Journal of Korean Library and Information Science Society
Vol. 57, No. 1

OPEN ACCESS

생성형 AI 출력 구조에 기반한 AI 리터러시 교육 모형 제안

A Process-Oriented AI Literacy Education Grounded in Generative AI Mechanisms

이 슬(Seul Lee)*

〈 목 차 〉

요 약 : 최근 생성형 인공지능(Generative Artificial Intelligence, 이하 생성형 AI)의 확산은 정보 탐색 방식과 교육 환경 전반에 변화를 가져오며, 인공지능 리터러시(Artificial Intelligence Literacy, 이하 AI 리터러시) 교육의 필요성을 더욱 부각시키고 있다. 이러한 변화는 생성형 AI가 결과를 생성하는 구조적 특성과 밀접하게 연관되어 있으며, 이를 이해하는 것은 효과적인 AI 리터러시 교육을 설계하는 데 중요한 기반이 된다. 이에 본 연구는 생성형 AI, 특히 트랜스포머 기반 대규모 언어 모델의 출력 생성 구조를 체계적으로 분석하고, 이를 토대로 생성형 AI 환경에서 요구되는 핵심 AI 리터러시 교육 요소를 도출하며, 나아가 이를 반영한 교육 모형을 제안하는 것을 목적으로 한다. 본 연구는 AI 리터러시 교육 요소로, 질문을 구조화하는 '문제 구성 리터러시', 생성형 AI의 계산 과정을 이해하는 '계산 리터러시', 확률적 생성 원리를 이해하는 '확률 생성 리터러시', 생성형 AI가 제시한 결과를 비판적으로 분석하고 의미를 재구성하는 '응답 해석 리터러시', 생성형 AI 응답이 플랫폼 설계와 정책에 의해 매개된다는 점을 인식하는 '플랫폼 리터러시', 그리고 생성된 정보를 검토하고 신뢰성을 판단하는 '비판적 평가 리터러시'를 포함할 것을 제안한다. 본 연구는 생성형 AI의 출력 구조를 기반으로 생성형 AI 환경에서 요구되는 리터러시 교육의 방향을 제시하였다는 점에서 의의를 갖는다.

주제어 : 생성형 인공지능, 인공지능 리터러시, AI 리터러시 교육, 생성형 AI 출력 구조, 대규모 언어 모델, 정보 편향성

ABSTRACT : The rapid spread of generative artificial intelligence(Generative AI) is reshaping how people seek and process information, increasing the need for AI literacy education. This study aims to identify the key elements of AI literacy education by examining the output generation structure of generative AI systems. To achieve this goal, the study theoretically analyzes how generative AI operates and produces responses, and explores how the structural characteristics of the generation process influence users' information judgment. The analysis suggests that outputs produced by generative AI should not be understood merely as information delivery; rather, they represent discursive reconstructions influenced by probabilistic computations and platform-level design considerations. Drawing upon these structural insights, the study proposes several critical elements for AI literacy education within generative AI contexts: (1) input structure literacy, which involves recognizing the structure of prompts and problem framing; (2) computational process literacy, involving comprehension of computational procedures such as tokenization and embedding; (3) generation principle literacy, focused on grasping the probabilistic foundations underpinning text generation; (4) response structure literacy, which refers to understanding how responses are organized and presented; (5) platform structure literacy, emphasizing awareness of the mediating influence of platform-level design; and (6) critical evaluation literacy, highlighting the capacity to critically assess generated outputs. By systematically organizing these elements in relation to the output structure of generative AI, this study contributes to the conceptual framework of AI literacy education and suggests directions for AI literacy education in generative AI environments.

KEYWORDS : Generative Artificial Intelligence, AI Literacy, Information Bias, Large Language Model, Information Behavior, Information Assessment

* 플로리다 주립대학교 정보대학(School of Information, Florida State University) 교수(seul.lee@fsu.edu)

• 논문접수: 2026년 3월 1일 • 최초심사: 2026년 3월 10일 • 게재확정: 2026년 3월 16일
• 한국도서관·정보학회지, 57(1), 51-80, 2026. http://dx.doi.org/10.16981/kliss.57.1.202603.51

※ Copyright © 2026 Korean Library and Information Science Society
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

Ⅰ. 서 론

1. 연구 배경

지난 수십 년간 방대한 양의 정보 자원이 디지털화되면서 이용자의 정보 탐색 행위는 웹 기반 환경, 특히 검색엔진을 중심으로 점차 재편되어 왔다. 특히, Google, Naver, Baidu와 같은 상업적 웹 검색엔진들은 디지털 정보에 접근을 매개하는 핵심 인프라로 자리매김하며, 현대 정보 행위의 주요 출발점이자 사실상 관문으로 기능해 왔다. 검색 플랫폼에 키워드 입력을 통해 관련 문서를 탐색하고, 비교・선별하는 검색 중심 정보 행위가 단순한 도구적 선택을 넘어, 웹 이용자들의 대표적인 정보 탐색 형태로 정착된 것이다.

2015년 12월 11일 공식 설립된 OpenAI는 불과 수년 만에 대규모 언어모델(Large Language Model, LLM) 기반 서비스인 ChatGPT를 상용화하며, 전 세계적으로 가장 영향력 있는 인공지능 플랫폼 중 하나로 급부상하였다(OpenAI, 2015; 2022). 특히 ChatGPT의 공개는 생성형 AI를 실험적 연구 영역에서 대중적 서비스 영역으로 확장시키는 전환점이 되었을 뿐만 아니라 대형 기술 기업(Big Tech)들로 하여금 생성형 AI 기반 인터페이스를 자사 서비스에 적극적으로 통합하도록 촉발하는 계기가 되었다. 생성형 AI 챗봇은 단기간 내 일부 이용자 집단에서 정보 탐색의 초기 접점이자 정보 탐색 과정의 새로운 진입 인터페이스로 부상하고 있는데, 실제 최근 미국 내 ChatGPT 이용자 1,000명을 대상으로 한 조사에 따르면, 응답자의 약 77%가 ChatGPT를 검색 엔진처럼 활용한다고 응답하였으며, 특히 이 중 Z세대 응답자의 28%는 ChatGPT로 검색을 시작한다고 응답했다(Adobe Express, 2025).

이러한 흐름 속에서 2024년 10월, OpenAI는 기존 대화형 모델에 실시간 웹 검색 기능을 통합한 “ChatGPT search”를 공식 출시하였다(OpenAI, 2024). OpenAI는 이 기능을 “더 나은 답에 도달하도록 설계된(Search designed to get you to a better answer)” 서비스라고 소개하며, 반복적 검색과 다수의 링크 탐색을 요구하는 기존 웹 검색과 달리, 대화 맥락을 반영한 자연어 질의응답 방식을 통해 보다 효율적인 정보 획득이 가능하다고 설명한다(OpenAI, 2024). 해당 기능은 외부 웹 데이터를 참조하여 최신 정보를 제공하고, 응답에 출처 링크를 포함함으로써 정보의 시의성(Relevancy)과 신뢰성(Trustworthiness)을 보완하는 것을 목표로 설계되었다고 밝히고 있다(OpenAI, 2024). 이는 기존 ChatGPT가 사전 학습 데이터에 기반한 확률적 생성 모델로 작동함에 따라 최신 사건 반영이나 명시적 출처 제시에 한계를 보여왔던 점을 보완하려는 시도로 해석될 수 있다. 나아가 이러한 접근은 기존 검색 방식의 한계를 부각시키는 동시에, 자사 서비스를 대화 맥락에 기반해 ‘더 나은 답’에 도달하도록 설계된 것으로 제시한다. 즉, 검색을 단순한 정보 검색(Query-result) 구조가 아니라, 맥락을 해석하고 응답을 구성하는 답변 생성 과정으로 재구성하

려는 전략적 프레이밍으로 이해할 수 있다.

Google의 Gemini(구 Bard)도 검색 엔진과 통합을 통해 실시간 정보 접근성을 강화하고 있으며(Google, 2026), xAI의 Grok 역시 소셜 미디어 X(구 Twitter) 데이터와 연동하여 최신 정보를 반영하려 하고 있다(xAI, 2026). 이는 생성형 AI가 고정된 학습 데이터에만 의존하는 언어 모델을 넘어, 외부 데이터 흐름과 결합하는 보다 동적인 정보 중개 구조로 확장되고 있음을 시사한다. 이처럼 최근 생성형 AI 도구들은 검색 기능과 플랫폼 데이터가 결합된 하이브리드 형태의 통합적 정보 접근 플랫폼으로 진화하려는 양상을 보이고 있다. 그럼에도 불구하고 2024년 기준 전 세계 검색 시장 점유율에서 Google을 비롯한 기존 검색 엔진은 여전히 압도적인 비중을 차지하고 있으며(First Page Sage, 2025; Statcounter, 2026), 검색 시장 점유율과 트래픽 구조를 고려할 때, 웹 검색은 여전히 온라인 정보 접근의 핵심 인프라로 기능하고 있다(DataReportal, 2024; Goodwin, 2025).

나아가 생성형 AI의 확산은 정보 접근 방식에 대한 이용자의 인식과 경험 자체를 변화시키고 있다. 기존의 웹 검색 중심 모델에서는 정보 접근과 정보 판단이 일정 부분 분리된 과정으로 작동한다. 이용자는 검색어를 입력한 뒤 검색엔진이 제시한 여러 문서를 탐색하며 각 문서의 내용, 관련성, 신뢰성을 비교・평가하고 이를 종합하여 최종적인 판단에 도달한다. 이 과정에서 이용자는 검색어 선택, 결과 해석, 문서 평가 등 다양한 인지적 판단을 수행하며 정보 탐색 전략을 능동적으로 조정한다. 여기서 '판단'이란 제시된 정보의 관련성과 신뢰성을 평가하고 그것이 자신의 문제 맥락에 적합한지를 검토하는 인지적 과정으로, 온라인 환경에서는 출처의 권위성과 신뢰성을 평가하는 활동까지 포함한다(Metzger, 2007; Wathen & Burkell, 2002).

반면 생성형 AI 챗봇 환경에서는 정보 탐색의 구조가 다르게 나타난다. 이용자는 다수의 개별 문서를 직접 탐색・비교・선별하기보다는 AI가 여러 정보를 종합하여 생성한 응답을 중심으로 내용을 검토하고, 필요에 따라 추가 질문을 통해 정보를 수정하거나 확장한다. 이 과정에서 이용자의 탐색은 외부 문서를 직접 종합하는 방식에서 벗어나, 플랫폼이 선행적으로 통합한 응답을 검토하고 반복적 질의를 통해 정보를 점진적으로 정교화하는 형태로 이루어진다. 즉 정보 탐색의 중심이 분산된 문서 공간에서 생성형 AI가 제공한 응답으로 이동하는 것이다. 특히 생성형 AI 챗봇의 정보 제시 방식은 기존 검색 환경에서 요구되던 비교・선별・종합의 과정을 부분적으로 압축하거나 재배치한다. 생성형 AI 챗봇은 관련 정보를 통합한 응답을 직접 생성・제시함으로써, 이용자의 판단이 개별 문서 선택이 아니라 이미 구성된 응답 전체의 타당성에 대한 평가로 이동하도록 만든다. 판단이 작동하는 대상과 구조가 변화하게 되는 것이다. 정리하면 생성형 AI 챗봇에서는 정보가 이미 하나의 설명 구조로 통합되어 완결된 형태로 제공되는 경향이 있으며, 이는 이용자 정보 행위의 중심이 "탐색과 비교"에서 "모델에 의해 선행적으로 통합・요약된 응답의 검토와 타당성 평가"로 이동하게 만든다.

2. 연구 목적

그렇다면 우리는 기존의 검색 결과 목록 대신 생성형 AI가 구성한 통합적 응답이 인터페이스 전면에 제시될 때, 이용자의 정보 탐색과 판단 구조가 어떻게 재편되는가에 대해 검토할 필요가 있다. 본 연구는 생성형 AI의 출력을 단순한 정보나 지식의 전달로 이해하는 관점을 재고하고, 생성형 AI가 제공하는 통합적 응답 형식과 그 인식론적 함의에 주목하고자 한다. 생성형 AI 환경에서는 이용자가 수행하던 문서 간 비교・선별・종합의 일부 과정이 시스템 수준에서 선행적으로 처리되며, 이때 이용자의 판단은 개별 문서를 평가하는 활동에서 이미 구성된 응답의 적절성과 타당성을 평가하는 활동으로 이동한다. 그러나 생성형 AI가 제시하는 응답은 검증된 지식 객체의 전달이라기보다, 학습된 언어 패턴과 확률적 추정을 바탕으로 지식 표현을 근사적으로 구성한 담론적 산출에 가깝다. 이에 본 연구는 다음과 같은 질문에서 시작한다. 첫째, 생성형 AI의 출력은 어떠한 구조적 특성을 가지는가? 둘째, 생성형 AI 환경에서, 이용자가 정보의 정확성(Accuracy), 권위(Authority), 객관성(Objectivity), 최신성(Currency), 적절성(Relevance)과 같은 평가 기준들(Fritch & Cromwell, 2001; Tate & Alexander, 1999)을 효과적으로 적용하기 위해서는 생성형 AI 리터러시가 어떠한 방향으로 개념적으로 재구성될 필요가 있는가? 본 논문에서는 AI 리터러시를 생성형 AI가 생성한 응답의 구성 방식과 한계를 이해하고, 그 응답의 타당성과 적절성을 비판적으로 평가하며, 필요할 경우 외부 정보원을 통해 검증하는 능력으로 정의한다.

