디지털경제사회연구 2026년 제4권 제1호, pp.81~117 ISSN 3059-0337
http://doi.org/10.23388/JDES.2026.4.1.003

인공지능 윤리 담론의 구조와 변화에 관한 연구: 토픽모델링 기반 언론기사 분석

유 성 열*

본 연구는 대규모 언론 텍스트를 활용하여 인공지능(AI) 윤리 담론의 구조적 특성과 시기별 변화를 분석하였다. 2022년과 2024년을 중심으로, 인공지능 윤리 관련 담론이 사회적으로 어떻게 형성되고 변화해 왔는지를 살펴보았다. 본 연구에서는 언론기사를 개인의 윤리의식을 직접 측정하는 자료가 아니라 공론장을 반영하는 담론 자료로 간주하였다. 분석 자료는 네이버 뉴스에서 수집한 2022년 6,337건, 2024년 14,612건의 기사로 구성되며, 중복 제거, 형태소 분석, 불용어 제거, 복합어 처리 등의 전처리를 수행하였다. 분석 방법으로는 키워드 빈도 분석, N-그램, TF-IDF, 그리고 Latent Dirichlet Allocation(LDA) 기반 토픽모델링을 활용하였으며, 토픽 수는 응집도를 기준으로 결정하였다. 분석 결과, 2022년의 담론은 규제와 위험 대응을 중심으로 비교적 통합적이고 중첩된 구조를 보인 반면, 2024년의 담론은 산업 활용, 거버넌스, 글로벌 규범 대응 등 보다 분화되고 전문화된 구조로 나타났다. 이는 인공지능 윤리 담론이 포괄적이고 규범적인 논의에서 벗어나 구체적이고 정책 중심의 논의로 전환되고 있음을 의미한다. 본 연구는 토픽모델링을 활용하여 인공지능 윤리 담론의 구조적 변화를 실증적으로 제시하였으며, 향후 정책 수립과 AI 거버넌스 설계에 대한 시사점을 제공한다.

주제어: *인공지능 윤리, 토픽모델링, 담론분석, 언론기사*

* 부산가톨릭대학교 컴퓨터정보공학과 교수, syyu@cup.ac.kr

1. 서 론

인공지능(AI)의 급격한 발전은 지난 10여 년간 사회 전반에 걸쳐 구조적 변화를 이끌어왔다. 특히 딥러닝과 생성형 AI의 등장은 언어·영상·음성 등 다양한 영역에서 인간 수준 혹은 그 이상의 성능을 발휘하며 산업·경제·문화 전반에 막대한 파급효과를 낳고 있다(Devlin, Chang, Lee, & Toutanova, 2019). 이러한 기술적 진보는 의료·교육·교통·금융 등 공공성과 사적 영역을 막론하고 긍정적 혁신을 촉진했지만, 동시에 사회적 우려와 윤리적 갈등을 동반하였다(Floridi & Cowls, 2019).

첫째, 데이터 프라이버시와 개인정보 보호문제다. 인공지능은 대규모 데이터에 기반해 학습하기 때문에 개인의 민감한 정보가 무단으로 활용될 위험이 크다(Zuboff, 2020). 딥페이크와 같은 기술은 허위 정보, 명예훼손, 범죄적 활용 가능성까지 내포하고 있으며, 이는 사회적 불신과 불안감을 증폭시킨다. 둘째, 공정성과 편향성 문제다. 알고리즘은 데이터의 편향을 그대로 재생산하거나 확대할 수 있으며, 이는 채용·대출·보험·사법 결정과 같이 사회적 형평성과 정의가 중요한 영역에서 불평등을 심화시킬 수 있다(Barocas & Selbst, 2016; Noble, 2018). 셋째, 책임성 및 설명가능성 문제다. 인공지능이 내린 의사결정의 근거를 인간이 이해하기 어렵다는 점에서, 결과에 대한 책임 주체가 불분명해지는 문제가 제기된다. 이는 법적·제도적 대응뿐 아니라 윤리적 신뢰 구축에도 심대한 영향을 끼친다(Mittelstadt, 2019). 넷째, 제도적·사회적 신뢰 기반의 취약성이다. 우리나라에서는 2025년 1월 "인공지능 기본법"이 국회를 통과하며 법제화 단계에 진입했지만, 법률의 실효성과 규제 범위에 대한 논란이 여전히 존재한다. 비판자들은 해당 법이 민간 자율 규범에 과도하게 의존하고 있으며, 국민들의 구체적인 우려와 윤리적 감수성을 충분히 반영하지 못하고 있다고 지적한다(홍종현, 2025). 이는 제도 설계와 사회적 신뢰 확보 사이의 간극을 잘 보여준다.

이러한 배경 속에서 본 연구의 목적은 국내 언론기사에 나타난 인공지능 윤리 관련 담론이 어떠한 주제를 중심으로 형성되고 있는지, 그리고 이러한 담론 구조가 시기별로 어떻게 변화하고 있는지를 실증적으로 분석하는 데 있다. 본 연구는 언론기사를 개인의 윤리의식을 직접 측정하는 자료로 보지 않고, 인공지능 윤리 이슈가 사회적으로 어떠한 방식으로 공론화되고 구조화되는지를 보여주는 담론 자료로 활용한다.

연구의 필요성은 다음과 같다. 첫째, 기존 연구들은 철학적 성찰이나 정책 제언 중심으로 이루어져 실증적 데이터 기반의 계량 분석은 부족했다(Hagendorff, 2020). 둘째, AI 윤리 논의는 국제적으로 폭넓게 이루어지고 있으나(Jobin, Ienca, & Vayena, 2019; OECD, 2019), 한국적 맥락을 반영한 체계적 분석은 드물다. 셋째, 언론은 사회적 어젠다를 형성하고 공론장의 방향을 결정짓는 주요 채널이므로, 언론 보도에 대한 계량적 분석은 사회적 인식과 정책적 논의 사이의 간극을 이해하는 데 매우 유효하다.

