

텍스트 네트워크 분석을 활용한 생성형 인공지능 연구주제 분석

구진희*, 김두리**, 강경희***
*목원대학교 소프트웨어교양학부
**건양대학교 간호학과
***건양대학교 치위생학과
e-mail: jhku@mokwon.ac.kr
kdr2015@konyang.ac.kr
dhkhkang@konyang.ac.kr

Analysis of generative artificial intelligence research topics using text network analysis

Jin-Hui Ku*, Doo Ree Kim**, Kyung-hee Kang***
*Division of Software Liberal Arts, Mokwon University
**College of Nursing, Konyang University
***Department of Dental Hygiene, Konyang University

요약

2022년 11월 생성형 인공지능의 등장 이후, 인공지능 및 생성형 인공지능 관련 연구가 급격히 증가하고 있다. 이 논문에서는 생성형 인공지능이 발표된 이후 2023년과 2024년에 게재된 국내 논문 중에서 논문제목 및 키워드에 '생성형 인공지능' 또는 '생성형 AI'를 포함하고 있는 논문 595편을 대상으로 텍스트 네트워크 분석을 수행하였다. 키워드의 연결 중심성 분석 결과, '교육', '이미지', '분석', '디자인', '학습', '기술', '서비스', 'ChatGPT' 등이 핵심 키워드로 나타났다. 이들 키워드의 커뮤니티 응집분석 결과에서는 다양한 분야 디자인을 위한 생성형 인공지능 활용 연구, 이미지 생성과 관련하여 생성형 인공지능 활용 연구, ChatGPT 활용 관련 연구 등 3개 그룹으로 구분되었다. 토픽 모델링 결과, 5개 토픽이 생성되었다. '교육' 키워드는 ChatGPT, 대학, 글쓰기 등으로 이뤄진 토픽과 창작, 저작, 중국 등의 토픽으로 연결성을 보였으며, 수업, 프로그램, 설계, 개발, 효과 등의 토픽과 연관이 있었다. '디자인' 키워드는 영상, 콘텐츠, 제작, 적용, 기술 등의 토픽과 이미지, 학습, 예술, 게임, 사례, 비교 등의 토픽과 연결성을 보였다. 그 외에 '서비스' 키워드는 대학생, 이용, 의도, 경험, 영향, 요인 등으로 토픽이 구성되었다.

행하였다.

연구 동향 분석은 특정 분야의 연구 경향과 특징을 파악하여 해당 분야의 발전 현황을 분석하고, 새로운 연구 주제를 발굴하며, 기존 연구의 한계를 극복하는 데 기여하고자 하는 것이 그 목표이다. 특히, 생성형 인공지능에 관한 연구 동향을 파악하고자 함은 최근 인공지능 및 생성형 인공지능에 관한 논문이 증가하고 있는 상황에서 연구 학문 분야의 분포 및 경향성을 파악할 수 있을 것으로 기대할 수 있다.

1. 서론

최근 거의 모든 정보기술은 인공지능으로 집약되고 있다. 특히, 2022년 11월 등장한 생성형 인공지능 기술은 인문학, 예술, 과학기술 등 분야를 막론하고 그 기술력에 큰 파장을 일으키고 있다. 기존의 인공지능 시스템이 주로 분류나 예측에 사용되었다면 생성형 인공지능은 훈련 데이터와 유사한 텍스트, 대화, 이야기, 이미지, 동영상, 음악 등 새로운 콘텐츠를 생성할 수 있는 기술이다.

