

텍스트 마이닝을 활용한 인공지능 윤리 연구 동향 분석

김미영*·박선주**

광주교육대학교 AI 융합교육전공*, 광주교육대학교 컴퓨터교육과**

요약

인공지능이 가지는 자율성은 인공지능이 단순한 기술이 아닌 윤리를 지닌 존재이어야 하는 당위성을 부여한다. 이는 인공지능 기술이 발전할수록 인공지능 윤리에 대한 다양한 문제가 제기되는 것을 통해 확인할 수 있다. 이에 본 연구는 국내 학위 논문 및 학술지에 게재된 논문을 대상으로 양적 연구를 통해 국내 인공지능 윤리 연구의 경향성을 파악한 후, 워드 클라우드 시각화, 토픽 모델링, 네트워크 분석을 통해 국내 인공지능 윤리 연구의 의미를 해석하였다. 그 결과 국내 인공지능 윤리 연구는 아직 시작 단계이며, 대부분의 연구가 일종의 해결책을 제시하는 방향으로 진행되었음을 알 수 있었다. 또한, 앞으로의 인공지능 윤리 연구는 사회 변화에 맞게 다양한 규제를 제시하는 방향으로 진행될 것으로 예상된다. 본 연구는 이러한 결과를 통해서 앞으로의 인공지능 윤리 연구 방향에 대한 시사점을 제안하고자 하는 시도로서 의의가 있다.

키워드 : 인공지능 윤리, 토픽 모델링, 네트워크 분석, 텍스트 마이닝, 워드 클라우드

Analysis of AI Ethics Research Trends Using Text Mining

Miyoung Kim* · Sunju Park**

Gwangju National University of Education majoring in AI convergence education*

Dept. of Computer Science Education·Gwangju National University of Education**

Abstract

The autonomy of artificial intelligence gives it justification that artificial intelligence should be an existence with ethics, not just technology. This can be confirmed through the fact that as artificial intelligence technology develops, various problems regarding artificial intelligence ethics are raised. Therefore, this study grasped the trend of domestic artificial intelligence ethics research through quantitative research on papers published in domestic academic papers and dissertations, and then interpreted the meaning of domestic artificial intelligence ethics research through word cloud visualization, topic modeling, and network analysis. As a result, it can be seen that domestic artificial intelligence ethics research is still in the beginning stage, and most of the research has been conducted in the direction of suggesting a kind of solution. In addition, future AI ethics research is expected to be conducted in the direction of presenting various regulations according to social changes. This study is meaningful as an attempt to suggest implications for the future direction of artificial intelligence ethics research through these results.

Keywords : Artificial intelligence ethics, Topic modeling, network analysis, text mining, word cloud

본 연구는 김미영의 2023년도 석사학위 논문(광주교육대학교 교육대학원)에서 발췌하여 정리하였음.

교신저자: 박선주(광주교육대학교 컴퓨터교육과)

논문투고 : 2022-12-23

논문심사 : 2022-12-29

심사완료 : 2023-01-13

1. 서론

2021년 페이스북은 사명을 메타로 바꾸고 메타버스 산업을 선점하기 위한 새로운 도전에 나섰다. 기업 입장에서 사명 변경은 복잡한 절차뿐만 아니라 막대한 경제적 손실까지 감내해야 하는 큰 도전이다. 그럼에도 불구하고 페이스북의 사명 변경은 이미지 세탁이라는 강한 비난을 받았다[18]. 이는 2018년 약 5,000만 명의 개인정보 유출 사건부터 2021년 청소년에게 유해한 영향을 미친다는 내부고발까지 이어지는 페이스북의 일련의 비도덕적 행위 때문이다.

급속하게 전개되는 4차 산업혁명으로 사회의 다양한 곳에서 기술 발달이 이루어지면서 우리는 기술의 발달과 윤리를 별개의 문제로 생각하는 경우가 많았다[15]. 하지만 4차 산업혁명의 핵심적인 기술인 인공지능에서는 이들의 상관관계를 더는 인정하지 않을 수 없다.

우리나라의 경우 챗봇 이루다 논란이 좋은 예가 될 수 있다. 실제 연인들이 나누는 대화 데이터 100억 건을 딥러닝을 통해 학습한 이루는 평범한 20대 여성의 컨셉으로 소개되어 많은 관심과 흥미를 일으켰으나 개인정보 침해와 혐오 발언 논란으로 출시 3주 만에 서비스가 종료되었다. 사람들과의 대화를 통해 말을 익히는 챗봇의 기능을 고려한다면 이루다는 우리의 윤리적 가치가 기계에 반영된 결과라고밖에 말할 수 없다. 이러한 인공지능의 특징으로 인공지능을 단순한 기술로 보지 않고 인공적 도덕 행위자(AMA, Artificial Moral Agent)로 인정하자는 주장이 늘고 있다. 이처럼 인공적 도덕 행위자라는 인공지능에 대한 새로운 시각 없이 인공지능을 단순히 기술로만 받아들인다면, 앞으로 우리가 겪을 윤리적 문제들은 해결 불가능한 일이 될 것이다[6]. 그러므로 앞으로의 인공지능 시대에는 기술뿐만 아니라 사회 전반에 걸친 모든 변화에서 윤리적 가치를 함께 논해야만 한다.

본 논문에서는 사회 전반에서 인공지능 윤리에 관한 연구 동향을 분석하고자 한다. 이를 위해 인공지능윤리 관련 국내 학술지 게재 논문 및 국내 학위 논문을 대상으로 양적 분석은 물론 토픽 모델링과 네트워크 분석을 포함한 질적 분석을 통해 인공지능윤리 연구 동향을 살펴보고, 앞으로의 발전 방향에 대한 시사점을 제안하고자 한다.

