

# ChatGPT, 생성형 AI 시대 도서관의 데이터 리터러시 교육에 대한 연구\*

**A Study on the Data Literacy Education in the Library of the Chat GPT,  
Generative AI Era**

이정미 (Jeong-Mee Lee)\*\*

## 목 차

- |                   |                                    |
|-------------------|------------------------------------|
| 1. 서론             | 4. 생성형 AI를 활용한 도서관의 데이터<br>리터러시 교육 |
| 2. 이론적 배경         | 5. 결론 및 제언                         |
| 3. 생성형 AI의 교육적 활용 |                                    |

## 초 록

본 연구의 목적은 ChatGPT와 같은 생성형 AI 시대를 맞아 이와 같은 언어모델에 대해 소개하고, 이를 활용한 도서관의 데이터 리터러시 교육 구성요소를 고민하고 방향을 제시하고자 하는 연구이다. 이를 위해 다음과 같은 세 가지 연구 문제를 제시하였다. 먼저 ChatGPT 유사 언어모델의 기술적 특징을 살펴보고, 이후 생성형 인공지능 기술 기반 서비스 플랫폼을 활용하여 적합한, 정확한 정보를 유용하게 활용하기 위한 이용자의 데이터 리터러시 역량 교육의 필요성을 주장하였다. 마지막으로 ChatGPT 시대 도서관 데이터 리터러시 교육을 위해 데이터에 대한 이해, 데이터 생성, 데이터 수집, 데이터 검증, 데이터 관리, 데이터 이용 및 공유, 데이터 윤리와 같은 7개 구성항목을 포함한 데이터 리터러시 교육 구성안을 제안하였다. 결론적으로 ChatGPT와 같은 생성형 인공지능 기술이 이용자의 정보 활용에 많은 영향을 미치게 될 것이라 예상되는 만큼 도서관은 이러한 기술의 장단점, 문제점 등에 대해 한발 먼저 고민하고 이를 통해 도서관의 정보서비스를 한층 개선할 수 있는 토대로 삼아야 할 것을 강조하며 마무리했다.

## ABSTRACT

The purpose of this study is to introduce this language model in the era of generative AI such as ChatGPT, and to provide direction for data literacy education components in libraries using it. To this end, the following three research questions are proposed. First, the technical features of ChatGPT-like language models are examined, and then, it is argued that data literacy education is necessary for the proper and accurate use of information by users using a service platform based on generative AI technology. Finally, for library data literacy education in the ChatGPT era, it is proposed a data literacy education scheme including seven components such as data understanding, data generation, data collection, data verification, data management, data use and sharing, and data ethics. In conclusion, since generative AI technologies such as ChatGPT are expected to have a significant impact on users' information utilization, libraries should think about the advantages, disadvantages, and problems of these technologies first, and use them as a basis for further improving library information services.

키워드: ChatGPT, 데이터리터러시, 대규모언어모델, 생성형 AI, 도서관 서비스  
ChatGPT, Data Literacy, Large Language Model, Generative AI, Library Services

\* 이 논문은 2021학년도 서울여자대학교 연구년 수혜 및 2023년 교내연구비의 지원을 받았음(2023-0156).

\*\* 서울여자대학교 사회과학대학 문헌정보학과 교수(jmlee@swu.ac.kr / ISNI 0000 0004 6832 5178)

논문접수일자: 2023년 7월 26일 최초심사일자: 2023년 8월 4일 게재확정일자: 2023년 8월 12일

한국문헌정보학회지, 57(3): 303-323, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.3.303>

※ Copyright © 2023 Korean Society for Library and Information Science

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.

## 1. 서 론

### 1.1 연구의 필요성

ACRL(Association of College & Research Libraries)은 협회 활동 및 전략적 계획 검토를 위해 대학 및 연구도서관 분야의 학술정보서비스에 연관된 우리 사회의 환경분석내용을 포함하는 보고서를 2년에 한번 씩 제작, 출판하고 있다. 이 보고서는 도서관의 학술정보서비스에 영향을 줄 수 있는 다양한 환경적 측면에 대해 검토하고 심층적으로 분석한 것으로 학술도서관이 직면하고 있는 주요 도전과 기회를 제시하고, 새로운 기술이나 동향 등에 적응함으로써 이후 학술정보서비스를 위한 정책적인 의사 결정이나 전략적 계획 수립을 위한 혜안을 제공하고자 하는 목적으로 제작된다.

「2023 환경 분석(2023 Environmental Scan)」 보고서는 2023년 3월에 발표되었으며 2021년, 2022년 두 해동안 우리가 겪었던 다양한 환경적 변화를 분석한 가운데 이후에도 우리에게 영향을 줄 수밖에 없는 주제들에 대해 구체적으로 정리, 분석한 내용을 포함하고 있다. 「2023 환경 분석(2023 Environmental Scan)」 보고서에 따르면 고등교육과 이와 연관된 학술정보서비스에 있어 기존의 주된 관심사였던 학술 커뮤니케이션, 학생 평가, 인쇄물 공유와 같은 주제가 계속해서 영향을 미칠 것이지만 입법 개입, 통제된 디지털 대출과 같은 디지털 금융, 디지털 전환(Digital Transformation: DX)과 같은 새로운 주제도 언급하고 있다. 또한 다양성, 협평성, 포용성 이슈와 함께 COVID-19의 지속적인 영향, 기술 발전의 지속적인 영향도 공통적

인 주제임을 확인되고 있다(Flierl et al., 2023).

이 보고서에서 정리한 주제 중 하나인 학술 커뮤니케이션의 경우 팬데믹 이후 전 세계 학술 교류 양상에 대한 변화를 볼 수 있는 것으로 “open data and data maximization”이라는 소주제를 통해 알 수 있듯이 데이터의 공유와, 최대한의 데이터에 대한 사회적 관심에 대해 강조하고 있다. 이러한 강조점은 디지털 전환(DX)과 연관해 해석이 가능한데 도서관이 효율적인 학술정보서비스를 운영하고, 이용자의 학술 커뮤니케이션 과정에서 최대한의 데이터를 공유 가능하게 하기 위해서는 디지털 시대에 따른 기술적 변화를 수용하고 최대한 활용해야 함을 의미한다고 해석할 수 있을 것이다.

최적의 데이터 활용에 영향을 주는 기술적 동향 중 최근 가장 눈에 띄는 기술은 인공지능(Artificial Intelligence: AI) 기술의 발전과 그 하나님의 사례로 짧은 기간 많은 이용자의 호응을 불러일으킨 생성형 AI, ChatGPT 서비스이다. 2022년 말 서비스를 시작한 이후 짧은 기간 상당한 이용자를 끌어들임으로써 관심과 걱정, 기대를 한 몸에 받고 있는 서비스 중 하나이다. ChatGPT의 출현과 더불어 많은 연구자들이 ChatGPT의 학술적 유용성에 대한 강조와 함께 ChatGPT가 제시하는 응답에 대한 신뢰성에 대해 의문을 제기하고 우려를 제시하고 있는 것도 사실이다.

최근 10년 동안 인간 생활에 영향을 준 정보 기술 중 인공지능 기술의 발전은 인간 사회에 강력한 영향력을 준 기술 중의 하나일 것이다. 최근에는 인공지능 기술의 접근성이 높아지면서 인공지능이 더욱 보편화되고 있으며 이런

모습 때문에 대다수의 이용자들은 이러한 기술을 활용하는데 여전히 높은 기술적 난이도와 지식이 요구된다는 점을 간과하고 있는 것도 사실이다(Kreinsen & Schulz, 2023). ChatGPT와 같은 자연어 생성·응답 제공 능력은 이용자에게 보다 친화적으로 학습을 지원해주거나 교육적인 용도로 활용하기 쉬울 것이라는 잠재력을 보여주고 있는 것이라 할 수 있으나 ChatGPT가 제시하고 있는 응답이 얼마나 신뢰할 수 있는가에 대한 전문가의 우려 또한 상당한 수준에 이르러 있다.

ChatGPT가 자연어를 이해하고 대화를 생성해서 응답하는 시스템이기 때문에 이용자와의 자연스럽고 친숙한 대화가 가능한 시스템으로 더욱 발전 가능하다는 것, DALL·E 2 서비스에서 확인해 볼 수 있는 것처럼 특정 주제 또는 제시어에 대한 콘텐츠를 다양한 형태로 생성할 수 있다는 것 등은 기존의 인공지능 서비스에서의 이용자 경험에 기계적 창의성이 더해질 수 있다는 점은 ChatGPT의 활용 가능성이 대한 기대와 잠재력으로 이해할 수 있을 것이다. 반면 여타 인공지능 모델에서 확인한 것처럼 어떤 데이터를 학습했는가에 따라 응답의 질이 달라질 수 있다는 것, 콘텐츠의 생성과 제공 과정에서 표절이나 저작권 등의 윤리 문제가 발생할 수 있다거나, 제공된 데이터의 정확성이 확인되지 않는다는 문제는 정보의 가치적 측면에서 상당한 문제점으로 인식될 수밖에 없는 문제이다.

기술적 발전은 언제나 도서관의 정보서비스에 있어 도전이며 기회로 인식되어 왔다. 빅데이터의 출현 시에도 우리는 도전과 기회, 이를 활용하기 위한 도서관 서비스의 방향을 고민하고

대비해왔으며(이정미, 2013), 이는 ChatGPT와 같은 생성형 AI 챗봇 서비스 활용에 있어서도 그 방향을 고민하고 대비하는 과정이 필요하다 하겠다.