따라서 본 연구는 2022년과 2024년의 국내 주요 언론에 보도된 인공지능 윤리 관련 기사를 수집하고, 토픽모델링 기법을 적용하여 담론 구조를 분석한다. 이를 통해 사회적 신뢰 기반 구축을 위한 제도 설계, 국민 수용성을 고려한 정책 형성, 윤리 담론의 학문적 확장에 기여할 수 있을 것이다. 또한, 본 연구는 정서적 반응을 직접 분석하기보다는, 언론기사에 나타난 주요 담론 주제와 그 변화 양상을 중심으로 인공지능 윤리 담론의 구조를 분석하는 데 초점을 둔다.

2. 이론적 배경 및 선행연구

인공지능 윤리 연구는 지난 10여 년간 급격히 성장해 왔으며, 국제적으로는 다수의 가이드라인, 원칙, 정책 프레임워크가 제시되었다. 그러나 이러한 논의는 종종 선언적 수준에 머물러 있으며, 실질적 실행 가능성이나 사회적 수용성과의 연결에 있어 한계를 지닌다. 본 연구에서는 인공지능 윤리 담론에 대한 기존 연구를 (1) 국제적 윤리 원칙 및 가이드라인, (2) 알고리즘 편향 및 미디어 담론 연구, (3) 국내 법·정책 및 연구 동향의 세 가지 측면에서 검토한다. 이러한 구분은 인공지능 윤리 논의가 규범적 원칙 수준에서 출발하여, 실제 사회적 담론과 정책 논의로 확장되는 구조를 반영한 것이다. 따라서 본 연구는 이러한 선행연구의 흐름을 토대로, 실제 공론장에서 형성되는 담론 구조를 실증적으로 분석할 필요성이 있다고 판단한다.

1) 국제적 AI 윤리 원칙의 형성

인공지능의 급속한 확산은 국제사회가 공통의 윤리적 기준과 규범을 수립할 필요성을 강하게 인식하게 만들었다. 이에 따라 여러 국제기구가 AI의 책임 있는 개발·활용을 촉진하기 위한 권고안을 제정하였으며, 이는 오늘날 각국의 정책과 법제화 과정에 중요한 준거로 기능하고 있다.

OECD는 2019년 세계 최초의 국제적 AI 권고안인 OECD AI Principles을 발표하고, 2024년 이를 보완·업데이트하였다. 이 원칙은 인간 중심성, 투명성, 안전성, 책임성을 핵심으로 하며, AI가 인권과 민주적 가치를 존중해야 함을 강조하고, 이후 G20 등에서 채택되며 글로벌 기준으로 자리잡았다(OECD, 2019).

유럽연합(EU)은 2019년 「신뢰할 수 있는 인공지능을 위한 윤리지침」을 통해 인간의 자율성, 안전성, 프라이버시, 공정성, 책임성 등을 포함한 7대 요건을 제시하였다(European Commission, 2019). 특히 '신뢰할 수 있는 AI' 개념을 중심으로 기술적 신뢰성과 윤리적 수용성을 동시에 강조하며, 이후 AI Act 논의의 기반을 형성하였다.

UNESCO는 2021년 「인공지능 윤리에 관한 권고안」을 통해 인권, 다양성, 포용성, 지속가능성을 핵심 가치로 제시하였다(UNESCO, 2021). 특히 글로벌 형평성과 개발도상국의 AI 역량 강화를 강조하며, 기존 OECD 및 EU 중심 논의를 보완하는 국제적 틀을 제공하였다.

종합하면, OECD·EU·UNESCO는 모두 AI의 책임성과 신뢰성을 강조하면서도 각기 다른 초점을 지닌다. OECD는 경제협력과 제도적 정책 프레임워크에, EU는 법제화와 규제 기반의 신뢰성에, UNESCO는 문화적 다양성과 글로벌 형평성에 무게를 두고 있다. 〈표 1〉은 국제기구의 인공지능 윤리 원칙을 비교하여 요약한 것이다.

〈표 1〉 국제기구의 인공지능 윤리 원칙 비교

구분	OECD(2019/2024)	EU(2019)	UNESCO(2021)
문서명	OECD Principles on AI(OECD Council Recommendation on AI)	Ethics Guidelines for Trustworthy AI	Recommendation on the Ethics of Artificial Intelligence
제정/발표	2019년 (2024년 4월 업데이트)	2019년	2021년
주관 기관	OECD	European Commission(High-Level Expert Group on AI)	UNESCO
주요 원칙/가치	– 인간 중심성 (human-centered) – 포용성(inclusive growth) – 투명성(transparency) – 견고성과 안전성 (robustness & safety) – 책임성(accountability) – 인권과 민주적 가치 존중	– 신뢰할 수 있는 AI (Trustworthy AI)를 위한 7대 요건: ① 인간 자율성 보장 ② 기술적 견고성 및 안전성 ③ 프라이버시와 데이터 거버넌스 ④ 투명성 ⑤ 다양성·비차별·공정성 ⑥ 사회적·환경적 복지 ⑦ 책임성	– 인권 존중 – 다양성·포용성 – 환경 지속가능성 – 평등·비차별 – 평화·민주주의 가치 강화 – 글로벌 남반구 역량 강화 및 국제 협력
특징	최초의 국제 AI 원칙, 경제·정책 프레임워크 강조, G20 등 국제 확산 기반	법적 규제와 연결된 지침, EU AI Act 입법의 토대	문화적 다양성·형평성 반영, 국제 개발 격차와 글로벌 협력 강조
영향력	각국 AI 거버넌스 및 국제 규제 논의의 기준점	유럽연합 AI 법제화 논의(특히 AI Act)의 근간	국제사회 윤리 가이드라인 확산, 개발도상국 지원 프레임워크 제공

학계에서도 인공지능 윤리에 대한 다양한 원칙 제안이 이루어졌다. 플로리디와 카울스(Floridi & Cowls, 2019)는 기존 가이드라인의 중복성과 체계 부족을 지적하며, 선행(Beneficence), 무해성(Non-maleficence), 자율성(Autonomy), 정의(Justice), 설명가능성(Explicability)의 다섯 가지 보편 원칙을 제시하였다. 특히 설명가능성은 알고리즘의 작동 원리 이해 가능성과 책임성(accountability)을 포함하는 개념으로, 이후 설명가능 인공지능(XAI) 연구 및 AI 거버넌스 논의의 핵심 기반이 되었다.