2. 이론적 배경

2.1. 선행연구 분석

국내 인공지능 윤리와 관련한 텍스트 마이닝 분석을 수행한 논문 또는 국내 인공지능 윤리 연구 동향을 분석한 기존의 논문들을 살펴보면 다음과 같다.

임미가(2021)는 2006년부터 2021년 9월까지의 국내 연구물 총 181편을 대상으로 연도별 출판 빈도 수, 연구 분야 영역, 분야별 주요 연구 내용을 체계적 문헌 고찰 방식으로 분석하였다. 그 결과, 국내 인공지능 윤리 연구는 양적 성장 속에서 주로 실천 윤리학 기반의 연구가 이루어진 것으로 나타났다. 또한, 영역별 주요 연구 내용을 철학적, 제도적, 교육적 측면에서 다음과 같이 제시하였다. 철학적 측면에서는 인공지능이란 존재와 지위에 관한 연구가 주를 이루었고, 제도적 측면에서는 인공지능 기술에 대한 규제와 그와 관련된 법제화 연구가 대부분이었으며, 교육적 측면에서는 인공지능 윤리의식을 함양할 수 있는 교육 방안을 모색하는 연구가 많은 것으로 나타났다[10]. 해당 연구는 대부분의 국내 전자 자료를 검색하여 자료의 양적 확보를 통해 결과의 신뢰성을 확보했으며, 인공지능 윤리 관련 영역별 연구 내용 분석을 통해 규범 윤리학 분야의 연구 필요성을 제안했다는 의의가 있다. 하지만 분석에 텍스트 마이닝 기법을 사용하지는 않았다.

유수진, 장운재(2021)는 대학생 83명이 작성한 5가지 인공지능 윤리 토론 주제에 대한 의견을 바탕으로 텍스트 마이닝을 수행하였다. 그 결과, 인공지능 미래에 대해서는 대부분 긍정적으로 보는 시각이 많았으며, 자율주행 자동차의 사고 시 책임은 차량 소유자에게 있다는 의견이 가장 높았다. 또한, 인공지능 발전의 역기능으로는 사생활 침해, 기술 악용 등을 꼽았으며, 얼굴 인식이 보편화된 미래에 대해서는 기술의 장점은 인정하나 발생할 수 있는 부작용을 최소화해야 한다는 중립적 의견이 과반수를 차지했다. 데이터 수집과 관련된 주제에서는 다양한 의견이 나와 사회적 합의가 필요함을 제시하였다[17]. 해당 연구는 텍스트 마이닝을 수행하여 고등교육에서 인공지능 윤리 교육에 대한 시사점을 제공하는 의의가 있으나 분석대상이 대학생 일부로 한정되어 국내 인공지능 윤리 연구 동향에 대한 분석은 이루어지

지 않았다.

김연후, 김하연(2021)은 총 216편의 논문을 대상으로 토픽 모델링과 키워드 네트워크 분석을 수행하였다. 그 결과, 도덕적 인공지능에 대한 6가지 토픽- 존재 및 지위, 자율 및 책임, 규제 및 가이드라인, 분야별 도입 및 활용, 개발 및 교육, 데이터 활용 및 법적 시스템-을 제시하였다[9]. 해당 연구는 토픽 모델링과 키워드 네트워크 분석을 사용하여 기존의 양적 분석과 달리 분석 결과의 객관성을 확보했다는 의의가 있다.

정도범, 유화선, 문희진(2022)은 2016년부터 2020년까지 국내 사회 과학 및 복합학 분야의 인공지능 키워드를 포함하는 논문 총 631편을 대상으로 키워드 분석을 수행하였다. 그 결과, 상위 6개 군집의 주요 이슈를 다음과 같이 제시하였다. 인공지능의 등장 배경과 기본적인 개념, 인공지능 시대 교육의 중요성, 알고리즘의 편향성과 개인정보보호, 인공지능 기반 창작물에 대한 법적 이슈, 인공지능 기반 서비스나 제품의 이용 의도, 인공지능 윤리 이슈 등이다[3]. 해당 연구는 텍스트 마이닝 중 키워드 네트워크 분석을 수행하였으나 인공지능을 키워드로 포함한 논문을 대상으로 분석하는 등 인공지능 윤리 연구에 대한 구체적인 분석은 이루어지지 않았다.

2.2 텍스트 마이닝

텍스트 마이닝(Text Mining)은 텍스트 데이터에서 가치와 의미가 있는 정보를 찾아내는 기법이다[8]. 오늘날 정보 대부분이 텍스트 형식으로 저장된다. 이렇게 공개된 정보에서 특정 주제에 대한 텍스트를 바탕으로 의미를 분석하고 경향성을 알아내는 방법으로 텍스트 마이닝 기법을 활용한다[19]. 본 논문에서 텍스트 마이닝은 2016년부터 2022년 8월까지 국내 학위 논문과 학술지에 게재된 인공지능윤리 관련 논문의 제목, 키워드, 국문 초록 등의 텍스트를 분석한 것을 의미한다.

2.3. 토픽 모델링

토픽 모델링(Topic Modeling)은 전체 내용물에서 일정한 패턴을 발견해 내는 알고리즘 기반 텍스트 마이닝의 한 형태이다[11]. 일종의 알고리즘으로 분석 대상 문

헌에서 단어들을 클러스터링하는 방식으로 주제를 추론하는 방법론이다[4][13]. 이 연구에서는 토픽 모델링 중 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 방법을 사용하였다.