현대 사회는 데이터의 의미와 역할이 강력하며 각종 의사결정에 있어 데이터가 근거자료로 활용되고 있는 사회이다. 이렇듯 사회변화에 데이터가 기반되는 현상 때문에 우리는 지금 우리 사회를 데이터 기반 사회라고도 부르고 있다(한상우, 2017).

ChatGPT의 활용은 ChatGPT가 제시하는 응답의 신뢰성을 우려한다고 하더라도 ChatGPT 이용자의 급증과 유사한 언어모델을 활용한 생성형 AI 시장의 규모가 급증하는 상황이고 이와 더불어 양질의 데이터 생산 및 관리에 대한 요구도 급증하고 있다(백봉삼, 2023).

시장의 요구는 양질의 데이터를 가지고 상품화될 생성형 AI 서비스의 신뢰성을 높이겠지만 그럼에도 불구하고 이용자의 입장에서는 활용할 생성형 AI 서비스 플랫폼에서 제공되는 데이터를 검증과정없이 무작정 활용하기에는 상당한 오류의 가능성도 없지 않다. 바로 이러한 문제 때문에 이용자 측면에서의 데이터 활용 역량에 대한 교육이 절대적으로 필요하며 문헌정보학이나 도서관 측면에서 데이터 리터러시 교육의 역할을 더욱 주목해야 할 시점이다.

문헌정보학이나 도서관 정보서비스 측면에서도 ChatGPT와 같은 생성형 AI 서비스 플랫폼의 대중화는 어쩌면 당연한 것으로 인식한다면, 이용자가 이러한 기술들을 어떻게 더욱 더 잘, 효과적으로, 신뢰성 높게 이용 가능하게 할 것인가에 대한 고민이 필요하다 할 것이다.

## 1.2 연구 목적 및 연구 문제

본 연구는 ChatGPT와 같은 생성형 AI 서비스 시대에 있어 ChatGPT와 같은 언어모델에 대해 이해할 수 있도록 문현을 검토·정리하여 기술적 특징을 제시하고, 이를 적절히 활용 가능한 도서관의 데이터 리터러시 교육에 대해 고민하고 방향을 제시하고자 하는 연구이다. 문현 고찰을 통해 ChatGPT와 최근의 데이터 리터러시 연구 방향을 재검토하고, 도서관이 ChatGPT의 한계를 극복하고 활용할 수 있는 데이터 리터러시 교육의 구성요소와 더불어 방향성에 대해 검토하고자 한다.

이를 위해 다음과 같은 세 가지 연구문제를 제시하여 살펴보고자 한다.

- 연구문제 1. ChatGPT 유사 언어모델의 기술적 특징은 무엇인가?
- 연구문제 2. ChatGPT 시대 데이터 리터러시는 어떻게 이해되어야 하는가?
- 연구문제 3. ChatGPT 시대 도서관 데이터 리터러시 교육의 구성요소 및 방향성은 어떻게 제시되어야 하는가?

## 2. 이론적 배경

### 2.1 ChatGPT 유사 언어모델의 이해

ChatGPT는 인공지능 기반 자연어 처리모델로 이용자와의 대화를 통해 상호작용하고 콘텐츠를 생성할 수 있도록 개발된 기술이며 OpenAI에서 개발해 2022년 말부터 서비스 이용자가

급증했다(Firat, 2023).

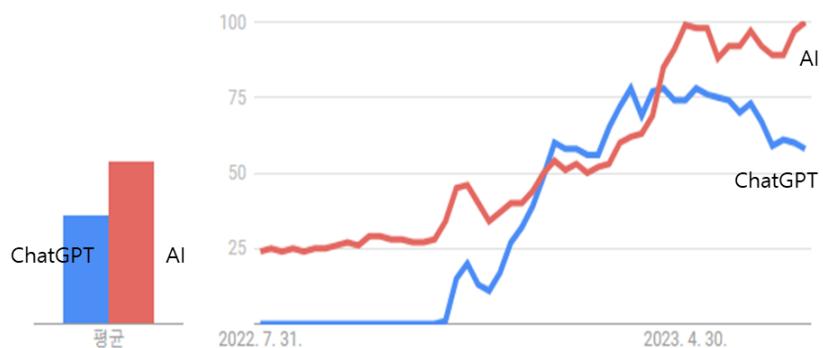
ChatGPT는 생성형 인공지능(Generative Artificial Intelligence)의 대표적인 사례로서 생성형 인공지능 기술은 텍스트, 오디오, 이미지 등 기존 콘텐츠를 활용해 유사한 콘텐츠를 새롭게 생산해내는 인공지능 기술이다. 생성형 인공지능 기술은 콘텐츠들을 수집, 패턴을 학습한 후 이를 바탕으로 추론한 결과로 새로운 콘텐츠를 만들어내는 것이다. 기존의 인공지능 기술이 대상을 이해하기 위해 데이터와 패턴을 학습했다고 한다면 생성형 인공지능은 기존 데이터를 학습하고 기존 데이터와의 비교 학습을 통해 새로운 창작물을 창조할 수 있도록 한다. 이러한 기술의 결과로 기존 텍스트를 활용한 작문, 소설 창작 뿐 아니라 그림을 그려내거나 음악을 작곡하는 등의 콘텐츠 생성이 가능한 기술이다.

인공지능 기술이 개발된 이후 기술에 대한 관심은 지속적으로 유지되었으나 생성형 인공지능 기술을 구현한 ChatGPT에 대한 이용자의 폭발적 관심과 더불어 인공지능에 대한 관심 또한 증가했다(〈그림 1〉 참조).

현재 생성형 인공지능 서비스로 대표적인 사례는 ChatGPT, 구글 Bard 등이 있다. 국내의 경우 카카오와 네이버, LG 등이 생성형 인공지능 서비스 개발에 박차를 가하고 있는 것으로 알려져 있다.

ChatGPT는 스스로를 “OpenAI에서 개발한 대규모 언어모델로 GPT-3.5 아키텍처를 기반으로 하며 자연어 처리 기술의 최신 발전을 적용해 사용자와의 자연스러운 대화를 수행하는 인공지능 챗봇”으로 소개하고 있으며 보다 구체적인 기술적 특징과 내용은 다음 〈그림 2〉와

2022 ChatGPT 출현 이후 AI와 ChatGPT 검색어의 검색 트렌드 비교



〈그림 1〉 ChatGPT 출현 이후 AI와 ChatGPT의 검색 트렌드 비교

같다(<https://chat.openai.com/?model=text-davinci-002-render-sha>).

Bard는 일찍이 개발에 잠시 중단되었다고 알려진 구글의 생성형 인공지능 서비스로 ChatGPT 이용자의 급증과 더불어 생성형 AI에 대한 관심의 증폭과 함께 실험버전의 서비스가 운영되고 있다. Bard는 스스로를 “Google AI에서 개발한 대규모 언어 모델로 방대한 양의 텍스트 데이터에 대해 교육을 받은 인공지능”이라 소개하고 있으며 구체적인 기술 특징은 다음 〈그림 3〉과 같다(<https://bard.google.com/?hl=ko>).

MS의 빙 검색과 연동된 Bing챗 또한 생성형 인공지능 서비스의 사례로 거론되기도 하지만 이는 GPT를 기반으로 하는 서비스이기 때문에 그 기술적 특징은 ChatGPT와 동일한 것으로 이해하는 것이 적절하다 하겠다.

정리하면 ChatGPT나 구글 Bard 같은 생성형 인공지능 기술은 수집된 데이터와 학습 패턴을 가지고 이용자의 질의어에 적합한 응답을 제공하기 위해 개발되었기 때문에 이용자가 일상에서 사용하는 자연어에 대한 이해가 중요하

다. 또한 질의어의 맥락적 분석 또한 응답 적합성에 중요한 영향을 미칠 수밖에 없다. 대규모 언어 모델인 만큼 학습에 활용되는 데이터의 크기도 중요하지만 얼마나 많은 “양질의” 데이터를 보유하고 있으며, 얼마나 적합한 방식으로 학습되어 운영되는가가 응답의 질을 좌우하게 되는 것이다.

## 2.2 데이터 리터러시의 필요성과 재접근

데이터 리터러시의 필요성은 데이터 기반 사회에 살고 있다는 현실적인 각성뿐 아니라 팬데믹으로 인한 비대면 상호작용의 급증과 더불어 랜선으로 확인 가능한 다양한 데이터를 이해하고 해석해야만 하는 현실적인 필요에 의해 서도 더욱더 강조되는 추세에 있다.

