키워드 네트워크 분석과 토픽 모델링을 활용한 AI 윤리연구의 핵심주제 분석

김문구, 박종현, 박민재*
한국전자통신연구원, *아주대학교

mkkim@etri.re.kr, stephanos@etri.re.kr *geoglove@ajou.ac.kr

Main research topics Al ethics using keyword network analysis and topic modelling

Moon-Koo Kim, Jong-Hyun Park, Min Jae Park*
ETRI, *Ajou Univ.

요 약

급격한 기술진화와 사회적 확산을 바탕으로 AI 윤리의 중요성과 선제적 대응 필요성이 글로벌 의제로 부각되고 있다. AI 윤리연구의 핵심 키워드를 분석하기 위해 키워드 네트워크 분석과 토픽모델링을 적용한 결과, AI 윤리의 주체와 대상, 기술적 이슈가 주요 키워드로 도출되었다. 클러스터링 분석 결과, 공공정책, 사회적 공정성, 개인 권익이 핵심 주제그룹을 형성하였다. LDA 분석결과, 공공 가치를 위한 정책과 의사결정(토픽 1), 사회적 책임을 위한 원칙과 규제(토픽 2), 편의에 대한 관리 책임(토픽 3), AI 윤리 강화를 위한 기술적 접근(토픽 4)이 핵심 토픽으로 도출되었다.

Ⅰ. 서 론

2010 년대 중반이후 AI 는 급격한 기술진화와 생태계 확장을 바탕으로 새로운 황금기를 맞이하고 있으며 생성형 AI 를 중심으로 개인, 공공, 기업에 거대한 변화와 혁신을 선도할 것으로 전망되고 있다[1]. AI 에 대한 기대와 동시에 위험에 대한 우려가 사회적 이슈로 부각되면서, 선제 대응으로 AI 윤리(AI ethics)가 정책적, 학술적, 기술개발의 주요 의제로 떠오르고 있다[2]. 본 연구에서는 키워드 네트워크 분석과 토픽 모델링을 활용하여 AI 윤리의 학술적 지식생산에서 핵심 토픽들을 체계화하였다. 본 연구결과를 종합하여 AI 윤리의 주제적 진화방향을 전망하고 연구활성화 방향을 제언하였다.

Ⅱ. 연구 방법과 절차

본 연구에서는 AI 윤리 연구의 핵심 주제를 분석하기 위해 전문 학술 DB 인 Web of Science 에서 다음 검색어를 통해 2023 년 12 월 주제어 검색을 실시하였다. TS=("artificial intelligence" NEAR/3 ethic*)OR(ai NEAR/3 ethic*). 논문이 본격적으로 발표된 2015 년 이후 2023 년까지 영문 저널에 발표된 논문으로 한정한 결과 633 개의 서지정보를 확보하였다. 저자 키워드를 대상으로 소문자 변환, 오탈자 정정, 용어가 혼용된 경우 표준화(예: machine learning 을 ml 로 변경), Porter stemming 의 데이터 전처리를 과정을 거쳤다. 이를 통해 최종적으로 1,114 개의 키워드를 확보하였다. 본 키워드를 대상으로 네트워크 분석과 LDA(Latent Dirichlet allocation)에 의한 토픽모델링을 적용하였으며, 공개 SW 인 Python 패키지(genism)를 활용하였다.

Ⅲ. 연구결과

AI 윤리는 2022 년(162 편, 25.6%), 2023 년(249 편, 39.3%)에 출판된 논문들이 대부분(64.9%)이었다.

총 70 개 국가가 AI 윤리 연구에 참여하였으며 미국(182 편, 28.8%), 영국(104 편, 16.4%), 독일(85 편, 13.4%) 순으로 연구를 선도하였다. AI 윤리연구는 이들 국가에 이어 호주, 네덜란드, 캐나다, 중국, 스페인, 이탈리아, 프랑스 순으로 최상위권을 형성하였다. 다른 학문분야와 다르게 중국은 6.5%를 점유하는 정도였으며 우리나라는 일본과 함께 2.1%로 20 위권으로 나타났다. 상위 20 권 국가의 대부분이 미국, 캐나다, 호주와 함께 EU 국가들이었다. 글로벌 연구 협업이 가장 활발한 국가는 미국으로 28 개 국가와 208 의 협업 링크가 발생한 것으로 나타났다.