2.4. 네트워크 분석

네트워크 분석은 연구대상을 네트워크 형태로 표현 및 분석하는 작업을 말한다. 네트워크는 현실 세계의 각종 시스템 간의 관계를 노드(점:개체)와 링크(선:관계)로 모델링한 결과로서, 이를 통해 노드와 노드 사이에 상호작용을 표현할 수 있다[12]. 본 논문에서는 네트워크 분석에서 가장 보편적으로 사용되는 기법인 연결정도 중심성, 근접 중심성, 매개 중심성, 고유벡터 중심성을 분석 지표로 사용하는 중심성 분석 기법을 사용한다.

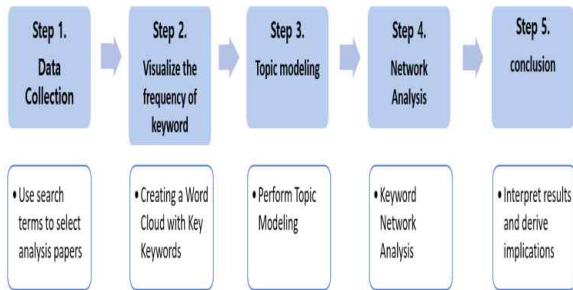
3. 연구방법

3.1. 연구대상

분석대상 연구 문헌은 국내 학술지 게재 논문 및 학위 논문 중 제목에 ‘인공지능 윤리’를 포함하는 논문으로 한정하였다. 국내 학술지 게재 논문 및 학위 논문을 선정하기 위해 한국교육학술정보원(KERIS)에서 제공하는 학술정보 서비스(<http://www.riss.kr>, 이하 RISS)의 데이터베이스를 이용하여 분석 시점인 2022년 8월 31일을 기준으로 검색한 결과 2016년부터 2022년 8월까지 총 192건의 논문이 수집되었다. 수집된 논문 중 분석에 필요한 자료인 키워드나 초록이 없는 논문(영문 초록이 있는 경우 국문 초록으로 변환함)과 중복된 논문 49건을 제외한 후 총 143건의 논문을 최종 분석대상으로 선정하였다.

3.2. 연구절차

본 연구는 선정된 국내 학술지 게재 논문 및 학위 논문을 대상으로 주피터 노트북과 파이썬을 활용하여 인공지능 윤리 연구 동향을 파악하였다. 구체적인 연구 절차는 (Fig. 1)과 같다.



(Fig. 1) Research Procedure

1단계 자료수집은 먼저, RISS에서 국내 학술지 게재 논문 및 학위 논문을 수집하였다. 상세 검색에서 제목에 ‘인공지능 윤리’를 포함하는 논문을 검색일 2022년 8월 31일 기준으로 발행연도 전체로 검색한 결과 2016년부터 2022년 8월까지 총 192건의 논문이 수집되었다. 웹 크롤링 후 각 논문의 제목, 발행 연도, 키워드, 국문 초록, 학술지, 저자 등의 정보를 추출하여 정리하였다.

2단계 키워드 출현 빈도 시각화 과정에서는 검색된 논문 중 키워드 또는 국문 초록이 없는 논문과 중복된 논문을 삭제하고 불용어를 설정하여 최빈 출현 단어 30개를 선정 이를 워드 클라우드를 활용하여 시각화하였다.

3단계 토픽 모델링 과정에서는 LDA 알고리즘을 사용하여 분석대상 143건의 논문에서 주요 토픽을 선정하였다. 토픽별 키워드를 통해 각 토픽의 주제를 파악하였다.

4단계 네트워크 분석과정에서는 분석 논문의 키워드를 추출하여 연결정도 중심성, 근접 중심성, 고유벡터 중심성 등을 분석 지표로 삼아 단어 사이의 관계를 파악하는 분석을 수행하였다.

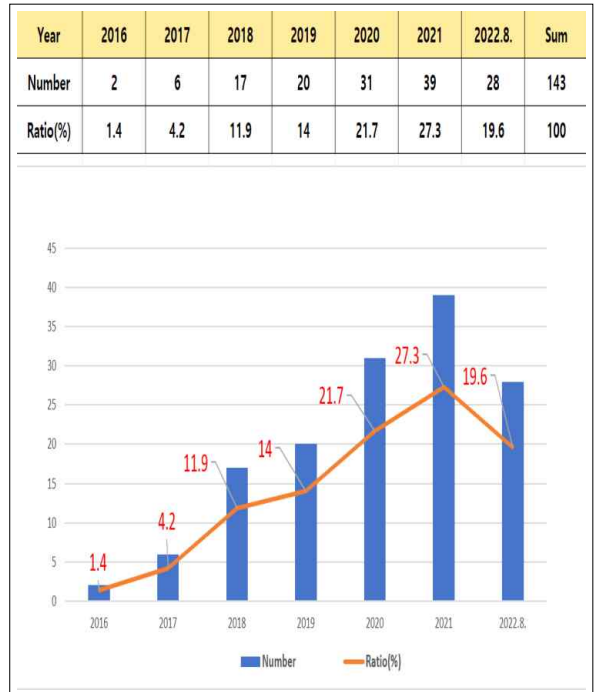
5단계 결론에서는 각 분석을 통해 얻은 결과를 해석하고 본 연구의 시사점을 도출하였다.