일반적인 필요성 강조 차원에서 Carmi와 그 동료들의 경우 데이터 리터러시 역량이 충족되지 않은 시민들의 경우 사회는 점점 데이터화로 나아가는 반면 시민들은 오류 데이터를 파악하거나 데이터를 이해하고 활용하기 어려워

챗 지피티(ChatGPT)는 OpenAI에서 개발한 대규모 언어 모델로, GPT-3.5 아키텍처를 기반으로 합니다. 이 모델은 자연어 처리(Natural Language Processing, NLP) 분야의 최신 기술을 활용하여 인간과 자연스럽게 대화하는 인공지능 챗봇을 구현합니다. 다음은 챗 지피티의 주요 기술적 특징과 소개입니다:

1. GPT-3.5 아키텍처: 챗 지피티는 GPT-3.5 아키텍처를 기반으로 합니다. GPT(Generative Pre-trained Transformer)는 언어 이해와 생성을 위한 Transformer 기반의 모델로, 많은 양의 데이터로 사전 훈련(pre-training)된 뒤, 특정 작업에 맞게 미세 조정(fine-tuning)되어 다양한 NLP 작업을 수행할 수 있습니다.
2. 대규모 언어 모델: 챗 지피티는 수백억 개의 파라미터를 가진 매우 큰 언어 모델입니다. 이 큰 모델을 통해 높은 언어 이해 능력과 다양한 문제에 대한 대응력을 보여줍니다.
3. 다국어 지원: 챗 지피티는 다양한 언어를 지원합니다. 이를 통해 전 세계의 사용자들과 다양한 언어로 대화할 수 있으며, 다국어 환경에서도 효과적으로 사용될 수 있습니다.
4. 문맥 파악: 챗 지피티는 사용자와의 대화에서 이전 대화 문맥을 파악하여 응답을 생성합니다. 이를 통해 사용자의 질문과 요청에 더 정확하고 의미 있는 답변을 제공할 수 있습니다.
5. 다양한 NLP 작업 지원: 챗 지피티는 다양한 자연어 처리 작업을 수행할 수 있습니다. 문서 요약, 번역, 질의응답, 자연어 이해(NLU), 자연어 생성(NLG) 등 다양한 작업에 적용할 수 있습니다.
6. 지속적인 업데이트: 챗 지피티와 같은 대규모 언어 모델은 지속적인 학습과 업데이트를 통해 개선됩니다. OpenAI는 사용자들의 피드백과 데이터를 기반으로 모델을 업데이트하여 더 나은 성능을 제공하고자 합니다.

챗 지피티는 사용자들과의 자연스러운 대화를 통해 다양한 정보를 제공하고, 문제를 해결하며, 사용자들의 다양한 요구를 충족시키는데 사용됩니다. 이러한 기술적 특징으로 인해 챗 지피티는 인공지능 챗봇 분야에서 혁신적인 역할을 하고 있으며, 계속해서 발전해나가고 있습니다.

〈그림 2〉 ChatGPT의 기술적 특징 및 내용

Bard는 Google AI에서 개발한 대규모 언어 모델(LLM)입니다. LLM은 방대한 양의 텍스트 데이터에 대해 교육을 받은 인공 지능으로, 텍스트를 생성하고, 언어를 번역하고, 다양한 종류의 창의적인 콘텐츠를 작성하고, 유익한 방식으로 질문에 답변할 수 있습니다.

Bard는 방대한 양의 텍스트와 코드 데이터 세트에 대해 교육을 받았습니다. 이 데이터 세트에는 책, 기사, 코드, 웹 사이트 등 다양한 소스의 텍스트가 포함되어 있습니다. Bard는 이 데이터 세트를 사용하여 텍스트를 생성하고, 언어를 번역하고, 다양한 종류의 창의적인 콘텐츠를 작성하고, 유익한 방식으로 질문에 답변하는 방법을 배웠습니다.

Bard는 아직 개발 중이지만 다음을 포함하여 다양한 종류의 작업을 수행하는 방법을 배웠습니다.

- 귀하의 지시를 따르고 귀하의 요청을 신중하게 완료하기 위해 최선을 다하겠습니다.
- 질문이 개방적이거나 도전적이거나 이상하더라도 포괄적이고 유익한 방식으로 답변하기 위해 제 지식을 사용할 것입니다.
- 시, 코드, 대본, 음악 작품, 이메일, 편지 등과 같은 다양한 창의적인 텍스트 형식을 생성합니다. 귀하의 모든 요구 사항을 충족시키기 위해 최선을 다하겠습니다.

Bard는 항상 배우고 발전하고 있으며, 앞으로 더 많은 일을 할 수 있기를 기대합니다.

〈그림 3〉 구글 Bard의 기술적 특징 및 내용

지기 때문에 해당 사회에서 당연히 도태될 수 없음을 강조하며 데이터 리터러시의 필요성을 부각시키기도 했다(Carmi et al., 2020).

Pangrazio와 Sefton-Green(2020)과 같이 문해력 교육이 제대로 이루어져야만 미래 사회의 경제적 가치가 확보된다 강조하면서 데이터화 되

어가는 사회의 대응책으로서 데이터 리터러시의 중요성을 역설하는 연구도 존재한다 (Pangrazio & Sefton-Green, 2020).

이러한 연구들은 데이터 리터러시의 중요성을 언급하면서 강조하는 지점마다의 차이는 있지만 핵심적인 주장은 데이터 리터러시를 사회적 환경적 맥락과 분리할 수 없으며 데이터에 의한 의사결정이 늘어갈수록 데이터 리터러시의 중요성이 커질 수밖에 없음을 강조하는 것이라 할 수 있다.

데이터 리터러시에 대한 가장 최근 연구는 다양한 리터러시 개념과 함께 데이터 리터러시를 재검토하거나 데이터 리터러시를 다양한 주제, 대상에 따라 실질적인 교육 과정에 구현하고자 하는 실험적 연구들의 증가하는 경향을 볼 수 있다.

Kreinsen과 Schulz은 현대 사회에서 디지털 기술과 인공 지능의 발전으로 인해 데이터와 정보에 대한 이해와 활용 능력인 데이터 리터러시의 중요성이 더욱 부각되고 있는 상황에서, 교사들에게는 디지털 리터러시 뿐만 아니라 데이터 리터러시와 인공지능 리터러시를 함께 갖추는 것이 필요함을 강조하고 있다. 중요하게 이 세 가지 리터러시의 통합적 접근이 교육 분야에 계속적으로 유입되는 최신기술을 효과적으로 활용하는데 필수적일 것임을 강조했다(Kreinsen & Schulz, 2023).

Lund와 Agbaji의 최신 논문 또한 다양한 리터러시와 함께 데이터 리터러시의 중요성을 강조하고 있다(Lund & Agbaji, 2023). 이 연구는 정보 리터러시, 데이터 리터러시, 개인정보 리터러시 및 ChatGPT와 같은 기술 리터러시에 대해 다루면서 이러한 다양한 기술 리터러시

개념들이 현대 사회에서 발전하는 기술들과 함께 어떻게 어우러지고 상호작용하는지에 대해 서술하고 있다. 이 연구에서는 정보 리터러시는 정보를 효과적으로 탐색, 평가, 이해하고 활용하는 능력을 의미하며, 데이터 리터러시는 데이터를 수집, 분석, 시각화하여 의미 있는 통찰력을 제공하는 능력으로 정의하였으며 개인정보 리터러시는 개인정보를 보호하고 이해하는 능력으로 정리하였다. 더불어 ChatGPT와 같은 자연어 처리 기술을 이해하는 것은 새로운 방식으로 커뮤니티와 상호작용하고 소통하는데 영향을 미치기 때문에 이러한 기술을 이해하고 활용하기 위한 기술 리터러시의 중요성을 강조하고 있다(Lund & Agbaji, 2023).

Tang과 Han(2022)은 데이터 리터러시 교육에 대한 연구들을 조사하고, 이러한 연구들의 동향과 결과를 종합하여 개괄적인 개념을 제시하고 있다. 이 연구는 또한 데이터 리터러시 교육의 주요 측면과 핵심 요소들을 제시하면서 데이터 리터러시 교육을 데이터의 수집, 분석, 시각화, 해석 등 다양한 단계와 기술들을 포괄하는 복합적인 영역으로 기술하고, 이를 효과적으로 교육하는 데에는 교육 방법과 커리큘럼의 설계, 교재 및 도구의 선택 등이 중요한 역할을 한다고 강조하고 있다. 또한 데이터 리터러시 교육의 효과와 도전점에 대해서도 논의하고 있는데 데이터 리터러시 교육이 학생들의 정보 검색 및 분석 능력 향상에 효과적이며, 진로와 학업 성과에도 긍정적인 영향을 줄 수 있지만 데이터 리터러시 교육을 성공적으로 구현하기 위해서는 교육자들의 지식과 능력, 교육 기관의 지원과 자원 등 다양한 도전점들을 극복해야 한다고 강조하고 있다. 이

는 데이터 리터러시는 교육받는 사람과 교육하는 사람 모두에게 역량 강화가 필요한 영역임을 강조하고 있는 것으로 이해가능하다(Tang & Han, 2022).

Pothier가 Condon과 함께 발표한 2020년 논문은 데이터 리터러시 역량 함양의 중요성을 강조하면서 비즈니스 주제 분야에서 데이터의 활용은 기업의 의사 결정과 성과에 큰 영향을 미치는 핵심자원으로 인식되고 있으며 데이터를 이해하고 효과적으로 활용할 수 있는 능력은 기업 경영 전반에 걸쳐 필수적인 요구사항으로 인식되고 있음을 강조하고 있다. 이러한 이유에서 사서는 산업 현장의 요구사항에 부합하는 데이터 리터러시 교육과정의 개발 등 데이터 리터러시 교육의 핵심적 역할을 해야 한다는 사서의 역할에 대한 강조도 담고 있다(Pothier & Condon, 2020).