또한 조뱅, 옌카, 그리고 바예나(Jobin, Ienca, & Vayena, 2019)는 84개 국가 및 기관의 윤리 지침을 비교 분석하여, 투명성, 공정성, 책임성, 프라이버시 보호가 공통된 핵심 원칙임을 도출하였다. 동시에 구체적 실행 메커니즘의 부족을 지적하며, AI 윤리의 글로벌 거버넌스 필요성을 강조하였다.

그러나 이러한 국제적 윤리 원칙은 주로 규범적 수준에서 제시된 것으로, 실제 사회

에서 어떠한 방식으로 수용되고 공론화되는지에 대한 실증적 분석은 상대적으로 부족하다. 즉, '어떤 원칙이 제시되었는가'에 대한 연구는 축적되어 있으나, '이러한 원칙이 사회적 담론에서 어떻게 재구성되고 있는가'에 대한 분석은 미흡한 실정이다. 이러한 점은 본 연구가 언론기사를 기반으로 인공지능 윤리 담론의 구조를 분석하고자 하는 중요한 근거가 된다.

2) 알고리즘 편향과 담론 연구

AI 윤리 논의에서 알고리즘 편향(Bias)은 핵심 주제 중 하나다. 바로카스와 셀브스트(Barocas & Selbst, 2016)는 빅데이터 활용이 사회적 불평등을 확대할 수 있음을 법학적 관점에서 분석하며, 특히 데이터 수집과 학습 과정에서의 편향이 '차별적 결과(Disparate Impact)'를 낳을 수 있음을 지적했다. 이 논문은 법·정책 차원에서 데이터 기반 의사결정의 규제 필요성을 강조한 대표적 연구다.

오닐(O'Neil, 2016)은 저서 "Weapons of Math Destruction"에서 알고리즘이 공정성을 해칠 뿐만 아니라, 불투명성과 확장성 때문에 사회 전반에 치명적 영향을 미칠 수 있음을 경고했다. 이 책은 대중적 차원에서 알고리즘 윤리 문제를 확산시키는 데 중요한 기여를 했다. 노블(Noble, 2018)의 "Algorithms of Oppression: How Search Engines Reinforce Racism." 역시 검색 알고리즘이 인종·성별 편견을 내포하고 있으며, 사회적 약자를 차별할 수 있음을 실증적으로 보여주었다.

언론 담론 분석 역시 AI 윤리 연구의 중요한 한 축이다. 매컴스와 쇼(McCombs & Shaw, 1972)의 '의제설정이론(Agenda-Setting Theory)'은 언론이 사회적 이슈의 중요성을 어떻게 규정하는지 보여준다. 이를 AI 맥락에 적용한 패스트와 호비츠(Fast & Horvitz, 2017)는 지난 수십 년간 대중의 AI 인식 변화를 분석하며, 2010년대 이후 언론 보도에서 '위험과 우려'의 프레임이 강화되었음을 보여주었다.

기존 연구들은 알고리즘 편향 문제나 언론 보도를 통해 인공지능에 대한 사회적 인식과 위험 인식을 분석하는 데 기여하였다. 그러나 이러한 연구는 특정 이슈나 사례에 집중하거나, 장기적 인식 변화에 초점을 두는 경우가 많아, 특정 시점에서 형성된 담론 구조를 체계적으로 분석하는 데에는 한계가 있다. 또한 다수의 연구가 질적 분석 또는 사례 중심 접근에 머무르고 있어, 대규모 텍스트 데이터를 활용한 계량적 담론

구조 분석은 상대적으로 부족하다. 이러한 한계를 보완하기 위해 본 연구는 토픽모델링을 활용하여 언론 담론의 구조를 정량적으로 분석하고자 한다.

3) 국내 연구 동향

국내에서도 최근 AI 윤리 논의가 활발히 전개되고 있다. 임미가(2021)는 국내외 연구를 종합하여 AI 윤리 담론의 체계적 문헌고찰을 수행하였고, 윤리적 원칙의 다층성과 실천적 한계를 지적했다. 박휴용(2022)은 교육학 관점에서 AI 윤리의 주요 쟁점을 정리하면서, 특히 학교 교육과정에 AI 윤리 교육을 포함할 필요성을 강조했다.

법·정책 연구에서는 홍종현(2025)이 "인공지능 기본법"과 개인정보보호법의 관계를 분석하며, 현행 법제가 민간 자율 규범에 과도하게 의존하고 있음을 비판했다. 이는 본 연구의 문제의식과 직접적으로 연결된다. 박미정(2024)은 데이터 프라이버시 보호를 위한 입법적 요구를 분석하며, 제도적 보완의 필요성을 제시했다.

국내 연구는 법·정책 및 교육적 측면에서 인공지능 윤리 논의를 확장해 왔으나, 여전히 규범적·이론적 논의에 머무르는 경향이 강하다. 특히 실제 사회에서 인공지능 윤리 이슈가 어떠한 담론 구조로 형성되고 있는지를 계량적으로 분석한 연구는 제한적이다. 따라서 본 연구는 국내 언론 데이터를 활용하여 인공지능 윤리 담론의 구조와 변화를 실증적으로 분석함으로써 기존 연구의 한계를 보완하고자 한다.

