

인공지능 윤리 인식에 대한 데이터 분석 및 시각화 연구 -대화형 인공지능 서비스 '이루다'를 중심으로-

이수련¹, 최은정^{2*}

¹서울여자대학교 정보보호학과 학부생, ²서울여자대학교 정보보호학과 부교수

A Data Analysis and Visualization of AI Ethics -Focusing on the interactive AI service 'Lee Luda'-

Su-Ryeon Lee¹, Eun-Jung Choi^{2*}

¹Student, Student, Department of Information Security, Seoul Women's University

²Associate Professor, Department of Information Security, Seoul Women's University

요약 사람을 대상으로 하는 인공지능 서비스가 증가하면서 인공지능에서도 윤리적 토대 위에서 이루어져야 한다는 사회적 요구가 증가하고 있다. 이러한 흐름에 따라 정부와 기업에서는 인공지능 윤리와 관련된 정책, 규범 등을 마련하고 있다. 합리적인 정책, 규범을 마련하기 위해서는 대중들이 가지고 있는 인식을 파악하는 것이 첫 번째 단계이다. 본 논문에서는 인공지능과 윤리에 대한 대중들의 인식을 파악하기 위해 소셜데이터와 뉴스 댓글을 수집하고 관심도 분석, 감성 분석, 담론 분석 수행 후 시각화하였다. 분석 결과, "인공지능 윤리"에 대한 관심도와 "인공지능" 호감도는 반비례하는 상관관계를 보여주었다. 담론분석 결과로, 가장 큰 이슈가 "개인정보 유출"이었고 학습 데이터의 오염 및 편향 문제와 컴퓨터로 만들어진 인공지능에게 법인격을 부여해야 하는지에 대한 담론도 보여주었다. 본 연구가 인공지능 윤리 규범, 정책을 마련할 때 대중들의 인식을 파악할 수 있는 자료로 활용될 수 있을 것이다.

주제어 : 인공지능, 인공지능 윤리, 이루다, 감성 분석, 빈도분석, CONCOR분석

Abstract As artificial intelligence services targeting humans increase, social demands are increasing that artificial intelligence should also be made on an ethical basis. Following this trend, the government and businesses are preparing policies and norms related to artificial intelligence ethics. In order to establish reasonable policies and norms, the first step is to understand the public's perceptions. In this paper, social data and news comments were collected and analyzed to understand the public's perception related to artificial intelligence and ethics. Interest analysis, emotional analysis, and discourse analysis were performed and visualized on the collected datasets. As a result of the analysis, interest in "artificial intelligence ethics" and "artificial intelligence" favorability showed an inversely proportional correlation. As a result of discourse analysis, the biggest issue was "personal information leakage," and it also showed a discourse on contamination and deflection of learning data and whether computer-made artificial intelligence should be given a legal personality. This study can be used as data to grasp the public's perception when preparing artificial intelligence ethical norms and policies.

Key Words : AI, Artificial Intelligence, AI ethics, BERT, CONCOR analysis

*This work was supported by a research grant from Seoul Women's University(2021-0235)

*Corresponding Author : Eun-Jung Choi(chej@swu.ac.kr)

Received December 31, 2021

Accepted February 20, 2022

Revised February 3, 2022

Published February 28, 2022

1. 서론

인공지능 기술이 나날이 발전하며 우리 생활에 큰 편리함을 가져다 주고 있다. 금융, 헬스케어, 사이버 보안 등 이제 인공지능 분야 기술이 활용되지 않는 분야를 찾기가 힘들다. 실제로 AI를 도입한 우리나라 기업의 87%가 AI 기술 도입이 긍정적이었다고 판단했으며, 정부도 한국판 뉴딜 종합계획을 2020년도에 발표하며 전 산업의 AI기술 도입과 확산을 도모하겠다고 밝혔다[1]. 이로 보아 향후에도 인공지능 기술 활용이 확산될 것으로 예상된다.