4. 연구결과

4.1. 국내 인공지능 윤리 논문 연구 동향 분석

국내 인공지능 윤리 연구 동향을 분석하기 위해 먼저, 연도별로 발행된 논문 현황을 알아보았다. 2016년부터

2022년 8월까지 연도별 인공지능 윤리 논문 게재 현황은 (Fig. 2)와 같다.

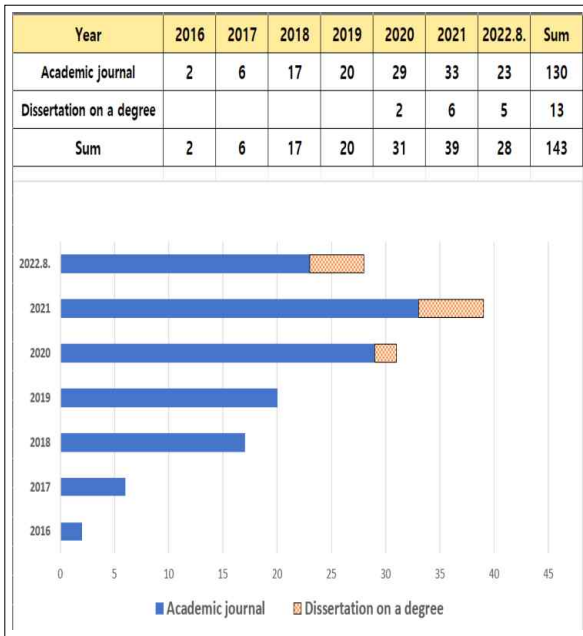


(Fig. 2) Annual Publication Status of Artificial Intelligence Ethics Research Papers

인공지능 윤리 관련 논문이 처음 게재된 것은 2016년임을 확인할 수 있다. 2016년은 구글이 개발한 알파고와 이세돌 9단의 세기의 대결이 펼쳐진 해이다. 소위 알파고 쇼크로 급속하게 발달하는 인공지능 기술에 대한 반성적 고찰이 시작된 것으로 보인다. 하지만 인공지능이라는 용어가 1950년대부터 사용되어 2000년 이후 급속도로 확산하였음을 고려하면 2016년에 첫 연구 논문이 발표되었다는 것은 인공지능 윤리 관련 연구가 매우 늦은 감이 있음을 보여준다. 이후, 2018년 관련 논문이 2배 이상 증가한다. 이는 인공지능 기술의 급속한 발달 때문으로 생각된다. 2018년 당시 발표된 논문의 내용을 살펴보면 주로 인공지능 로봇, 자율시스템, 대화형 인공지능, AI 치명적 자율 무기 등 고도화된 인공지능 기술과 관련된 것으로 기술의 급격한 발달로 사회 전반에서 인공지능 윤리에 대한 문제가 제기되고 있음을 알 수 있다. 또한, 2018년은 과학기술정보통신부와 한국정보화

진흥원에서 지능정보화사회 윤리 가이드라인을 발표한 시기이기도 하다. 이러한 정부의 정책도 논문 발행 편수 증가에 영향을 미쳤을 것이다. 2020년 이후 현재까지 인공지능 윤리 연구 논문의 발행 수는 증가 추세이다. 이는 전 세계적으로 코로나 19의 영향을 받아 원격수업과 재택근무가 증가하면서 인공지능 기술의 대중화로 다양한 디지털 윤리에 대한 사회적 문제의식이 높아졌기 때문으로 생각된다.

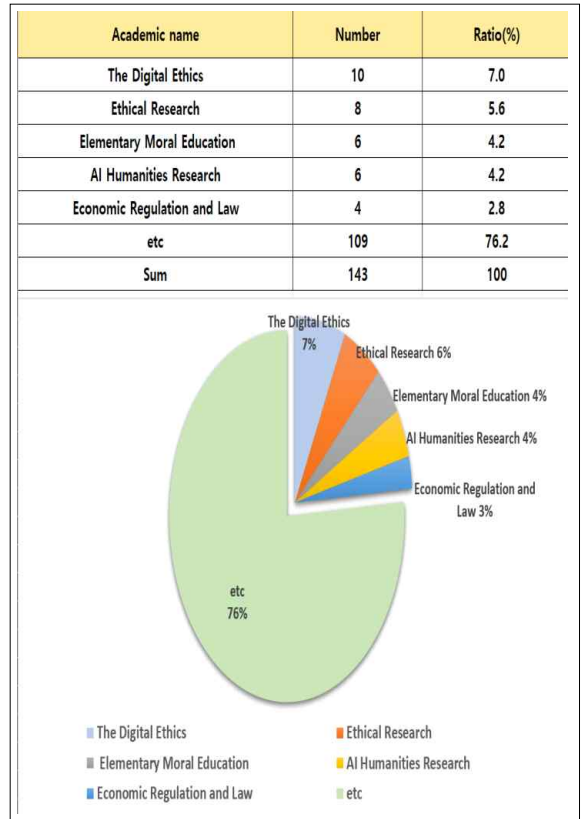
다음 연도별 학위 및 학술지 게재 논문 수를 분석한 결과는 (Fig. 3)과 같다.



(Fig. 3) Number of academic journals and dissertations by publication year

2016년 인공지능 윤리 관련 논문이 학술지널에 소개된 후 2019년까지 그 발행 편수는 계속 증가하고 있음을 알 수 있다. 또한, 2020년부터는 학위 논문에서도 인공지능 윤리 관련 연구가 나타나기 시작하며 점점 증가하는 추세를 볼 수 있다. 학위 논문으로까지 연구 문제가 넘어왔다는 것은 인공지능 기술의 발달에 따라 제기되는 윤리적 문제의 심각성이 커졌으며, 코로나 19로 인해 국민 대다수가 디지털 환경에 노출되면서 인공지능 윤리에 대한 사회적 관심이 높아졌기 때문이다.