4) 연구의 필요성

이상의 선행연구를 종합하면, 인공지능 윤리 논의는 국제적 차원에서 다양한 원칙과 가이드라인이 제시되고, 알고리즘 편향 및 미디어 담론 연구를 통해 사회적 인식과 위험 요인이 분석되어 왔으며, 국내에서도 법·정책 및 교육적 측면에서 논의가 확장되고 있다. 그러나 이러한 연구들은 규범적 원칙 제시 또는 개별 이슈 분석에 집중되어 있어, 실제 공론장에서 형성되는 담론의 구조와 변화 양상을 체계적으로 분석하는 데에는 한계를 지닌다. 특히 대규모 언론 데이터를 기반으로 인공지능 윤리 담론의 구조를 정량적으로 분석한 연구는 상대적으로 부족하며, 시기별 변화 양상을 비교한 실증 연

구 역시 제한적이다.

따라서 본 연구는 2022년과 2024년 국내 언론 기사 데이터를 기반으로 토픽모델링을 적용하여 인공지능 윤리 담론의 구조와 시기별 변화 양상을 분석하고자 한다. 이를 통해 기존 연구가 충분히 다루지 못한 '공론장 기반 담론 구조 분석'이라는 측면에서 학문적 기여를 도출하고자 한다. 또한 본 연구는 기존 연구와 달리 대규모 언론 데이

〈표 2〉 기존 연구 요약

연구자	연구 주제	주요 기여	한계
Floridi & Cowls (2019)	AI 윤리 원칙 통합	5대 원칙(선행, 무해성, 자율성, 정의, 설명가능성) 제시 → 국제적 통일성 확보	원칙은 제시했지만 실행 메커니즘 부족
Jobin, Ienca & Vayena (2019)	전 세계 84개 AI 윤리 가이드라인 비교	국제적으로 공통 원칙(투명성, 공정성, 책임성, 프라이버시) 도출	실행 수준에서의 차이, 지역별 특수성 반영 부족
OECD (2019)	AI 원칙 발표	인류 중심 AI, 공정성 · 투명성 · 책임성 강조, 회원국 정책 반영	권고 수준, 강제력 미약
IEEE (2019)	Ethically Aligned Design	개발자 관점에서의 설계 원칙 제시, 기술표준 기반 제공	현실 적용에 필요한 구체성 부족
Barocas & Selbst (2016)	빅데이터와 차별	데이터 편향 → 차별적 결과 도출 가능성 규명	법학적 접근에 국한
O'Neil (2016)	『Weapons of Math Destruction』	알고리즘이 불평등 심화, 사회적 위험 고발	대중서로 과학적 정밀성 부족
Noble (2018)	『Algorithms of Oppression』	검색 알고리즘의 인종 · 성별 편향 실증	특정 플랫폼(Google) 사례 중심
Fast & Horvitz (2017)	대중의 AI 인식 변화	언론 보도 분석, AI 위험 프레임 강화 확인	장기간 추세에 집중, 국가별 차이 분석 부족
임미가 (2021)	AI 윤리 연구 문헌고찰	국내외 연구 동향 체계적 정리	실증 분석 부족
박휴용 (2022)	교육에서의 AI 윤리	AI 윤리 교육 필요성, 교과 반영 강조	교육적 실행 모델 부족
홍종현 (2025)	인공지능 기본법 분석	개인정보보호법과의 관계 분석, 제도적 한계 지적	법학적 관점에 치중
박미정 (2024)	데이터 프라이버시 입법 분석	제도 보완 필요성 제시	입법적 · 규범적 논의에 한정

터를 기반으로 토픽모델링을 활용하여 인공지능 윤리 담론의 구조를 실증적으로 분석하고, 시기별 변화를 비교한다는 점에서 차별성을 가진다.

3. 연구 방법

본 연구는 인공지능 윤리 담론의 변화를 분석하기 위해 2022년과 2024년 두 시점의 데이터를 중심으로 한다. 2022년은 인공지능 윤리 논의가 국내에서 본격적으로 사회적 의제화되며 다양한 가이드라인과 정책적 제안이 등장한 시점으로, 초기 담론 형성을 파악하기에 적합하다. 반면 가장 최근인 2024년은 인공지능 기본법 논의, 딥페이크 규제, 개인정보 · 프라이버시 보호와 같은 구체적 제도적 장치가 활발히 추진된 시점으로, 담론의 제도화와 성숙 과정을 드러내는 데 중요한 의미를 지닌다. 이에 따라 2022년과 2024년을 비교하는 것은 인공지능 윤리 담론이 초기 의제화 단계에서 제도적 정착 단계로 전환되는 과정을 파악하기 위한 합리적인 분석 전략이라 할 수 있다.

또한 2023년은 담론적 전환이 뚜렷하게 나타나기보다는 2022년의 흐름이 유지되며 2024년의 제도화로 이어지는 과도기적 성격을 보인다. 따라서 2023년을 별도로 분석하는 것은 연구의 초점을 분산시킬 수 있다고 보이며, 2022년과 2024년을 양극점으로 삼아 비교하는 것이 담론 변화의 핵심적 특징을 보다 명확히 드러내는 것이 더 적합하다고 판단하였다. 이러한 연구 목적을 달성하기 위해 본 연구에서는 토픽모델링 기법을 활용하여 언론 담론의 구조를 분석하였다.

본 연구에서는 텍스트 데이터 분석을 위해 Python과 텍스톰(Textom)을 병행하여 활용하였다. 먼저 데이터 전처리와 형태소 분석, 토픽모델링(LDA), coherence score 계산은 Python 환경(KoNLPy, Gensim, Scikit-learn)을 기반으로 수행하였다. 한편, 키워드 빈도 분석, N-그램 분석, TF-IDF 분석 등 기초 통계 분석과 일부 시각화는 텍스톰을 활용하였다.