사람에 대한 데이터를 기반으로 사람들을 대상으로 하는 인공지능 서비스가 증가하면서 인공지능에서도 윤리 문제를 생각하지 않을 수 없게 되었다. 특히, 2021년에는 인공지능 챗봇 “이루다” 사건으로 인공지능 윤리가 큰 이슈로 떠올랐다. “이루다”는 연인들이 나누는 100억 건의 대화를 딥러닝으로 학습하여 제공하는 챗봇서비스이다. “이루다” 사건으로 개발사는 개발단계에서 개인정보 수집 및 처리 과정 중 개인정보 보호법을 위반하여 과징금 5550만원과 과태료 4780만원이 부과되었다[2]. 이루다는 개인정보 유출 뿐만 아니라, 서비스 출시 후 사회적 약자를 대상으로 한 혐오와 차별적인 발언을 대화에서 쏟아내며 또다른 논란을 빚었다[3]. 이러한 사회적 이슈로 인해 정부에서는 “사람이 중심이 되는 인공지능을 위한 신뢰할 수 있는 인공지능 실현전략”을 발표하고 국내의 기업들도 윤리 원칙 준수를 위한 인공지능 개발 환경 구축을 위하여 노력을 시작하고 있다[4,5].

국가에서 어떠한 정책을 세우기 위해서는 대다수 국민이 공감할 수 있는 정책 대안을 모색해야 한다. 이를 위해 국민들의 다양한 의견을 적절히 수렴하는 과정이 필요하다[6]. 기업에서 사업을 기획할 때에도 소비자의 인식을 사전에 파악하는 작업은 필수이다. 본 논문에서는 SNS와 인터넷 기사 속 댓글을 수집하여 인공지능과 인공지능 윤리에 대한 인식 변화를 살펴본다. 2016년 1월부터 2021년 10월까지의 “인공지능”과 “인공지능 윤리”의 내용을 담은 기사에 달린 다음, 네이버의 댓글과 두 키워드를 언급한 트윗을 수집한다. 수집된 데이터에 대해 다양한 분석 및 시각화 기법을 이용하여 대중들의 인식과 변화를 확인한다.

2. 관련연구

2.1 담론분석 : CONCOR Analysis

빅데이터 속에서 담론을 나누기 위해서는 클러스터링 분석을 해야한다. 클러스터링 분석 알고리즘은 *clauset-newmanmoore*, *girvan-newman*, *CONCOR* 등이 있다. 본 논문에서는 네트워크 시각화 시 직관적으로 군집을 파악할 수 있으며, 특정 주제에 대한 인식을 파악하는 용도로 가장 널리 사용되고 있는 *CONCOR*를 활용할 것이다. *CONCOR*는 구조적 등위성을 통해 연결관계의 형태가 유사한 것끼리 군집화하는 알고리즘인데, 쉽게 말해 연결관계 패턴의 유사도를 블록으로 분할할 수 없을 때까지 수행하여 집단을 구분하는 방법이다[7,8]. 예를 들면 Fig.1의 노드B, C, D는 공통적으로 A와 연결되어 있지만 노드 B, C는 노드 E, F, G와 연결되어 있고 D, H와는 연결되어 있지 않다. 따라서 B, C, E, F, G가 한 그룹, D, H가 한그룹으로 묶이는 것이며 공통된 주제를 가지고 있다[9]. 이러한 특성으로 *CONCOR* 분석은 특정 주제에 대한 인식을 파악하는 데에 많이 사용되며 본 논문에서는 “이루다”를 키워드로 하여 인공지능 윤리에 대한 인식을 파악하는 데에 *CONCOR* 분석을 이용할 것이다.