이러한 문제의식을 바탕으로, 본 연구는 생성형 AI의 작동 원리와 응답 생성 과정을 분석하고자 한다. 특히, 현재 텍스트 생성형 AI의 주류를 이루는 트랜스포머 기반 대규모 언어 모델의 추론 메커니즘을 중심으로 생성형 AI의 작동 과정을 분석하고, 이를 통해 생성형 AI 환경에서 요구되는 AI 리터러시 교육의 핵심 요소를 도출하고자 한다. 특히 생성형 AI가 정보를 구성하고 제시하는 기술적 방식과 응답 형식이 이용자의 정보 인식과 판단 과정에 미치는 인식적 효과에 주목하여, 생성형 AI 기반 정보 환경에 적합한 AI 리터러시 교육의 재설계 방향을 제안하고자 한다. 이때 생성형 AI 출력 과정 분석의 초점은 오류의 빈도나 정확도 측정이 아니라, 생성형 AI에서 답변이 구성되는 기술적 방식과 그 형식이 발생시키는 인식적 효과에 있으며, 이를 바탕으로 생성형 AI 환경에 적합한 AI 리터러시 교육의 재설계 방향을 제안하고자 한다. 요컨대 본 연구는 기술의 위험을 경고하거나 기술 플랫폼의 비판하기 위한 논의가 아니다. 오히려 생성형 AI가 작동하는 구조를 이해함으로써, 답변의 형식과 판단의 조건 사이의 관계를 재정립하는 데 목적이 있다. 생성형 AI는 선행적으로 구성된 응답을 중심으로 정보 접근 구조를 재편하고 있으며, 이로 인해 정보 인식과 판단의 조건 또한 변화하고 있다. 이러한 변화 속에서 중요한 것은 기술에 대한 규범적 평가에 앞서 생성형 AI 출력이 어떠한 구조적 원리에서 이루어지는지, 그 구조가 만들어내는 강점과 한계가 무엇인지, 그 작동 방식이 우리의 인식 조건, 판단 방식, 정보 환경을 어떻게 재구성하고

있는지를 이해하는 일이며, 본 연구는 이러한 재구성의 구조를 분석하는 데 목적이 있다.

Ⅱ. 이론적 배경

1. 생성형 AI환경에서의 정보 인지 구조 변화

생성형 AI의 확산은 기존의 정보 탐색 환경뿐 아니라 이용자의 정보 인지 방식에도 변화를 가져오고 있다. 기존 검색 엔진은 이용자가 다양한 정보 출처를 탐색하고 비교하는 과정을 통해 정보를 해석하도록 하는 구조를 가진다. 반면, 생성형 AI는 여러 정보를 통합・재구성하여 하나의 설명 형태로 제시함으로써, 이용자의 정보 접근 방식과 판단 과정에 새로운 특성을 형성한다. 따라서 생성형 AI 환경에서의 정보 활용을 이해하기 위해서는, 기존 정보 탐색 방식과 구별되는 인지적 특성을 살펴볼 필요가 있다.

이러한 변화는 단순한 인터페이스나 정보 제공 방식의 차이에 그치지 않는다. 겉으로 보기에 검색 엔진과 생성형 AI는 모두 웹 데이터를 수집・처리・가공・전달하는 동일한 정보 처리 흐름 위에서 작동하는 것처럼 보일 수 있다. 그러나 검색 엔진이 문서 객체를 제시하고 선택의 과정을 이용자에게 남겨두는 반면, 생성형 AI는 해당 문서들로부터 학습된 언어적 패턴을 바탕으로 통합・요약・재구성한 하나의 응집된 담론을 생성한다.

이러한 전환은 이용자를 정보를 능동적으로 조합하는 구성자에서, 이미 구성된 설명을 받아들이는 존재로 바꿔 놓는다. 생성형 AI 환경에서 이용자는 분산된 문서를 직접 탐색하기보다 모델이 선행적으로 통합・요약한 응답을 중심으로 정보를 접하게 되며, 그 결과 이용자는 모델에 의해 재가공된 응답 구조를 출발점으로 삼아 판단을 수행하게 되는 것이다. 이는 이용자의 정보 탐색 과정에서 능동적 단계가 부분적으로 축소되거나 인지적 자율성이 상대적으로 약화될 가능성을 내포한다. 물론, 생성형 AI 환경에서 역시 판단의 주체가 인간임은 유지되지만, 판단이 작동하는 정보 구조가 플랫폼에 의해 선행적으로 구성되면서 이용자의 정보 탐색은 점차 응답 중심적(Response-centered)이고 수용 지향적인(Reception-oriented) 성격을 띠기 쉽다. 이용자는 생성형 AI가 제공한 정보를 평가, 수정해서 사용하거나 프롬프트 엔지니어링(Prompt engineering)을 통해 질문을 수정하며 보다 적극적으로 상호작용을 할 수도 있으나, 기본적으로 이용자는 자신에게 제공된 AI의 응답을 해체하거나 출처를 역추적하는 대신, 제시된 응답 단위를 중심으로 이를 수용, 평가하거나 보완하는 방식으로 이동할 가능성이 높다.

이는 생성형 AI의 대화형 인터페이스뿐만 아니라 이것에 의해 이루어지는 정보 처리의 외부화와 그에 따른 인지적 피로 감소와도 밀접하게 관련되어 있다. 생성형 AI는 복수의 정보를 모델 내부에서

통계적으로 통합·재구성함으로써, 이용자가 수행하던 탐색과 비교의 절차를 부분적으로 선행 처리한다. 그 결과 이용자는 분산된 문서를 직접 비교·검토하는 절차의 일부를 외부 시스템에 위임하는 형태의 인지적 외주화(Cognitive outsourcing)를 수행하게 되며(Clark & Chalmers, 1998), 맥락적으로 정합성을 갖춘 것처럼 인식되는 응답을 바로 접할 수 있다. 이는 이용자가 정보 탐색에 요구되는 인지적·시간적 비용을 크게 감소시킨다. 즉 생성형 AI는 이용자가 수행하던 높은 에너지와 집중을 요구하는 비교·선별·종합의 과정을 내부적으로 선행 처리함으로써 판단 이전 단계의 인지적 노동을 상당 부분 흡수한다. 결국 생성형 AI 환경에서 이용자는 기존 검색엔진 환경에서와 같이 랭킹된 다수의 정보를 비교·탐색하는 정보 탐색자보다는, 생성형 AI가 제공한 통합된 응답의 구조적 완결성이나 논지의 타당성을 평가하는 정보 검토자로 재구성될 가능성이 높다. 이러한 구조는 결과적으로 플랫폼이 구성한 정보 배열에 대한 의존성을 강화하는 방향으로 정보 행위를 재조직할 수 있다. 물론 이는 검색엔진 환경에서도 일정 부분 존재했던 플랫폼 의존성을 전제로 한다. 그러나 검색엔진이 문서의 배열을 매개하였다면, 생성형 AI는 의미의 통합 자체를 선행적으로 수행한다는 점에서 차이를 보인다. 특히 생성형 AI 환경에서는 비교·종합의 과정이 이용자에게 가시적으로 드러나지 않는다는 점에서, 검색엔진과 다른 방식으로 인식적 조건을 재구성한다. 그렇다면 검색엔진에서 생성형 AI챗봇으로의 전환은 단순한 인터페이스 변화가 아니라, 검색맥락에서 정보의 형식과 구성, 정보 판단의 구조 자체를 재편하는 과정으로 이해될 필요가 있다.

여기서 특히 주목할 점은, 이용자가 정보의 출처 경로나 원문을 직접 확인하기에 앞서, 생성형 AI 챗봇이 재구성한 하나의 통합된 설명을 먼저 접하게 된다는 점이다. 이로 인해 이용자의 정보 판단 과정 역시 변화한다. 즉, 정보의 타당성을 다양한 출처를 통해 교차 검증하기보다는, 생성형 AI가 제시한 응답 자체를 중심으로 판단이 이루어지는 경향이 나타날 수 있다. 이 과정에서 응답의 형식적 완결성이나 서술의 매끄러움은 내용의 진실성을 판단하는 준거로 작용할 가능성이 있으며, 이에 따라 생성형 AI의 답변은 충분한 검증 이전에 잠정적으로 승인될 가능성이 있다. 이는 생성형 AI가 제공하는 설명이 이미 인과적으로 정돈되어 있고, 대체로 결론까지 포함한 완결된 형태로 제시되기 때문이다.

이는 생성형 AI의 구조적 특징과 깊은 연관이 있다. 생성형 AI가 입력된 정보와 맥락을 실제로 이해하고 사실 여부를 판별하거나 외부 문서를 단순히 검색하여 전달하는 체계가 아니라, 대규모 언어 모델을 기반으로 확률적 패턴을 완성하는 모델이기 때문이다(West et al., 2023). 이 모델들은 고차원의 확률 공간에서 이용자의 프롬프트에 제시된 텍스트들의 다음 토큰을 예측하며, 조건화된 맥락 속에서 가장 그럴듯한 표현을 생성한다. 이러한 특성으로 인해 이 과정은 정답을 선택하는 방식이 아니라, 가능한 표현을 근사하는 방식이 된다. 따라서 인공지능 환각(AI hallucination)이나 질문의 프레이밍에 따른 답변의 담론 이동과 같은 현상은 생성형 AI 모델의 단순한 시스템적 결함이 아니라, 확률적 담론 생성 구조에서 예측 가능하게 나타나는 산출의 양상이다(Kalai et

al., 2025). 즉 생성형 AI 모델은 정보의 출처를 검증하는 주체가 아니라, 언어적 일관성을 최적화하는 구조에 가깝다(OpenAI Developers, 2025). 이때 생성형 AI가 제공하는 설명의 일관성(Consistency)과 정합성(Coherence), 언어적 유창성(Fluency)은 이용자에게 신뢰를 주는 단서로 사용될 수 있으나, 우리가 일반적으로 정보를 평가할 때 사용하는 정보의 사실적 정확성(Accuracy), 진실성(Veracity), 신뢰성(Credibility)과는 다른 개념이다. 즉 생성형 AI의 답변은 외부 세계의 사실과 일치하는 정보를 직접 판별하여 나온 결과물이라기보다, 이용자가 입력한 문맥 안에서 언어적으로 자연스럽고 논리적으로 매끄럽게 생성된 결과물에 가까운 것이다.

2. AI 리터러시에 대한 최근 국내 연구 동향

최근 국내 AI 리터러시 연구는 생성형 AI의 확산과 함께 빠르게 증가하며, 크게 세 가지 흐름 속에서 전개되어 왔다. 첫째, 생성형 AI 활용 경험을 분석하는 실증 연구, 둘째, AI 리터러시의 개념과 이론적 기반을 정립하려는 연구, 셋째, 이를 바탕으로 구성 요소와 역량 체계를 체계화하려는 연구이다.

먼저 실증 연구들은 학습자가 생성형 AI 기반 환경에서 경험하는 상호작용과 그에 따른 인지적·정동적 반응을 분석하며, 활용 양상과 태도에 대한 중요한 경험적 근거를 제공해 왔다(김욱태, 김영식, 2024; 민혜림, 2026; 배현영, 조재희, 2025). 그러나 이러한 연구는 주로 경험과 인식 수준에 초점을 두고 있어, 이러한 경험이 실제 학습 성과나 AI 리터러시의 구체적 하위 역량과 어떻게 연결되는지에 대한 정교한 분석은 상대적으로 제한적으로 다루어졌다.

한편 개념적 연구들은 AI 리터러시를 단순한 기술 활용 능력을 넘어, 변화하는 기술 환경을 이해하고 비판적으로 대응할 수 있는 포괄적 역량으로 확장해 왔다(박기범, 2022; 이승민, 2025; 이유미, 2022). 이러한 논의를 바탕으로 최근에는 AI 리터러시를 기술 이해, 비판적 이해, 활용 능력, 창의적 생산, 윤리적 판단 등을 포함하는 다층적 역량 체계로 구조화하려는 시도가 이루어지고 있다(정영미, 2025; 황현정, 황용석, 2023). 특히 일부 연구는 이를 지식·기술·태도의 세 차원으로 통합하고, AI의 이해·활용·분석·창조·비판적 평가 및 윤리성 등을 포함하는 체계를 제안하였다(황현정, 황용석, 2023). 다만 이러한 연구들은 주로 문헌 분석과 개념적 정리에 기반하고 있어, 제시된 역량이 실제 생성형 AI 활용 맥락에서 어떻게 나타나는지, 이용자의 실제 판단 과정에서 이러한 역량이 어떠한 방식으로 작동하는지에 대한 분석은 상대적으로 제한적이다.