1) 데이터 수집 방법

본 연구에서 언론기사를 분석 대상으로 선정한 이유는 언론이 특정 이슈를 선별 · 강

조하는 과정을 통해 사회적 관심과 공론장의 의제를 형성하는 주요 매개체이기 때문이다. 따라서 언론기사는 개인의 윤리의식을 직접 측정하는 자료는 아니지만, 인공지능 윤리와 관련된 사회적 담론이 어떠한 방식으로 형성되고 확산되는지를 보여주는 대표적인 공론장 자료로서 의미를 지닌다.

데이터는 네이버 뉴스 플랫폼을 기반으로 수집하였다. 네이버는 한국 내 가장 큰 뉴스 포털로, 종합일간지와 IT 전문지 기사를 통합적으로 제공한다는 장점이 있다. 네이버 오픈 API와 웹 크롤링 방식(BeautifulSoup, Selenium, JSON API 요청 등)을 병행하여, 지정된 기간과 키워드를 기준으로 기사를 체계적으로 수집하였다. 수집 과정에서 중복 기사, 광고성 기사, 제목만 동일한 재송출 기사 등을 필터링하여 데이터의 일관성을 확보하였다. 수집한 데이터셋에는 기사 제목, 원문 텍스트, 작성일, 언론사, 기사 원본 URL이 포함되었다.

데이터셋 구축을 위해 본 연구는 두 단계의 검색 조건을 설정하였다. 첫째, 모든 기사에는 "인공지능" 또는 "AI"라는 용어가 반드시 포함되도록 하였다. 이는 연구 주제가 인공지능과 윤리 담론을 다루는 만큼, 분석의 기본 단위가 되는 코퍼스가 인공지능 관련 기사임을 보장하기 위한 것이다. 둘째, 윤리와 규제, 사회적 수용성을 나타내는 11개의 세부 키워드를 설정하여, 이들 중 하나 이상이 함께 포함된 기사를 수집 대상으로 삼았다. 구체적으로는 "가이드라인", "개인정보", "공정성", "규제", "기본법", "데이터보호", "딥페이크", "알고리즘편향", "윤리", "책임", "프라이버시"를 선정하였다.

이 키워드는 선행연구(Jobin et al., 2019; Floridi & Cowls, 2019; OECD, 2019)에서 반복적으로 제시된 국제적 AI 윤리 원칙(공정성, 투명성, 책임성, 프라이버시 보호 등)을 반영함과 동시에, 한국 사회의 특수한 법 · 정책적 맥락(예: AI 기본법, 개인정보보호법)을 포괄하기 위해 선정되었다. 또한 생성형 AI 확산으로 최근 이슈가 된 딥페이크를 포함하여, 전통적 윤리 쟁점과 신흥 이슈를 동시에 담도록 구성하였다.

이들 키워드는 인공지능 윤리 담론의 주요 축을 포괄한다는 점에서 연구적으로 타당하다고 판단한다. "가이드라인"과 "기본법"은 제도적 장치와 법적 규범을 의미하며, "규제"와 "책임"은 거버넌스 및 책임소재 문제를 반영한다. "개인정보", "데이터보호", "프라이버시"는 데이터 시대의 핵심 윤리 이슈로, 특히 생성형 AI 확산에 따라 민감하게 부각된 주제이다. "공정성"과 "알고리즘편향"은 AI 시스템의 결과가 사회적 불평등을 재생산할 수 있다는 우려를 담아내며, "윤리"는 전반적 규범 논의를 포괄하는 개념이다.

마지막으로 "딥페이크"는 기술 발전에 따른 사회적 위험 사례를 대표하는 키워드로서 선정하였다.

따라서 최종적으로 본 연구의 데이터 수집은 "인공지능" 또는 "AI"라는 포괄 조건과 위의 11개 세부 키워드 중 최소 하나의 동시 출현을 만족하는 기사들로 한정하였다. 이는 인공지능 기술의 발전 과정에서 제기되는 다양한 윤리 · 규제 · 사회적 수용성 문제를 포괄하면서도, 불필요하게 광범위한 비관련 기사의 포함을 방지하기 위함이다. 〈표 3〉은 데이터 수집을 위한 키워드를 요약한 것이다.

〈표 3〉 데이터 수집을 위한 키워드 검색 구조

구분	검색식 구조
1차 조건 (포괄적 범주)	("인공지능" OR "AI")
2차 조건 (윤리·정책 관련 키워드)	("가이드라인" OR "개인정보" OR "공정성" OR "규제" OR "기본법" OR "데이터보호" OR "딥페이크" OR "알고리즘편향" OR "윤리" OR "책임" OR "프라이버시")
최종 검색식	("인공지능" OR "AI") AND("가이드라인" OR "개인정보" OR "공정성" OR "규제" OR "기본법" OR "데이터보호" OR "딥페이크" OR "알고리즘편향" OR "윤리" OR "책임" OR "프라이버시")

수집을 위한 세부 조건으로 사진 및 동영상 뉴스, 보도자료, AI 자동생성기사를 모두 제외하고 지면기사로 발행된 기사만을 포함하였다. 텍스트 분석을 진행하고자 하므로 사진, 동영상 위주의 기사는 제외한 것이며, 데이터의 양이 많아 분석의 편의를 위해, 담론 지형에 대해 대표성을 가질 것으로 보이는 주요 언론사(지면이 있는 언론사)로 수집 대상을 제한하였다. 또한, 기사 중, 네이버 기사 분류에서 연예 뉴스(m.entertain.naver.com) 또는 스포츠 뉴스(m.sports.naver.com)에 게시된 기사도 연구주제와의 관련성을 고려하여 분석에서 제외하였다. 또한 공공기관 또는 기업의 보도자료를 그대로 게시한 기사는 제외하였으며, 관공서 및 기업의 인사와 관련된 기사 역시 분석 대상에서 제외하였다.