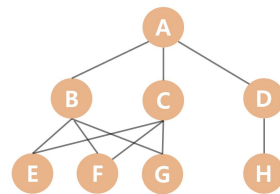


Fig. 1. CONCOR Analysis Graph

2.2 감성분석 : BERT Model

데이터를 분석하기 전에 전처리 단계에서 각 단어를 벡터로 처리하는 임베딩(Embedding)과정을 거친다. *Word2Vec*와 *FastText* 등의 임베딩 방식이 있으며, *BERT*도 임베딩 방식 중 하나이다. *BERT*(Bidirectional Encoder Representations from Transformers)는 2018년 구글이 공개한 사전 훈련 언어모델이다[10]. *BERT*는 랜덤하게 마스크 처리된 토큰들을 컨텍스트를 고려하여 예측하여 학습하는 방식(Masked Language Model)과 함께 다음 문장이 이전 문장 뒤에 자연스럽게 연결되는지 예측하는 방법(next-sentence prediction)을 이용한다[11]. 이러한 특성으로 다른 임베딩 모델보다 획기적인 성능을 보여주며 *BERT*는 2018년에 공개된 이후 NLP분야에서 널리 사용되는 모델이 되었다. 본 논문에서는 *BERT*모델을 이용하여 감성분석을 수행하여 대중들의 “인공지능”에 대한 호감도를 파악한다.

2.3 데이터 시각화

데이터 시각화(Data Visualization)는 정보를 시각적으로 표현하는 기술로 대규모 데이터를 색채, 통계, 이미지 등을 활용해 직관적으로 전달하는 방법을 말한다. 데이터 시각화는 베이스가 되는 데이터의 유형에 따라 종류가 나뉜다. 빅데이터 양 자체를 보여주는 정보형과 설득을 목적으로 하는 설득형이 있다. 데이터 시각화의 대표적 툴로 엑셀, R, Gephi, Tableau가 있으며, 바차트, 3D 차트, 히트맵, parallel 그래프, 네트워크 그래프 등 다양한 방식의 그래프가 있다[11-14]. 본 논문에서는 담론분석의 결과를 UCINET[15] 도구를 이용하여 네트워크 그래프로 시각화한다.

2.4 인공지능 윤리 인식 연구

현재까지 인공지능과 인공지능 윤리에 대한 인식 파악을 위해 다양한 연구 및 조사가 이루어졌다. 2019년도에는 4차산업혁명위원회가 “4차산업혁명 대국민 인식 조사”를 수행했다. 4차산업 혁명 기술에 대한 인식, 4차산업 혁명과 관련된 정부 정책에 대한 인식을 주로 설문하였다. 이 조사를 보완하여 2021년에는 4차산업혁명 관련 법적책 인식조사 연구가 이루어졌다. 시민들이 정책 결정 과정이 민주적으로 이루어지길 원하는 것을 확인할 수 있었다[16]. 다만 두 조사 모두 인공지능 혹은 인공지능 윤리에 대해 초점이 맞추어진 것이 아닌 4차산업 기술을 주제로 진행된 조사라 인공지능과 인공지능 윤리 인식의 구체적인 변화를 파악하기는 어렵다.

한편 2021년도에는 인공지능 윤리에 대한 인식 조사 연구가 이루어졌는데, 살펴본 두 연구 모두 설문을 이용한 방법으로 진행되었다는 점에서 한계가 있다[17,18]. 설문지법, 인터뷰 등과 같은 조사를 위한 설문은 조사 참여자들이 일반 대상의 여론을 대표할 수 있는가에 대한 의문과 함께 조사 설계자의 편견이 개입될 수 있다는 단점이 있다[19].

본 논문에서는 SNS 데이터와 인터넷 기사 댓글을 수집하여 2016년부터 2021년까지의 인공지능과 인공지능 윤리에 대한 인식변화를 살펴본다.

3. 연구방법

3.1 데이터 수집

데이터는 트위터, 네이버, 다음의 세 플랫폼에서 크롤

링을 통해 수집한다. 트위터는 기간 내에 특정 키워드가 언급된 트윗을 크롤링하며, 네이버와 다음은 기간 내의 특정 키워드가 언급된 기사에 달린 댓글을 크롤링한다.

첫 번째 데이터셋은 인공지능에 대한 인식변화를 파악하기 위해 “인공지능”을 키워드로 수집한다. 트위터에서는 2016년 1월부터 2021년 10월까지 “인공지능”이 언급된 트윗을 수집한다. 네이버와 다음에서 역시 2016년 1월부터 2021년 10월까지 게시된 기사 중 본문에 “인공지능”이 언급된 기사의 댓글을 수집한다. 두 번째 데이터셋은 인공지능 윤리에 대한 인식변화를 알아보기 위한 것으로 “인공지능”과 “윤리”를 키워드로 데이터를 수집하였다. 2016년 1월부터 2021년 10월까지 두 키워드가 동시에 언급된 트윗을 수집했으며, 네이버와 다음에서도 “인공지능”과 “윤리”가 본문에 동시에 언급된 기사의 댓글을 크롤링을 수행하였다. 세 번째는 인공지능과 인공지능 윤리에 중요한 변화를 가져온 키워드인 “이루다”를 키워드로 데이터를 크롤링하여 데이터셋을 생성하였다.