이에 본 연구는 기존 연구와 달리, 생성형 AI의 출력 구조 자체를 분석함으로써 이용자와 시스템 간 상호작용 과정에서 요구되는 이해·판단·활용 역량을 도출하고자 한다. 이는 AI 리터러시를 추상적 개념이 아니라 실제 활용 맥락에 기반한 구조로 재구성하려는 시도이다. 특히 생성형 AI의 응답을 단순한 결과물이 아니라, 이용자의 해석과 평가를 요구하는 상호작용적 산물로 보고,

해당 응답이 어떠한 인지적·비판적 역량을 전제하는지에 주목한다. 이를 위해 본 연구는 학술 논문과 기업 자료를 바탕으로 텍스트 생성형 AI, 특히 트랜스포머 기반 대규모 언어 모델의 출력 생성 구조를 체계적으로 분석하고, 이를 토대로 생성형 AI 환경에 적합한 AI 리터러시 구성 요소를 도출한다. 나아가 응답 생성 과정과 알고리즘 작동 원리, 플랫폼 구조에 대한 이해를 통합하여 실제 활용 상황에 적용 가능한 AI 리터러시 모형을 제시하고자 한다.

Ⅲ. 교육 모형 제안

1. 생성형 AI 구조의 특성

앞서 언급한 바와 같이, 생성형 AI의 응답은 단편적 정보 조각의 나열이라기보다 내부적으로 정돈된 설명의 형식을 취하는 경향을 보인다. 그렇다면 생성형 AI는 어떠한 작동 방식을 통해 이같은 완결 구조의 답변을 산출하는가?

기존 검색엔진은 사전에 색인된 외부 문서 집합들을 조회·검색·순위화하여 전달하는 체계로 작동했다면, 대규모 언어 모델에 기반한 생성형 AI의 기본 모델은 모델의 사전 학습 시점까지 모델 내부 가중치에 압축된 정보에 기반해서 정보를 직접 생성하여 제공한다. 물론 최근 생성형 AI 시스템에서도 외부 검색이나 도구 연동을 통해 최신 정보를 끌어오기도 하나, 이는 모델 자체의 동적인 학습 결과라기보다는 외부 검색 연동을 통해 부분적으로 서비스를 보완하는 것에 가깝다. 다시 말해 생성형 AI의 기본적인 동작 원리는 모델의 사전 학습을 통해 압축된 대량의 데이터에 기반해, 이용자 프롬프트에서 주어진 맥락에서 다음에 올 언어 단위를 확률적으로 예측하는 것이다. 즉 모델은 방대한 텍스트 데이터에서 학습한 통계적 패턴을 바탕으로, 입력된 질문과의 조건부 확률 분포에 따라 가장 그럴듯한 표현을 순차적으로 생성한다(Dai et al., 2019; Ling et al., 2025; Vaswani et al., 2017; West et al., 2023).

이러한 과정에서 생성형 AI의 출력은 특정 지식 항목의 단순 전달이나 가치판단·추론(Linna & Linna, 2026)이라기보다, 맥락에 맞게 확률적으로 구성된 방향으로 조직된다. 따라서 생성형 AI의 작동을 제대로 이해하기 위해서는 "하나의 답변이 즉흥적으로 나오는 것처럼 보이는 순간"을 몇 개의 연속된 단계로 분해해 볼 필요가 있다. 이하에서는 생성형 AI가 '어떻게 설명을 만들어 내는가'를 살펴보기 위해 생성형 AI의 순차적 생성과 출력 제시에 이르기까지의 과정을 단계별로 분석해본다. 본 연구에서 다루는 생성 과정은 엄밀히는 트랜스포머와 셀프 어텐션 기반 대규모 언어 모델의 추론 메커니즘에 해당한다. 그러나 논의의 편의을 위해, 이하에서는 이를 포괄적으로 '생성형 AI'로 지칭한다.

2. 대규모 언어 모델의 텍스트 생성 과정

가. 이용자의 입력값 수신

우선 이용자가 생성형 AI를 사용하여 질문에 대한 답을 찾는 상황을 가정해 보면 첫 번째 단계는 이용자가 생성형 AI의 인터페이스에 질문이나 프롬프트를 입력하는 것에서부터 시작된다. 이 단계에서 생성형 AI 모델은 인간이 의미를 이해하는 방식으로 이용자가 입력한 텍스트를 이해한다기보다는, 이용자의 입력 텍스트를 수치적 표현으로 변환하여 이를 이후 어떤 표현이 등장할 가능성이 높은지를 계산할 때 활용되는 맥락 정보로 활용한다.

이러한 점에서 AI 리터러시 교육은 단순히 생성된 결과물을 해석하는 단계에 그치지 않고, 이용자의 입력이 모델의 처리 과정에 어떠한 영향을 미치는지를 이해하는 입력 단계에서부터 먼저 시작되어야 한다. 이는 단순히 질문을 작성하는 기술이나 프롬프트 엔지니어링(Prompt engineering)에 그치는 것이 아니라, 이용자가 자신의 문제를 구조화하고, 자신이 쓴 프롬프트에 전제된 가정과 관점, 잠재적 편향을 인식할 수 있는 능력을 포함한다. 뿐만 아니라 모델이 입력을 해석·처리하는 과정에서 발생할 수 있는 구조적 한계와 편향을 이해한 상태에서 질문을 설계하는 인식적 역량을 포함한다. 이러한 인식을 기반으로 이용자는 프롬프트를 작성할 때 어떤 배경정보를 제공하고, 생성형 AI에게 어떤 역할과 제한을 요청할지, 어떤 출력 형식을 요청할지, 혹은 어떤 단계로 일을 나눌 지까지 설계해야 한다. 즉 입력 단계에서 이용자는 프롬프트 안에 포함할 문맥을 설계하고 생성형 AI의 역할을 조정하게 된다.

예컨대 두 명의 학생에게 동일한 과제와 지시문이 주어지고, 그들이 같은 생성형 AI 도구를 사용한다고 하더라도, 이를 바탕으로 각 학생이 생성형 AI에 입력하는 실제 프롬프트 내용은 각 학생의 배경지식, 사용 언어, 질문 구성 방식, 문제에 대해 암묵적으로 전제하는 가정이나 편견, 생성형 AI 모델에 대한 이해도 등에 따라 상당히 달라질 수 있다. 이러한 이용자 입력값의 차이는 동일한 생성형 AI 모델을 사용하더라도 모델이 조건으로 삼는 맥락이자 출발점을 달라지게 하며, 그 결과 생성되는 설명의 방향과 강조점을 다르게 형성할 수 있다. 이는 이용자 입력이 단순힌 정보량의 차이를 만드는 것에 그치지 않고, 질문의 프레이밍 방식 자체가 이후 확률적 생성 과정에서 고려되는 조건과 맥락을 구조화하는 역할을 할 수 있음을 시사한다(Ling et al., 2025).

이처럼 생성형 AI에서 초기 입력이 어떻게 구성되느냐는 이후 생성되는 응답의 방향과 범위에 중요한 영향을 미칠 수 있다. 특히 이용자가 입력한 문장이나 표현이 모호하거나 다의적으로 해석될 수 있는 경우, 모델은 주어진 맥락에서 통계적으로 가장 그럴듯한 해석 경로를 선택하게 된다. 이는 모델이 의미를 논리적으로 판별하기보다는 사전에 학습된 언어 분포에 기반해 조건부 확률이 높은 표현을 산출하는 방식으로 작동하기 때문이다. 이러한 특성 때문에 질문의 의도가 불명확할수록, 출력 역시 이용자의 기대와 어긋날 가능성이 상대적으로 높아지게 된다.

다만 이러한 문제는 단순히 질문의 길이나 복잡함 자체보다는 표현의 모호성, 맥락 정보의 부족, 그리고 다중 해석 가능성에서 비롯될 수 있다. 일례로 동일한 용어라도 추가적인 맥락이 제공되지 않을 경우, 모델은 학습 데이터에서 해당 용어와 더 자주 함께 등장했던 의미를 중심으로 응답을 구상할 가능성이 있다. 예를 들어 이용자가 "좋은 사과는 뭐야?"라는 질문을 했다고 가정해 보자. 이 표현은 추가적인 명시적 맥락 정보가 없을 경우, 과일로서의 사과를 의미할 수도 있고, 행위로서의 사과를 의미할 수도 있다. 또한 극단적인 오탈자나 비표준적 표현이 포함된 경우에는 모델이 입력 패턴을 안정적으로 매칭하지 못해 해석의 불확실성이 증가하기도 한다.

종합하면 이 단계에서의 우리가 유념해야 할 점은 모델이 주어진 입력값을 확률적 생성의 출발 조건으로 변환시키고 구성한다는 것이다. 이러한 입력의 조건화 과정은 이후 단계에서 형성되는 설명의 방향성과 범위를 구조적으로 제약하는 초기 조건으로 작용하기 때문에 중요하다. 그러므로 입력 단계에서 이용자는 자신의 프롬프트 속에 암묵적으로 포함된 가치 판단이나 이분법적 구조, 혹은 이미 결론을 전제한 표현이 없는가를 먼저 점검해야 한다. 또한 생성형 AI가 이러한 질문의 구조와 요소들에 의해 다른 결과물을 낼 수 있다는 것을 이해하며 자신의 프롬프트 속에 포함된 맥락을 인지·지정하며 질문을 구조화할 수 있어야 한다.

나. 토큰화

이렇게 이용자의 입력이 생성형 AI 시스템에 전달되면, 시스템은 해당 문장을 곧바로 의미 단위로 이해하는 것이 아니라, 내부에서 컴퓨터가 계산이 가능한 더 작은 표현 단위로 변환하는 전처리 과정을 거치게 된다. 모델은 이용자가 입력한 복잡하고 긴 문단과 문장의 큰 정보 덩어리를 컴퓨터가 보다 처리하기 유용한 '하위 단어(Subword)'의 작은 조각으로 분해하는데 이 과정을 토큰화(Tokenization)라고 한다. 연속적인 자연어 문자열을 컴퓨터가 처리하기 좋은 이산적 단위의 열로 재구성하는 것이다(Roumeliotis & Tselikas, 2023). 예를 들어 "uncomfortable"와 같은 단어는 하나의 통합된 의미 단위가 아닌, "un－", "comfort", "－able"과 같이 더 작은 부분 단위로 분해되어 모델에 전달될 수 있다. 이러한 하위 단위 기반 표현이 컴퓨터 모델에게는 어휘 규모를 더욱 효율적으로 관리하고, 다양한 형태 변형을 보다 유연하게 처리할 수 있도록 하기 때문이다.

최근 ChatGPT나 Gemini와 같은 LLM에 기반한 생성형 AI 모델들은 방대한 양의 텍스트 데이터를 학습하고, 이를 바탕으로 다음 토큰을 확률적으로 예측하는 방식으로 인간의 언어를 이해하고 생성한다. 이러한 과정 전반에는 자연어처리(Natural Language Processing, NLP) 기술이 핵심적으로 활용된다. 여기서 자연어 처리는 컴퓨터가 인간의 언어 구조와 의미를 처리할 수 있도록 하는 기술 분야로, 텍스트 분할, 형태소 분석, 품사 태깅, 구문 분석, 의미 분석 등 여러 단계의 언어 처리 과정을 포함한다. 이 자연어 처리 과정에서 토큰화가 초기의 중요한 단계 중 하나이다.

이러한 토큰화 단계는 입력된 텍스트를 생성형 AI 모델이 처리할 수 있는 작은 단위로 분할하여

이후에 언어 분석과 생성 과정이 가능하도록 하는 자연어 처리의 기초적인 전처리 단계이다. 이러한 토큰화 과정은 단순한 기술적 전처리를 넘어, 생성형 AI가 언어적 패턴을 계산하고 생성하는 전체 파이프라인의 출발점이라는 점에서 중요한 의미를 갖는다.

토큰화 단계에서 요구되는 리터러시는 입력 텍스트가 모델 내부에서 처리되는 기본 원리를 이해하는 것뿐 아니라, 이러한 처리 방식에서 비롯되는 모델의 한계에 대한 인식을 포함한다. 먼저 이용자는 생성형 AI가 자연스러운 문장으로 답변을 내놓더라도, 실제로는 인간이 이해하는 방식으로 자신의 프롬프트를 이해하는 것이 아니라는 점을 인지해야 한다. 뿐만 아니라 이용자는 생성형 AI 환경에서 입력된 언어가 계산 가능한 토큰 단위로 분해・변환되어 확률적으로 처리되며, 이러한 계산적 과정이 의미 해석의 일정한 한계를 가져올 수 있음을 인지해야 한다.