다음 학술지별 발행 논문 수를 분석한 결과는 (Fig. 4)와 같다.



(Fig. 4) Number of published papers by academic journal

총 143편의 논문이 95개 학술지에 게재되었다. 이는 사회 전반에서 인공지능 윤리 연구가 다양하게 논의되는 긍정적인 해석도 가능하겠으나, 실제 대부분의 학술지별 발행 논문의 수가 1~2편으로 매우 적으며, 상위 5개 학술지 중 발행 논문 비율이 전체의 10%를 넘는 곳이 없다는 것은 아직 관련 연구가 미비한 상황에서 인공지능 윤리 연구에 대한 중추적 역할을 수행하는 곳이 없다는 의미로 볼 수 있다. 인공지능 윤리가 무엇인지 정의하고, 모든 사람이 지녀야 할 공통적인 인공지능 윤리 원칙에 관한 연구를 주도할 전문적인 연구 기관이 필요함을 알 수 있다.

4.2. 키워드 출현 빈도 시각화 결과 분석

분석대상 143건의 논문에서 키워드와 국문 초록을 가지고 국내 인공지능 윤리 연구에 나타난 단어의 출현 빈도를 분석하였다. 해당 분석을 통해 추출한 총 22,953개의 명사 중 불용어를 제외한 후 총 14,781개의 단어를 추출하였다. 이 중 출현 빈도가 높은 상위 30개의 단어 별 빈도수는 <Table 1>과 같고 이를 워드 클라우드로 표현한 것은 (Fig. 5)와 같다.

<Table 1> Appearance Frequency The frequency of the top 30 words

Ranking	Word	Frequency	Ranking	Word	Frequency
1	기술	267	15	기준	71
2	연구	263	17	시스템	69
3	교육	212	17	사용	69
4	문제	149	19	의식	66
5	사회	111	20	가능성	64
6	로봇	110	20	관점	64
7	책임	108	22	분야	63
8	원칙	98	23	관계	62
8	데이터	98	24	윤리학	61
8	규제	98	25	시대	60
11	개발	97	26	발전	57
12	활용	87	26	인식	57
13	적용	79	28	제시	56
13	정책	79	29	가치	54
15	자율	71	30	영향	52

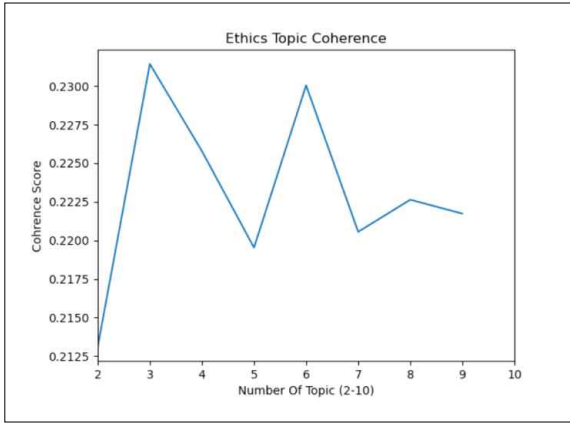


(Fig. 5) Keyword Appearance Frequency Word Cloud Visualization

출현 빈도가 100회 이상인 단어로는 기술, 연구, 교육, 문제, 사회, 로봇, 책임 등이 있다. 이는 그동안의 인공지능 윤리 연구의 방향을 어느 정도 가늠할 수 있게 해준다. 출현 빈도가 높은 단어를 통해 143건의 분석대상 논문에서 인공지능 윤리는 보편적 원칙으로서의 윤리보다는 인공지능이라는 하나의 기술 발달 과정에서 파생한 문제 해결 과정 또는 그 책임을 묻는 방법으로 사용되었음을 알 수 있다. 원칙, 규제, 개발, 활용, 등의 단어가 다음으로 빈도수가 높은 것도 이를 뒷받침한다. 단어 사이의 관계를 좀 더 명확하게 파악하기 위해 토픽 모델링 분석을 수행했다.

4.3. 토픽 모델링 결과 분석

인공지능 윤리와 관련된 최종 분석 논문 143편의 텍스트 데이터를 전처리한 후, 연구 동향 분석에 널리 사용되는 LDA 기반의 토픽 모델링을 수행하였다. 최적 토픽 개수를 찾기 위해 2부터 9까지 1씩 증가시켜가며 코헤런스 값을 확인한 결과 (Fig. 6)과 같이 최적화된 토픽 수는 3으로 결정하였다.



(Fig. 6) Coherence value for selecting optimal number of topics

LDA 기반 토픽 모델링을 적용하여 3개의 토픽을 찾았고, 이들 토픽 간의 거리를 pyLDAvis를 활용하여 시각화한 결과는 (Fig. 7)과 같다.