2) 수집한 데이터 분량

본 연구는 상기 키워드를 활용하여 네이버 뉴스 플랫폼으로부터 인공지능 윤리 관련

기사를 수집하였다. 그 결과, 2022년에는 총 6,337건의 기사가 확보되었으며, 데이터 용량은 약 12.4MB에 해당한다. 반면 2024년에는 총 14,612건, 약 26.5MB의 데이터가 수집되어, 기사 건수와 데이터 용량 모두에서 2022년에 비해 약 두 배 이상 증가하였다. 이러한 차이는 인공지능 기술의 사회적 확산과 더불어 윤리·규제 담론이 활발히 전개되었음을 보여준다(〈표 4〉).

즉, 2022년 데이터셋은 인공지능 윤리 담론이 막 본격적으로 의제화되던 시점의 초기적 성격을 반영하고 있으며, 2024년 데이터셋은 제도적 논의와 구체적 규제안이 집중적으로 다루어진 시기의 담론을 담고 있다. 따라서 두 시점의 데이터 규모 차이는 단순한 양적 증가를 넘어, 사회적 담론의 확산과 성숙도를 보여주는 중요한 지표라 할 수 있다.

〈표 4〉 연도별 데이터 수집 규모

연도	기사 건수	데이터 용량
2022년	6,337건	12.4MB
2024년	14,612건	26.5MB

3) 데이터 전처리

수집된 뉴스 기사 데이터는 원자료(raw data)의 특성상 불필요한 요소와 잡음을 포함하고 있기 때문에, 본격적인 분석에 앞서 체계적인 전처리 과정을 거쳤다. 전처리는 크게 중복 기사 제거, HTML 태그 및 특수문자 정제, 형태소 분석, 불용어 제거, 그리고 복합어 결합 처리의 단계로 이루어졌다.

첫째, 중복 기사 제거는 동일한 기사가 서로 다른 경로(포털 재송출, 중복 인용 등)로 수집된 경우를 걸러내기 위한 과정이다. 기사 제목과 본문 텍스트의 유사도를 비교하여 일정 수준 이상 동일한 문서들을 중복으로 간주하였으며, 대표 문서만 남겨 최종 코퍼스의 순도를 높였다. 이 과정을 통해 불필요한 데이터 부풀림을 방지하고 분석 결과의 왜곡을 최소화하였다.

둘째, HTML 태그 및 특수문자 정제는 기사 본문 내에 포함된 광고 문구, 하이퍼링크, 줄 바꿈 태그(〈br〉, 〈p〉) 등 자연어 분석과 무관한 요소를 제거하는 작업이다. 또

한 본문에 혼입된 불필요한 특수기호(예: “※”) 역시 일괄적으로 제거하였다. 이를 통해 분석 대상이 되는 텍스트만을 순수하게 확보하였다.

셋째, 형태소 분석은 한국어의 언어적 특성을 고려한 필수적인 전처리 단계이다. 본 연구에서는 형태소 분석기(MeCab-ko)를 사용하여 문장을 형태소 단위로 분절하고, 각 단어에 품사 태그를 부착하였다. 한국어는 조사, 어미 변화가 풍부하여 원형 단어를 식별하는 과정이 중요하므로, 이를 통해 “인공지능/NNG[1)]”, “규제/NNG”, “책임/NNG”와 같은 기본 단위의 분석 가능한 토큰을 확보할 수 있었다.

넷째, 불용어 제거는 의미 해석에 기여하지 않는 단어들을 필터링하는 과정이다. 예를 들어 “그리고”, “그러나”, “등”, “통해”, “대한”과 같은 일반적 접속어나 조사성 단어는 문서 내 출현 빈도가 높지만 주제 파악에는 기여하지 않는다. 이러한 불용어를 제거함으로써 토픽모델링에서 의미 있는 핵심 단어들이 더 잘 드러날 수 있도록 했다. 불용어 사전은 표준 불용어 리스트를 기본으로 하되, 본 연구의 기사 데이터 특성을 반영해 일부 단어를 추가·수정하였다.

마지막으로 본 연구에서는 복합어 결합 처리를 수행하였다. 이는 형태소 분석 과정에서 의미적으로 하나의 개념을 구성하지만, 기계적으로는 두 개 이상의 단어로 분절되는 경우를 보정하기 위함이다. 예를 들어, “인공지능”이라는 개념이 “인공/NNG + 지능/NNG”으로 분리되거나, “딥페이크”가 “딥/NNG + 페이크/NNG”로 나뉘는 경우가 대표적이다. 이러한 경우 단어를 개별적으로 분석할 경우 본 연구의 주제와 직접 관련된 핵심 개념이 희석되거나 토픽모델링에서 분산되어 나타날 우려가 있다. 따라서 도메인 사전(domain-specific dictionary)을 구축하여 해당 단어들을 하나의 토큰으로 병합하였다. 이 과정을 통해 분석에 필요한 주요 개념 단어들이 보다 안정적으로 추출될 수 있었으며, 동시에 토픽모델링에서 주제어로의 응집도를 높이는 효과를 얻을 수 있다.

1) NNG: Non-inflected Noun(일반명사)

4. 연구 결과

본 연구에서는 수집된 기사 코퍼스를 대상으로 일련의 텍스트 마이닝 기법을 적용하여 인공지능 윤리 담론의 구조적 특징과 변화를 다각도로 분석하였다. 분석 절차는 크게 네 가지 단계로 구성된다. 첫째, 키워드 분석을 통해 연도별 기사에서 빈번하게 출현하는 핵심 용어를 도출하고, 이를 기반으로 담론의 주요 쟁점을 파악하였다. 둘째, N-그램(N-gram) 분석을 실시하여 단일 단어 수준을 넘어, 단어 간 결합 패턴을 탐색함으로써 담론 내에서 자주 함께 언급되는 개념과 의미망을 규명하였다. 셋째, TF-IDF 분석을 수행하여 단순 빈도와 달리 특정 시점의 기사에서 상대적으로 두드러진 핵심어를 식별하고, 이를 통해 연도별 담론의 차별적 특징을 도출하였다. 넷째, LDA 토픽모델링 기법을 적용하여 문서 집합이 형성하는 잠재 주제 구조를 도출하고, 주제 간의 관계 및 연도별 변화 양상을 시각적으로 확인하였다.