인공지능 관련 연구는 2016년 알파고 이후 급격히 증가한 것으로 나타난다[1-4]. 본 연구에서는 2023년부터 급격히 증가하고 있는 생성형 인공지능의 연구 동향 및 세부 분야를 탐색하기 위해 텍스트 중심성 분석, 토픽 모델링 기법 등을 활용하여 2023년~2024년 2년 동안의 연구 주제 분석을 수

2. 연구방법

2.1 데이터수집 및 전처리

본 연구에서는 생성형 인공지능에 관한 최근 연구동향을 파악하기 위해 텍스트 네트워크 분석을 활용한 연구 주제를 분석하였다. 연구를 위해 KCI 등재 저널을 대상으로 키워드

‘생성형 인공지능’, ‘생성형 AI’로 2023년 1월 1일~2024년 12월 31일까지 출판된 논문 595편을 추출하였고, 논문제목, 저자, 학술지명, 발행연도, 주제어를 정리하였다. 분석 도구는 Netminer 4.5와 Excel 2024 프로그램을 사용하였다. 텍스트에서 유의미한 특징을 추출하는 방법으로는 특정 단어가 문서 내에서 얼마나 자주 등장하는지(Term Frequency), 전체 문서에서 얼마나 희귀한지(Inverse Document Frequency)를 고려하여 단어의 중요도를 계산하는 TF-IDF(Term Frequency-Inverse Document Frequency)방법에 기반하여 텍스트의 유의미한 특징을 추출하였다.

595개 논문을 대상으로 전처리과정에서 동의어(Thesaurus words)와 제외어(Exception words)를 처리하였고, 2음절 미만, 빈도수 50 이하인 키워드를 제외한 결과 최종 74개의 키워드를 분석하였다.

3. 연구결과

3.1 빈도분석

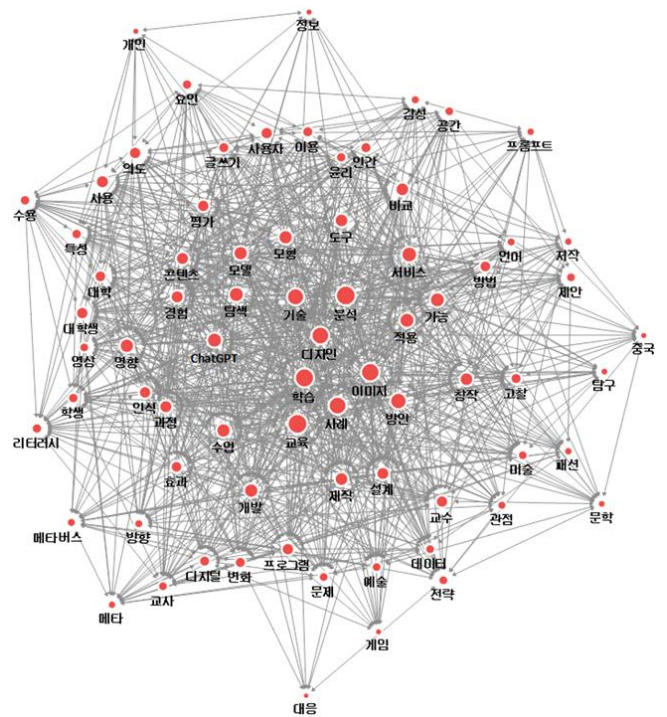
본 연구에서는 생성형 인공지능과 관련한 595편의 논문에서 전처리과정을 통해 분석에 사용할 74개의 핵심 키워드를 추출하였다. [표 1]은 핵심 키워드에 대한 빈도분석 결과이다.

[표 1] 주제어 빈도분석

No.	Keywords	Frequency
1	교육, 이미지, 디자인, 분석, 방안, 학습	100 이상
2	ChatGPT, 사례, 기술, 개발, 영향	90~99
3	탐색, 서비스, 가능	80~99
4	수업, 제작, 창작, 적용, 모델	75~79
5	경험, 인식, 설계, 비교, 고찰	71~73
6	평가, 이용, 대학, 글쓰기, 과정, 콘텐츠	69~70
7	프로그램, 의도, 예술, 방향, 대학생	65~68
8	효과, 저작, 문제, 도구	60~64
9	중국, 요인, 사용, 교수, 공간	57~59
10	영상, 프롬프트, 윤리, 디지털, 데이터	50~55

3.2 연결 중심성 분석

[그림 1]은 주제어 연결 중심성 지수를 시각화하여 나타낸 것이다. ‘교육’, ‘이미지’, ‘분석’, ‘디자인’, ‘학습’ 등이 핵심 주제어로 분석된 것을 알 수 있다. 그 외에 ‘리터러시’는 인공지능 리터러시를 다룬 연구가 가장 많았고, 디지털 리터러시, 데이터 리터러시, 미디어 리터러시 등 인공지능과 관련한 리터러시 연구들이 있었다.