(Fig. 7) Intertopic Distance Map (IDM)

토픽 간 거리 지도(IDM: Intertopic Distance Map)

에서 원의 크기는 해당 토픽이 전체 글에서 차지하는 비중을 나타내며, 원의 크기가 클수록 높은 빈도수의 단어들로 구성되어 있다[5][16]. 세 원의 크기가 거의 비슷한 것으로 보아 143건의 전체 분석 논문을 아우르는 메인 토픽은 없다고 생각할 수 있으며, 세 가지 토픽이 모두 비슷한 비중을 차지함을 알 수 있다. 또한, 원의 위치는 각 토픽 간의 상호 관련성을 나타낸다. 토픽 간의 거리가 멀수록 주제가 뚜렷하게 구분되며 토픽 간의 거리가 가깝거나 겹칠 경우는 비슷한 주제를 나타내는 것이다[7][14]. (Fig. 7)에 따르면 토픽 1, 2, 3은 모두 거리가 먼 것을 확인할 수 있다. 이는 세 가지 토픽 모두 각각의 주제를 가지고 있으며 서로 연구 주제가 중복되지 않는 것을 말한다.

토픽별 구체적인 주제를 알아보기 위해 λ 값을 0.1로 조정하고 공통 키워드를 제외한 각 토픽을 구성하는 상위 5개의 단어를 확인한 결과는 <Table 2>와 같다.

<Table 2> Key Keywords by AI Ethics Research Topic

Topic Number	keyword($\lambda=0.1$)	Common Keywords($\lambda=1$)
Topic 1	사회, 가능, 필요, 원칙, 로봇	기술, 연구 문제, 교육
Topic 2	자율, 필요, 책임, 인식, 사용	
Topic 3	인간, 사회, 데이터, 가능, 개발	

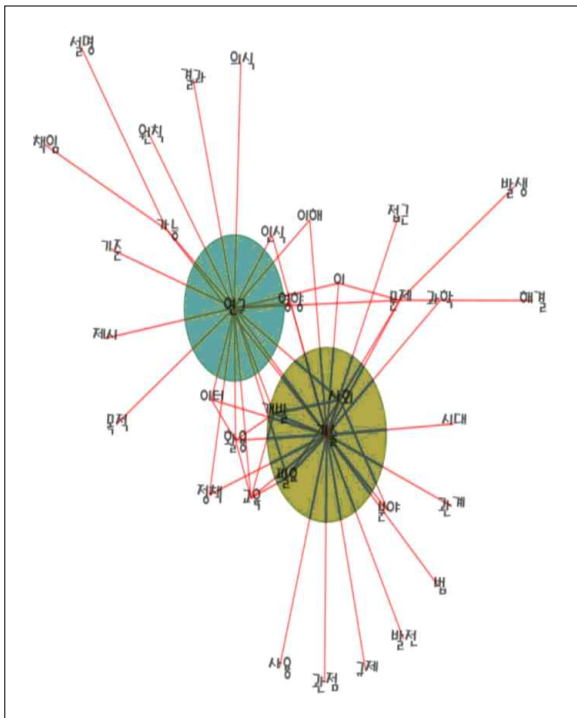
세 가지 토픽의 공통된 키워드는 기술, 연구, 문제, 교육이다. 이는 인공지능 기술의 발달에 따른 다양한 윤리적 문제의 대두로 인공지능 윤리가 교육되어야 한다는 사회적 필요성을 뜻하는 것으로 생각된다. 인공지능 윤리 연구 및 교육에 대한 필요성이 사회 전반에 걸쳐 요구되는 것을 알 수 있다.

토픽별 키워드를 살펴보면, 먼저 λ 값을 0.1로 조정된 토픽 1만의 주요 키워드는 사회, 가능, 필요, 원칙, 로봇이다. 다른 토픽들과 거리가 먼 것을 고려할 때 사회에서 사용되는 인공지능 기술을 갖춘 로봇 사용의 원칙과 관련된 내용으로 판단된다. 토픽 2만의 주요 키워드는 자율, 필요, 책임, 인식, 사용이다. 이는 인공지능 기술 사용의 자율성과 책임 부여를 강조하는 내용으로 판단

된다. 토픽 3만의 주요 키워드는 인간, 사회, 데이터, 가능, 개발이다. 인공지능 기술의 기반이 되는 데이터의 중요성과 사용 규제 개발과 관련된 내용으로 판단된다.

4.4. 네트워크 분석

키워드 사이의 관계를 분석하기 위해 지지도 (support) 값이 0.01 이상인 단어를 대상으로 네트워크 분석을 시행한 결과는 (Fig. 8)과 같다.



(Fig. 8) Visualize Network Analysis

143건의 분석 논문에서 기술과 연구가 핵심 노드이며, 데이터, 활용, 개발, 필요, 교육, 정책이 위의 두 가지 노드 사이에 들어가 있는 것을 확인할 수 있다. 이를 통해 인공지능 기술과 연구의 주요 내용이 데이터 활용, 교육, 관련 정책의 개발과 관련되어 있음을 알 수 있다. 노드에 해당하는 각 키워드의 중요성을 판별하기 위해 출현 빈도 상위 30개 단어의 중심성(Centrality) 값을 구했으며 그 결과는 <Table 3>과 같다.