“인공지능” 키워드로 크롤링한 데이터셋은 데이터셋 I로 하고 “인공지능, 윤리” 키워드로 크롤링한 데이터셋은 데이터셋II로 하고 “인공지능, 이루다” 키워드로 크롤링한 데이터는 데이터셋III으로 한다.

3.2 데이터 분석

데이터는 데이터셋 I 21만 348건, 데이터셋II 9782건, 데이터셋III 6129건으로 총 22만 6259건을 수집했으며, 이 데이터를 바탕으로 수행할 분석은 다음과 같다. 먼저 데이터셋 I의 감성분석 중 호감도분석을 중심으로 진행한다. 감성분석은 구글에서 개발한 BERT(Bidirectional Encoder Representations from Transformers)를 이용한다. 네이버 영화 리뷰 데이터셋 15만건과 수집한 데이터 중 “인공지능”에서 1000건, “인공지능 윤리”에서 500건, “인공지능 이루다”에서 100건을 레이블링하여 훈련 데이터를 구성한다. 총 15만 1600개의 훈련 데이터와 5만개의 테스트 데이터로 모델을 학습시켰다. 테스트 데이터셋으로 측정된 모델의 정확도는 85.3% 수준이었다.

다음은 데이터셋 I 속 윤리를 언급한 횟수를 체크하여 전체에서 윤리를 언급한 데이터의 비율을 정리하여 “인공지능” 데이터셋의 감성분석 결과와 비교한다. 이를 통해 윤리가 인공지능 호감도와 연관성이 있는지 확인한다.

다음으로는 키워드에 대한 각 년도의 담론 변화를 살펴보기 위해 데이터셋 I 과 데이터셋II으로 빈도분석을 진행한다. 먼저 문장을 형태소 단위로 쪼개어 카운트하

고, 최상위 빈도 50개를 추출한다. 각 년도마다의 달라지는 관심도 파악을 위해 최상위 빈도 50개 중 모든 해에 등장하는 키워드들은 제외하고 최종 최상위 빈도 25개를 워드클라우드를 시각화한다.

마지막으로 “인공지능 윤리”가 대중들로 하여금 직접적인 언급하게 되었던 계기가 된 “이루다” 키워드에 대해 담론 파악을 위한 분석을 수행한다. 데이터셋Ⅲ으로는 CONCOR 분석을 진행한 후 UCINET을 활용하여 네트워크 그래프로 시각화한다. 이때 각 키워드의 노드는 각 문서에서 해당 키워드가 갖는 TF-IDF값에 따라 크기를 다르게 한다. “인공지능 이루다”를 두고 만들어진 구체적인 담론과 인공지능 윤리에 대한 대중들의 의견을 파악한다.