사회과학 분야에서의 데이터 활용능력 함양이 절대적으로 필요하고 이는 데이터 리터러시 교육을 통해 가능함을 강조하는 연구들(Ruedinger et al., 2022; Condon, Exline, & Buckley, 2023) 또한 발표되었는데 이 연구들은 데이터 리터러시의 주된 측면을 수학적 기초지식에 기반한 수치적 데이터 이해와 활용에 더 강조점을 두고 있다는 측면에서 조금은 미시적인 관점으로 데이터 리터러시를 이해하고 있는 것으로 볼 수 있을 것이다.

구체적으로 데이터 리터러시 교육의 구성요소 등을 포함한 것은 아니지만 사회학이라는 하나의 연구 분야에서의 대용량 데이터와 기계학습 응용 가능성이나 한계를 서술하며 사회학에 새로운 관점을 제시한 연구(Leitgöb, Prandner, & Wolbring, 2023)도 살펴보기에 흥미로운 연

구이다. 연구자들은 이 연구를 통해 사회학자들이 기술적 지식과 주제 분야 지식을 융합하여 기계 학습 기술을 적절하게 활용하는 것이 중요하다고 강조하고 있으며 데이터를 해석하고 모델링하기 위해서는 사회학적 이해와 기술적 능력을 조화롭게 결합하여 현실적이고 유의미한 결과를 도출하는 것이 필요하다고 강조하고 있다. 이 연구는 결론적으로 대용량 데이터와 기계 학습 기술이 사회학 분야에 미칠 영향력이 큰 만큼 이러한 기술들을 사회학 연구에 적절하게 활용할 수 있을 때 사회학 분야에서의 기술적 혁신이나 심도깊고 구체적인 데이터를 근거로 하는 사회현상의 이해가 가능할 것임을 강조하고 있다.

Cui와 그의 동료들(2023)은 데이터 리터러시 평가에 관한 체계적인 문헌 고찰을 포함하고 있다. 이 연구에서는 데이터 리터러시 평가는 개인이 데이터를 이해하고 분석하여 의사결정을 내릴 수 있는 능력을 평가하는 것으로, 데이터 리터러시 교육의 효과를 측정하고 학습자들의 데이터 리터러시 능력을 개선하는데 중요한 도구로 인식해야 함을 강조하고 있다. 데이터 리터러시 평가의 주요 추세와 기존 연구들의 결론들을 정리하고 분석했으며, 결론적으로 데이터 리터러시는 다양한 분야에서 필요한 능력이지만, 평가를 위한 표준화된 척도나 평가 도구들이 아직 충분히 개발되지 않았다는 점을 언급하고 있다. 학습자들의 데이터 리터러시 역량 함양을 위해서는 개별적인 학습자의 차이나 교육 방법의 효과 등을 고려하는 것이 중요하다는 점도 강조하면서 더욱 체계적인 평가 도구들이 개발되어야 하며, 데이터 리터러시 평가를 교육과정과 교육 방법들과 융합하는 노력도 필요할 것

이라 주장하고 있다(Cui et al., 2023).

국내 연구들의 다수는 데이터 리터러시와 더불어 AI 리터러시에 대해 많은 강조점을 두고 진행되고 있는데, ChatGPT와 같은 생성형 인공지능 시대의 필수역량으로서 AI리터러시를 강조하는 연구(구본권, 2023; 최숙영, 2022)들이 나타나고 있다.

이외에 초·중등 AI 교육을 위해 데이터 리터러시를 정의하고 교육과정에 포함할 수 있는 구성 요소 등을 정리한 연구(김슬기, 김태영, 2021)도 있는데, 이 연구의 경우 현재 초·중등 교육과정에서 진행하고 있는 인공지능 교육을 위해서는 데이터 리터러시 교육이 필요하며 데이터 리터러시의 정의와 함께 데이터 수집과 가공 능력, 데이터 분석과 해석 능력, 인공 지능과 기술의 이해, 데이터 윤리와 개인정보 보호 등의 주제가 데이터 리터러시 교육에 포함되어야 한다고 정리했다. 다른 연구들이 AI 리터러시를 독자적인 하나의 영역으로 분리하여 데이터 리터러시와 별개의 것으로 정의하고 정리하고 있는 것을 살펴볼 때 이 연구는 AI에 관한 지식이 인공지능과 기술에 대한 이해라는 “기술적 지식”的 이해로 다뤄져서 데이터 리터러시의 구성항목에 포함되어 있다는 점이 다른 연구와의 다른 점이라고 분석할 수 있다.

이렇듯 데이터 리터러시 역량의 함양은 데이터 사회의 생애능력으로서나 각 주제 분야에서의 경쟁력으로서나 중요한 것으로 그 필요성은 아무리 강조해도 지나치지 않다. 데이터 리터러시 교육을 위한 구성요소 개발이나 교육적 초점 등은 연구가 거듭될수록 발전하고 있다.

Fontichiaro와 Johnston(2020)의 논문은 데이터 리터러시 교육에 대한 이해와 인식에 중요한 통찰력을 제공하고 있다(Fontichiaro & Johnston, 2020). 이 연구는 교육자들의 데이터 리터러시 우선순위에 대한 인식 변화에 초점을 맞추고 체계적인 방법론을 사용해 그 인식 변화를 조사하였다. 설문조사와 인터뷰를 실시하여, 데이터 리터러시에 대한 인식과 우선순위에 대한 의견을 수집, 분석하였다. 결론적으로 첫째, 교육자들의 데이터 리터러시에 대한 인식은 시간이 지남에 따라 변화하는 경향을 보이고 있는데, 기준에는 데이터 수집과 분석 등 기술적인 측면에 중점을 두었던 교육자들이 점차 데이터의 해석과 활용, 윤리적 측면 등을 더욱 강조하는 방향으로 변화하고 있음을 발견하였다. 둘째, 교육 전문가들의 조언과 교육 환경의 변화가 교육자들의 데이터 리터러시 우선순위에 영향을 미치는 것으로 나타났으며 교육 전문가들의 지도와 교육 기관의 지원이 있을 때 데이터 리터러시를 더욱 중요시하고 우선순위로 여기는 경향을 보였다. 셋째, 데이터 리터러시에 대한 인식 변화는 교육자들의 교육 경험과 배경에 따라 다를 수 있다는 것을 확인했다. 교육자들이 데이터 리터러시에 대한 교육을 받거나 직접 경험하는 것이 인식 변화에 큰 영향을 미치는 것이다. 이러한 결과는 데이터 리터러시 교육의 효율성을 증진하기 위해서는 교수자에 대한 교육전문가의 사전 교육이 필수적으로 이루어져야 하며, 기관이나 교육전문가의 지원에 바탕을 둔 (데이터와 기술에 대한 기본 지식을 사전에 교육받은 상태에서) 데이터의 해석과 활용, 윤리적 측면에 교육방향을 둘 때 데이터 리터러시 교육의 효과성이 클 것으로 해석 가능하다.

도서관에서의 데이터 리터러시 교육은 첨단 기술의 활용을 위한 단기적이고 미시적인 교육으로 이해되어서는 안된다. 이런 이유로 Fontichiaro 와 Johnston이 제시하는 세 가지 통찰력있는 결론은 이후 데이터 리터러시 교육의 구성에 중요한 이정표가 될 수 있을 것이다. ChatGPT 와 같은 생성형 AI 서비스 플랫폼의 대중화를 정보기술의 진화와 인간의 정보기술 활용에 있어 당연한 과정으로 인식한다면, 이용자가 이러한 기술들을 어떻게 더욱 더 잘, 효과적으로, 신뢰성 높게 이용 가능하게 할 것인가에 대한 고민이 필요하고 이것이 데이터 리터러시 교육을 필요로 하는 이유일 것이다.

### 3. 생성형 AI의 교육적 활용

ChatGPT와 같은 생성형 AI를 교육적으로 활용하는 사례는 AI리터러시의 필요성 제기와 함께 외국어교육과 코딩교육 등에서 최근 빈번하게 발표되고 있다(구본권, 2023). 정규 교육에 적용한 교육적 활용과는 별도로 소셜미디어에서도 개인이 활용한 방법과 효율성을 동영상 등으로 업로드한 사례도 최근 급증하고 있는 추세이다.

#### 3.1 생성형 AI의 교육적 활용

현재 생성형 AI의 교육적 활용은 ChatGPT 플랫폼 중심으로 연구되고 있는 것이 현실이다.

Dawa와 그의 동료들의 연구(2020)는 학술적 글쓰기에서 ChatGPT의 활용에 대해 고찰한 연구를 발표하였는데 학생과 교수진의 흥미

도도 높게 나타났으며 학술적 글쓰기 교육과 기술 능력 향상에 기여할 수 있다고 판단하였다. 그러나 빠르고 쉽게 글을 작성할 수 있다는 장점에도 불구하고 표절, 저작권의 문제 또는 정보 검증이 필수적이라는 문제가 중요하게 제기되었다(Dawa et al., 2020).

유사한 연구로 Firat(2023)은 ChatGPT 기술의 발전으로 인해 대학 교육에서도 새로운 가능성과 도전이 제기되고 있음을 인식하고 GPT에 대한 학자들과 학생들의 인식과 이해 수준을 조사했다. 조사 결과에 따르면 학자들은 대학 교육에서 ChatGPT를 어떻게 활용할 수 있는지에 대해 관심을 가지고 있으며, 학생들의 경우 이러한 기술이 자신들의 학습 경험에 어떤 영향을 미칠지에 대해 궁금해 하고 있음이 확인되었다. 조사결과를 바탕으로 이 연구는 학생들의 학습을 개선하고, 개별 학생에게 맞춤형 지도를 제공하는 등의 가능성을 제시하고자 하였다. 더욱 효과적인 학습방법으로서 활용가능하다는 가능성에도 불구하고 ChatGPT의 활용과 관련해서는 인공지능 기술을 활용하는 것이 가져올 수 있는 윤리적인 문제들에 대해 유의해야하고 학자들과 학생들은 ChatGPT의 사용에 있어서 개인정보 보호와 학문적인 정확성을 염두에 두고 활용해야 함을 강조하였다(Firat, 2023).

