial Intelligence)는 대학에서도 보편적으로 광범위하게 활용되고 있으며, 이는 대학생의 학습 활동 전반에서도 예외가 아니다. 대학생들은 정보 탐색, 글쓰기 보조, 아이디어 생성, 요약, 번역, 코딩, 과제 수행 등의 다양한 학습과업에 생성형 AI를 활용하고 있다. 이는 학습 효율성과 생산성을 높인다는 장점이 있다. 그러나 생성형 AI를 교육적으로 활용하는 것이 비단 긍정적 효과만 있는 것은 아니다. 생성형 AI에 대한 과잉의존은 학습자의 의사결정, 비판적 사고, 분석적 사고와 같은 인지적 역량을 약화시킬 수 있다[1]. 그리고 생성형 AI를 학습 증진보다 과업 부담을 줄이는 수단으로 사용하고 있다는 오용 양상도 보고된 바 있다[2]. 생성형 AI 의존이 높을수록 비판적 사고가 낮아지고, 인지적 피로가 이 관계를 매개할 수 있다는 연구결과 또한 생성형 AI 활용이 학습자의 사고 과정에 미치는 영향을 보다 면밀히 검토해야 함을 뒷받침한다[3].

이 같은 현상은 생성형 AI 활용의 핵심이 더 이상 사용 여부나 빈도에만 국한되지 않고 질적인 측면도 검토할 필요가 있음을 시사한다. 즉, 동일하게 생성형 AI를 활용하더라도 학습자가 생성형 AI를 과업 수행의 대체 수단으로 사용하는지, 사고 확장과 산출물 점검의 도구로 사용하는지에 따라 교육적 의미는 달라질 수 있다. 이러한 생성형 AI 활용의 질적인 측면은 생성형 AI 고도활용(advanced use)이라는 관점에서 구체화 될 수 있다. 생성형 AI 고도활용은 단순히 AI를 자주 사용하거나 프롬프트를 능숙하게 작성하는 것만을 의미하지 않으며, 프롬프트 작성 전략과 더불어 생성된 산출물을 비판적으로 검증하는 전략이 균형을 이루는지를 확인할 수 있는 변인이다[4-6]. 생성형 AI의 최근 사용 맥락에서는 적절한 도구 선택 및 효과적인 프롬프트 작성 능력과 함께 AI가 생성한 산출물의 정확성, 편향, 환각, 윤리적 위험을 검토하는 능력도 요구되기 때문에 학습자의 생성형 AI 고도활용 수준 및 패턴을 점검할 필요가 있다[4].

한편, 이러한 생성형 AI 사용맥락은 특히 공학계열에서 더욱 면밀하게 검토되어야 한다. 공학계열 학생들은 전공학습 과정에서 문제해결, 설계, 자료 분석, 코딩, 실험 보고서 작성 등 생성형 AI를 활용할 수 있는 과업을 빈번하게 경험하기 때문이다[7]. 동시에 공학적 산출물은 정확성, 재현가능성, 근거 기반 판단, 안전성과 같은 기준을 요구하므로 AI가 생성한 결과를 그대로 수용하는 방식의 활용은 학습성과뿐 아니라 현장 의사결정 역량 형성에도 부정적 영향을 줄 수 있다[7, 8]. 따라서 공학계열 대학생의 생성형 AI 활용을 분석할 때는 단순한 사용 능력보다 프롬프트 작성 전략과 비판적 검증이 어떻게 결합되어 있는지, 그리고 이러한 활용 양상이 AI 리터러시 수준과 어떠한 관계를 가지는지 살펴볼 필요가 있다. 여기서 AI 리터러시는 단순히 AI 도구를 사용할 줄 아는 능력을 넘어 AI의 산출물을 이해하고 평가하며 책임 있게 활용하는 역량으로 정의된다[4, 9]. 따라서 생성형 AI의 교육적 활용을 논의하기 위해서는 단순한 활용 경험보다는 학습자가 AI를 이해하고 적절히 활용하며, 산출물을 평가하고 책임 있게 사용할 수 있는지를 종합적으로 확인할 수 있는 AI 리터러시를 함께 고려해야 한다.

기존에 제시된 선행연구를 살펴보면 주로 AI 리터러시의 개념과 측정도구 개발[4, 9-11], 공학계열 대학생의 생성형 AI 활용과 비판적 사고[7], 프롬프트 리터러시 및 프롬프트 작성 전략[6, 12], AI 과잉의존 및 윤리적 우려[1-3]를 각각 독립적으로 다루는 데 그쳤다. 그러나 생성형 AI 고도활용의 하위전략을 프롬프트 작성 전략과 비판적 검증으로 구분하여 두 전략이 균형적으로 발달하는지를 분석한 연구는 이루어지지 않았다. 두 전략은 모두 생성형 AI 활용 과정에서 요구되는 역량이지만 동일하게 발달한다고 가정하기는 어렵다. 따라서 프롬프트 작성 전략은 높은데 비판적 검증은 상대적으로 낮은 학습자, 혹은 두 전략이 모두 높은 학습자와 같은 다양한 유형의 학습자가 존재할 수 있다. 이에 따라 생성형 AI 고도활용을 프롬프트 작성 전략과 비판적 검증 수준의 결합 양상에 따라 유형별로 나누어 살펴볼 필요가 있다.