4. 연구결과

4.1 호감도와 관심도와와의 상관관계 분석

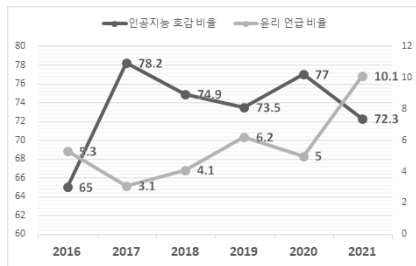


Fig. 2. “AI” positive percentage, “Ethic” mentioned percentage graph

BERT모델을 이용한 데이터셋 I의 감성분석 결과 인공지능에 대한 긍정비율은 2016년도부터 2021년까지 각각 65%, 78.2%, 74.9%, 73.5%, 72.3%였다. 한편 데이터셋 I에서 “윤리” 키워드가 언급된 빈도는 2016년도부터 2021년까지 각각 5.3%, 3.1%, 4.1%, 6.2%, 10.1%였다. 이 결과를 그래프로 나타내면 Fig.2와 같다. 인공지능 윤리에 대한 언급 비율이 전년도와 비교해 높아지면, 인공지능 데이터셋의 긍정 비율은 전년도에 비해 낮아짐을 확인할 수 있다. 이로써 인공지능 윤리에 대한 관심도와 인공지능에 대한 호감도가 높아지는 해에는 인공지능 윤리에 대한 관심도가 낮아지고 인공지능에 대한 호감도가 떨어지는 해에는 인공지능 윤리에 대한 관심도가 높아지며 반비례하는 관계가 있음을 알 수 있었다. 이와 같은 결과는 인공지능과 인간의 공존에 대한 사람들

의 견해를 보여주는 것으로 4.2절의 키워드 빈도분석에서 자세히 다룬다.

4.2 키워드 빈도 분석

데이터셋 I의 키워드 빈도 분석 결과는 Fig.3과 같다. 2016년부터 2021년까지 공통적으로 “인간과의 비교” 구조가 나타나고 있고 시기를 네 단계로 나눌 수 있었다.



Fig. 3. Wordcloud visualisation of “AI”

첫 번째 시기는 2016년으로 알파고를 계기로 인공지능이 대중들에게 본격적으로 알려지기 시작한 해이다. “이세돌”과 “알파고”의 바둑 대전이었으나 의미가 단순하지 않았다. 인공지능 “알파고”가 세계 최고의 바둑 실력을 가진 인간 “이세돌”을 이길 수 있는가, 즉 인공지능이 인류를 뛰어넘을 수 있는가의 의미가 담긴 매치였다. 두 번째는 2017년부터 2018년까지로 인간과 인공지능이 “일자리를 두고 대립하는 의견으로 인공지능이 노동 시장에서 인간을 대체할 수 있는지에 대한 이슈를 보여주었다. 세 번째는 2019년부터 2020년까지로 인간과 인공지능이 대립 관계를 보여주다 2019년에 처음으로 “활용”, “가능” 등 인공지능 활용과 관련된 키워드가 등장했고, 상위 빈도 키워드 중 미래지향적 키워드가 다수 등장했고 이 흐름은 2020년까지 이어졌다. 네 번째는 2021년이다. 2021년에는 처음으로 “문제”키워드가 상위 키워드로 등장했으며, 이는 상위 키워드인 “이루다”가 큰 영향을 끼친 것으로 보인다. 또한 “감정”, “사랑”이 상위 키워드로 등장하며 2021년에는 인공지능과 인간의 단순한 대립이 아닌, 인간과 인공지능을 나눌 수 있는 근본적인 것에 대한 담론이 만들어진 것을 볼 수 있다.

2021년부터 담론의 흐름이 달라진 것은 데이터셋Ⅱ에서도 확인할 수 있다. 결과는 Fig.4와 같으며 변화를 세 단계로 나눌 수 있다. 첫 번째는 2016년부터 2018년까지

지로 인공지능이 신기술로써 인식되는 단계이다. 인공지능에 대해 윤리와 관련된 키워드보다는 알파고, 가상현실 속 성추행, 블록체인과 같이 각 년도의 대표 신기술 키워드와 함께 인공지능이 언급된 것을 확인할 수 있다. 두 번째는 2019년도에서 2020년도까지이다. 이때는 국가주도의 인공지능 기술 발전과 관련된 키워드가 상위 빈도 키워드에 등장한다. 2020년도에 “윤리기준”, “사람 중심”의 키워드가 등장했으나 두 키워드는 국가가 주도한 인공지능 발전 키워드에 속한다. 즉 2020년까지는 인공지능 윤리와 관련한 담론이 대중들 사이에서 직접 만들어진 것이 아닌, 국가주도의 인공지능 발전 정책 소식에 인공지능 윤리가 함께 언급된 것이다. 세 번째는 2021년도이다. 2021년도에는 “혐오”, “개인정보”, “차별”, “인권”, “편향” 등 대부분 키워드가 윤리와 직접 맞닿아 있는 것을 확인할 수 있다. “이루다”가 주 키워드가 되어 “혐오”, “개인정보”, “차별”, “인권”, “편향” 등의 키워드가 상위 빈도 키워드에 등장하며 인공지능 윤리에 대해 대중들 사이에서 본격적으로 담론이 만들어졌음을 알 수 있다. 2021년도의 “이루다”로 시작된 인공지능 윤리 담론이 어떻게 형성되었는지 자세하게 파악하기 위하여 “인공지능 이루다”로 담론분석을 진행했고 결과는 4.3에 자세히 서술했다.