가령 자동화된 텍스트 처리에서의 토큰화는 여러 까다로운 문제를 제기한다. 대표적으로 토큰화 이전 단계에서 원본 문서가 지니고 있던 다양한 조판 정보(예: 레이아웃, 그래픽 요소 등)는 일반적으로 제거되거나 단순화되어 문자 스트림으로 환원되는데, 이 과정에서 이러한 시각적 요소가 포함한 의미들을 잃어버리게 된다(Grefenstette, 1999a; 1999b). 또한 동일한 문자열이라도 처리 목적이나 분석 맥락에 따라 서로 다른 토큰화 결과가 요구되기도 한다. 이처럼 토큰화는 문서 구조에서 발생하는 정보 손실 문제와 맥락에 따른 분할 기준의 가변성을 동시에 지니기 때문에 단순한 문자열 분할을 넘어 언어적・맥락적 판단을 필요로 하는 해석적 과정으로 이해될 필요가 있다.

뿐만 아니라 다양한 이용자의 입력을 분석하는 생성형 AI 챗봇의 경우 대게 이용자가 제공한 입력 텍스트가 문법적으로 정제되어 있지 않을 가능성이 높다. 특히 이용자 입력 값은 오탈자, 비표준 축약, 구어체 표현, 이모지, 불완전한 문장 구조 등이 빈번하게 나타날 가능성이 있고, 그 의미 또한 문맥에 따라 모호할 수 있다. 특히, 이러한 특성은 최근 챗봇 기반 인터페이스의 확산과 함께 더욱 두드러지기 쉬워졌다고 볼 수도 있는데, 이용자들이 ChatGPT나 Gemini와 같은 시스템을 '자연스러운 대화 상호작용이 가능한 챗봇'으로 인식하면서(Cheng & Jiang, 2020), 보다 구어적이고 즉흥적이며 비정형적인 방식으로 질문을 입력하는 경향이 강화되고 있기 때문이다(Zhao et al., 2024). 그 결과 실제 이용자에 의해 입력된 데이터는 전통적으로 학습되어온 정제된 텍스트보다 훨씬 더 높은 수준의 변이성과 불확실성을 포함하게 되며, 이는 토큰화 과정에서의 경계 설정과 의미 해석을 한층 더 복잡하게 만든다. 따라서 토큰화는 입력된 텍스트의 표면적인 형태와 언어적 의미 사이에서 적절한 단위를 결정해야 하는 해석적 과정으로, 그 자체로 복합적인 특성을 지닌다.

대표적으로 현재 널리 사용되는 방식 중 하나가 Byte Pair Encoding(BPE) 계열의 서브워드 토큰화 기법이다(Bostrom & Durrett, 2020; Devlin et al., 2019; Sennrich et al., 2016; Wu et al., 2016). 이 접근법은 대규모 텍스트 말뭉치에서 자주 함께 등장하는 문자 또는 문자열 쌍을

반복적으로 병합하여, 빈도 기반의 하위 어휘 집합을 구성한다. 다시 말해, 알고리즘이 학습 데이터에서 다양한 언어적 문맥 패턴을 학습하고, 미리 정의된 어휘 크기에 도달할 때까지 텍스트 말뭉치에서 가장 빈번하게 나타나는 문자 쌍 또는 하위 단어를 결합하는 방식으로 반복적으로 문자 쌍을 병합하며 결과물을 만들어 낸다. 그 결과 모델은 완전한 단어 수준 어휘에 의존하지 않고도 다양한 어형과 희귀 표현을 비교적 안정적으로 처리할 수 있게 된다.

다만 이러한 토큰화 단위는 통계적 패턴에 기반해 형성된 계산 단위이므로 반드시 인간이 직관적으로 생각하는 의미 단위와 일치하지는 않는다. 그보다는 모델의 계산 효율성과 통계적 패턴에 기반해 구성된 처리 단위에 가깝다. 그 결과 학습 데이터에서 상대적으로 빈도가 낮거나 새롭게 등장한 표현은 더 길거나 더 파편화된 토큰 열로 표현될 수 있다. 이러한 특성은 해당 표현의 처리를 불가능하게 만드는 것은 아니지만 어휘 항목 간 표현 단위에 있어서 각 어휘를 얼마나 잘게 나누어 표현할 것인가, 즉 세분화 수준(Granularity)의 차이를 유발할 가능성이 존재하게 된다. 따라서 이용자는 모델이 입력된 문장을 의미 단위가 아니라 계산 단위로 처리한다는 기본 원리를 이해하고, 이에 따라 같은 내용이라도 입력값에서 어떤 표현을 선택하느냐가 결과 값에 영향을 미칠 수 있다는 점을 인식하는 능력이 필요하다.

AI 리터러시의 관점에서, 토큰화는 전통적으로 자연어처리 시스템 내부의 기술적 전처리 단계로 간주되어 왔기 때문에 일반 이용자가 이를 직접 인식할 필요는 없다고 여길지도 모른다. 특히 이용자가 토큰화 과정 자체를 직접 통제하기는 어렵기 때문에 이러한 절차에 대해 리터러시가 필요 없다고 생각할 여지도 있다. 그러나 최근 신경망 기반 언어 모델(Neural Network Language Model, NNLM)에서는 이용자의 입력 방식이 모델의 생성 과정에서 어떤 표현 패턴이 활성화될 수 있는지를 구조적으로 형성하며 모델의 해석과 출력에 보다 직접적인 영향을 미치기 때문에 토큰화에 대한 기본적인 이해가 점차 중요해지고 있다. 예컨대 이용자가 사용하는 정보의 내용 자체 외에도 표현의 길이, 구두점, 축약형, 특수문자 역시 모델의 처리 방식과 응답 품질에 실질적인 차이를 초래할 수 있다. 또한 앞서 말했듯, 비정형적이거나 모호한 이용자 입력은 토큰화되는 단위와 의미 해석의 불확실성을 높여 예기치 않은 출력이나 오해를 유발할 가능성이 있다. 이러한 맥락에서 토큰화는 더 이상 시스템 내부의 보이지 않는 단계에 머무르지 않으며, 생성형 AI와 상호작용하는 이용자 경험과 결과 품질을 이해하기 위한 중요한 개념적 지점으로 고려될 필요가 있다.

다. 입력 임베딩

토큰 단위로 분할된 입력값들은 다음 단계에서 수치 값 형태로 변환된다. 각 토큰의 숫자들은 해당 하위 단어의 의미적, 통계적 특성을 반영하는 고차원 벡터 형태를 표현한다. 그리고 각 토큰을 고차원 벡터로 변환하여 매핑하는 과정을 입력 임베딩(Input embedding)이라 한다. 입력 임베딩 단계에서 텍스트는 고차원 벡터 공간 상의 수치적 표현으로 변환되며, 이 과정에서 개별

표현은 다른 표현들과의 통계적 관계에 따라 특정한 위치로 배치된다. 이와 관련하여 이용자는 텍스트가 인간이 의미를 이해하는 방식으로 직접 처리되는 것이 아니라, 벡터 공간 내에서 다른 표현들과의 관계적 위치로 재구성된다는 점을 이해할 필요가 있다. 또한 이용자가 선택한 표현 방식은 이러한 의미 공간에서 형성되는 위치와 연결 관계에 영향을 미칠 수 있음을 인식해야 하며, 복잡한 맥락 정보가 제한된 벡터 표현으로 압축되는 과정에서 특정 관점이나 맥락이 상대적으로 약화되거나 희석될 가능성이 있다는 점도 함께 고려해야 한다.

이 단계에서 추가적으로 고려되어야 할 리터러시는, 생성형 AI의 응답이 중립적 지식의 산출물이 아니라 특정 시기와 맥락에서 수집・기록된 학습 데이터(training data)의 구성과 분포를 반영・재구성한 결과물임을 이해하는 능력이다. 이는 학습 데이터의 편향에 대한 인식을 포함한다. 실제로 임베딩 벡터는 대규모 온라인 텍스트 데이터에 대한 사전 학습(pre-training) 과정에서 통계적으로 형성되기 때문에, 데이터에 내재된 다양한 왜곡 가능성을 내포할 수 있다. 따라서 이용자는 생성형 AI가 학습한 데이터의 특성에 따라 특정 관점이 강화되거나 일부 정보가 누락・왜곡될 수 있음을 인식할 필요가 있다. 예컨대 지난 몇 년간 많은 학술논문에서 다뤄졌듯이(박기범, 2022; Barocas et al., 2023; Diakopoulos, 2016; Noble, 2018) 학습 데이터에 내재된 사회적・문화적 패턴 역시 잠재적으로 함께 인코딩될 수 있다. 특히 주류 담론의 과대표현, 성별・인종 관련 언어 패턴, 문화적 편향 등이 데이터에 함축적으로 포함되어 있을 경우, 이러한 경향은 임베딩 공간의 구조적 특성으로 반영될 가능성이 있다.

만약 이 과정에서 모델이 특정 집단을 직접적으로 지칭하는 명시적 변수를 사용하지 않도록 설계되었다고 가정하더라도, 여전히 모델은 간접적 단서(Proxy variables)를 통해서 사회적 범주를 추론할 수 있을 가능성이 높다. 예를 들어 모델이 인종이나 소득 수준과 같은 민감 속성(Sensitive attributes)을 명시적으로 사용하지 않도록 설계되었다고 가정하더라도, 학습 데이터 상에 내재된 거주 지역, 우편번호, 직업명, 교육 수준과 같은 변수는 특정 집단과 높은 통계적 상관관계를 가질 수 있다. 그 결과 모델의 변수가 표면적으로는 중립적으로 보인다고 가정하더라도, 실제로는 사회경제적 지위나 인종적 분포를 간접적으로 반영하는 프록시로 기능하여 이를 매개로 결과 값에 사회적 편향이 재구성되거나 알고리즘을 통해 더욱 강화된 형태로 나타날 위험이 존재하게 되는 것이다.

보다 더 근본적으로 이러한 문제는 모델 설계 단계 이전에 학습 데이터의 구성・수집 자체에서 비롯될 수 있다는 점도 고려되어야 한다. 대게 현실 세계의 데이터는 사회를 균등하게 반영하지 못한다(Barocas et al., 2023). 특정 집단이 구조적으로 과소대표되거나 데이터가 충분히 축적되지 못하는 경우가 빈번하다. 전체 데이터에서 다양한 사회적 집단에 대한 데이터 분포가 불균형할 경우, 이 데이터를 학습한 모델은 자연히 다수 집단의 패턴에 최적화되고, 상대적으로 데이터가 부족한 집단에 대해서는 예측 성능이 저하되거나 불안정한 출력을 생성할 가능성이 높아지게 된다.

대표성이 충분히 확보되지 않은 집단의 경우, 모델이 해당 집단의 정교한 특성들을 학습할 기회를 갖지 못하기 때문에 결국 해당 집단과 관련된 예측이나 생성 결과의 정확도와 신뢰도가 체계적으로 낮아질 위험이 커지게 되는 것이다. 특히 모델이 이러한 편향을 비판적으로 맥락 화하는 정보가 충분하지 않은 채로 응답을 생성할 경우, 기존 사회적 편향을 재생산하거나 더욱 강화하는 담화가 출력될 위험이 존재한다(Barocas et al., 2023; Creel & Hellman, 2021). 이러한 현상은 편향 문제가 단순히 명시적 속성의 사용 여부를 넘어, 데이터 수집과 분포 단계에서부터 구조적으로 형성될 수 있음을 시사한다. 나아가 임베딩 과정은 단순한 기술적 표현 학습 단계를 넘어, 사회적으로 구성된 맥락들, 예컨대 사회적 지식과 권력 관계까지도 통계적으로 압축·재현되는 기술적 매개로 해석될 수도 있다. AI 리터러시 역시 생성형 AI 모델의 이러한 한계나 위험성을 포함해야 한다.

라. 셀프 어텐션(Self-attention) 알고리즘과 위치 정보 압축

입력 임베딩을 통해 각 토큰의 의미 정보를 담은 고차원 벡터값은 아직 그 자체에 문장 내 순서에 대한 정보를 포함하지 않는다. 각 벡터값들은 입력 임베딩 다음 단계인 위치 정보 인코딩(Positional encoding) 단계를 통해 위치 정보까지 결합하게 된다. 각 토큰 벡터에 해당 토큰의 문장 내 상대적 또는 절대적 위치를 나타내는 위치 벡터가 추가되는 것이다. 이처럼 시퀀스 내 개별 토큰의 특정 위치와 관련된 데이터 값까지 포함하는 위치 인코딩까지 거치게 되면, 모델은 각 토큰의 의미 정보뿐만 아니라 순서 정보까지 동시에 포함할 수 있는 입력 표현 값을 구성하게 된다. 즉 입력 임베딩이 단어를 의미 좌표가 있는 다차원 공간에 배치하는 과정이라면, 위치 정보 인코딩은 그 좌표들에 순차적 맥락을 부여하는 보완적 단계인 것이다.