<Table 3> Centrality values for the top 30 words of frequency of appearance

	degree centrality	eigenvector centrality	closeness centrality	betweenness centrality
기술	1	0.410042769	0.504094836	0.036935268
연구	1	0.377246955	0.504094836	0.036935268
문제	1	0.247561976	0.504094836	0.036935268
교육	0.965517241	0.236278176	0.495097485	0.035911189
사회	1	0.255114158	0.504094836	0.036935268
가능	1	0.166563753	0.504094836	0.036935268
필요	1	0.201457503	0.504094836	0.036935268
개발	1	0.240827476	0.504094836	0.036935268
로봇	0.931034483	0.177529586	0.485869703	0.035617835
활용	0.931034483	0.126918654	0.485869343	0.035397634
책임	1	0.203218579	0.504094836	0.036935268
원칙	0.965517241	0.116866071	0.495097485	0.035911189
사용	0.965517241	0.146206853	0.495114222	0.036058604
시대	1	0.149275122	0.504094836	0.036935268
규제	1	0.130711776	0.504094836	0.036935268
데이터	0.896551724	0.112274177	0.476287707	0.034878376
자율	1	0.15605315	0.504094836	0.036935268
관계	0.931034483	0.094210904	0.485937644	0.036243021
관점	1	0.121462745	0.504094836	0.036935268
기준	0.965517241	0.115927134	0.495114222	0.036058604
제시	1	0.11518223	0.504094836	0.036935268
의식	0.965517241	0.137302871	0.495114222	0.036058604
분야	0.793103448	0.095835542	0.44546758	0.033779149
인식	0.931034483	0.147938368	0.485837668	0.035362782
이해	0.965517241	0.138948623	0.495168722	0.036571513
규범	0.75862069	0.065047689	0.434356744	0.032711772
발전	0.896551724	0.121848177	0.476199819	0.034310744
영향	0.896551724	0.112416218	0.476320392	0.035435044
가치	0.965517241	0.09606925	0.495114222	0.036058604
시스템	0.965517241	0.14224304	0.495097485	0.035911189

연결정도 중심성(Degree centrality, Cd)은 특정 노드에 직접 연결된 링크의 수를 의미한다. 연결정도 중심성 값이 높은 노드는 주로 네트워크의 중심에 위치하는 경향이 있다[1][3]. 그러나 단순히 연결이 많다고 중요한 노드인 것은 아니다. 중요한 노드와의 연결이 많을수록 핵심 노드가 되는데 이를 나타내는 것이 바로 고유벡터 중심성(Eigenvector centrality, Ce)이다[12]. <Table 3>에서 기술, 연구 등의 단어가 고유벡터 중심성, 연결정도 중심성 등 모든 중심성 값이 높음을 확인할 수 있으며 이를 통해, 이 단어들이 그동안의 인공지능 윤리 연구의 핵심 키워드였음을 다시 한번 확인할 수 있다. 매개 중심성(Betweenness centrality, Cb)과 근접 중심성(Closeness centrality, Cc)이 높은 단어는 새로운 연구 주제 탐색 시 활용할 수 있다[2]. 이 값이 높은 단어는

기술, 연구 외에 책임, 필요, 시대, 규제, 개발, 제시 등이 다. 이러한 중심성 값을 바탕으로 앞으로의 연구는 전체적으로 기존 연구의 틀을 유지하면서 시대의 흐름에 맞는 적절한 책임과 규제를 제시하는 방식으로 이루어질 것으로 예측된다.

5. 결론

본 연구는 텍스트 마이닝을 통해 국내 논문에 게재된 인공지능 윤리 연구 동향을 알아보고 다음과 같은 시사점을 제안하고자 한다.

첫째, 우리나라의 인공지능 윤리 연구는 인공지능 기술 발달에 비해 아직 초기 단계이다. 이는 인공지능 윤리 연구 논문이 처음 나온 시기가 늦었고, 현재까지도 관련 논문의 수가 많지 않다는 점, 관련 논문을 주도적으로 발행하는 곳이 없다는 점에서 내린 결론이다. 그러나 인공지능 기술이 그 어떤 기술보다 빠른 속도로 발달하는 것과 인간에게 미치는 영향력을 고려한다면, 인공지능 윤리 연구는 인공지능 기술 연구와 병행되어 이루어져야 한다. 이러한 점에서 최근 교육부에서 발표한 교육분야 인공지능 윤리 원칙(2022.8.)은 매우 주목할 만하다. 기존의 개발자 중심의 윤리 기준을 벗어나 교육 현장에서 적용이 가능한 인공지능 개발과 활용 지침을 최초로 제시했기 때문이다[20]. 해당 가이드라인에 관한 후속 연구뿐만 아니라 교육 현장에서 타당성과 실효성을 검토 후 사회 전반에서 활용될 수 있도록 지속적인 모니터링과 개선 작업이 요청된다.

둘째, 기존의 인공지능 윤리 연구 방향은 기술 발달로 대두된 윤리 문제를 개별적인 하나의 현상으로 보고 이를 해결하는 데 중점을 두고 이루어졌다. 이는 워드 클라우드를 통해 알아본 최빈 출현 단어와 토픽 모델링에서 세 가지 토픽이 모두 같은 비중으로 주제가 겹치지 않은 점, 그리고 네트워크 분석에서 고유 중심성과 연결정도 중심성이 높은 단어들을 통해 내린 결론이다. 챗봇 이루다 논란의 경우처럼 인공지능 기술은 우리의 윤리적 가치가 반영되어 다시 사회 전반에 걸쳐 변화를 일으킨다. 즉, 개별적인 하나의 문제란 존재하지 않는다. 인공지능 윤리는 특정 기술의 사용자 또는 개발자들끼리 논의해야 할 문제가 아니라 사회 구성원 모두가 논의하여 얻어낸 보편적 가치를 바탕으로 이루어져야 할

것이다. 이를 위해서는 소통과 협력이 그 무엇보다도 우선되어야 한다. 즉, 인공지능 윤리 정책 관련 전문가 논의뿐만 아니라 주기적이고 지속적인 간담회, 공청회 등을 통해 사회 전반의 다양한 의견을 수렴하는 작업이 필요하다.