Fig. 4. Wordcloud visualisation of “AI Ethic”

4.3 CONCOR 분석

먼저 데이터셋Ⅲ에서 각 단어의 TF-IDF값을 추출하고 상위 100개의 단어를 선별했다. TF-IDF는 문서빈도와 역문서 빈도를 곱한 값이다. d를 문서 속 총 단어, t가 TF-IDF 값 추출 대상 단어, D를 문서 전체의 수라고 할 때, $tf(t,d)$ 는 d를 분모로 한 t 개수의 비율이고, $idf(t,D)$ 전체 문서의 수를 t를 포함한 문서로 나눈 뒤 로그를 취한 값이며 전체 공식은 Fig.5와 같다. 결과는 Fig.6과 같으며 Fig.6의 각 노드의 텍스트 크기는 TF-IDF값이며 Fig.7부터 Fig.10에서 자세하게 보여주고 있다.

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

Fig. 5. Formula of TF-IDF

Fig.6에서 나타나듯이 4개의 담론이 형성되었다. 4개의 담론은 Fig.7에서 Fig.10까지 나눠서 보여주고 있고 Table 1에 내용을 중심으로 보여주고 있다. Table 1에 보면 4개의 담론은 각각 개인정보 유출, AI에 인격 부여, AI 차별, 데이터 학습을 주제로 만들어졌다. 먼저 A담론은 Fig.7에서 보여주는 것과 같이 개인정보 유출이 주제인 담론이며 노드의 텍스트 크기가 다른 담론에 비교해 크게 나타나 개인정보 유출문제가 가장 중요하게 다루어졌음을 알 수 있었다. B담론은 Fig.8에서 보여주는 것과 같이 컴퓨터로 만들어진 캐릭터 즉 인공지능에 인격을 부여하는 것을 주제로 만들어진 담론이다. 성희롱, 성폭력, 살인과 같이 극단적인 범죄가 인공지능을 대상으로 했을 때도 적용해야 하는지에 대한 문제이다. C담론은 Fig.9에서 보여주는 것과 같이 인공지능이 재생산하는 차별에 대한 문제이다. 동성애(자), 성소수자와 같은 소수자에 대한 차별, 혐오 발언을 한 것으로 만들어진 담론인데, 이 담론은 D담론과도 이어진다. D담론은 Fig.10에서 보여주는 것과 같이 데이터 학습의 한계에 관한 담론이다. 좋지 않은 데이터(쓰레기 데이터)로 학습할 때 한계가 있다는 내용이다.

결국, 학습 데이터 수집 과정 중 개인정보 처리, 학습 데이터의 오염 및 편향, 인공지능의 법인격 부여가 인공지능 윤리와 관련하여 대중들의 많은 관심을 받은 이슈로 볼 수 있다.

4.4 분석 종합 결과

감성분석을 바탕으로 인공지능의 호감도와 인공지능 윤리의 언급도 사이에 유의미한 상관 관계가 있는 것을 확인했다. 또한 키워드 빈도 분석을 통하여 인공지능에 대한 대중들의 인식은 네 단계에 걸쳐 바뀐 것을 확인했다. 인공지능 윤리에 대한 인식 또한 세 단계에 걸쳐 바뀐 것을 확인할 수 있었는데, 특히 2021년 이루다 사건을 계기로 인공지능 윤리에 대한 대중들의 관심이 크게 늘어난 것을 알 수 있었다. 더 구체적인 인식 파악을 위해 CONCOR분석을 이용했다. 그 결과로 개인정보처리, 학습 데이터 편향, 인공지능에 대한 법인격 부여가 인공지능 윤리 문제 안에서도 대중들 사이에서 뜨겁게 논의된 담론임을 알 수 있었다.