AI 윤리 연구에서 기술이나 도메인 용어를 제외하고 TF-IDF 측면에서 상대적 중요성을 지니는 AI 윤리 관련 키워드는 <표 1>과 같이, 교육, 책임, 정부, 도덕, 인간, 사회성, 프라이버시, 규제, 지속성, 정책, 공공, 설명가능, 정보제공, 리스크, 편의로 나타났다. AI 윤리의 주체 또는 대상, AI 윤리의 기술적 이슈가 주를 형성하였다.

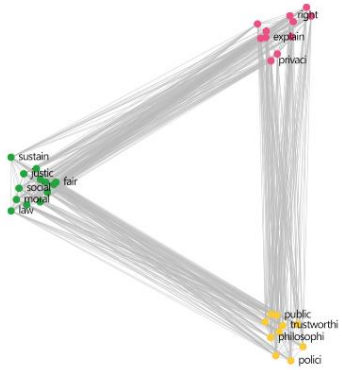
<표 1> AI 윤리연구에서 TF-IDF 측면에서 핵심 키워드

키워드	DF	TF-IDF	키워드	DF	TF-IDF
ethic	369	189.12	regul	26	85.40
educ	42	131.09	sustain	20	82.81
respons	47	127.81	polici	23	79.32
govern	45	125.07	public	23	76.14
moral	26	109.80	explain	21	71.80
human	35	104.60	inform	19	67.27
social	29	99.99	risk	16	67.17
privaci	26	85.40	bia	16	63.64

상위 AI 윤리 키워드를 대상으로 Leiden 방법을 적용한 클러스터링을 수행한 결과, (그림 1)과 같이 총 3 개 차원으로 군집화되었다.

- 그룹 1(노란색)은 공공 차원에서 윤리의 기반이 되는 철학적 공감대를 형성하고 투명성을 확보하기 위한 정책을 강화하는 연구가 주요 주제군을 형성하였다.
- 그룹 2(녹색)는 사회적 차원에서 공정과 정의, 지속성을 확보하기 위해 도덕과 법 체계를 강화하는 연구가 주된 테마가 되었다.
- 그룹 3(적색)은 개인의 권리를 강화하기 위해 프라이버시를 강화하고 AI 에 대한 인간의 통제권 확보를 위해 설명가능성을 제고하는 연구주제가 주요 연구그룹을 형성하였다.

핵심 키워드를 대상으로 네트워크 중심성(centrality)을 파악한 결과, (표 2>와 같이 나타났다. Degree centrality, Closeness Centrality, Betweenness centrality, Eigenvector centrality 에서 공통적으로 윤리, 정부, 책임, 프라이버시, 인간, 사회성이 최상위권을 형성하였다. 이들 키워드들은 다른 키워드와 연결되어 AI 윤리 연구의 핵심 허브 주제 역할을 수행하였다.



(그림 1) AI 연구주제의 핵심 주제어 클러스터링 결과

<표 2> AI 윤리연구에서 핵심 키워드의 네트워크 중심성

키워드	Degree	Closeness	Betweenness	Eigenvector
ethic	1.0000	1.0000	0.1126	0.3019
respons	0.7813	0.8205	0.0520	0.2509
educ	0.3750	0.6154	0.0044	0.1418
govern	0.8125	0.8421	0.0598	0.2610
human	0.6875	0.7619	0.0407	0.2211
moral	0.2500	0.5714	0.0040	0.0808
social	0.6563	0.7442	0.0292	0.2196
privaci	0.7500	0.8000	0.0355	0.2544
regul	0.5938	0.7111	0.0250	0.1996
polici	0.5625	0.6957	0.0169	0.1998
sustain	0.3438	0.6038	0.0027	0.1371
public	0.5938	0.7111	0.0177	0.2105
explain	0.6250	0.7273	0.0252	0.2113
inform	0.5625	0.6957	0.0188	0.1986
risk	0.5000	0.6667	0.0118	0.1791

LDA 에 의한 토픽 모델링을 수행하기 위해 본 연구는 Hoffman 외(2010) 등에서 제시한 방법을 적용하여 최적의 토픽수로 4 개를 결정하였으며 기술적 키워드를 제외한 윤리 키워드를 중심으로 분석한 결과는 <표 3>과 같았다[3].