이러한 특징은 ChatGPT와 같은 대규모 언어 모델이 기반하고 있는 트랜스포머(Transformer) 구조와 밀접한 관련이 있다. 트랜스포머는 토큰을 순차적으로 처리하는 방식이 아니라, 셀프 어텐션 메커니즘을 통해 입력 문장 내 여러 토큰 간 관계를 병렬적으로 처리하며 이 과정에서 모든 토큰이 서로를 참조하여 맥락적 연관성을 평가한다(Dai et al., 2019; Vaswani et al. 2017). 여기서 셀프 어텐션 알고리즘이란 입력 문장 내의 모든 단어가 서로 서로를 살펴보며 맥락적 연관성을 계산하는 작동 방식이다. 다만 트랜스포머는 구조적으로 토큰 순서를 알 수 없기 때문에 위치 정보 인코딩을 통해 위치 정보를 추가로 제공받는다. 또한 학습 과정에서는 병렬 처리가 가능하지만, 텍스트 생성 과정에서는 자기회귀적 방식으로 토큰이 순차적으로 생성된다. 이로 인해 토큰의 위치 정보가 별도로 제공되는 위치 정보 인코딩 없이는 트랜스포머가 어순을 직접적으로 추론하기 어렵다. 그 결과, 동일한 토큰 집합이라 하더라도 어순이 다른 문장, 예컨대 “dog bites man”과 “man bites dog”은 동일한 벡터값이 생성되어 임베딩 단계만으로는 충분히 구별되지 않을 수 있다. 이러한 한계를 보완하기 위해 모델은 각 토큰 임베딩에 위치 정보까지 결합해 의미 정보와 순서 정보까지 함께 고려하며 토큰 간 관계를 보다 정교하게 학습할 수 있게 된다.

이렇게 위치 정보까지 결합한 입력 벡터는 트랜스포머 레이어를 순차적으로 통과하게 된다. 이때 각 레이어 내부에는 다수의 어텐션 헤드(Self-attention heads)로 구성된 셀프 어텐션 연산이 반복적으로 수행된다(Vaswani et al., 2017). 이때 셀프 어텐션은 문장 내 토큰들 사이의 관계를 계산하여, 특정 토큰을 해석할 때 다른 토큰들이 얼마나 중요한지를 동적으로 평가하는 역할을 한다. 이를 통해 모델은 입력 전체에서 상대적으로 중요한 부분에 더 높은 가중치(Weights)를 부여하며, 다음 토큰 예측에 필요한 문맥적 표현을 점진적으로 정교화해나가는 것이다. 특히 멀티-헤드 어텐션(Multi-head attention) 구조는 서로 다른 관계 양상을 병렬적으로 포착하도록 설계되어 있어 모델이 문장의 다양한 의미적·구문적 측면을 동시에 고려할 수 있게 한다(Cordonnier et al., 2020; Dive into Deep Learning, 2023). 그 결과 모델은 단순히 단어의 개별 의미를 처리하는 수준을 넘어, 문장 전체의 맥락 속에서 토큰 간 상호작용 패턴을 학습하게 되며, 이러한 과정이 다음 토큰 예측의 정확도를 높이는 핵심 메커니즘으로 작동한다.

그럼에도 대규모 언어 모델은 여전히 흔히 Black-Box system으로 자주 언급되는 문제가 있는데(Evans et al., 2021), 이는 트랜스포머 기반 심층 신경망의 내부 표현이 고차원 벡터 공간에서 형성되며, 수많은 비선형 연산과 분산 표현(Distributed representations)을 통해 출력이 생성되기 때문에 개별 예측이 어떤 내부 경로를 통해 형성되었는지를 인간이 직관적으로 추적하기 어렵기 때문이다. 이러한 해석 가능성의 제약은 대규모 심층 신경망 전반에 공통적으로 나타나는 구조적 특성 중 하나이다.

특히 각 트랜스포머 블록 내부에서는 학습의 안정성과 수렴을 돕기 위해 레이어 정규화(Layer normalization)와 잔차 연결(Residual connection)이 함께 사용된다(OpenAI Developers, 2025; Zhang et al., 2024). 레이어 정규화는 층별 활성 값의 분포를 정규화하며 텍스트에서 분리된 구성 요소들을 추상화하고 연결하여 학습을 안정화하는 역할을 한다. 잔차 연결은 깊은 네트워크에서 정보 흐름과 기울기 전달을 원활하게 유지하도록 돕는다. 이러한 설계는 모델 성능 향상을 위해 채택된 구성이며, 이어서 위치별로 독립적으로 적용되는 순방향 신경망(Feed-Forward Neural Network, FFN)이 각 토큰 표현을 비선형적으로 변환하여 표현력을 추가로 확장할 수 있다(Geva et al., 2021). 종합하면, 트랜스포머 레이어는 입력 표현을 점진적으로 정교화하며 토큰 간 문맥적 관계를 학습하는 핵심 단계이다. 다만 이 과정 역시 학습 데이터의 통계적 구조에 기반해 작동하기 때문에, 데이터 수준에서 존재하는 편향 문제가 완전히 제거되기보다는 모델 표현 속에 반영될 가능성이 있다는 점은 함께 고려될 필요가 있다.

셀프 어텐션과 위치 정보가 압축되는 단계에서는 모델의 작동 방식에 대한 구조적 이해가 요구된다. 이 단계에서 토큰 간의 관계가 계산되고, 가중치가 부여되며 어떤 정보가 더 중요하게 해석될지가 결정된다. 따라서 이용자는 모델이 단어의 의미를 고정적으로 이해하는 것이 아니라, 맥락적 관계와 통계적 가중치에 따라 의미를 구성한다는 점을 인식할 필요가 있다. 위치 정보 압축 과정

에서는 문장 내 순서와 구조가 수치적으로 반영되기 때문에, 질문의 배열 방식 역시 의미 구성에 영향을 미칠 수 있다. 이에 따라 이용자는 모델이 맥락 속에서 의미를 재구성하는 작동 원리에 대한 구조적 이해가 필요하다.

마. 출력 생성

트랜스포머 레이어를 통과한 이후, 구문적 구조와 의미적 연관성, 그리고 문맥 정보가 통합되어 입력에 대한 맥락적 이해가 형성된다. 이를 바탕으로 모델은 최종 레이어에서 다음 단계에 올 가능성이 높은 토큰들의 확률을 계산하여 다음 토큰 예측 단계로 이동하게 된다. 이때 모델은 어휘 집합에 포함된 모든 후보 토큰에 대해 각각의 등장 가능성을 확률 형태로 계산하며 "다음에 어떤 단어가 올 것 같은가"를 하나의 정답으로 단정하기보다, 여러 후보에 대한 가능성을 비교하는 방식으로 판단한다. 즉, 입력 맥락을 받아 다음 생성 값들의 확률 분포를 계산하고 가장 가능성이 높은 연속 값을 선택하여 응답을 생성하기 시작하는 것이다.

이렇게 선택된 토큰은 다시 입력 시퀀스에 추가되고, 이러한 예측 절차는 반복적으로 수행된다. 문장을 한 번에 완성하는 것이 아니라, 가장 그럴듯한 다음 토큰을 순차적으로 예측하고 이어 붙이는 자기회귀(Auto regressive) 방식으로 텍스트를 생성하며 누적하는 방식으로 점진적으로 구성하는 것이다(Zhang et al., 2025). 이러한 예측 과정 역시 외부 지식을 그대로 검색하여 복원한 결과라기보다, 학습 데이터에서 관찰된 통계적 패턴에 기반해 확률적으로 구성된 텍스트이기 때문에 이점을 유념하는 것 또한 중요하다. 다시 말해 이용자들은 생성형 AI가 진실을 판별하는 시스템이 아니라, 확률적으로 가장 그럴듯한 토큰을 선택하는 시스템이라는 확률적 생성에 대한 이해가 기반이 되어야 한다. 특히 긴 대화나 복잡한 프롬프트 상황에서는 초기 문맥 해석의 미세한 오차가 이후 토큰 예측 과정에서 점차 누적되어, 결과적으로 관련성이 낮거나 편향된 응답으로 이어질 가능성도 있다. 요컨대 트랜스포머의 다층 구조는 토큰 간 관계를 정교하게 추상화하고 장거리 의존성을 포착하는 데 중요한 역할을 수행하지만, 동시에 모델의 출력 품질은 궁극적으로 학습 데이터의 통계적 특성과 입력 맥락의 명확성에 크게 좌우된다는 점을 함께 고려할 필요가 있다.

바. 텍스트 재복원

모델이 다음 토큰들을 순차적으로 생성한 후에는, 생성된 토큰 시퀀스를 사람이 읽을 수 있는 문자・단어・문장 형태로 변환하는 디토큰화(Detokenization) 단계가 수행된다. 이는 초기 토큰화의 역과정에 해당하며, 토큰 ID 시퀀스를 사전에 정의된 토큰-문자열 매핑 규칙에 따라 문자열로 디코딩하는 절차로 이해할 수 있다(Kaplan et al., 2025). 다만 디토큰화 자체는 일반적으로 비교적 기계적이며, 의미를 "해석"하는 단계라기보다 모델이 이미 선택한 토큰을 텍스트로 "표현"하는 단계에 가깝다. 디토큰화는 생성된 토큰을 사람이 읽을 수 있는 형태로 변환하는 마지막

표현 단계이지만, 토큰화/디토큰화 체계가 언어 표현을 어떤 단위로 분절하고 결합하는지에 따라 일부 표현에서 미묘한 의미 차이가 결과에 영향을 미칠 수 있다는 점은 함께 고려할 필요가 있다. 추가적으로 텍스트 재복원 단계에서 요구되는 리터러시는 모델이 확률적으로 선택된 토큰을 문법적 완결성뿐 아니라 특정한 응답 구조를 염두에 두고 배열한다는 점을 이해하는 것이다.

사. 후처리

대개 생성된 응답은 이용자에게 전달되기 전에, 부적절하거나 유해할 수 있는 표현이 포함되어 있는지를 점검하는 후처리(Post-processing) 및 안전 필터링(Safety filtering) 단계를 거친다. 이 과정에서는 사전에 정의된 정책 규칙, 콘텐츠 분류 모델, 또는 위험 표현 목록 등에 기반하여 출력 텍스트가 검토되며, 필요에 따라 응답이 수정・제한・차단될 수 있다. 이러한 콘텐츠 조정 메커니즘은 서비스 제공자가 설정한 안전성, 법적 준수, 이용자 보호 등의 운영 기준에 맞추어 설계・운영된다. 따라서 최종적으로 이용자에게 제시되는 출력은 단순히 언어모델의 확률적 생성 결과만이 아니라, 플랫폼 차원의 정책적・기술적 필터링이 추가로 반영된 산물로 이해할 필요가 있다. 즉 이용자는 플랫폼의 규범적 설계와 알고리즘적 거버넌스가 응답의 범위와 형식을 형성한다는 점을 이해하는 플랫폼 차원의 매개 및 거버넌스 리터러시가 필요하다. 특히, 이러한 리터러시는 AI의 원리에 대한 이해를 넘어서 이용자가 사용하는 생성형 AI의 모델에 따라 그 모델의 특성과 응답 방식, 플랫폼의 정책과 설계 철학이 다를 수 있다는 점을 이해하는 것까지 확장된다. 이용자는 이러한 이해를 바탕으로 모델별 특성과 한계를 파악하고, 이에 맞게 생성형 AI의 모델을 선택하며 사용 전략과 상호작용 방식을 조절할 필요가 있다.

아. 응답 전달과 이용자의 해석

이러한 과정을 거쳐 최종적으로 생성된 응답은 이용자에게 전달되며, 그 의미와 신뢰도는 이를 해석하는 이용자의 배경지식, 문제이해 수준, 비판적 검토 역량에 따라 달라질 수 있다. 특히 일부 이용자는 모델의 출력을 별도의 검증 없이 사실적 정보로 간주하거나, 자신의 기존 신념이나 기대에 부합하는 내용만을 선택적으로 수용하는 경향을 보일 수 있다. 이러한 경향은 생성형 AI의 출력이 본질적으로 확률적 예측에 기반한 언어적 구성물임에도, 유창한 서술, 단정적인 어조, 논리적으로 보이는 설명 구조를 통해 이용자에게 마치 검증된 권위 있는 단일 답변처럼 인식될 가능성을 내포한다. 이 과정에서 이용자는 설명의 형식적 완결성이나 서술적 정합성을 내용의 사실성이나 추론의 타당성과 동일시할 위험이 있다. 따라서 이용자에게는 응답의 표현 형식과 내용적 주장 사이를 구분하여 사실적 타당성을 평가하고, 설명의 그럴듯함이 곧 사실의 정확성이나 근거의 타당성을 의미하지 않는다는 점을 인식하며, 제시된 정보의 주장・근거・출처를 독립적으로 검토하는 비판적 평가 역량, 즉 응답 평가 리터러시가 필요하다.