이 네 가지 분석 절차를 통합적으로 고찰함으로써, 본 연구는 인공지능 윤리 담론의 빈도적 특성(어떤 단어가 얼마나 언급되는가), 연관성 패턴(어떤 단어들이 함께 언급되는가), 맥락적 중요성(어떤 단어가 특정 시점에 상대적으로 중요한가), 주제 구조(담론이 어떠한 토픽으로 조직되는가)를 다층적으로 조망하였다. 이를 통해 2022년과 2024년 두 시점 사이에서 윤리 담론이 어떻게 확장·심화되었는지를 입체적으로 제시할 수 있었다.

1) 키워드 빈도 분석

키워드 출현빈도는 분석 대상 기사 전반에서 특정 단어가 얼마나 자주 언급되었는지를 정량화한 지표다. 〈표 5〉의 '출현빈도(건)'는 해당 연도 코퍼스 전체에서 해당 단어가 등장한 총 횟수를 뜻하며, '비율(%)'은 전체 토큰 대비 해당 단어가 차지하는 상대적 비중을 나타낸다. 따라서 상위권에 위치한 단어일수록 그 시기 담론에서 중심 의제이거나 배경 개념으로 반복적으로 호출되었음을 의미한다. 반대로 순위가 하락하거나 상위 30위에서 제외된 단어는 해당 시기 담론에서 상대적 주목도가 떨어졌음을 시사한다.

(1) 2022년 키워드의 특징

최상위는 '기술'(1위), '기업'(2위), '산업'(3위)으로, 기술-기업-산업 축이 담론의 정중앙을 형성했다. 이어 '디지털'(4위), '정부'(5위), '데이터'(6위)가 뒤따라 디지털 전환 · 정부 역할 · 데이터 논의가 동시에 활발했음을 보여준다. 그 다음으로 '서비스'(7위), '반도체'(8위), '사업'(9위), '규제'(10위)로, 상용화· 제조 인프라(반도체) · 사업화 · 초기 규제 논의가 함께 언급되었다.

그 다음 순위로는 '개발'(11위), '시장'(12위), '지원'(13위), '투자'(14위), '교육'(15위) 등이 연속적으로 나타나 육성 · 지원 · 투자 · 인력 양성 같은 전환(transition) 준비 의제가 두드러졌다. 20위권에는 '인공지능'(21위), '미국'(22위), '경제'(23위), '대표'(24위), '활용'(25위), '세계'(26위), '정책'(28위), '국가'(29위), '미래'(30위)가 포함되어, 개념·계획·활용 담론이 비교적 넓게 분포했음을 알 수 있다.

〈표 5〉 연도별 키워드 출현 빈도(상위 30 키워드)

출현 순위	2022년			2024년		
	키워드	출현빈도(건)	비율(%)	키워드	출현빈도(건)	비율(%)
1	기술	24259	1.014	기업	38622	0.794
2	기업	20809	0.870	기술	37363	0.768
3	산업	15020	0.628	미국	23600	0.485
4	디지털	13834	0.578	산업	22405	0.461
5	정부	12973	0.542	시장	19762	0.406
6	데이터	12457	0.521	개발	19413	0.399
7	서비스	10809	0.452	규제	19383	0.399
8	반도체	10597	0.443	데이터	19367	0.398
9	사업	10321	0.431	투자	19138	0.393
10	규제	9499	0.397	한국	18925	0.389
11	개발	9307	0.389	인공지능	18865	0.388
12	시장	9300	0.389	반도체	18655	0.384
13	지원	9214	0.385	정부	18433	0.379
14	투자	9072	0.379	서비스	18212	0.374
15	교육	8706	0.364	사업	16444	0.338
16	혁신	8496	0.355	지원	16351	0.336
17	플랫폼	8312	0.347	글로벌	15532	0.319
18	분야	8283	0.346	중국	14431	0.297
19	한국	8274	0.346	세계	14311	0.294
20	정보	8205	0.343	필요	14201	0.292

출현 순위	2022년			2024년		
	키워드	출현빈도(건)	비율(%)	키워드	출현빈도(건)	비율(%)
21	인공지능	8137	0.340	혁신	14006	0.288
22	미국	8109	0.339	정보	13861	0.285
23	경제	8098	0.338	경제	13268	0.273
24	대표	8008	0.335	분야	13158	0.271
25	활용	7621	0.319	정책	13057	0.268
26	세계	7331	0.306	디지털	12984	0.267
27	필요	7090	0.296	국가	12882	0.265
28	정책	6657	0.278	계층	12379	0.255
29	국가	6577	0.275	금융	11968	0.246
30	미래	6540	0.273	문제	11487	0.236

(2) 2024년 키워드의 특징

최상위가 '기업'(1위), '기술'(2위)로 순위가 교차했고, '미국'(3위)이 상위권으로 크게 상승했다(2022년 22위 → 2024년 3위). '산업'(4위), '시장'(5위)이 뒤따르며 시장·경쟁 지향성이 상위로 부상했다. '개발'(6위), '규제'(7위), '데이터'(8위), '투자'(9위), '한국'(10위)이 상위권에 자리한다. 특히 '규제'는 2022년 10위에서 2024년 7위로 순위가 상승했고, '시장'도 12위에서 5위로 상승하여 집행·운영 및 시장 맥락의 중요성이 강화되었음을 보여준다.

'인공지능'(11위)은 2022년 21위에서 상위권으로 상승했고, '반도체'(12위)는 상위권을 유지했다. '정부'(13위)와 '서비스'(14위)는 여전히 중요하지만 상대 순위는 하락(각각 5위 → 13위, 7위 → 14위)했다. 그 다음은 '지원'(16위), '글로벌'(17위), '중국'(18위), '세계'(19위), '필요'(20위)가 포함된다. 2022년 상위 30위에 없던 '글로벌'과 '중국'이 2024년에 새롭게 진입했고, '필요'는 27위에서 20위로 상승했다.