본 연구는 공학계열 대학생을 대상으로 생성형 AI 고도활용 유형에 따른 AI 리터러시 수준을 분석하고자 한다. 즉, 생성형 AI 고도활용의 하위변인인 프롬프트 작성 전략과 비판적 검증의 균형성 및 불일치 양상을 탐색하고 이러한 유형에 따라 AI 리터러시 전체 및 하위요인 수준이 어떻게 달라지는지를 분석하고자 한다. 이에 따른 연구문제는 다음과 같다.

연구문제 1. 공학계열 대학생의 생성형 AI 고도활용에서 프롬프트 작성 전략과 비판적 검증의 균형 양상 혹은 불일치 양상은 어떻게 나타나는가?

연구문제 2. 생성형 AI 고도활용 유형에 따라 AI 리터러시 전체 및 하위요인 수준에는 어떠한 차이가 있는가?

2. 이론적 배경

2.1 AI 리터러시의 개념 및 하위요인

AI 리터러시는 AI 기술을 사용할 수 있는 기능적인 능력뿐만 아니라 AI의 작동 원리를 이해하고 목적에 맞게 활용하며, AI가 산출한 결과를 비판적으로 평가하고 윤리적인 측면을 고려하여 책임 있게 사용할 수 있는 역량으로 논의된다[4, 9]. 이러한 AI 리터러시는 AI 기술이 대학생의 학습이나 정보 탐색, 문제해결, 의사결정 과정 등 여러 방면에서 활용됨에 따라 고등교육 학습자가 갖추어야 할 기초적인 역량으로 확장·논의되고 있다[9, 11, 13]. 따라서 AI 리터러시는 단일한 지식이나 기능으로 환원되기보다는 인지적 이해, 실제적 활용, 평가적 판단, 윤리적 성찰을 포함하는 다차원적인 역량이라고 볼 수 있다[10, 11].

이와 관련하여 임철홍 외[10]는 AI 리터러시를 AI 이해, AI 활용, AI 평가, AI 윤리의 하위요인으로 구성하여 학습

자의 AI 리터러시를 측정하였다. Yan et al.[4] 역시 AI 시스템에 대한 이해와 활용뿐만 아니라 산출물 검증, 위험 평가, 윤리적 판단이 AI 리터러시에서 중요하게 고려해야 하는 요소라고 주장하였다. 이는 AI 리터러시가 단순히 AI를 사용할 수 있는지의 문제가 아니라, AI의 특성과 한계를 이해하고, 이를 바탕으로 AI 산출물을 평가하며, 그 결과를 책임 있게 활용할 수 있는 복합적인 역량임을 의미한다.

이러한 논의에 따라 AI 리터러시는 AI 이해, AI 활용, AI 평가, AI 윤리의 하위요인으로 구성된 다차원적 역량으로 볼 수 있다. AI 이해는 AI의 기본 개념과 작동 방식에 대한 이해를 의미하고, AI 활용은 학습이나 문제해결 과정에서 AI를 목적에 맞게 사용할 수 있는 능력을 의미한다. AI 평가는 AI가 산출한 정보나 결과의 타당성 및 신뢰성을 판단하는 능력이며, AI 윤리는 AI 활용 과정에서 발생할 수 있는 책임, 공정성, 편향, 개인정보, 사회적 영향 등을 고려하는 역량을 의미한다.

2.2 생성형 AI 고도활용 및 하위전략

생성형 AI의 교육적 활용이 늘어나면서[13], 단순한 사용 경험이나 사용 빈도만으로는 학습자의 AI 활용 수준을 충분히 설명하기 어려워졌다[4, 9]. 생성형 AI는 사용자의 입력 방식과 상호작용 전략에 따라 산출물의 방향, 범위, 정확성, 유용성 등이 달라질 수 있기에 학습자가 어떠한 방식으로 과업을 구조화시켜 산출물을 생성 및 검토하는지가 중요하다[5, 6]. 이러한 맥락에서 Yan et al.[4]은 생성형 AI 활용 역량을 단순한 사용 여부가 아니라 프롬프트 작성, 산출물 검증, 윤리적 위험 평가 등 고차적 활용 요소를 포함하는 다차원적 개념으로 제시하였다. 이에 따라 생성형 AI 고도활용은 생성형 AI를 학습 및 문제해결 과정에서 전략적으로 활용하고, 그 산출물을 비판적으로 점검하는 활용 양상으로 이해할 수 있다[4]. 이를 바탕으로 본 연구에서는 생성형 AI 고도활용을 프롬프트 작성 전략과 비판적 검증의 두 하위변인으로 구성하였다.

프롬프트 작성 전략은 생성형 AI로부터 원하는 산출물을 얻기 위해 과업의 목적, 조건, 맥락, 산출물의 형식과 수준을 구체화하는 전략적 행위를 의미한다[5, 6]. 생성형 AI는 동일한 주제라도 사용자가 어떤 방식으로 질문하고 조건을 제시하는가에 따라 산출물의 질이 달라질 수 있으므로, 프롬프트 작성은 단순한 입력 행위가 아니라 AI와의 상호작용을 조정하는 핵심 전략으로 볼 수 있다[6]. 이에 대해 Knoth et al.[6]은 프롬프트 엔지니어링이 목표지향적 생성형 AI 활용을 가능하게 하는 중요한 전략임을 강조하였으며, 권이은과 이용준[12] 또한 대학생의 프롬프트 유형은 AI 리터러시 수준에 따라 달라질 수 있다고 하였다.