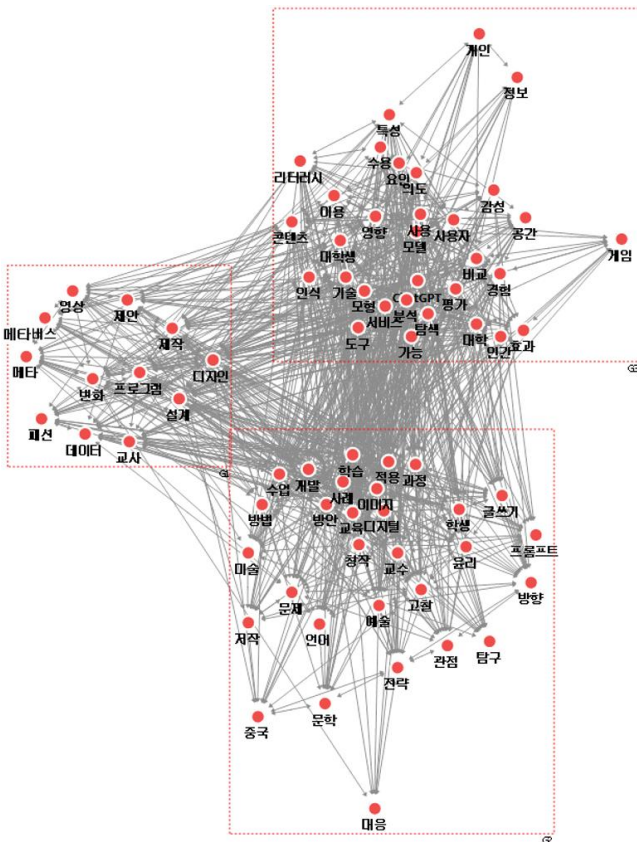
[그림 1] 연결 중심성 분석 시각화

[그림 2]는 주제어 간 관계성을 살펴보기 위해 응집성이 높은 주제어를 집단화한 것으로 커뮤니티 응집분석 결과를 나타내고 있다. 3개의 그룹으로 구분되었으며, 중심성 지수 0.4~0.5 이상의 상위 20개 키워드를 [표 2]에 제시하였다.

[표 2] 연결 중심성 분석 결과

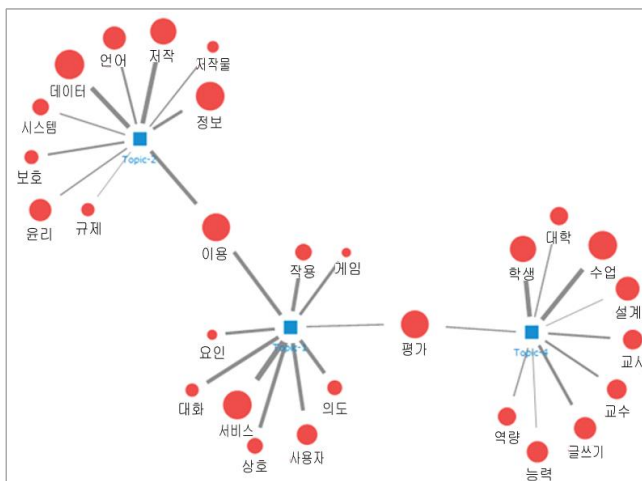
No.	Keywords	Connectivity centrality	Groups
1	디자인	0.712329	1
2	제작	0.493151	1
3	설계	0.465753	1
4	프로그램	0.438356	1
5	교육	0.794521	2
6	학습	0.767123	2
7	이미지	0.739726	2
8	사례	0.69863	2
9	방안	0.643836	2
10	적용	0.561644	2
11	개발	0.534247	2
12	수업	0.520548	2
13	분석	0.794521	3
14	기술	0.657534	3
15	서비스	0.589041	3
16	가능	0.575342	3
17	ChatGPT	0.561644	3
18	탐색	0.561644	3
19	영향	0.547945	3
20	모델	0.534247	3