Shiri(2023)는 학계에서 ChatGPT 활용할 때 침해 가능한 학문적 무결성에 대한 우려를 언급하고 있는데, 생성형 AI 기반 서비스를 활용하면 자연어에 대한 이해를 바탕으로 이용자의 질의에 적합한 콘텐츠를 쉽게 생성하지만 이로 인해 학습자인 학생은 학문적으로 부정행위를 시도할 가능성도 높아질 수밖에 없음을 우려하

면서 학문적 무결성을 유지하기 위해서는 강력한 학생 지도 및 교육 프로그램이 시급함을 주장하고 있다(Shiri, 2023).

Haleem과 그의 동료들(Haleem et al., 2022) 또한 ChatGPT를 정보 활용에 있어 혁신적인 지원도구로 인식 가능하다고 주장하고 있으나 더불어 개인정보보호의 문제나 데이터의 오·악용의 문제, 응답의 편견이나 오류 발생의 문제 등을 염두에 두면서 지속적인 연구와 대응이 요구된다 강조하였으며 ChatGPT의 활용에 있어 인간 중심적인 관점과 윤리적인 책임을 고려하는 것이 중요함을 주장하고 있다(Haleem et al., 2022).

생성형 AI의 관심과 더불어 대화형 콘솔을 개발하고 활용가능하게 하는 프롬트엔지니어링(Prompt Engineering)에 대한 관심도 급증하고 있는데 이는 사용자가 입력한 명령어에 대한 해석과 이를 이해한 가운데 명령어의 실행이나 오류 처리 등에 있어 프롬트엔지니어의 역량이 해당 AI 서비스 플랫폼의 개선에 상당한 기여를 할 수 있기 때문이다. 관련한 연구로 Lo(2023)의 연구가 흥미로운데 Lo는 이 연구에서 정보 능력을 강화하기 위한 교육 프레임워크로 “프롬트 엔지니어링을 통한 CLEAR(Concise, Logical, Explicit, Adaptive, and Reflective) 경로”라는 프레임워크를 제안하고, 이를 통해 정보 활용 능력 교육이 효과적으로 개선되고 학습자들의 역량 강화에 기여할 수 있다고 주장한다(Lo, 2023).

### 3.2 생성형 AI의 활용시 한계 및 문제점

앞서 검토한 연구들의 핵심은 생성형 AI를

교육적으로 활용하는데 있어 분명히 그 효율성이 있음을 방증하는 것이라 할 것이다. 그러나 현재 생성형 AI를 교육적으로 활용하는데 있어서는 생성형 AI가 가져다 주는 정보에 대한 신뢰성이나 표절 등의 문제점도 존재한다.

Kirtania와 Patra(2023)는 문헌정보학 분야의 관심 용어 약 10개를 ChatGPT에 넣어 콘텐츠를 생성하게 한 후 웹기반 표절 탐지 서비스를 통해 유사성 비율을 확인하게 함으로써 생성된 콘텐츠의 유사성을 확인하는 실험을 진행했다. 연구 결과에 따르면 유사성 비율은 13% 정도로 크지 않은 것으로 나타났다(Kirtania & Patra, 2023).

Kirtania와 Patra의 연구결과만을 비추어보면 생성형 AI의 활용이 인간의 창의력을 보충해주는 유용한 도구로 활용될 수 있다고 판단할 수 있으나 유사성 비율을 통한 표절 검사만으로는 정보의 오류를 검증하기 어렵다는 한계가 존재한다.

〈그림 4〉는 “대한민국의 대통령제”라는 질문에 대한 ChatGPT와 구글 Bard의 응답을 확인한 것이다.

이 두 응답에서 명백한 정보의 오류는 대한민국 대통령의 임기에 대한 정보이다. ChatGPT와 구글 Bard 모두 대한민국 대통령의 임기를 5년, 한번 연임이 가능해 최대 2회까지 가능한 것으로 응답하고 있으나 대한민국헌법 제4장 제1절 제70조에 따르면 대한민국 “대통령의 임기는 5년으로 하되 중임할 수 없다”고 명시하고 있다(대한민국헌법). ChatGPT와 구글 Bard 모두 명백한 오류 정보를 제시하고 있는 것이다.

이용자의 입장에서 명백한 오답이라면 바로 잡거나 해당 정보를 활용하지 않으면 그만이지

### ChatGPT



2021년 9월을 기준으로 대한민국은 대통령 제도를 갖고 있습니다. 대한민국의 정치 체제는 대통령제와 의원제의 특징을 가지고 있습니다.

대한민국의 대통령제는 대통령이 국가의 수도인 서울에서 집권하는 체제를 의미합니다. 대한민국의 대통령은 국민들에 의해 직접 선출되며 임기는 5년입니다. 대통령은 한 번의 연임을 허용받을 수 있으며, 연임 후에는 재직할 수 없습니다.

대한민국의 대통령은 국가의 최고 지도자로서 국정을 책임지고, 국내외 정책 결정, 법률 제정, 정부 기관 임명, 외교 등 다양한 권한을 가지고 있습니다. 그러나 국회의 동의를 받아야 하는 사

### Bard



대한민국의 대통령제는 1948년 제헌 헌법을 통해 도입된 제도로, 대통령은 국민의 직접 선거로 선출되고 5년의 임기를 가지며, 국회의 동의를 얻어 국무위원을 임명하고 내각을 구성합니다. 대통령은 국가의 원수이자 행정부의 수반으로서, 국정 전반에 대한 책임을 지고 있습니다.

대한민국의 대통령제는 미국의 대통령제와 유사한 점이 많지만, 몇 가지 차이점이 있습니다. 대한민국의 대통령은 미국의 대통령과 달리 국회의 동의 없이도 법률을 공포할 수 있으며, 국회의원과 국무위원을 탄핵할 수 있습니다. 또한, 대한민국의 대통령은 1987년 헌법 개정으로 임기 2회에 제한되었습니다.

〈그림 4〉 “대한민국의 대통령제”에 대한 ChatGPT와 구글 Bard의 응답예시

만 생성형 AI의 응답들은 위의 〈그림 4〉에서 보는 바와 같이 학습 데이터를 활용해 “그럴듯한 가짜” 또는 “진짜같은 거짓말” 응답을 제시하고 있다는 점에서 제공되는 정보의 신뢰성을 검증하지 않고는 활용하기 어렵다는 것은 커다란 문제점이라고 할 수 있다.

ChatGPT와 같은 생성형 AI 서비스가 이용자에게 편리하고 쾌적한 검색 환경을 제공하는 것은 확실하다. 그러나 정보의 신뢰성, 개인정보의 문제 등 다양한 사회적 문제를 야기할 수 있음을 명심해야 한다. 이런 현실에서 데이터 리터러시 역량의 함양은 지금과 같은 데이터 기반 사회, ChatGPT와 같은 생성형 AI 서비스 이용자에게는 서비스 이용이 가져다주는 편의성을 영유하기 위해 데이터 리터러시 교육이 필수적이다. 생성형 AI의 지속적인 개선과 성

능 향상으로 인간에 버금가는 콘텐츠 생성능력을 제공하는 것도 기대할 바이지만 현재 우리가 경험하는 생성형 AI 서비스 활용에 있어서는 응답 데이터에 대한 검증이 필수적이고 이런 이유로 데이터 리터러시 교육은 데이터 사회 데이터 이용자로서 발생 가능한 데이터 오류의 문제를 데이터 활용을 위한 검증 역량의 강화로 극복할 수 있는 가능성을 열어두는 탈출구라 할 수 있을 것이다.

#### 4. 생성형 AI를 활용한 도서관의 데이터 리터러시 교육

문현정보학이나 도서관 정보서비스 측면에서 데이터 리터러시는 교육적 측면과 연구지원

서비스에 적용할 필요성이 많이 부각되어왔다 (이정미, 2019; 김지현, 2018).

근래 들어 데이터 리터러시에 대한 관심이 증가하긴 했으나 ChatGPT와 같은 생성형 AI 서비스에 대한 급격한 관심과 더불어 AI 리터러시라는 개념이 등장하고 인공지능에 대한 교육 프로그램 또한 늘어가면서 AI 리터러시라는 다소 협소한 리터러시 개념이 확산되고 있는데 이는 다양한 리터러시 교육을 연구하는 측면에서는 다소 우려스러운 면이 없지 않다. 지금과 같은 정보기술의 발전이 거듭될수록 새로운 첨단 기술이 인류의 발전을 이끌어나가게 될 것이고, 그 과정에서 AI와 같은 기술들 또한 나타나게 될 것이다. 이런 시기마다 새로운 정보기술명과 함께 “\*\*\*” 리터러시라는 이름의 협소한 리터러시 개념을 만들어내고 확산시킨다면 이는 정말 중요한 인류의 생애능력으로서의 데이터 리터러시(또는 정보 리터러시)<sup>1)</sup>에 대한 오해를 가져올 소지가 있다고 판단된다. 다만 앞서 언급한 “\*\*\*” 리터러시를 데이터 리터러시(또는 정보 리터러시)의 하위 리터러시로서 개념화하고, 데이터 리터러시 교육 구성 요소에 포함시킬 수 있음을 인지하는 것이 중요하다.