비판적 검증은 생성형 AI가 산출한 정보나 결과를 그대로 수용하지 않고, 그 정확성, 근거, 출처, 논리적 타당성, 편향, 윤리적 위험을 점검하는 전략을 의미한다[4, 8]. 이는 생성형 AI가 문장 수준에서는 자연스럽고 설득력 있는 산출물을 생성할 수 있으나, 실제로는 부정확한 정보, 존재하지 않는 출처, 편향된 해석, 맥락에 부합하지 않는 결과를 제시할 수 있기 때문이다[1, 4, 14]. 따라서 비판적 검증은 AI 활용에 대한 단순한 주의나 태도가 아니라, 산출물의 신뢰성과 타당성을 판단하는 실천적 전략으로 이해할 필요가 있다.

2.3 AI 리터러시와 생성형 AI 고도활용의 관계

앞서 살펴본 바와 같이 AI 리터러시는 AI 전반에 대한 이해, 활용, 평가, 윤리적 판단을 포괄하는 다차원적 역량이다[4, 9-11]. 반면 생성형 AI 고도활용은 생성형 AI 사용 측면에서 나타나는 구체적인 활용 전략에 초점을 둔다. 즉, AI 리터러시가 AI 전반에 대해 학습자가 무엇을 이해하고, 어떻게 평가하며, 어떠한 윤리적 판단을 할 수 있는지를 살펴보는 역량이라면, 생성형 AI 고도활용은 생성형 AI와 상호작용하는 과정에서 학습자가 어떻게 프롬프트를 구성하고, 생성된 산출물을 어떻게 검토하는지를 설명하는 개념이다[4-6, 12]. 예를 들어, AI 리터러시에서 AI 평가는 산출물의 신뢰성을 판단하는 인지적·역량적 수준(perceived competency)을 측정하는 반면, 생성형 AI 고도활용의 비판적 검증은 생성형 AI 사용 과정에서 나타나는 구체적인 행동 전략(behavioral strategy)이라는 점에서 구분된다[4, 10]. 따라서 두 개념 모두 AI 활용 역량과 관련되지만, AI 리터러시는 AI에 대한 포괄적 역량 수준을, 생성형 AI 고도활용은 생성형 AI 사용 시 전략적 활용 양상을 설명한다는 점에서 구분된다. 그러므로 생성형 AI를 많이 활용한다고 하더라도 그것이 곧 AI 리터러시 수준이 높다는 것을 의미하지는 않기에 두 변인 간의 관계를 들여다볼 필요가 있다.

3. 연구 방법

3.1 연구대상

본 연구는 수도권 소재 공학계열 중심 T 대학의 재학생을 대상으로 진행되었다. 자료는 2026년 2월 4일부터 13일까지 약 열흘간 온라인 Google 설문을 통해 수집되었다. 설문 참여자에게는 연구목적과 개인정보 수집 여부, 자료의 사용 범위 등을 사전에 안내하였다. 그리고 수집된 자료 중 불성실 응답, 중복 응답, 결측 응답 등을 제거하여 최종 143명의 응답자료를 분석에 활용하였다. 연구대상의 일반적 특성인 성별, 학년, 소속대학에 대한 자세한 결과는 Table 1과 같다.

3.2 연구 대상

본 연구의 대상은 초등학교 교사 122명으로, 초등교사 온라인 커뮤니티를 통해 자발적으로 참여 의사를 밝힌 교사들이다. 사전 설문에는 총 122명이 참여하였으며, 일반적인 특성은 Table 1과 같다.

Table 1. General Characteristics of Research Participants

Characteristic	Category	*N*	%
Gender	Male	85	59.4
	Female	58	40.6
Grade	Freshman	13	9.1
	Sophomore	27	18.9
	Junior	34	23.8
	Senior or above	69	48.3
College/ Faculty	Software	30	21.0
	IT-Semiconductor Convergence	30	21.0
	Smart Mechanical Convergence	27	18.9
	Advanced Convergence	35	24.5
	Business Data Science	12	8.4
	Design Engineering	9	6.3

Table 2. Measurement Variables and Reliability

Variable	Subfactor	Representative Item	No. of Items	α
AI literacy	AI understanding	I can explain the advantages and disadvantages of AI.	3	.882
	AI use	I can use AI to get the results I need.	3	.745
	AI evaluation	I can explain the risks that arise from the use of AI systems.	3	.717
	AI ethics	I can ethically judge the results or actions derived by AI.	3	.689
Advanced Generative AI use	Prompting strategy	I refine my prompts based on the answers provided by Gen AI.	4	.647
	Critical verification	I verify the sources of information generated by AI tools.	3	.859

3.2 연구도구

이 연구에서 활용된 변인은 AI 리터러시와 생성형 AI 고도활용이다. AI 리터러시는 AI 이해, AI 활용, AI 평가, AI 윤리의 네 가지 하위요인으로 구성되며, 임철홍 외[10]와 Yan et al.[4]의 선행연구에 기반하여 문항을 구성하였다. 생성형 AI 고도활용은 Yan et al.[4]의 선행연구를 바탕으로 프롬프트 작성 전략과 비판적 검증의 두 하위요인으로 구성하였다. 모든 문항은 내용타당도 확보를 위해 교육학(교육측정 및 평가, 교육공학 등) 박사 5인의 검토를 거쳐 연구의 목적에 맞게 수정 및 번안되었으며, Likert 5점 척도로 측정하였다.