- 토픽 1 은 정부, 도덕, 인간, 교육이 공통 키워드를 형성하는 가운데, 정책, 공공, 가치, 의사결정이 토픽 1 에만 특화된 키워드로 나타났다. 공공 가치를 위한 정책과 의사결정이 핵심 토픽을 형성하였다.
- 토픽 2 는 교육, 사회적, 정부, 인간, 도덕, 책임이 공통 키워드를 형성하였으며 규제, 원칙, 지속성, 도전이 특화 키워드로 도출되었다. 사회적 책임을 위한 원칙과 규제가 핵심 주제로 나타났다.
- 토픽 3 은 책임, 교육이 공통 키워드였으며 편의, 책무성, 혁신이 특화 키워드로 나타났다. AI 혁신에 따른 편의발생과 관리 책임이 핵심을 형성하였다.
- 토픽 4 는 사회성, 책임, 정부, 교육이 공통 키워드인 가운데, 평가, 설명가능성, 정보제공, 프라이버시가 특화 키워드로 나타났다. AI 윤리에 해당하는 전반적인 기술과 관련된 이슈들이 핵심 키워드를 형성하였다.

<표 3> AI 윤리연구의 핵심 토픽

키워드	Topic 1	Topic 2	Topic 3	Topic 4
공통 키워드	govern moral human educ	educ social govern human moral respons	respons educ	social respons govern educ
특화 키워드	polici public valu decis	regul principl sustn challeng	bia account innov	assess expln inform privaci

*토픽 확률을 기준으로 키워드 순서를 배정

Ⅳ. 결론: AI 윤리 연구의 활성화 방향

본 연구결과를 바탕으로 AI 윤리 연구의 활성화 방향을 제언하면 다음과 같다.

첫째, AI 윤리 연구에서 본격적인 지식성장과 영향력 확대가 필요하다. 2021 년 이후 양적 성장이 나타나지만 AI 기술진화나 확산에 비해 절대적으로 부족한 편이다. 특히 우리나라를 비롯하여 중국, 일본, 인도 등 아시아 국가들의 활발한 지식생산과 연구협력이 강화될 필요가 있다.

둘째, AI 연구주제의 다양성을 통해 연구의 외연확장이 필요하다 키워드 클러스터링과 토픽 모델링을 통해 분석한 AI 연구주제는 개인, 공공, 사회 차원에서 기술적 이슈(책임성, 투명성, 설명가능성), 정책적 이슈(규제)에 집중되어 있는 상황이다. AI 로 인해 동인되는 다양한 윤리적 이슈를 발굴하고 이에 따른 정책적 아젠다와 기술적 솔루션을 제시하는 다양한 범위의 연구가 확대될 필요가 있다.

끝으로, AI 윤리가 적용되는 학문범위를 확대할 필요가 있다. 현재 AI 윤리는 인문학, 사회과학에서 규범적 차원에서 접근을, 공학에서 기술적 대응으로 접근하고 있는 상황이다. AI 윤리가 적용되는 범위는 의학을 비롯하여 거의 모든 학문분야에 해당될 잠재력이 높다. 다양한 학문분야에서 특화된 주제를 다루며 이를 학제적 또는 융합적으로 확대하는 것이 필요하다.

참 고 문 헌

[1] Galindo, L., Perset, K., & Sheeka, F. (2021). An overview of national AI strategies and policies. OECD.

[2] ETRI-KISTEP, 디지털 역기능 전망과 대응방향, 2022.

[3] Hoffman, Matthew & Blei, David & Bach, Francis. (2010). “Online Learning for Latent Dirichlet Allocation”. Advances in Neural Information Processing Systems. 23. 856-864.