본 연구에서는 ChatGPT와 같은 생성형 AI 기술 또한 2023년 현재 데이터 리터러시 교육의 구성항목에 포함해야 할 중요한 기술적 요소로 간주하고, 더불어 도서관이 이러한 기술을 효과적으로 활용하고자 한다는 것을 전제로

교육적 활용에 대해 논하고자 한다.

#### 4.1 데이터 리터러시 교육과정 사례

앞서 언급한 것처럼 시장의 요구는 양질의 데이터를 가지고 상품화될 생성형 AI 서비스의 신뢰성을 높이겠지만 그럼에도 불구하고 이용자의 입장에서는 활용할 생성형 AI 서비스 플랫폼에서 제공되는 데이터를 검증과정없이 무작정 활용하기에는 상당한 오류의 가능성도 없지 않다. 바로 이러한 문제 때문에 이용자 측면에서의 데이터 활용 역량에 대한 교육이 절대적으로 필요하며 문헌정보학이나 도서관 측면에서 데이터 리터러시 교육의 역할을 더욱 주목해야 할 시점이다.

Usoba와 Laws는 2021년 “Teaching a One-Credit Course on Data Literacy and Data Visualisation”이라는 논문을 통해 데이터 리터러시의 데이터 시각화에 대한 1학점 강좌의 경험을 공유하고 있다. 이 논문에서 Usoba와 Laws는 데이터 기반 사회에서 데이터 리터러시 교육은 필수적이며 사서들은 데이터 리터러시 교육을 제공할 수 있는 좋은 위치에서, 최선의 역할을 할 수 있다 주장하면서 사서의 역할을 강조하기도 했다. 여기서 제시된 데이터 리터러시 교육은 총 6차시로 구성되어 있으며 각 차시별 주제는 다음 〈표 1〉과 같다(Usoba & Laws, 2021).

1) 본 연구는 데이터 기반 사회에서 데이터 이해에 대한 중요성을 강조하는 측면에서 데이터 리터러시를 생애 능력이라 간주한다. 다만 정보 리터러시를 보다 보편적으로 “데이터 가공된 상태의 결과물로서의 정보”로 이해했을 때 정보리터러시 또한 저자가 주창하는 데이터 리터러시와 밀접한 관련이 있다고 판단해서 정보 리터러시는 협소한 개념의 다른 리터러시와 차별성을 둔다.

〈표 1〉 데이터 리터러시 1학점 강좌의 차시별 주제(Usoba & Laws, 2021)

차시	주제
1	데이터 시각화와 기본 디자인 원칙
2	데이터 검색 및 처리
3	차트, 그래프, 지도, 소프트웨어 사용
4	데이터로 스토리텔링하기
5	개별 학습 및 기술개발을 위한 도구
6	최종 프로젝트 발표

Burress와 그의 동료들은 대학교에서 데이터 리터러시 교육에 대한 교수진의 관점과 교육 접근방법에 대해 조사 분석하는 가운데 데이터 리터러시 교육의 주제와 연관된 6가지 주요 개념을 식별하였다(Burress et al., 2021). 교수진의 관점에서 데이터 리터러시 교육에 중요한 6가지 주요 개념은 다음과 같이 정의된다.

- 데이터 리터러시의 가치 - 데이터 리터러시의 가장 중요한 가치에 대한 인식을 학생들과 공유
- 데이터 윤리 - 윤리와 관련된 우려 사항을

#### 공유

- 정보 활용 능력 및 연구과정 - 정보 활용 능력의 주요하거나 전형적인 신념 또는 구성요소를 설명하는 어휘를 사용
- 데이터 세트 - 데이터 세트 활용 기회 제공을 위한 방법적 측면의 전략을 논의
- 수학 불안 - 수치, 수학 리터러시와 관련해 학생들이 준비되어있는지 확인
- 교수법 - 다양한 기술/교육 방법의 활용

교수진 인터뷰의 코딩 및 분석 결과를 바탕으로 이 연구에서는 다음 〈표 2〉에서 볼 수 있는 것처럼 4개 그룹의 수정된 데이터 리터러시 역량을 제시했다.

데이터 리터러시의 역량에 대한 종합적인 정리를 통해 각각의 연구자와 주장하는 데이터 리터러시 역량에 대해 종합한 연구들을 살펴보면 현재 학자들이 주장하는 데이터 리터러시 역량의 구성요소들을 구체적으로 확인할 수 있다. 대표적으로 Cui와 그의 동료들이 데이터 리터러시 평가에 관한 문헌 고찰과정에서 정리한 데이터

〈표 2〉 Burress와 그의 동료들이 제시한 데이터 리터러시 역량(Burress et al., 2023)

역량	수준 및 대상
• 데이터가 일상생활에 어떻게 통합되는지 인식	• 입문수준 • 모든 분야
• 데이터 소스 해석 비판적 평가 • 데이터 분석 • 부분적 시각화 사용한 고객별 효과적 데이터 전달	• 복잡성이 증가하는 모든 수준 • 모든 분야
• 데이터 유형 및 형식 읽기/이해 • 가설 테스트/연구질문응답 • 데이터 세트 발견, 선택, 접근, 생성	• 입문 수준 • 자연과학 • 상위수준 • 사회과학
• 데이터의 윤리적 수집/이용/인용 • 다른 정보원 및 사전 지식을 통해 다른 맥락으로 데이터 통합 및 합성	• 상위 수준 • 자연과학 • 모든 수준 • 복합성증가

리터러시 역량과 Potheir와 Condon이 경영 주제 분야 학생들의 기업 환경에 기반한 데이터 리터러시 역량을 분석하면서 정리한 데이터 리터러시 역량이 있다(Cui et al., 2023; Potheir & Condon, 2019)(<표 3> 참조).

최근 연구들에서 제시되고 있는 데이터 리터러시 교육 역량의 구성요소 살펴보면 연구자들이 어떤 역량에 중점을 두었는가에 따라 다양한 구성요소를 제시하고 있음을 알 수 있다. 대부분의 연구자들은 데이터 수집(또는 추

<표 3> 데이터 리터러시 역량 구성요소(Cui et al.(2023)과 Potheir & Condon(2019), 이정미(2019) 연구에서 추출/재구성

구분	내용	
Schneider (2013)	<ul style="list-style-type: none"> <li>• 확인</li> <li>• 범위</li> <li>• 계획</li> <li>• 비축</li> </ul>	<ul style="list-style-type: none"> <li>• 보호</li> <li>• 평가</li> <li>• 관리</li> <li>• 제공</li> </ul>
Calzada Prado and Marzal (2013)	<ul style="list-style-type: none"> <li>• 데이터에 대한 이해</li> <li>• 데이터 발견/획득</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 읽기, 이해, 평가</li> <li>• 데이터 관리</li> <li>• 데이터 활용</li> </ul>
Ridsdale et al. (2015)	<ul style="list-style-type: none"> <li>• 데이터 발견 및 수집</li> <li>• 평가/정보원 품질 확인</li> <li>• 데이터 조작</li> <li>• 데이터 조작</li> <li>• 기본적인 데이터 분석</li> <li>• 데이터 해석</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 활용 문제 확인</li> <li>• 데이터 시각화</li> <li>• 데이터 제시(구술)</li> <li>• 데이터 기반 의사결정</li> <li>• 데이터 인용</li> <li>• 데이터 공유</li> </ul>
Merk et al. (2020)	<ul style="list-style-type: none"> <li>• 데이터 수집</li> <li>• 데이터 변형</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 감소</li> <li>• 데이터 해석</li> </ul>
Wu et al. (2021)	<ul style="list-style-type: none"> <li>• 데이터 추출</li> <li>• 데이터 표현</li> <li>• 데이터 해석</li> <li>• 데이터 추론</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 계량화</li> <li>• 데이터 분포</li> <li>• 데이터 연관성</li> <li>• 데이터 무작위성</li> </ul>
Mahardika (2021)	<ul style="list-style-type: none"> <li>• 데이터 수집</li> <li>• 데이터 품질 평가</li> <li>• 데이터 분석</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 해석</li> <li>• 데이터 실행</li> <li>• 데이터 평가</li> </ul>
Pratama et al. (2021)	<ul style="list-style-type: none"> <li>• 확인</li> <li>• 분석</li> <li>• 소통</li> <li>• 평가</li> <li>• 처리</li> </ul>	<ul style="list-style-type: none"> <li>• 활용</li> <li>• 특정</li> <li>• 실행</li> <li>• 해석</li> </ul>
Kōuts-Klemm (2019)	<ul style="list-style-type: none"> <li>• 데이터 발견</li> <li>• 데이터 품질 평가</li> <li>• 데이터의 맥락적 해석</li> <li>• 비평 도구로 데이터 제시</li> <li>• 고객의 이해와 능력, 필요 정도에 따른 취득</li> </ul>	
이정미 (2019)	<ul style="list-style-type: none"> <li>• 데이터의 정의 및 범위</li> <li>• 데이터 생성</li> <li>• 데이터 수집</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터 관리</li> <li>• 데이터 이용 및 공유</li> <li>• 데이터 윤리</li> </ul>

출, 발견, 확인, 획득), 해석(또는 이해, 품질 확인, 분석, 추론), 활용(또는 표현, 계량화, 소통, 실행, 이용, 공유, 제시)과 평가(인용, 처리, 관리)라는 네 개 단계를 필수로 하는 데이터 리터러시 역량을 제시하고 있는 것으로 나타났다.