아울러 측정도구의 내적일관성을 확인하기 위해 Cronbach's α를 산출하였다. Cronbach's α는 .6 이상을 최소 허용 가능한 수준으로 해석할 수 있다는 Hair et al.[15]의 기준을 참고하였다. 그 결과 AI 리터러시 전체의 신뢰도는 .848, 생성형 AI 고도활용 전체는 .782로 나타났다. 다만 프롬프트 작성 전략의 Cronbach's α는 .647로 다른 하위요인에 비해 상대적으로 낮게 나타났다. 하지만 프롬프트 작성 전략 문항이 포함된 Yan et al.[4]의 활용(usage)변인에서도 Cronbach's α가 .68로 나타났다는 점을 고려하고, 확인적 요인분석 결과 합성신뢰도(CR)이 .746으로 나타났다는 점을 종합하였을 때 탐색적 분석에 활용할 수 있는 수준으로 판단하였다. 측정도구에 대한 상세한 내용과 하위 변인의 신뢰도 검정 결과는 다음 Table 2와 같다.

이어서 측정도구의 타당도를 추가로 검토한 결과 AI 리터러시의 4요인 모형은 CFI=.996, RMSEA=.023, 생성형 AI 고도활용의 2요인 모형은 CFI=1.000, RMSEA=.008로 나타나 양호한 구성타당도를 보였다. 표준화 요인부하량은 AI 리터러시가 .582~.982, 생성형 AI 고도활용이 .596~.933이었으며, 합성신뢰도(CR)는 각각 .735~.915와 .746~.902로 나타나 수렴타당도도 수용 가능한 수준이었다. 또한, 전체 하위요인 간 HTMT가 .90 미만으로 나타나 판별타당도 역시 전반적으로 확보된 것으로 판단하였다[16].

3.3 연구절차 및 분석방법

자료 분석을 위해 활용된 프로그램은 R version 4.4.2이다. 자료 전처리에는 readxl과 dplyr, 신뢰도 및 기술통계 분석에는 psych, 대응표본 t-test에는 stats, 잔차 기반 분류에는 base와 stats, Welch ANOVA와 Games-Howell 사후검정에는 rstatix 패키지를 활용하였다.

본격적인 분석에 앞서 수집된 자료의 결측치와 이상치를 확인한 후 데이터 분석에 투입한 뒤 AI 리터러시와 그 하위요인, 생성형 AI 고도활용과 그 하위요인의 평균 및 표준편차, 신뢰도 계수를 산출하였다. 그리고 각 하위변인의 측정문항이 3~4개의 수준인 점을 고려하여 확인적 요인분석을 수행하였다. 구성타당도는 CFI와 RMSEA, 수렴타당도는 표준화 요인부하량과 합성신뢰도(CR), 판별타당도는 HTMT를 산출하여 .85의 엄격한 기준과 .90의 완화된 기준을 토대로 검토하였다[16].

다음으로 대응표본 t-test를 실시하여 프롬프트 작성 전략과 비판적 검증의 평균 수준을 비교하였다. 그리고 생성형 AI 고도활용 유형화를 위해 잔차 기반 분류를 수행하였다.

잔차 기반 분류는 한 변인의 수준을 기준으로 다른 변인의 기대수준을 추정하고, 실제 관측값과 예측값의 차이

를 통해 상대적 과잉이나 결핍 양상을 파악하는 접근법이다[17]. 이는 연구의 목적이 프롬프트 작성 전략과 비판적 검증의 단순 평균 혹은 총점에 따른 유형화가 아니라, 프롬프트 작성 전략 수준과 비판적 검증 수준의 상대적 양상을 살펴보는데 있기 때문에 잔차 기반 분류를 통해 유형을 분류하였다. 이를 위해 프롬프트 작성 전략과 비판적 검증 점수를 표준화한 뒤에 비판적 검증을 종속변인, 프롬프트 작성 전략을 독립변인으로 하는 단순 회귀모형으로 설정하였다. 프롬프트 작성 전략이 AI 활용에 대한 선행 행위이며[5, 6], 비판적 검증은 산출물 생성 이후에 이루어지는 후속 행위이기 때문이다[4, 8]. 이는 두 전략 간의 인과적 선후관계를 가정한 것이 아니라, 프롬프트 작성 전략 수준을 고려했을 때 비판적 검증이 상대적으로 어느 정도 동반되는지를 파악하기 위한 분석적 설정이다. 따라서 실제 비판적 검증 점수와 회귀모형에 의해 예측된 비판적 검증 점수의 차이(잔차)를 활용하여 프롬프트 작성 전략 수준 대비 비판적 검증의 상대적 높낮이를 판단하였다[18].

구체적으로 프롬프트 작성 전략은 5점 Likert 척도 기준 동의 수준에 해당하는 4점을 기준으로 4점 이상이면 높은 수준, 4점 미만이면 낮은 수준으로 구분하였다. 비판적 검증은 회귀모형에서 산출된 잔차의 부호를 기준으로 잔차가 0 이상이면 비판적 검증이 기대 수준 이상으로, 잔차가 0 미만이면 기대 수준보다 낮은 것으로 해석하였다. 이에 따라 프롬프트 작성 전략의 절대 수준과 비판적 검증의 상대적 수준을 조합하여 네 가지 생성형 AI 고도활용 유형을 분류하였다.