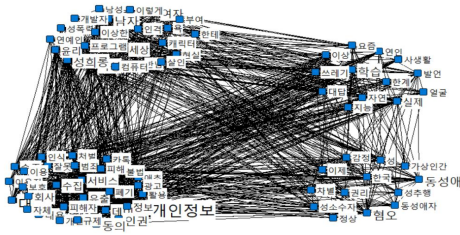


Fig. 6. Visualization results of CONCOR analysis



Fig. 7. Group A of CONCOR analysis



Fig. 8. Group B of CONCOR analysis



Fig. 9. Group C of CONCOR analysis

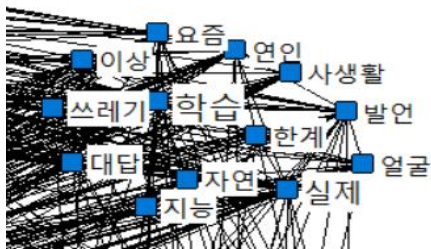


Fig. 10. Group D of CONCOR analysis

Table 1. Topic and keywords of each cluster

	Topic	Keywords
A	Personal information leakage.	“Personal information”, “Leakage”, “Protection”, “Contents”, etc.
B	Giving personality to AI	“Character”, “Personality”, “Giving”, “Murder”, “Ethic”, etc.
C	AI discrimination	“Homosexuality”, “Discrimination”, “Sex minority”, “Hate”, “Sexual harassment”, etc.
D	Limitations of data learning	“Learning”, “Garbage”, “Limitation”, “Actuality”, “Intelligence”, etc.

5. 결론

인공지능 윤리에 대한 인식변화를 살펴보기 위해 SNS 데이터와 인터넷 댓글을 수집하였고 다양한 분석을 진행했다. 결과적으로, 인공지능 윤리에 대한 관심도가 인공지능에 대한 호감도와 관련이 있는 것을 알 수 있었다. 특히 2019년, 2020년도에는 국가주도의 인공지능 윤리 담론에서 2021년에는 이루다 사건을 계기로 대중들이 직접 인공지능 윤리에 대한 담론을 이어간 것을 볼 수 있었다. 2021년에는 “개인정보 유출”, “학습 데이터의 오염 및 편향 문제”, “컴퓨터로 만들어진 인공지능에 법인격을 부여”에 대한 담론도 존재했다. 2016년 이세돌과 알파고의 바둑대전으로 인공지능에 대한 대중의 인식이 본격화되면서 2020년까지는 끊임없이 인공지능이 인간의 능력을 능가하는 존재, 일자리를 위협하는 존재 혹은 인간의 삶을 더 윤락하게 해줄 극단적 존재로 인식하는 경향이 있었다. 그러던 중 발생한 “이루다”사건으로 인공지능이 결국 사람을 위해 존재해야 한다는 인식이 중요하게 나타났다.

현재 기업과 정부에서는 인공지능의 신뢰성과 윤리에 대한 논의를 이어가고 있다. 정책이나 규범을 만들기에 앞서 주제에 대한 사회적 인식을 파악하는 것은 중요하다. 본 연구를 통해 인공지능과 인공지능 윤리에 대해 대중들의 인식 변화를 알 수 있었고 특히 2021년에 “이루다”사건으로 뜨거워진 흐름 속에서 구체적으로 어떤 담론이 형성되었는지 파악할 수 있었다. 본 연구가 인공지능 윤리 규범, 정책을 마련할 때 대중들의 인식을 파악할 수 있는 자료로 활용될 수 있을 것이다.