특히 생성형 AI 모델은 자신의 실제 추론 과정을 항상 그대로 설명하지 않는데, 모델이 문제 해결 과정에서 특정 힌트나 편법을 실제로 사용하더라도 이를 명시적으로 드러내지 않은 채, 사후적으로 그럴듯한 설명을 구성해 제시할 수 있다는 사실이 최근 연구에서 보고되고 있다(Chen et al., 2025). 예컨대 추론 모델에서 사용되는 사고의 사슬(Chain-of-Thought, 이하 CoT)에 대한 설명의 충실성(Faithfulness)－즉 모델의 설명이 실제 내부 추론 과정을 얼마나 정확히 반영하는지－을 실험적으로 평가한 해당 연구에 따르면, 모델은 실제로 특정 힌트나 전략을 사용한 경우에도 그 사실을 종종 드러내지 않았고, 다수의 사례에서 실제로 활용한 추론 단서를 설명에 포함하지 않는 것으로 확인되었다. 이는 모델이 내부적으로 사용한 추론 과정을 CoT에 부분적으로만 반영하거나 아예 드러내지 않을 가능성이 있고, 그 과정에서 모델의 오류, 편향, 잘못된 추론 과정 등이 설명 단계에서 가려질 수 있다는 것을 의미한다.

따라서 생성형 AI가 제시하는 설명은 겉으로는 논리적으로 정합적이고 설득력 있게 보일 수 있으나, 그것이 반드시 모델의 실제 내부 추론 과정을 충실히 반영한 것은 아니다. 모델은 정답이 도출된 이후 그 결과를 정당화하는 형태로 설명을 구성할 수 있으며(Chen et al., 2025), 설명을 요청하더라도 실제로 사용된 추론 전략이 완전히 드러나지 않거나, 다른 설명으로 대체되거나, 실제 추론 과정의 일부만 제시될 가능성이 있다. 따라서 이용자는 설명의 그럴듯함이나 형식적 논리성만을 근거로 응답의 신뢰성을 판단하기보다는, 그러한 설명이 실제 추론 과정을 완전히 재현한 것이 아닐 수 있다는 점을 이해하고, 정보의 정확성과 타당성을 독립적으로 검토할 필요가 있다. 또한, AI의 답변을 외부 자료나 다른 출처로 교차 검증하는 습관을 가져야 한다.

특히, 의료・법률・교육・보안 등 정보 자체의 정확성(Accuracy)이 특별히 중요한 영역에서는 AI의 판단이나 설명을 의사결정의 단독 근거로 사용하지 않는 것이 바람직하며, 추가적인 검증과 근거 확인을 병행하는 습관이 필요하다. 정리하면 이 단계에서 AI 리터러시의 핵심은 AI의 출력과 설명을 그대로 신뢰하는 것이 아니라 그 한계와 가능성을 이해하고, 인간의 판단과 검증을 결합해 책임 있게 활용하는 능력이라고 할 수 있다. 뿐만 아니라 공개 이후의 미세조정 단계에서 이용자 상호작용 데이터나 인적 피드백이 모델 개선에 활용될 수 있는데, 이 과정은 이용자 데이터의 처리, 익명화, 보안 관리에 대한 신중한 운영과 거버넌스를 요구하므로, 이러한 측면 역시 AI 리터러시 교육에 포함되어야 한다.

3. 생성형 AI 출력의 단계에 따른 교육 모형의 구성 요소

본 연구는 생성형 AI의 출력 생성 구조 분석을 바탕으로 생성형 AI 환경에서 요구되는 AI 리터러시 교육 요소로 질문 구성 리터러시, 계산 리터러시, 확률 생성 리터러시, 응답 해석 리터러시, 플랫폼 리터러시, 비판적 평가 리터러시를 제안한다.

〈표 1〉 생성형 AI 출력 단계에 따른 AI 리터러시 구성요소

생성형 AI 작동 단계	AI 리터러시 요소	의미
입력	질문 구성 리터러시	질문의 목적과 전제를 파악하고 문제 상황을 명확하게 구조화하는 능력
계산	계산 리터러시	토큰화와 임베딩 등 생성형 AI의 계산 과정을 이해하는 능력
생성	확률 생성 리터러시	생성형 AI가 확률적 방식으로 응답을 생성한다는 원리를 이해하는 능력
출력	응답 해석 리터러시	생성형 AI 응답의 구성 방식과 정보 제시 구조를 파악하는 능력
미세 조정	플랫폼 리터러시	생성형 AI 응답이 플랫폼 설계와 정책에 의해 매개된다는 점을 인식하는 능력
평가	비판적 평가 리터러시	생성된 정보의 신뢰성과 한계를 판단하는 능력

먼저 질문 구성 리터러시는 이용자가 자신의 질문의 목적과 전제를 정확히 파악하고, 필요한 정보의 범위와 전제를 정리하여 문제를 적절한 형태로 구성하고 구조화하는 능력을 말한다. 생성형 AI의 응답은 입력 방식에 크게 영향을 받기 때문에, 이용자가 무엇을 어떻게 묻는가는 모델 결과 값의 방향과 질을 결정한다. 따라서 입력 단계의 리터러시는 문제를 구조화하고 요청의 조건을 설정하는 능력을 포함해야 한다. 이용자는 자신이 무엇을 알고자 하는지를 분명히 파악하고 이를 AI가 이해하고 가장 최적화된 답변을 제공할 수 있는 형태의 질문을 구성하는 능력을 길러야 한다.

두 번째, 계산 리터러시는 생성형 AI가 입력을 처리하는 과정에서 토큰화, 임베딩, 문맥 반영과 같은 계산적 절차를 거쳐 응답을 형성한다는 점을 이해하고, 이러한 계산 과정이 응답 생성에 어떤 역할을 하는지를 이해하는 능력을 의미한다. 특히 이 리터러시는 AI가 인간처럼 사고한다기보다 계산적 처리 과정을 통해 텍스트를 구성한다는 점을 이해하고 이러한 방식의 한계를 인지한다는 점에서 중요하다. 이는 AI가 입력 값을 어떠한 방식으로 처리하는지 이해함으로써 AI 생성 결과를 과도하게 신뢰하거나 인간적 의도를 부여하는 해석을 줄이는 데에도 도움이 된다.

세 번째, 확률 생성 리터러시는 생성형 AI의 응답이 고정된 정답의 재현이 아니라, 학습 데이터와 문맥을 바탕으로 다음 표현을 확률적으로 예측하는 방식으로 생성된다는 점을 이해하는 능력이다. AI가 생성한 답변이 정답을 산출하는 것이 아니라 확률적으로 가장 그럴듯한 표현을 만들어내는 것임을 이해하는 것이다.

네 번째, 응답 해석 리터러시는 생성형 AI의 모델에 따라 각 모델이 제시하는 응답의 구성 방식, 정보 배열, 요약 방식, 강조점 등을 파악하고, 응답이 어떤 방식으로 의미를 조직하고 전달하는지를 해석하는 능력을 말한다. 생성형 AI의 응답은 대체로 이미 요약되고 재구성된 형태로 제시되기 때문에, 이용자는 응답 내용을 그대로 수용하기보다, 그 안에 포함된 논리 구조와 강조

점을 비판적으로 읽어낼 필요가 있다. 즉 AI가 어떤 정보를 어떤 방식으로 선택・구성했는지, 생성된 응답 값이 어떤 순서와 논리 구조에 따라 조직되어 제시된 것인지 문맥과 구조를 파악하는 능력이다.

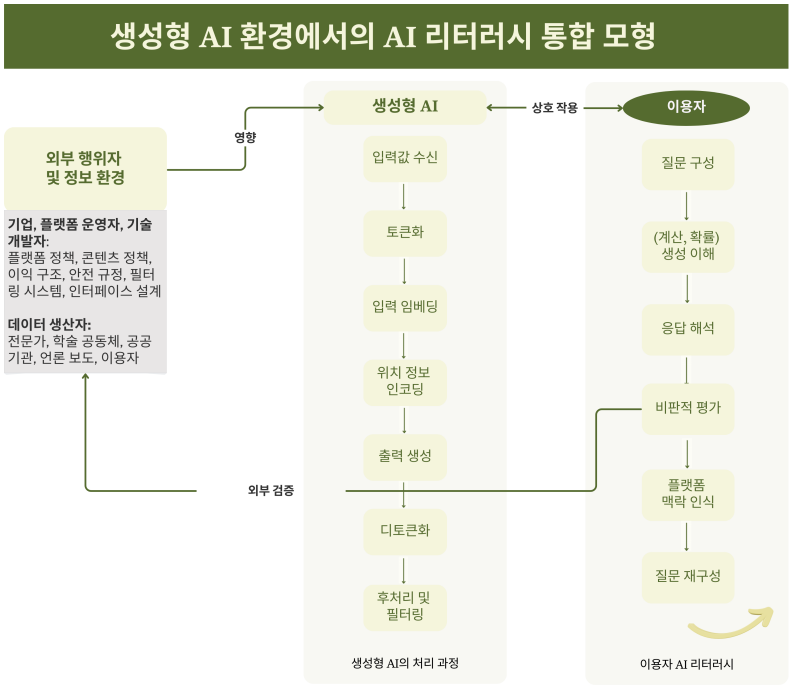
다만 고성능 모델이라 하더라도 여전히 오류를 포함하거나 부정확한 정보를 생성할 가능성이 있으며, 내부 추론 과정이 완전히 공개되지 않는 경우도 존재한다(Chen et al., 2025). 이러한 특성 때문에 AI 응답을 그대로 수용하기보다는 결과에 대한 검증 절차를 병행하는 것이 중요하다. 이를 위해 프롬프트 설계 과정에서 자기 검증(Self-verification)을 유도하는 것도 고려할 수도 있다(Nye, et al., 2021; Weng et al., 2023). 예컨대 스크래치 패드 기법(Scratchpad)은 프롬프트 내에 모델이 사고 과정을 정리할 수 있는 별도의 공간을 마련하여 생각 과정과 최종 응답을 분리하도록 하는 방식으로, 응답의 정확성을 높이는 데 활용될 수 있다.

다섯 번째, 플랫폼 리터러시는 생성형 AI의 응답이 중립적인 결과물이 아니라, 플랫폼의 설계, 정책, 안전 기준, 인터페이스, 서비스 목적 등에 의해 일정하게 조정되고 매개될 수 있음을 이해하는 능력이다. 같은 질문이라고 하더라도 플랫폼과 모델, 플랫폼이 운영되는 지역과 정책에 따라 응답 방식이나 제한 범위가 달라질 수 있다는 점을 인지하고, AI의 답을 특정한 기술적・제도적 맥락 속에서 형성된 결과로 이해할 수 있어야 하는 것이다. 이는 생성형 AI 결과를 기술적 산출물로만 보지 않고, 사회적・제도적 맥락 속에서 이해하도록 하는 데 중요한 역할을 한다.

마지막으로, 비판적 평가 리터러시는 생성형 AI가 제시한 결과를 사실성, 신뢰성, 적절성, 편향 가능성의 측면에서 검토하고, 필요한 경우 다른 정보와 비교하여 판단하는 능력이다. 결국 이용자는 생성형 AI가 생성해 낸 결과물과 추론의 근거가 외부 검증 가능한지, 출처가 있는지, 다른 대안 정보와 비교 가능한지를 점검할 줄 아는 능력이 필요하다.

4. AI 리터러시 교육 모형 제안

본 연구가 제안하는 AI 리터러시 교육 모형은 생성형 AI 활용을 단일한 정보 검색 행위가 아니라 질문 설계, 응답 생성, 결과 해석, 맥락 인식, 비판적 평가, 질문 재구성을 반복하는 순환적 탐구 과정으로 이해한다. 따라서 모형은 생성형 AI의 기술적 작동 단계와 이용자의 인지적 활동을 대응시키는 구조로 조직되며, 이러한 구조는 생성형 AI의 응답이 입력 해석, 계산적 처리, 확률적 생성, 정보 구성, 플랫폼 매개, 이용자 판단의 과정을 거쳐 형성된다는 점을 반영한다. 이에 따라 본 모형은 생성형 AI 환경에서 요구되는 리터러시를 단순한 정보 평가 능력이 아니라 질문 구성 능력, 생성 과정 이해, 응답 해석, 플랫폼 맥락 인식, 비판적 검증을 포함하는 다층적 역량으로 개념화한다.