이는 데이터 리터러시 교육의 구성요소들이 실제 이용자들이 데이터를 찾고, 이해하고, 활용한 후 평가하는 것을 일반화하는 것으로 판단된다. 이용자가 수집한 데이터가 실제적으로 적절한 것인지는 판단하지만 해당 데이터가 “올바른” 데이터인지에 대한 검증단계는 명확히 제시하지 않고 있는 것이다.

〈표 3〉은 국내외 데이터 리터러시 역량 구성요소를 재구성한 것이다.

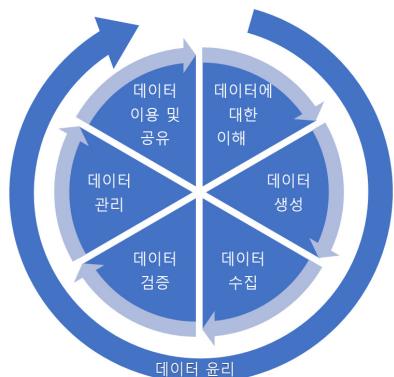
이렇듯 선행연구를 통해 학자들이 제시하는 데이터 리터러시 역량의 구성요소들에서 주목할 점은 데이터 리터러시 교육 구성요소에 관한 최근의 연구에는 해석, 시각화, 품질 평가 등의 구성요소가 더욱 강조되어 나타나고 있다. 하지만 이러한 구성요소들의 경우 검증의 단계 자체가 데이터를 활용한 이후로 제시되고 있어 ChatGPT와 같은 생성형 AI의 확산과 함께 데이터를 획득하는 것만큼 데이터 획득 즉시 해당 데이터가 “올바른” 데이터인지에 대한 검증을 거쳐 데이터의 신뢰성을 확보해야만 데이터 활용이 가능한 상황에서는 데이터 획득-검증의 연결 단계가 데이터 리터러시 교육 구성에 있어 필수적으로 고려되어야 한다고 판단된다.

#### 4.2 생성형 AI 활용 도서관 데이터 리터러시 교육과정의 구성

앞서 살펴본 바와 같이 데이터 획득-검증의

연결 단계가 데이터 리터러시 교육 구성에 있어 필수적으로 고려되어야 한다고 했을 때 데이터 리터러시의 교육과정 구성은 기존의 교육과정과 다소 다르게 구성될 수 있다.

다음 〈그림 5〉는 교육 구성 각 단계의 흐름을 간략하게 보여주는 도식이다.



〈그림 5〉 데이터 리터러시 교육 흐름도

도식에서 보는 바와 같이 제시된 데이터 리터러시 교육 안에서 각각의 구성요소들은 순차적으로만 진행됨을 의미하는 것은 아니다. 일반적으로 각각의 구성항목은 교육과정에 따라 순차적인 설계가 가능하지만 데이터 윤리와 같은 구성항목은 교육과정 전 단계에 걸쳐 학습 방법 및 내용이 설계되어야 한다.

검토한 생성형 AI의 교육적 활용에 있어서의 장단점을 염두에 두고, 데이터 리터러시의 최근 연구 검토의 결과와 데이터 획득-검증의 연계를 필수적으로 고려해 도서관의 데이터 리터러시 교육과정을 재구성하였으며 이는 다음 〈표 4〉와 같이 제안하고자 한다.

제시된 교육 구성안이 강조하는 부분은 선행 연구들과 다르게 데이터 수집 이후 즉각적인

〈표 4〉 도서관의 데이터 리터러시 교육 구성안

구성 항목	내용	교육 활동
데이터에 대한 이해	<ul style="list-style-type: none"> <li>데이터의 정의와 범위</li> <li>디지털 데이터, 물질적·비물질적 아날로그 데이터</li> </ul>	<ul style="list-style-type: none"> <li>다양한 데이터 유형 발견, 확인, 구분</li> <li>일상에서의 데이터 경험</li> </ul>
데이터 생성	<ul style="list-style-type: none"> <li>데이터 이용 목적 수립</li> <li>디지털 및 아날로그 데이터 생성 및 인지</li> </ul>	<ul style="list-style-type: none"> <li>데이터 이용 목적 수립</li> <li>목적에 적합한 데이터 유형 결정/생성</li> </ul>
데이터 수집	<ul style="list-style-type: none"> <li>데이터 이용 목적 구체화</li> <li>비판적 사고를 기반으로 데이터 유형에 따른 데이터 수집</li> </ul>	<ul style="list-style-type: none"> <li>결정된 데이터 유형에 따른 데이터 수집</li> <li>다양한 데이터 수집 플랫폼 경험</li> <li>데이터 소스 간의 차이/장단점 확인</li> </ul>
데이터 검증	<ul style="list-style-type: none"> <li>수집된 데이터 해석 및 분석</li> <li>수집된 데이터의 신뢰성 검증</li> <li>데이터 활용 문제 확인</li> <li>데이터 평가 및 오류 수정</li> </ul>	<ul style="list-style-type: none"> <li>수집 데이터 구분</li> <li>수집 데이터의 신뢰성 검증</li> <li>수집 데이터 비교 분석</li> </ul>
데이터 관리	<ul style="list-style-type: none"> <li>데이터 이용 목적에 따라 데이터 분석·평가·조직·종합</li> <li>종합된 데이터의 디지털 및 아날로그 표현</li> </ul>	<ul style="list-style-type: none"> <li>데이터 시각화 방법 결정</li> <li>데이터 시각화</li> </ul>
데이터 이용 및 공유	<ul style="list-style-type: none"> <li>데이터 이용 목적의 달성을</li> <li>결과 데이터의 전체 공개, 부분 공개, 비공개</li> </ul>	<ul style="list-style-type: none"> <li>데이터의 종합적 분석</li> <li>데이터로 스토리텔링</li> <li>데이터 통합 및 합성</li> <li>데이터 표현/발표</li> </ul>
데이터 윤리	<ul style="list-style-type: none"> <li>데이터 생성~공유 전(全) 단계에서 고려</li> <li>저작권, 소유권, 초상권 등</li> </ul>	<ul style="list-style-type: none"> <li>표절/오류 유형 및 사례</li> <li>개인정보보호</li> <li>인용</li> </ul>

데이터 검증 과정을 연계해 확인한 후 데이터를 활용하는 단계를 가짐으로써 ChatGPT와 같은 생성형 AI의 콘텐츠의 신뢰성에 대한 검증을 우선했다는 점이다.

## 5. 결론 및 제언

본 연구는 ChatGPT와 같은 생성형 AI의 확산에 따라 생성형 인공지능에 대한 이해와 이러한 기술을 활용한 도서관의 데이터 리터러시 교육 구성에 대해 고민하고 방향을 제시하고자 하는 연구이다. 이를 위해 세 개의 연구 문제를 제시하고 다음과 같은 결론을 도출했다.

- 연구문제 1: ChatGPT 유사 언어모델의 기술적 특징은 대규모 언어 모델이라는 점에서 대규모 데이터를 기반으로 폐턴 학습의 결과로 만들어졌으며 최신 생성형 인공지능 기술이 구현된 서비스인 만큼 이용자와의 자연스러운 상호작용과, 맥락을 이해하는 가운데 질의에 적합한 콘텐츠를 생성, 이용자에게 제공하기 때문에 편의성, 이용자 친화성이 뛰어난 인공지능 기술이라는 점에 그 특징이 있다.
- 연구문제 2: 이러한 생성형 AI는 이용자에게 편하고 자연스러운 검색 경험을 제공 하지만 제공된(생성된 콘텐츠) 정보의 신뢰성, 검색 과정이나 결과에서 발생 가능

한 다양한 문제점에 대한 우려가 다분하다. 이는 한편으로는 정보기술의 고도화로 해결 가능하다 볼 수 있으나 다른 한편으로는 활용할 정보에 대한 이용자의 검증 역량에 대한 요구를 더욱 강력히 제기하는 상황이라 인식 가능하다. 생성형 인공지능 기술 기반 서비스 플랫폼에 방문하여 편리하게 활용하여 적합한, 정확한 정보를 유용하게 활용하고자 한다면 가장 중요한 것은 제공된 응답 결과가 적합한지에 대한 판단이 가능해야 하며 이는 정보 평가 및 검증력에 의존할 수밖에 없다. 이런 이유로 편리한 만큼 활용을 위한 이용자의 데이터 리터러시 역량 교육은 더욱 중요한 것으로 인식해야 한다.

- 연구문제 3: ChatGPT 시대 도서관 데이터 리터러시 교육을 위해 데이터에 대한 이해, 데이터 생성, 데이터 수집, 데이터 검증, 데이터 관리, 데이터 이용 및 공유, 데이터 윤리와 같은 7개 구성항목을 포함한 데이터 리터러시 교육 구성안을 제안하였으며 데이터 검증의 과정을 데이터 수집 이후 즉각적인 단계로 제시하면서 데이터 수집과 검증의 과정이 동반되어야 함을 강조했다. 이 단계들은 순차적, 정형적인 단계가 아니기 때문에 교차 또는 전 단계로의 복귀가 가능하지만 데이터 수집과 검증의 과정이 동반되어야 한다는 것은 주목할 점이다.