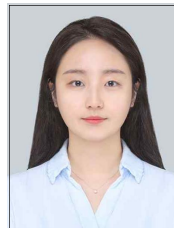
REFERENCES

[1] K. H. Kim. (2021). *Current Status and Implications of*

- Artificial Intelligence (AI) Introduction by Major Industries*. Jincheon : KISDI
- [2] Personal Information Protection Commission, (2021). *Case of corrective action for violation of the Personal Information Protection Act*. no 2021-007-072
- [3] K. S. Kwang. (2021. 03). The problem left by the chatbot "IruDa" in our society: Mounting a human rights manual on artificial intelligence. *Culture and science*, 105, 183-198.
- [4] Ministry of Science and ICT. (2021.05.13.). *The government has come up with a reliable AI realization strategy centered on people*. <https://www.korea.kr/news/policyNewsView.do?newsId=148887381>
- [5] NAVER. (2021.02). *AI ethical rules*. <https://www.naver.com/value/aiCodeEthics>
- [6] W. Y. Lee. (2011). Seoul National University Law. *The Law Research Institute Seoul National University*, 52(4), 125-168
UCI : G704-002133.2011.52.4.003
- [7] Y. H. Kim. (2020. Oct). Understanding and application of social network analysis technology. *KIPA Research Forum*, 34, 58-68
- [8] J. H. Lee, J. M. Moon & Y. S. Jang. (2017). Analysis of 2018 PyeongChang Olympic keywords using social network big data analysis. *Korean Journal of Sport Management*, 22(6), 73-89
DOI : 10.31308/KSSM.22.6.5
- [9] W. G. Kang, E. S. Ko, H. R. Lee & J. N. Kim. (2018). A Study of the Consumer Major Perception of Packaging Using Big Data Analysis -Focusing on Text Mining and Semantic Network Analysis-. *Journal of the Korea Convergence Society*. 9(4). 15-22
DOI : 10.15207/JKCS.2018.9.4.015
- [10] J. Devlin, M. W. Chand, K. Lee, K. Toutanova. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [11] J. S. Lee. (2013). A Study on Visualizing Method and Expression of Information Design for Big Data. *Journal of Basic Design & Art*, 14(3), 259-269
UCI : G704-001069.2013.14.3.026
- [12] G. S. Choe, Y. G. Ham & S. H. Kim. (2013) Bigdata Visualization. *KSCI Review*, 21(1), 33-43
- [13] Friendly, M. (2008). *A brief history of data visualization*. In Handbook of data visualization (pp. 15-56). Springer, Berlin, Heidelberg.
DOI : 10.1007/978-3-540-33037-0_2
- [14] Veale, M. & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97-112.
DOI : 10.9785/cr-2021-220402
- [15] Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- [16] J. H. Park. (2021). An Analysis of Perception on Law and Policy regarding the 4th Industrial Revolution. *Korean Journal of Law & Society*, 66, 175-204
DOI : 10.33446/KJLS.56.3
- [17] Y. H Ko & C. S Leem. (2021). The Influence of AI Technology Acceptance and Ethical Awareness towards Intention to Use. *Journal of Digital Convergence*, 19(3), 217-225
DOI : 10.14400/JDC.2021.19.3.217
- [18] G. S. Kim & Y. J. Shin. (2021). A Cross-Sectional Study of Artificial Intelligence Ethics Awareness - Preparation for Climate Change Education Using Artificial Intelligence -. *Journal of Energy and Climate Change Education (JECCE)*, 11(1), 27-36
DOI : 10.22368/ksecc.2021.11.1.27
- [19] D. G. Kim & S. Y. Shin. (2021). Comparing the Results of Big-Data with Questionnaire Survey. *Journal of the Korea Institute of Information and Communication Engineering*, 20(11), 2027-2032
DOI : 10.6109/jkiice.2016.20.11.2027

이 수 련(Lee, Su Ryeon)

[학생회원]



- 2018년 3월 ~ 현재 : 서울여자대학교 정보보호학과
- 관심분야 : 빅데이터, 인공지능, 개인 정보보호
- E-Mail : dolpong@swu.ac.kr

최 은 정(Choi, Eun Jung)

[정회원]



- 1997년 2월 : 서울여자대학교 컴퓨터학과(이학사)
- 2000년 2월 : 서울여자대학교 대학원 컴퓨터학과(이학석사)
- 2005년 8월 : 서울여자대학교 대학원 컴퓨터학과(이학박사)
- 2006년 3월 ~ 현재 : 서울여자대학교 정보보호학과 교수
- 관심분야 : 빅데이터, 인공지능, 악성코드, 개인 정보보호
- E-Mail : chej@swu.ac.kr