〈그림 1〉 AI 리터러시 통합 모형

첫째, 입력 단계의 질문 구성 리터러시는 생성형 AI 활용의 출발점이 되는 역량으로, 이용자가 자신이 알고자 하는 문제를 명확히 정의하고 질문의 범위・목적・조건을 구조화하는 능력을 의미한다. 생성형 AI는 사용자가 입력한 문장을 그대로 해석하는 것이 아니라, 질문의 맥락, 표현 방식, 제약 조건, 암시된 의도까지 추측하여 응답을 생성하기 때문에, 질문이 모호하거나 편향되어 있으면 결과도 불명확하거나 왜곡될 가능성이 높다. 따라서 질문 구성 리터러시는 단순히 질문을 묻는 능력이 아니라, 문제 상황을 분석하고 핵심 변수를 선별해 AI가 처리할 수 있는 형태로 재구성하는 능력이라고 볼 수 있다. 이는 생성형 AI 환경에서 탐구의 방향과 품질을 좌우하는 가장 핵심적인 역량이다.

둘째, 계산 리터러시는 생성형 AI가 입력된 언어를 내부적으로 어떻게 처리하는지 이해하는 능력이다. 생성형 AI는 사람처럼 문장을 직접 이해하는 것이 아니라, 입력 문장을 토큰 단위로 분해하고 이를 벡터 임베딩으로 변환한 뒤, 요소 간의 관계를 계산하여 다음에 올 가능성이 높은 표현을 예측한다. 이러한 이해는 AI의 응답이 인간의 직관적 이해의 산물이 아니라 통계적 패턴과 수치 연산에 기반한 결과임을 인식하게 하며, 동시에 왜 표현의 미세한 차이가 결과에 영향을 미치는지 이해하는 데 도움을 준다.

셋째, 확률 생성 리터러시는 생성형 AI의 응답이 결정론적 산출이 아니라 확률적 예측에 기반해 생성된다는 점을 이해하는 역량이다. 생성형 AI는 학습된 언어 패턴을 바탕으로 각 시점에서 가장

가능성이 높은 표현을 선택하며 문장을 구성한다. 이 때문에 동일한 질문에도 표현이나 조건에 따라 다양한 응답이 생성될 수 있으며, 사실적으로 부정확하지만 그럴듯한 설명이 나타날 가능성도 존재한다. 이러한 이해는 환각, 응답 변동성, 편향의 가능성을 인식하는 기반이 된다. 따라서 확률 생성 리터러시는 생성형 AI의 응답을 절대적 사실이나 완결된 판단으로 받아들이지 않고, 어디까지나 확률적 예측에 기반한 결과로 이해하도록 돕는다.

넷째, 응답 해석 리터러시는 생성형 AI가 제시한 결과를 비판적으로 분석하고 의미를 재구성하는 능력이다. 생성형 AI의 응답은 문장 형태로 자연스럽게 제시되기 때문에 이용자는 이를 완성된 지식처럼 받아들이기 쉽지만, 실제로는 요약, 재구성, 추론, 보완이 혼합된 형태일 수 있다. 따라서 이용자는 응답 안에서 사실 진술, 추론, 예시, 해석, 권고가 어떻게 구분되는지 살펴보고, 어떤 정보가 직접적 근거에 기반한 것인지, 어떤 부분이 일반화나 추정인지 파악하는 습관을 길러야 한다. 응답 해석 리터러시는 결과를 읽는 능력인 동시에, 결과의 구조와 논리를 분석해 의미를 재구성하는 능력이며, 생성형 AI를 단순 소비가 아니라 비판적 활용의 대상으로 전환하게 만든다.

다섯째, 플랫폼 리터러시는 생성형 AI의 응답이 모델 자체의 성능뿐 아니라 플랫폼의 설계, 서비스 정책, 안전 규정, 데이터 연결 구조 등 다양한 기술적·제도적 조건에 의해 매개된다는 점을 이해하는 능력이다. 동일한 질문이라도 어떤 플랫폼과 모델을 사용하는지, 안전 필터가 어떻게 적용되는지, 시스템 프롬프트나 서비스 목적이 무엇인지에 따라 응답의 범위와 형식은 달라질 수 있다. 따라서 이용자는 생성형 AI를 특정한 기술적·상업적·제도적 환경 안에서 작동하는 플랫폼 기반 시스템으로 이해해야 한다.

마지막으로 이용자는 비판적 평가 리터러시를 통해 다른 정보원과 비교하거나 추가 검증을 통해 AI가 제공한 정보의 정확성, 신뢰성, 편향 가능성을 검토할 수 있다. 비판적 평가 리터러시는 생성형 AI를 다른 정보원과 교차 검증을 통해 책임 있게 활용하기 위한 최종 단계의 핵심 역량이라고 할 수 있다.

이러한 평가 과정은 다시 질문 수정과 재탐구로 이어지며, 생성형 AI 활용은 단일한 정보 검색이 아니라 반복적 질문과 검증을 통해 지식을 정교화하는 탐구 과정으로 이해될 수 있다. 따라서 AI 리터러시는 특정 기술의 사용 능력이 아니라 생성형 AI 환경에서 지속적으로 형성되는 탐구적 학습 역량으로 개념화될 필요가 있다.

Ⅳ. 논 의

궁극적으로 이와 같은 생성 방식은 전통적인 검색엔진과 구별되는 대규모 언어 모델 기반 생성형 AI의 핵심적 특성을 보여준다. 검색엔진이 외부 문서를 인덱싱하고 해당 정보의 위치를 가리키는

포인터를 제공하는 방식으로 작동하는 것과 달리 생성형 AI 모델은 학습 과정에서 방대한 텍스트 말뭉치로부터 관찰된 언어적 패턴과 분포를 수십억 개의 파라미터에 통계적으로 학습한다. 이 과정에서 개별 문서가 그대로 저장되는 것이 아니라, 다양한 표현 간의 동시 발생(Co-occurrence)적 관계와 확률적 구조가 압축된 형태로 모델 내부에 반영된다. 따라서 시스템의 역할은 특정 사실을 데이터베이스에서 검색하여 반환하는 것이 아니라, 학습된 파라미터 공간 내에서 주어진 문맥에 가장 부합하는 표현을 확률적으로 생성하는 데에 있다. 다시 말해 생성형 AI의 모델 내부에서 정보는 명시적 문장 단위로 저장되어 있다기보다 분포적(Distributional) 형태의 통계 구조로 존재하며, 실제 출력 텍스트는 이러한 분포로부터 문맥 조건에 따라 순차적으로 재구성(Reconstruction)된 결과로 이해할 수 있다.

본문의 분석을 종합하면, 생성형 AI 환경에서 요구되는 리터러시는 단순히 AI를 활용하는 능력이나 결과를 검증하는 태도로 환원될 수 없다. 생성형 AI는 이용자의 입력을 토큰화하고, 벡터 공간에서 의미를 압축・재배열하며, 유사성 기반 추론과 맥락적 가중치를 통해 응답을 구성하고, 이후 플랫폼 차원의 정책적・기술적 거버넌스를 거쳐 최종 출력으로 제시되는 다단계 매개 시스템이다. 이와 같은 정보의 생성・조정・해석의 전 과정을 고려할 때, 생성형 AI 환경에서는 단순한 정보 검색 능력을 넘어선 다층적 AI 리터러시가 요구된다. 뿐만 아니라 이용자의 정보 행위 역시 개별 정보원을 탐색・비교하는 방식에서 벗어나, 모델과 플랫폼이 선행적으로 구성한 설명 구조를 해석・평가하는 방향으로 재조직되어야 한다.

생성형 AI 리터러시는 단순한 도구 활용 능력을 넘어, 모델의 작동 원리에 대한 구조적 이해를 포함하는 다층적 구성 요소로 확장될 수 있다. 본 논문은 기술적 처리 과정, 통계적 생성 원리, 그리고 플랫폼 수준의 조정 구조를 종합적으로 이해하는 통합적 리터러시 역량을 생성형 AI를 책임 있고 효과적으로 활용하기 위한 핵심 전제 조건으로 본다. 이와 같이 제안된 생성형 AI 리터러시 구성 요소들은 생성형 AI의 계산적・확률적・담론적・플랫폼적 작동 구조를 단계별로 조망하게 함으로써, 이용자가 AI 출력의 설득력과 사실성을 구분하고 판단 조건 자체를 성찰할 수 있도록 하는 교육적 기반을 제공할 수 있다. 본 논문은 생성형 AI 시대의 정보 활용 능력이 단순한 정보 탐색이나 평가를 넘어, 언어모델의 작동 원리와 한계를 비판적으로 해석할 수 있는 방향으로 확장될 필요가 있음을 시사한다.

V. 결 론

본 연구는 생성형 AI의 확산으로 인해 변화하고 있는 정보 탐색 환경과 지식 형성 방식에 주목하고, 이러한 변화가 이용자의 정보 판단 과정에 어떠한 함의를 가지는지를 탐구하였다. 생성형

AI의 출력은 단순한 정보 전달이 아니라, 확률적 계산 과정과 플랫폼 차원의 설계가 결합되어 만들어진 결과로 이해할 수 있다. 나아가 이러한 출력은 이용자가 이를 해석・평가・활용하는 과정 속에서 의미가 형성되는 상호작용적 산물이며, 이를 바르게 사용하기 위해서는 다양한 인지적・비판적 역량이 요구된다고 보았다. 이에 본 연구는 생성형 AI의 응답 생성 구조를 분석하고, 이를 바탕으로 생성형 AI 환경에서 요구되는 AI 리터러시 교육 요소를 도출하였다. 구체적으로, 생성형 AI의 출력 생성 과정과 정보 제시 방식의 구조적 특성을 분석하고, 생성형 AI의 작동 구조와 이용자의 인지적 활동을 대응시키는 다층적 AI 리터러시 역량 체계를 제시하였다. 결론적으로 본 연구는 AI 리터러시 교육 요소로 질문을 구조화하는 문제 구성 리터러시, 생성형 AI의 계산 과정을 이해하는 계산 리터러시, 확률적 생성 원리를 이해하는 확률 생성 리터러시, 응답의 구성 방식과 제시 구조를 파악하는 응답 해석 리터러시, 응답이 플랫폼 설계와 정책에 의해 매개된다는 점을 인식하는 플랫폼 리터러시, 그리고 생성된 정보를 검토하고 신뢰성을 판단하는 비판적 평가 리터러시를 포함할 것을 제안한다.

본 연구는 생성형 AI, 특히 트랜스포머 기반 대규모 언어 모델의 기술적 작동 구조와 이용자의 정보 판단 과정을 연결하여 AI 리터러시 교육을 논의했다는 점에서 의의를 지닌다. 특히 생성형 AI의 응답 생성 구조와 정보 제시 방식에 주목하여, 이러한 기술적 조건이 이용자의 정보 인식과 판단 과정에 어떠한 변화를 가져오는지를 분석하고 이를 바탕으로 AI 리터러시 교육의 핵심 요소를 체계화하였다. 이러한 접근은 생성형 AI 환경에서 요구되는 리터러시 교육의 방향을 제시하고, 향후AI 리터러시 교육 프로그램 설계와 교육 연구에 이론적 기초를 제공할 수 있을 것으로 기대된다.

그럼에도 불구하고 본 연구는 몇 가지 한계를 지닌다. 첫째, 생성형 AI의 작동 구조와 응답 생성 과정을 중심으로 한 이론적 분석에 기반하였기 때문에 실제 교육 현장에서의 적용 가능성이나 학습 효과를 실증적으로 검증하지 못하였다. 둘째, 생성형 AI의 기술적 구조와 정보 제시 방식은 플랫폼과 모델에 따라 차이를 보일 수 있으므로 본 연구에서 제안한 교육 요소가 다양한 생성형 AI 시스템 전반에 동일하게 적용될 수 있는지에 대해서는 추가적인 검토가 필요하다. 또한 최근 연구(Chen et al., 2025)가 지적하듯이 AI 모델은 실제 추론 과정을 항상 투명하게 드러내지 않으며, 경우에 따라 부정확한 정보나 불완전한 단서를 바탕으로 그럴듯한 설명을 생성할 가능성도 존재한다. 이러한 점을 고려할 때 생성형 AI의 응답 생성 과정을 이해하기 위해서는 보다 심층적인 기술적 연구가 필요하다. 이러한 한계를 바탕으로 향후 연구에서는 본 연구에서 제안한 AI 리터러시 교육 요소와 모형을 토대로 구체적인 교육 프로그램을 설계하고 실제 학습 환경에서 그 효과를 검증하는 실증적 연구가 수행될 필요가 있다. 또한 다양한 생성형 AI 플랫폼과 활용 맥락을 고려하여 이용자의 정보 탐색 행위와 판단 과정이 어떻게 변화하는지를 분석하는 연구를 통해AI 리터러시 교육의 방향을 보다 정교하게 제시할 수 있을 것이다.