마지막으로 도출된 생성형 AI 고도활용 유형에 따른 AI 리터러시 수준 차이를 검증하기 위해 Welch ANOVA를 수행하였다. 이는 유형 간 표본 수가 동일하지 않고, 등분산 가정을 충족시키지 못할 경우를 고려하여 선택하였다[18]. 이때, 집단 간 차이가 유의한 경우 Games- Howell 사후검정을 추가 실시하여 구체적으로 어떤 유형 간의 차이가 나타나는지 확인하였다. Games-Howell 사후검정방법은 집단 간 분산이 다르거나 표본 수가 동일하지 않은 상황에 적합한 사후검정 방법이다[19].

4. 연구 결과

4.1 변인별 기술통계 및 생성형 AI 고도활용 하위전략 차이

AI 리터러시 전체 및 하위변인, 생성형 AI 고도활용 전체 및 하위변인에 대한 기술통계 결과는 Table 3과 같다. 분석 결과, AI 리터러시 전체 평균은 3.73으로 나타났으며, 하위요인별로는 AI 이해 3.16, AI 활용 3.97, AI 평가 3.81, AI 윤리 3.98로 나타났다. 생성형 AI 고도활용 전체 평균은 4.06이었으며, 하위요인인 프롬프트 작성 전략은 4.14, 비판적 검증은 3.95로 나타났다.

변인별로 살펴보면 AI 리터러시에서는 AI 이해가 3.16으로 가장 낮았으며, 생성형 AI 고도활용에서는 비판적 검증의 수준이 3.95로 프롬프트 작성 전략보다 낮았다. 모든 하위변인을 종합하여 살펴보면 AI 이해가 가장 낮은 수준으로 나타났다. 이는 공학계열 대학생들이 코딩, 설계, 자료분석 및 과제 수행 등의 과정에서 생성형 AI를 실용적으로 활용하는 경험은 축적하고 있지만[7, 20], AI의 작동원리와 장단점, 기술적 한계에 대한 개념적 이해는 활용 수준보다 충분히 형성되지 않았을 가능성을 보여준다.

Table 3. Descriptive Statistics of Variables

Variable	*M*	*SD*
AI literacy(overall)	3.73	.652
AI understanding	3.16	1.22
AI use	3.97	.725
AI evaluation	3.81	.828
AI ethics	3.98	.715
Advanced Generative AI use(overall)	4.06	.583
Prompting strategy	4.14	.532
Critical verification	3.95	.907

이어서 공학계열 대학생들의 생성형 AI 활용 측면에 있어 프롬프트 작성 전략과 비판적 검증이 균형을 보이는지를 확인하기 위해 대응표본 t-test를 수행하였다(Table 4). 그 결과 두 변인 간 차이는 p=.007 수준에서 유의하였다. 즉, 공학계열 학생들의 프롬프트 작성 전략 수준(M=4.14)은 비판적 검증 수준(M=3.95)보다 높은 것으로 나타났다.

Table 4. Paired t-test results

Comparison	*t*	*df*	*p*
Prompting strategy − Critical verification	2.745	142	.007

4.2 생성형 AI 고도활용 유형화 결과

다음으로 두 변인의 평균 차이만으로는 설명할 수 없는 개인의 상대적 결합 양상을 확인하기 위해 잔차 기반 분류를 수행하였다. 유형화 결과 생성형 AI 고도활용 유형은 전반적 중저활용형(A), 고프롬프트-검증결핍형(B), 저프롬프트-검증우세형(C), 전반적 고활용형(D)의 네 가지로 나타났다(Table 5). 가장 많은 분포를 보인 유형은 전반적 고활용형(D)이 57명(39.9%)이었고, 고프롬프트-검증결핍형(B)이 46명(32.2%)으로 그 뒤를 이었다. 전반적 중저활용형(A)은 25명(17.5%), 저프롬프트-검증우세형(C)은 15명(10.5%)으로 비교적 낮은 분포를 보였다.

Table 5. Paired t-test results

Profile	*N*	%	*PS* *M(SD)*	*CV* *M(SD)*	Residual
A	25	17.5	3.55(.228)	2.72(.636)	-.90(.671)
B	46	32.2	4.41(.362)	3.56(.659)	-.64(.635)
C	15	10.5	3.38(.376)	4.42(.462)	1.11(.570)
D	57	39.9	4.38(.357)	4.67(.342)	.62(.317)

Note. A: Low-to-moderate use, B: High-prompting/ verification-deficit, C: Low-prompting/verification-dominant, D: High use

4.3 생성형 AI 고도활용 유형화 결과에 따른 AI 리터러시 차이

도출된 생성형 AI 고도활용 유형에 따른 AI 리터러시 전체 및 하위변인 수준에 차이가 있는지를 검증하기 위해 Welch ANOVA를 실시하였다(Table 6).

그 결과 AI 리터러시 전체와 각 하위변인별 집단 간 차이가 유의한 것으로 나타났다. 이에 따른 Games- Howell 사후검정 결과 AI 리터러시 전체에 대해 고프롬프트-검증결핍형(B)집단과 전반적 고활용형(D)집단은 각각 중저활용형(A)집단과 저프롬프트-검증우세형(C)집단에 비해 높은 AI 리터러시 수준을 가지고 있는 것으로 나타났다.

AI 리터러시의 하위변인별로 살펴보면 AI 이해에서만 p=.01 수준에서 생성형 AI 고도활용 유형별 차이가 유의하였으며, 나머지 하위변인에 대해서는 p<.001 수준에서 유의한 집단별 차이가 나타났다.