[그림 2]의 G1 그룹은 다양한 분야에서 디자인을 위한 생성형 인공지능 활용 연구로 파악되었고, G2 그룹은 이미지 생성과 관련하여 생성형 인공지능을 활용하는 교육, 방안, 학습, 사례, 적용 등의 연구가 주류를 이루었다. G3 그룹은 ChatGPT 활용 관련 연구로 서비스, 교육 분야의 활용 외에도 중심성 지수 0.3 이상 키워드를 살펴보면, 대학생 및 사용자 인식, 글쓰기 등 대학 교육에서의 활용, 리터러시 등에 관련한 연구들이었다.

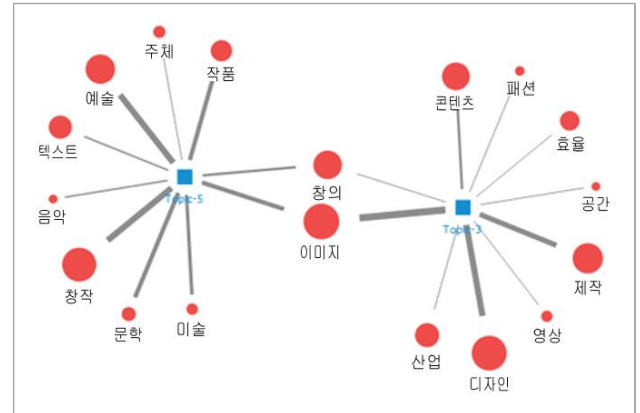


[그림 2] 커뮤니티 응집분석 결과

[그림 3]은 핵심 키워드에 대한 토픽 모델링 결과이다. 토픽 모델링을 위한 튜닝을 통해 5개의 토픽이 선정되었다.



(A) 토픽 1, 2, 4



(B) 토픽 3, 5

[그림 3] 주제어 토픽 모델

토픽1과 토픽2는 ‘이용’을 중심으로 연결성을 보였으며, 토픽1과 토픽4는 ‘평가’를 중심으로 연결성을 보였다. 토픽3과 토픽5는 ‘이미지’와 ‘창의’를 중심으로 연결성을 보였다.

4. 결론

RISS 학술관계분석서비스에 따르면, 2023년 ~ 2024년 최근 2년간 많이 연구된 키워드는 인공지능, 딥러닝, 머신러닝, 텍스트마이닝, COVID-19 등으로 나타났다. 이같이 인공지능 관련 연구는 생성형 인공지능의 등장 이후 급격히 증가하였으며 특히, 생성형 인공지능에 관한 연구는 2023년과 2024년 600여 편의 KCI 등재 논문이 출판된 것으로 나타났다. 생성형 인공지능은 ChatGPT를 비롯하여 교육 분야, 이미지 생성, 디자인, 서비스 분야에서 그 가능성을 탐색하는 연구들이 발표된 것으로 파악되었다.

참고문헌

- [1] 박홍진, “인공지능, ‘기계학습’, ‘딥 러닝’ 분야의 국내 논문 동향 분석”, 한국정보전자통신기술학회논문지, 제 13권 4호, pp. 283-292, 2020년.
- [2] 이정선, “생성형 AI 기술의 진화방향과 정책과제”, 정보통신정책연구, 24-01, pp. 1-31, 2024년.
- [3] 박정호, “국내 인공지능교육 연구 동향 분석”, 에듀테크인먼트연구, 제 3권 2호, pp. 51-62, 2021년.
- [4] 문건두·김경재, “소셜 네트워크 분석과 토픽 모델링을 활용한 설명 가능 인공지능 연구 동향 분석”, JITAM, 제 30권 1호, pp. 53-67, 2023년.