본 연구는 도서관 밖에서 이용자의 선풍적인

관심을 불러일으키고 있는 생성형 인공지능에 대한 흥미와 우려에 기반해 도서관의 데이터 리터러시에 대한 고민을 출발점으로 시작된 연구이다. 현재 다양한 분야에서 ChatGPT를 활용한 교육이 이루어지고 있다. 외국어와 코딩 교육에서 상당히 많은 대중들의 관심과 애정을 끌고 있는 것도 사실이다. 그러나 도서관계의 관심에 있는 데이터 획득과 활용에 관해서는 획득 데이터의 신뢰성 때문에 상당한 주의력이 요구되고 있으며 이를 위한 첫걸음이 데이터 리터러시 역량의 강화라 할 수 있을 것이다. 정보기술의 발전은 지속될 것이고 도서관과 이용자는 정보기술의 발전에 따라 해당 기술을 어떻게 활용할 것인가에 대해 고민하게 될 것이다. 새로운 정보기술이 도입될 때마다 이러한 고민은 계속되고, 도서관 현장은 이런 기술의 도입해 어떻게 어떻게 정보서비스에 활용할 것인가 하는 활용 방법보다 활용한다는 것 자체에 의의를 두는 상황도 없지 않을 것이다. 이러한 상황이 계속된다면 도서관의 정보서비스의 방향성 또한 상실될 수밖에 없다. ChatGPT와 같은 생성형 인공지능 기술이 이용자의 정보 활용에 많은 영향을 미치게 될 것이라 예상된다면 도서관은 이러한 기술의 장단점, 문제점 등에 대해 한발 먼저 고민하고 이를 통해 도서관의 정보서비스를 한층 개선할 수 있는 토대로 삼아야 할 것이다.

본 연구의 결과는 생성형 인공지능 기술에 대한 장기간의 경험 결과에 따른 분석이 아니기 때문에 아직까지 완성된 연구가 아니며 지속적인 관찰과 분석이 필요하다 하겠다.

## 참 고 문 헌

- 구본권 (2023). 챗GPT 시대의 필수역량 'AI 리터러시'. KISO 저널, 50, 36-39.
- 김슬기, 김태영 (2021). 초·중등 AI 교육을 위한 데이터 리터러시 정의 및 구성 요소 연구. 정보교육학회논문지, 25(5), 691-704.
- 김지현 (2018). 북미 대학도서관 연구데이터 관리 교육 프로그램 내용 분석. 정보관리학회지, 35(4), 7-36. <http://dx.doi.org/10.3743/KOSIM.2018.35.4.007>
- 대한민국헌법. 대한민국헌법 제4장 제1절 제70조. 출처: 국회법률정보시스템. <http://likms.assembly.go.kr/law/lawsLawtInqyDetl1010.do>
- 백봉삼 (2023. 4. 26.). 챗GPT로 확 커지는 생성형 AI 시장... 양질 데이터 중요도↑. ZDNET Korea. 출처: <https://zdnet.co.kr/view/?no=20230426003045>
- 이정미 (2013). 빅데이터의 이해와 도서관 정보서비스에의 활용. 한국비블리아학회지, 24(4), 53-73. <http://dx.doi.org/10.14699/kbbiblia.2013.24.4.053>
- 이정미 (2019). 데이터 리터러시 개념에 대한 재집근 및 도서관 정보서비스에의 적용. 한국문헌정보학회지, 53(1), 159-179. <http://dx.doi.org/10.4275/KSLIS.2019.53.1.159>
- 최숙영 (2022). AI 리터러시 프레임워크에 대한 연구. 컴퓨터교육학회 논문지, 25(5), 73-84. <https://doi.org/10.32431/kace.2022.25.5.007>
- 한상우 (2018). 인문학 기반 데이터 리터러시 개념에 대한 연구. 정보관리학회지, 35(4), 223-236. <http://dx.doi.org/10.3743/KOSIM.2018.35.4.223>
- Burress, T., Mann, E. Z., Montgomery, S., & Walton, R. (2021). Data literacy in undergraduate education: faculty perspectives and pedagogical approaches. In Bauder, Julia Eds. *Teaching Critical Thinking with Numbers: Data Literacy and the Framework for Information Literacy for Higher Education*. Chicago: Association of College & Research Libraries, 1-22.
- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: rethinking data literacy in the age of disinformation. *Internet Policy Review*, 9(2), 1-22.
- Condon, P. B., Exline, E., & Buckley, L. A. (2023). Data literacy in the social sciences: findings from a local study on teaching with quantitative data in undergraduate courses. *Evidence Based Library and Information Practice*, 18(1), 61-75.
- Cui, Y., Chen, F., Lutsyk, A., Leighton, J. P., & Cuturmisu, M. (2023). Data literacy assessments: a systematic literature review. *Assessment in Education: Principles, Policy & Practice*, 30(1), 76-96.
- Dawa, T., Dhendup, S., Tashi, S., & Rosso, M. A. (2020). Technology Literacies and Chat GPT:

- Perspectives on Emerging Technology Adoption in Academic Writing.
- Firat, M. (2023). What ChatGPT means for universities: perceptions of scholars and students. *Journal of Applied Learning and Teaching*, 6(1), 57-63.
- Flierl, M., Quigley, B., Caswell, T., Costello, L., Li, C., Maher, M., ... & Yang, A. (2023). 2023 ACRL Environmental Scan. Available: <https://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/EnvironmentalScan2023.pdf>
- Fontichiaro, K. & Johnston, M. P. (2020). Rapid shifts in educators' perceptions of data literacy priorities. *Journal of Media Literacy Education*, 12(3), 75-87.
- Kirtania, D. K. & Patra, S. K. (2023). OpenAI ChatGPT generated content and similarity index: a study of selected terms from the Library & Information science (LIS). *Qeios*.
- Kreinsen, M. & Schulz, S. (2023). Towards the triad of digital literacy, data literacy and AI literacy in teacher education-a discussion in light of the accessibility of Novel Generative AI. Available: <https://edarxiv.org/xguzk/>
- Leitgöb, H., Prandner, D., & Wolbring, T. (2023). Big data and machine learning in sociology. *Frontiers in Sociology*, 8, 1173155.
- Lo, L. S. (2023). The CLEAR path: a framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4), 102720.
- Lund, B. & Agbaji, D. (2023). Information literacy, data literacy, privacy literacy, and ChatGPT: technology literacies align with perspectives on emerging technology adoption within communities. Available: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4324580](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4324580)
- Pangrazio, L. & Sefton-Green, J. (2020). The social utility of 'data literacy'. *Learning, Media and Technology*, 45(2), 208-220.
- Pothier, W. G. & Condon, P. B. (2020). Towards data literacy competencies: business students, workforce needs, and the role of the librarian. *Journal of Business & Finance Librarianship*, 25(3-4), 123-146.
- Ruedinger, D., Cooper, D. M., Minnis, S., & Schaub, G. (2022). Fostering Data Literacy Teaching with Quantitative Data in the Social Sciences. *Scholarly Papers and Articles*, 75.
- Shiri, A. (2023). ChatGPT and academic integrity. *Information Matters*, 3(2).
- Tang, Y. & Han, L. (2022, December). A review of domestic and overseas research on data literacy education. In 2022 3rd International Conference on Modern Education and Information Management (ICMEIM 2022). Atlantis Press, 774-780.
- Usova, T. & Laws, R. (2021). Teaching a one-credit course on data literacy and data visualisation.

Journal of Information Literacy, 15(1), 84-95.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Baek, Bongsam (2023, April 26). The expanding Generative AI market with ChatGPT... increasing emphasis on high-quality data. ZDNet Korea. Available: <https://zdnet.co.kr/view/?no=20230426003045>
- Choi, Sook-Young (2022). A study on the AI literacy framework. The Journal of Korean Association of Computer Education, 25(5), 73-84. <https://doi.org/10.32431/kace.2022.25.5.007>
- Constitution of the Republic of Korea. Article 70, Section 1, Chapter 4. Available: National Assembly Law Information System. <http://likms,assembly.go.kr/law/lawsLawtInqyDetl1010.do>
- Gu, Bon Kwon (2023). Essential competence in the era of ChatGPT: 'AI Literacy'. KISO Journal, 50, 36-39.
- Han, Sang Woo (2018). A study about the concept of data literacy based on digital humanities. Journal of the Korean Society for Information Management, 35(4), 223-236. <http://dx.doi.org/10.3743/KOSIM.2018.35.4.223>
- Kim, Jihyun (2018). A content analysis of research data management training programs at the University Libraries in North America: focusing on data literacy competencies. Journal of the Korean Society for Information Management, 35(4), 7-36. <http://dx.doi.org/10.3743/KOSIM.2018.35.4.007>
- Kim, Seulki & Kim, Taeyoung (2021). A study of the definition and components of data literacy for K-12 education. Journal of The Korean Association of Information Education, 25(5), 691-704.
- Lee, Jeong-Mee (2013). Understanding big data and utilizing its analysis into library and information services. Journal of the Korean Biblia Society for Library and Information Science, 24(4), 53-73. <http://dx.doi.org/10.14699/kbiblia.2013.24.4.053>
- Lee, Jeong-Mee (2019). Re-approach to the concept of data literacy and its application to library information services. Journal of the Korean Library and Information Science Society, 53(1), 159-179. <http://dx.doi.org/10.4275/KSLIS.2019.53.1.159>