AI 이해에서는 고프롬프트-검증결핍형(B)집단이 저프롬프트-검증우세형(C)집단보다 높은 수준을 보였고, AI 활용에서는 AI 리터러시 전체와 동일하게 고프롬프트-검증결핍형(B)집단과 전반적 고활용형(D)집단이 각각 중저활용형(A)집단과 저프롬프트-검증우세형(C)집단보다 높은 것으로 나타났다. 그리고 AI 평가에서는 고프롬프트-검증결핍형(B)집단과 전반적 고활용형(D)집단이 중저활용형(A)집단보다 높았다. AI 윤리에서는 중저활용형(A)집단이 타 집단에 비해 가장 낮은 수준을 보였다.

Table 6. Differences in AI Literacy by Advanced Generative AI Use Profiles

Variable	AI literacy (overall)	AI understanding	AI use	AI evaluation	AI ethics
A *M(SD)*	3.19 (.642)	2.68 (1.321)	3.41 (.722)	3.23 (.875)	3.43 (.691)
B *M(SD)*	3.90 (.570)	3.40 (1.117)	4.14 (.556)	3.99 (.692)	4.05 (.700)
C *M(SD)*	3.22 (.656)	2.47 (1.104)	3.13 (.862)	3.20 (1.265)	4.07 (.715)
D *M(SD)*	3.97 (.499)	3.36 (1.183)	4.29 (.472)	4.08 (.527)	4.14 (.633)
Welch F	13.66	4.18	16.56	8.32	6.63
df2	46.275	48.797	44.129	43.062	47.895
p	<.001	.010	<.001	<.001	<.001
Post-hoc	B,D>A,C	B>C	B,D>A,C	B,D>A	B,C,D>A

Note. A: Low-to-moderate use, B: High-prompting/ verification-deficit, C: Low-prompting/verification-dominant, D: High use

5. 결론 및 제언

본 연구는 공학계열 대학생을 대상으로 생성형 AI 고도활용 유형에 따른 AI 리터러시 전체 및 하위변인 별 어떠한 차이가 나타나는지 살펴보기 위한 목적으로 수행되었다. 이를 위해 생성형 AI 고도활용을 프롬프트 작성 전략과 비판적 검증의 결합 양상으로 보고, 두 하위변인을 잔차 기반 분류방법으로 유형화하였다. 그리고 유형별 AI 리터러시 수준 차이를 검증하였다. 이에 따른 주요 결론 및 제언은 다음과 같다.

첫째, 공학계열 대학생의 생성형 AI 고도활용 변인 (프롬프트 작성 전략, 비판적 검증) 간 유의한 차이(=2.745)가 나타났다. 분석 결과 공학계열 대학생들은 생성형 AI를 사용하는 데 있어 비판적 검증보다 프롬프트 작성 전략을 더 높은 수준으로 활용하고 있었다. 이는 공학계열 학생들이 생성형 AI를 적극적으로 활용하고 있다고 보고한 Fan et al.[20]의 연구와 동일한 맥락의 결과이다. 또한, Martin-Moncunill & Alonso[8]가 주장한 바와 마찬가지로 학생들은 생성형 AI를 활용할 때 원하는 산출물을 얻기 위해 질문을 조정하거나 조건을 제시하는 전략은 적극적으로 펼치지만, 생성된 산출물의 정확성, 근거, 출처, 편향 등을 점검하는 과정은 상대적으로 덜 수행하고 있음을 보여준다. 그러므로 생성형 AI 활용 교육에서는 프롬프트 작성법을 교육하는 것과 더불어 생성된 산출물을 평가하고 검토하는 과정도 중요하게 다루어질 필요가 있다[21]. 생성형 AI 고도활용 유형 분류 결과 전반적 고활용형 다음으로 고프롬프트-검증결핍형에 해당하는 학생들이 많았다는 점에서도 생성형 AI 정보 검증 교육의 중요성은 더욱 높다고 할 수 있다.

둘째, 학습자들의 생성형 AI 고도활용 수준은 프롬프트 작성 전략과 비판적 검증 양상에 따라 상이한 유형으로 나뉘었다. 즉, 잔차 기반 분류 분석 결과 전반적 중저활용형, 고프롬프트-검증결핍형, 저프롬프트-검증우세형, 전반적 고활용형의 네 가지 유형으로 도출되었다. 여기서 주목할 점은 프롬프트 작성 전략이 높았던 학생 모두가 동일한 패턴을 보이지는 않았다는 점이다. 프롬프트 작성 전략이 높았던 학생 중에서도 일부는 비판적 검증도 함께 높은 유형으로, 일부는 비판적 검증이 상대적으로 낮은 유형으로 분류되었다. 이는 프롬프트를 능숙하게 작성하는 능력이 산출물을 검증하는 능력으로까지 이어지지는 않을 수 있음을 의미한다[2, 4].

셋째, 생성형 AI 고도활용 유형에 따른 AI 리터러시 전체 및 하위변인 수준 간에 유의한 차이가 나타났다. AI 리터러시 전체 수준에서는 전반적 고활용형과 고프롬프트-검증결핍형이 전반적 중저활용형과 저프롬프트-검증우세형보다 높은 수준을 가진 것으로 나타났다. 이는 프롬프트 작성 전략을 적극적으로 활용하는 학생들이 AI 리터러시 전체 수준에서도 상대적으로 높은 자기인식 수준을 보일 수 있음을 의미한다[4, 6]. 다만 두 유형은 AI 평가를 포함한 전반적 AI 리터러시 수준이 유사하게 높았음에도 생성형 AI 활용의 비판적 검증에서는 서로 다른 수준을 보였다. 이는 AI 산출물을 평가할 수 있다는 역량 인식이 실제 생성형 AI 활용 과정

에서 출처, 근거 및 정확성을 검증하는 전략으로 반드시 이어지는 것은 아님을 보여준다.