셋째, 앞으로의 인공지능 윤리 연구는 그동안의 큰 틀은 벗어나지 않으면서 사회의 변화에 맞게 다양한 규제를 제시하는 쪽으로 진행될 것으로 예측한다. 이는 키워드 네트워크 분석에서 매개중심성과 근접중심성이 높은 단어들을 바탕으로 내린 결론이다. 하지만 우리 사회에서 인공지능을 사용하는 범위가 급격하게 넓어지는 것을 고려한다면, 매번 새로운 규제를 적용하는 것은 거의 불가능에 가까운 일이다. 이는 기술 변화의 속도와 인간의 적응력 사이의 큰 차이 때문이다. 그러므로 앞으로의 인공지능 윤리는 사후적 규제보다 예방적 규범의 역할을 할 수 있도록 인공지능 윤리에 대한 인식의 전환이 필요하다. 챗봇 이루다 논란이 우리에게 시사하는 바는 현재 우리에게 가장 필요한 것은 인공지능 윤리라는 또 다른 이름의 규제가 아닌 우리 스스로 윤리적 인간이 되어야 한다는 것이다. 인공지능 윤리를 만드는 것은 결국 인간의 윤리이므로 인공지능 기술 개발 및 활용 단계에서 도구로서의 윤리가 아닌 원칙으로서의 윤리를 적용해야 할 것이다.

참고문헌

- [1] Han, J. Y. & Shin, Y. J. (2020), Analysis of research trends related to artificial intelligence education: Keyword network Crack Analysis, Artificial Intelligence Research Paper, 1(2), 20-33.
- [2] Han, S. Y. & Kim, T. J. (2022), Analysis of domestic artificial intelligence education research trends: Using LDA techniques Focusing on the analysis of useful keywords and topics, a learner-centered curriculum education study, 281-294
- [3] Jeong, D. B., Yoo, H. S. & Moon, H. J. (2022), A study on the major issues of artificial intelligence through keyword analysis: Focusing on KCI-listed academic journals in the field of social science, Journal of the Korean Society of Contents, 22(7), 1-9.

- [4] Jeong, D. H. & Joo, H. S. (2018), Method of deriving interdisciplinary convergence technology through content analysis based on topic modeling, *Journal of Information Management*, 35(3), 77-100.
- [5] Kim, M. H. & Kim, T. J. (2021), Topic analysis of successive agricultural managers using news big data. *Journal of the Korean Digital Content Society*, 22(7), 1143-1152.
- [6] Kim, M. J. (2017), The need for AI ethics and domestic and international trends, *Information and communication*, 34(10), 45-54.
- [7] Kim, S. A. (2021), Direction and tasks of technical education according to the analysis of research trends related to artificial intelligence education in elementary and secondary schools using topic modeling, *Journal of the Korean Society of Technology Education*, 106-124
- [8] Kim, Y. H. (2021), Analysis of research topics related to artificial intelligence education: Focusing on network text analysis, *Education innovation research*, 197-217
- [9] Kim, Y. H., & Kim, H. Y. (2021), Analyze trends in moral artificial intelligence research using topic modeling and keyword network analysis, *Moral Education Research* 33(4), 115-141.
- [10] Lim, M. K. (2021), Systematic literature review on the study of artificial intelligence ethics, *Journal of ethics*, 135(1), 47-66.
- [11] Min, Y. H., Kim, J. Y., & Park, O. N. (2021), Topic modeling and keyword network analysis Analysis of trends in cultural content research, *Social Science Research*, 32(2), 113-131.
- [12] Park, G. Y. (2019), Introduction to network analysis techniques using networkx.
- [13] Park, J. H., & Song, M. (2013), Analysis of domestic literature informatics research trends using topic modeling, *Journal of Information Management*, 30(1), 7-32.
- [14] Park, J. H. (2021), Analysis of domestic artificial intelligence education research trends. *Education Research*, 3(2), 51-62.
- [15] Park, Y. H. (2022), *DQ Digital Intelligence*. GimmYoungSa Press.
- [16] Seo, H. J., & Kwon, S. A. (2019), Analysis of research published in domestic journals under the theme of education and artificial intelligence (1985-2018): Application of language network analysis, *Artistic Humanities Social Convergence Multimedia Paper*, 9(11), 209-217.
- [17] Yoo, S. J., & Jang, Y. J. (2021), A study on the ethics of artificial intelligence among college students through text mining, *Journal of the Society of Information Education*, 25(6), 947-960.
- [18] Chosun Ilbo. The worst crisis Facebook, changed its name to 'Meta' and 'Image Laundry'. https://www.chosun.com/economy/tech_it/2021/10/29/O2EZ5NSXT5HOTAD5IRYTS5FQSM/
- [19] Korea Information and Communication Technology Association Information and Communication Terminology Dictionary, Text mining. http://terms.tta.or.kr/dictionary/dictionaryView.do?word_seq=057929-2
- [20] Ministry of Education, Artificial intelligence, use it safely in the field of education, <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=92297>

저 자 소 개



김 미 영

2002년 전남대학교 사범대학
화학교육과(학사)

2003 ~ 현재 중등교사 근무

관심분야: 인공지능, 인공지능 교육,
융합교육

e-mail: chemi20@naver.com



박 선 주

1995 전남대학교 전산통계학과 (이학
박사)

2003 George Mason University 객원
교수

1996~현재 광주교육대학교 컴퓨터교
육과 교수

관심분야 : 컴퓨터교육, SW교육, AI,
앱개발, 빅데이터

e-mail : sjpark@gnue.ac.kr