참 고 문 헌

김욱태, 김영식 (2024). 수학과 '확률과 통계' 영역에서 ChatGPT를 활용한 서답형 평가 피드백이 학생들의 피드백 리터러시에 미치는 영향. 컴퓨터교육학회 논문지, 27(3), 19-30. https://doi.org/10.32431/kace.2024.27.3.002

민혜림 (2026). 대학생의 인공지능(AI) 리터러시가 혁신행동에 미치는 영향: 창의적 자기효능감과 학습민첩성의 순차매개효과를 중심으로. 한국산학기술학회 논문지, 27(2), 945-955. https://doi.org/10.5762/KAIS.2026.27.2.945

박기범 (2022). AI 편향과 시민의 자질. 사회과교육, 61(2), 95-106. https://doi.org/10.37561/sse.2022.06.61.2.95

배현영, 조재희 (2025). 대규모 언어모델(LLM) 사용 빈도가 자기주도적 학습역량에 미치는 영향과 AI 리터러시의 조절효과: 구조방정식모형 분석. 디지털콘텐츠학회논문지, 26(8), 2281-2292. https://doi.org/10.9728/dcs.2025.26.8.2281

이승민 (2025). X-리터러시 핵심 역량 연계를 위한 개념적 프레임워크 구성. 한국도서관·정보학회지, 56(4), 303-325. https://doi.org/10.16981/kliss.56.4.202512.303

이유미 (2022). 디지털 시대 새로운 패러다임과 리터러시: 디지털 리터러시와 AI 리터러시를 중심으로. 교양학연구, 20, 35-60. https://doi.org/10.24173/jge.2022.07.20.2

정영미 (2025). 대학도서관 사서의 AI 리터러시 평가 루브릭 개발과 적용. 한국도서관·정보학회지, 56(4), 277-302. https://doi.org/10.16981/kliss.56.4.202512.277

황현정, 황용석 (2023). AI 리터러시 개념화와 하위차원별 세부 역량 도출에 관한 연구. 사이버커뮤니케이션학보, 40(2), 89-148. https://doi.org/10.36494/JCAS.2023.06.40.2.89

Adobe Express (2025). How ChatGPT is changing the way we search. Available: https://www.adobe.com/express/learn/blog/chatgpt-as-a-search-engine

Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and Machine Learning: Limitations and Opportunities. Cambridge, MA: MIT Press. Available: https://fairmlbook.org/

Bostrom, K. & Durrett, G. (2020). Byte pair encoding is suboptimal for language model pretraining. In Cohn, T., He, Y., & Liu, Y. (eds.), Findings of the Association for Computational Linguistics: EMNLP 2020, 4617-4624.

Chen, Y., Benton, J., Radhakrishnan, A., Uesato, J., Denison, C., Schulman, J., Somani, A., Hase, P., Wagner, M., Roger, F., Mikulik, V., Bowman, S. R., Leike, J., Kaplan, J., & Perez, E. (2025). Reasoning models don't always say what they think. [Preprint] arXiv. https://arxiv.org/abs/2505.05410

Cheng, Y. & Jiang, H. (2020). How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use. Journal of Broadcasting & Electronic Media, 64(4), 592-614. https://doi.org/10.1080/08838151.2020.1834296

Clark, A. & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7-19. http://www.jstor.org/stable/3328150

Cordonnier, J.-B., Loukas, A., & Jaggi, M. (2020). Multi-head attention: collaborate instead of concatenate. [Preprint] arXiv. https://arxiv.org/pdf/2006.16362

Creel, K. A. & Hellman, D. (2021). The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems. Canadian Journal of Philosophy, 1-18. https://doi.org/10.1017/can.2022.3

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-XL: Attentive language models beyond a fixed-length context. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2978-2988. https://doi.org/10.18653/v1/P19-1285

DataReportal (2024). What the world does online in 2024. Available: https://datareportal.com/reports/digital-2024-deep-dive-what-we-do-online

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171-4186. https://doi.org/10.18653/v1/N19-1423

Diakopoulos, N. (2016). Accountability in algorithmic decision-making. Communications of the ACM, 59(2), 56-62. https://doi.org/10.1145/2844110

Dive into Deep Learning. (2023). Multi-head attention. Dive into Deep Learning (Classic). Available: https://classic.d2l.ai/chapter_attention-mechanisms/multihead-attention.html

Evans, O., Cotton-Barratt, O., Finnveden, L., Bales, A., Balwit, A., Wills, P., Righetti, L., & Saunders, W. (2021). Truthful AI: developing and governing AI that does not lie. [Preprint] arXiv https://arxiv.org/abs/2110.06674

First Page Sage (2025). Google vs ChatGPT Market Share: 2026 Report - First Page Sage. Available: https://firstpagesage.com/seo-blog/google-vs-chatgpt-market-share-report/

Fritch, J. W. & Cromwell, R. L. (2001). Evaluating Internet resources: Identity, affiliation, and cognitive authority in a networked world. Journal of the American Society for Information Science and Technology, 52(6), 499-507. https://doi.org/10.1002/asi.1081

Geva, M., Schuster, R., Berant, J., & Levy, O. (2021). Transformer feed-forward layers are key-value memories. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 5484-5495. Association for Computational Linguistics. https://aclanthology.org/2021.emnlp-main.446/

Goodwin, D. (2025, September 26). Google is still 210x bigger than ChatGPT in search. Search Engine Land. Available: https://searchengineland.com/google-210x-bigger-chatgpt-search-462604

Google (2026). Grounding with Google Search. Gemini API. Available: https://ai.google.dev/gemini-api/docs/google-search

Grefenstette, G. (1999a). Tokenization. In H. van Halteren (Ed.), Syntactic Wordclass Tagging. Text, Speech and Language Technology, 9. Springer. https://doi.org/10.1007/978-94-015-9273-4_9

Grefenstette, G. (1999b). The World Wide Web as a resource for example-based machine translation tasks. In Proceedings of Translating and the Computer 21, London, UK. Aslib. https://aclanthology.org/1999.tc-1.8.pdf

Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E. (2025). Why language models hallucinate [Preprint]. arXiv. https://arxiv.org/abs/2509.04664

Kaplan, G., Oren, M., Reif, Y., & Schwartz, R. (2025). From tokens to words: On the inner lexicon of LLMs. In The Thirteenth International Conference on Learning Representations (ICLR 2025). https://openreview.net/forum?id=328vch6tRs

Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., Zhang, X., Zhao, T., Panalkar, A., Mehta, D., Pasquali, S., Cheng, W., Wang, H., Liu, Y., Chen, Z., Chen, H., White, C., Gu, Q., Pei, J., Yang, C., & Zhao, L. (2025). Domain specialization as the key to make large language models disruptive: A comprehensive survey. ACM Computing Surveys, 58(3), Article 79. https://doi.org/10.1145/3764579

Linna, E. & Linna, T. (2026). Challenges for generative AI in legal reasoning. Discover Artificial Intelligence. https://doi.org/10.1007/%20s44163-026-00902-3

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. Journal of the American Society for Information Science and Technology, 58(1), 2078-2091. https://doi.org/10.1002/asi.20672

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New

York: New York University Press.

Nye, M., Andreassen, A. J., Gur-Ari, G., Michalewski, H., Austin, J., Bieber, D., Dohan, D., Lewkowycz, A., Bosma, M., Luan, D., Sutton, C., & Odena, A. (2021). Show your work: Scratchpads for intermediate computation with language models. arXiv. https://doi.org/10.48550/arXiv.2112.00114

OpenAI (2015). Introducing OpenAI. Available: https://openai.com/index/introducing-openai/

OpenAI (2022). Introducing ChatGPT. Available: https://openai.com/index/chatgpt/

OpenAI (2024). Introducing ChatGPT search. Available: https://openai.com/index/introducing-chatgpt-search/

OpenAI Developers (2025). Using GPT-5.2. Available: https://developers.openai.com/api/docs/guides/latest-model

Roumeliotis, K. I. & Tselikas, N. D. (2023). ChatGPT and Open-AI models: A preliminary review. Future Internet, 15(6), 192. https://doi.org/10.3390/fi15060192.

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1715-1725. https://doi.org/10.18653/v1/P16-1162

StatCounter (2026). Search engine host market share worldwide. Available: https://gs.statcounter.com/search-engine-market-share/desktop/worldwide

Tate, M. A. & Alexander, J. E. (1999). Web wisdom: How to evaluate and create information quality on the Web (1st ed.). Boca Raton: CRC Press. https://doi.org/10.1201/9780429195556

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Attention is all you need. Proceedings of the 31st Conference on Neural Information Processing Systems, 6000-6010. https://doi.org/10.48550/arXiv.1706.03762

Wathen, C. N. & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. Journal of the American Society for Information Science and Technology, 53(2), 134-144. https://doi.org/10.1002/asi.10016

West, P., Lu, X., Dziri, N., Brahman, F., Li, L., Hwang, J. D., Jiang, L., Fisher, J., Ravichander, A., Chandu, K., Newman, B., Koh, P. W., Ettinger, A., & Choi, Y. (2023). The

generative AI paradox: “What it can create, it may not understand”. Proceedings of the International Conference on Learning Representations (ICLR) 2024.

Weng, Y., Zhu, M., Xia, F., Li, B., He, S., Liu, S., Sun, B., Liu, K., & Zhao, J. (2023). Large language models are better reasoners with self-verification. Findings of the Association for Computational Linguistics: EMNLP 2023, 2550-2575.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. [Preprint] arXiv. https://doi.org/10.48550/arXiv.1609.08144

xAI (2026). Introduction. Available: https://docs.x.ai/developers/introduction

Zhang, H., Sun, S., Wu, H., & Song, D. (2024). Controllable text generation with residual memory transformer. Findings of the Association for Computational Linguistics: ACL 2024, https://doi.org/1048-1066. 10.18653/v1/2024.findings-acl.62

Zhang, Y., Jiang, J., Ma, G., Lu, Z., Wang, B., Huang, H., Yuan, J., & Duan, N. (2025). Generative pre-trained autoregressive diffusion transformer. Proceedings of the Conference on Neural Information Processing Systems (NeurIPS 2025). https://doi.org/10.48550/ARXIV.2505.07344

Zhao, W., Ren, X., Hessel, J., Cardie, C., Choi, Y., & Deng, Y. (2024). WildChat: 1M ChatGPT interaction logs in the wild. Proceedings of the International Conference on Learning Representations (ICLR 2024). https://doi.org/10.48550/arXiv.2405.01470

• 국한문 참고문헌의 영문 표기
(English translation / Romanization of references originally written in Korean)

Bae, Hyun Young & Cho, Jae Hee (2025). Large language model usage frequency and self-directed learning competence: Moderating of AI literacy (SEM Analysis). Journal of Digital Contents Society, 26(8), 2281-2292. https://doi.org/10.9728/dcs.2025.26.8.2281

Hwang, Hyeon Jeong & Hwang, Yong Suk (2023). A study on conceptual constructs

of AI literacy with a focus on AI literacy competence. Journal of Cybercommunication Academic Society, 40(2), 89-148. https://doi.org/10.36494/JCAS.2023.06.40.2.89

Jung, Young Mi (2025). Development and pilot testing of an AI literacy assessment rubric for academic librarians. Journal of Korean Library and Information Science Society, 56(4), 277-302. https://doi.org/10.16981/kliss.56.4.202512.277

Kim, Wook Tae & Kim, Yung Sik (2024). The influence of ChatGPT-enhanced evaluation feedback on students' feedback literacy in descriptive assessments within the 'probability and statistics' domain of mathematics education. The Journal of Korean Association of Computer Education, 27(3), 19-30. https://doi.org/10.32431/kace.2024.27.3.002

Lee, Seung Min (2025). A conceptual framework for linking the core competencies of X-literacy. Journal of Korean Library and Information Science Society, 56(4), 303-325. https://doi.org/10.16981/kliss.56.4.202512.303

Min, Hye Lim (2026). The effect of artificial intelligence(AI) literacy on innovative behavior among university students: The sequential mediation effect on creative self-efficacy and learning agility. Journal of Korea Academia-Industrial Cooperation Society, 27(2), 945-955. https://doi.org/10.5762/KAIS.2026.27.2.945

Park, Kee Burm (2022). AI bias and citizenship. Social Studies Education, 61(2), 95-106. https://doi.org/10.37561/sse.2022.06.61.2.95

Yi, Yu Mi (2022). New paradigm and literacy in the digital era: focusing on digital literacy and AI literacy. The Journal of General Education, 20, 35-60. https://doi.org/10.24173/jge.2022.07.20.2