넷째, 저프롬프트-검증우세형은 전체 유형 중 가장 낮은 비율을 차지하였지만, 생성형 AI 고도활용의 불균형 양상을 보여준다는 점에서 주목할만한 유형이다. 이 유형의 AI 리터러시 수준을 함께 살펴보면 비판적 검증과 AI 윤리 수준은 상대적으로 높았지만, 프롬프트 작성 전략, AI 이해, AI 활용 및 AI 리터러시 전체 수준은 상대적으로 낮게 나타났다. 이는 AI 산출물을 신중하게 검토하고 윤리적 위험을 고려하려는 태도가 형성되어 있더라도 생성형 AI를 적극적으로 활용하거나 과업의 목적에 맞게 프롬프트를 구성하는 역량까지 함께 높아지지는 않는다는 것을 보여준다. 따라서 해당 유형의 학생에게는 기존의 검증지향적 태도를 유지하도록 하면서 과업 구조화, 프롬프트 구체화 및 AI와의 반복적 상호작용을 지원하는 교육이 제공될 필요가 있다.

마지막으로 AI 리터러시와 그 하위변인을 살펴보면 AI 윤리의 경우 예상과 다른 패턴이 도출되었다. 먼저 저프롬프트-검증우세형은 AI 리터러시 전체와 AI 활용에서는 상대적으로 낮은 수준을 보였으나, AI 윤리에서는 전반적 중저활용형보다 유의하게 높은 수준을 보였다. 이는 생성형 AI 활용에서 AI 산출물에 대한 신중한 태도나 검증지향 인식이 윤리적 판단과 연결될 수 있음을 시사한다. 하지만 고프롬프트-검증결핍형에서는 비판적 검증이 상대적으로 낮은 유형임에도 AI 윤리 수준이 낮게 나타나지 않았다.

AI 윤리와 같이 책임 있는 활용, 공정성, 편향, 개인정보 보호 등 사회적으로 바람직한 판단이 요구되는 영역에서는 응답자가 자신의 실제 수행 수준보다 바람직한 방향으로 응답할 가능성, 즉 사회적 바람직성 편향(social desirability bias)을 배제하기 어렵다[22]. 따라서 본 연구에서 나타난 AI 윤리 수준은 학습자가 인식한 윤리적 판단 수준을 보여주는 결과로 이해해야 하며, 이것이 실제 산출물 검증, 출처 확인, 오류 탐지, 윤리적 의사결정 수행을 충분히 담보한다고 보기는 어렵다. Hornberger et al.[11] 또한 AI 리터러시 측정에서 자기보고식(self-reported) 평정이 실제 지식이나 수행역량과 차이를 보일 수 있음을 지적하였다. 그러므로 후속연구에서는 자기보고식 AI 윤리 문항과 AI 산출물의 오류 식별, 출처 검증, 편향 판단, 윤리적 딜레마 상황에 대한 의사결정 과제 등 실제 역량 수행기반 평가를 토대로 AI 윤리 인식과 실제 윤리적 수행 간의 관계를 검토할 필요가 있다.

본 연구는 다음과 같은 한계를 갖는다. 첫째, 연구대상이 수도권 소재 특정 대학의 공학계열 중심 학생으로 제한되므로 연구결과를 전체 대학생이나 모든 전공계열로 일반화하는 데에는 한계가 있다. 둘째, AI 리터러시와 생성형 AI 고도활용은 자기보고식 설문자료를 통해 측정되었으므로 본 연구에서 나타난 수준은 실제 수행역량 그 자체라기보다 학습자가 인식한 역량 수준으로 해석될 필요가 있다. 따라서 후속연구에서는 다양한 전공계열과 표본을 대상으로 실제 프롬프트 작성 과제, AI 산출물 오류 탐지, 출처 검증 등의 수행기반 평가를 병행하여 생성형 AI 활용역량을 측정할 필요가 있다. 한편, 본 연구의 표본에서 4학년이 상대적으로 높은 비율을 차지한 점은 프롬프트 작성 전략의 전체 평균을 끌어올리는 데 영향을 미쳤을 가능성이 있다. 학년별 응답 양상을 확인한 결과, 1학년(M=3.67)은 2~4학년(2학년: M=4.19, 3학년: M=4.16, 4학년: M=4.20)보다 프롬프트 작성 전략 수준이 상대적으로 낮았으나, 2학년부터는 학년별 차이가 크지 않았다. 이는 프롬프트 작성 전략이 학년 상승에 따라 지속적으로 증가한다기보다는, 전공수업과 프로젝트 등 생성형 AI 활용 경험이 본격화되는 시기에 일정 수준까지 형성될 가능성을 시사한다.

그럼에도 국내에서는 아직까지 생성형 AI 고도활용 변인을 바탕으로 학습자의 생성형 AI 활용 유형을 구분하고, 유형에 따른 AI 리터러시 수준 차이를 실증적으로 분석한 연구는 찾을 수 없었다. 따라서 본 연구는 공학계열 대학생의 생성형 AI 활용 유형을 분석하고, 이를 바탕으로 생성형 AI 교육에서 비판적 검증의 중요성을 제언하였다는 점에서 의의가 있다.

참고문헌

[1] Zhai, C., Wibowo, S., & Li, L. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, 28. https://doi.org/10.1186/s40561-024-00316-7

[2] Yeung, L., Chow, D., Wong, P., & Lau, S. (2026). In ChatGPT they trust: A study of students' perceptions and misuse of ChatGPT in higher education. *AI and Ethics, 6*, 11. https://doi.org/10.1007/s43681-025-00855-w

[3] Tian, J., & Zhang, R. (2025). Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. *Acta Psychologica, 260*, 105725. https://doi.org/10.1016/j.actpsy.2025.105725

[4] Yan, W., Liu, Y., Mamaeva, V., Dong, F., Tao, G., Li, R., & Yang, H. (2026). Generative AI literacy: Scale development and its influence on privacy protection behaviors and information verification behaviors. *Telecommunications Policy, 50*, 103117. https://doi.org/10.1016/j.telpol.2025.103117

[5] Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2025). Generative AI literacy: Twelve defining competencies. Digital Government: *Research and Practice, 6*(1), 13. https://doi.org/10.1145/3685680

[6] Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

[7] Han, O. (2024). Generative AI and critical thinking in engineering education: A study on application of prompt engineering. *Journal of Engineering Education Research, 27*(6), 38-47. https://doi.org/10.18108/jeer.2024.27.6.38
[8] Martín-Moncunill, D., & Alonso Martínez, D. (2025). Students' trust in AI and their verification strategies: A case study at Camilo José Cela University. *Education Sciences, 15*(10), 1307. https://doi.org/10.3390/educsci15101307
[9] Chen, K., Tallant, A., & Selig, I. (2025). Exploring generative AI literacy in higher education: Student adoption, interaction, evaluation and ethical perceptions. *Information and Learning Sciences, 126*(1/2), 132-148. https://doi.org/10.1108/ILS-10-2023-0160
[10] Im, C., Cho, Y., & Jeong, D. (2024). A study on developing of AI literacy scale for college students. *Journal of Information Technology and Architecture, 21*(3), 285-296. https://doi.org/10.22865/jita.2024.21.3.285
[11] Hornberger, M., Bewersdorff, A., Schiff, D., & Nerdel, C. (2025). Development and validation of a short AI literacy test (AILIT-S) for university students. *Computers in Human Behavior: Artificial Humans, 5*, 100176. https://doi.org/10.1016/j.chbah.2025.100176
[12] Kwon, E., & Lee, Y. (2025). A study on university students' prompt types in generative AI: Focusing on the discussion of prompt literacy. *The Korean Journal of Literacy Research, 16*(5), 115-144. https://doi.org/10.37736/KJLR.2025.10.16.5.04
[13] Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., et al. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLOS ONE, 20*(2), e0315011. https://doi.org/10.1371/journal.pone.0315011
[14] Holzmann, U., Anand, S., & Payumo, A. (2025). The ChatGPT Fact-Check: Exploiting the limitations of generative AI to develop evidence-based reasoning skills in college science courses. *Advances in Physiology Education, 49*, 191-196. https://doi.org/10.1152/advan.00142.2024
[15] Hair, J., Hult, G., Ringle, C., Sarstedt, M., Danks, N., & Ray, S. (2021). Evaluation of reflective measurement models. *In Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. (pp. 75-90). Springer. https://doi.org/10.1007/978-3-030-80519-7_4
[16] Henseler, J., Ringle, C., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8
[17] Gnas, J., Urban, J., Feuchter, M., & Preckel, F. (2024). Socio-emotional experiences of primary school students: Relations to teachers' underestimation, overestimation, or accurate judgment of their cognitive ability. *Social Psychology of Education, 27*, 2417-2454. https://doi.org/10.1007/s11218-024-09915-1
[18] Delacre, M., Leys, C., Mora, Y., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology, 32*(1), 13. https://doi.org/10.5334/irsp.198
[19] Games, P., & Howell, J. (1976). Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. *Journal of Educational Statistics, 1*(2), 113-125. https://doi.org/10.3102/10769986001002113
[20] Fan, L., Deng, K., & Liu, F. (2025). Educational impacts of generative artificial intelligence on learning and performance of engineering students in China. *Scientific Reports, 15*, 26521. https://doi.org/10.1038/s41598-025-06930-w
[21] Oates, A., & Johnson, D. (2025). ChatGPT in the classroom: Evaluating its role in fostering critical evaluation skills. *International Journal of Artificial Intelligence in Education, 35*, 1793-1824. https://doi.org/10.1007/s40593-024-00452-8
[22] Ling, Y., Kale, A., & Imas, A. (2026). Underreporting of AI use: The role of social desirability bias. Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, *Barcelona, Spain, 62*, 1-22. https://doi.org/10.1145/3772318.3791073

이대영
· 2018년 건국대학교 유아교육과(교육학사)
· 2021년 건국대학교 교육공학전공(교육학석사)
· 2024년 건국대학교 교육공학전공(교육학박사)
· 2024년~현재 한국공학대학교 전공교육혁신센터 연구교수
관심분야 : AI, 빅데이터분석, 에듀테크, 핵심역량
2vs0@tukorea.ac.kr