하위권에서는 '혁신'(21위), '정보'(22위), '경제'(23위), '분야'(24위), '정책'(25위), '디지털'(26위), '국가'(27위), '계층'(28위), '금융'(29위), '문제'(30위)가 위치한다. 2022년 4위였던 '디지털'이 2024년에는 26위로 하락했고, 2022년 상위 30위에 없던 '계층', '금융', '문제'가 신규 진입했다.

(3) 2022년과 2024년 키워드 비교

2022년과 2024년 키워드 변화의 특징은 다음과 같이 요약할 수 있다.

- 상위권의 교체: 2022년 1위였던 '기술'이 2024년에는 2위로 내려가고, '기업'이 1위로 올라섰다. 또한 '미국'은 22위에서 3위로 급상승하여 대외 · 지정학적 요소의 부상을 보여준다.
- 시장·집행 키워드의 상승: '시장'(12위 → 5위), '규제'(10위 → 7위), '투자'(14위 → 9위), '지원'(13위 → 16위, 상위권 유지) 등 운영 · 거버넌스 및 경쟁 환경을 가리키는 단어의 상대적 비중이 상승 혹은 상위권을 지속했다.
- 지정학·글로벌 관련어의 진입/상승: 2024년에 '글로벌'(17위), '중국'(18위)이 새로 상위 30위에 들어왔고, '미국'은 최상위로 상승했다. 반면 2022년 26위였던 '세계'는 2024년에도 상위 30위에 남아 있으나(19위) '글로벌'이라는 표현이 별도 키워드로 추가 상위권을 차지했다.
- 개념·전환 키워드의 약화: 2022년 4위였던 '디지털'이 2024년에 26위로 내려갔고, 2022년 상위 30위에 있었던 '플랫폼'(17위), '교육'(15위), '대표'(24위), '활용'(25위), '미래'(30위) 등은 2024년 상위 30위에서 제외되었다. 이는 2024년 담론이 개념적 전환·준비 단계에서 시장·경쟁·거버넌스 중심의 실행 단계로 이동하였다고 해석할 수 있겠다.
- 핵심 기술·산업 축의 지속성: '산업'(3위 → 4위), '반도체'(8위 → 12위), '데이터'(6위 → 8위), '개발'(11위 → 6위) 등은 상위권 내에서 위치를 바꾸며 여전히 높은 빈도를 유지한다. 특히 '개발'의 상승(11위 → 6위)은 구체적 개발·도입 활동의 비중이 커졌음을 보여준다.
- 정책·정부의 상대적 후퇴: '정부'는 5위에서 13위, '정책'은 28위에서 25위로 상대 순위가 낮거나 소폭 상승에 그쳤다. 반면 '규제'는 상위권으로 상승하여, 2024년에는 정책 일반보다 규제 · 집행이 더 빈번히 호명되는 경향을 확인할 수 있다.
- 사회·금융 이슈의 가시화: 2024년 하위권에 '계층'(28위), '금융'(29위), '문제'(30위)가 새롭게 등장했다. 이는 2024년 담론에서 사회적 분배/계층, 금융 부문, 운영상의 문제 인식이 표면화되었음을 시사한다.

결론적으로, 2022년 상위권은 기술-기업-산업-디지털-정부-데이터로 대표되는 전환·

준비 프레임이 강했고, 2024년에는 기업-기술-미국-산업-시장-개발-규제-데이터-투자-한국이 상위권을 구성하여 시장·경쟁·집행(규제·개발)과 지정학적 요소가 동시에 강화되었다. 이는 〈표 5〉에서 보여주는 순위 이동과 신규/제외 키워드의 변화만으로도 확인되는 담론의 중심 이동이라고 해석할 수 있다.

(4) 키워드 변화의 시각화

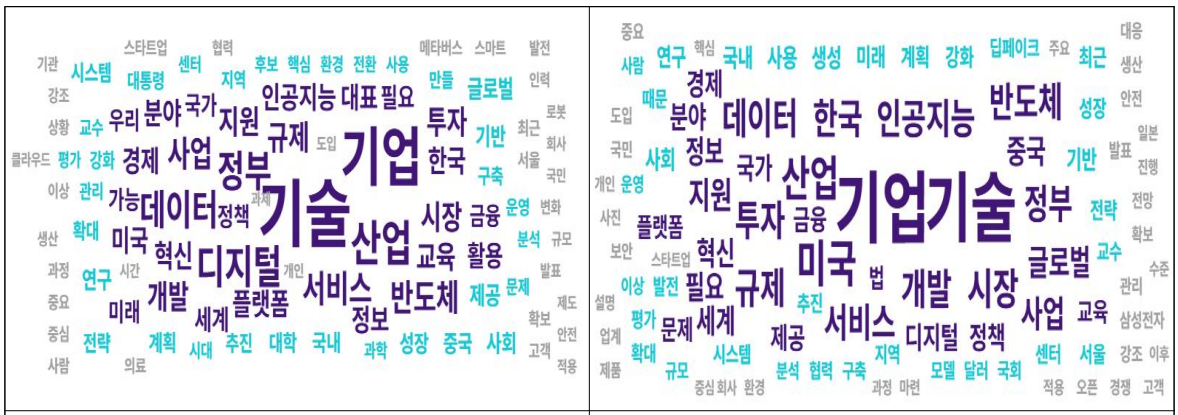
[그림 1]은 2022년과 2024년의 인공지능 윤리 관련 기사에서 출현 빈도가 높은 상위 100개 키워드를 기반으로 작성된 워드클라우드이다. 워드클라우드는 텍스트 데이터의 핵심 단어와 그 상대적 중요도를 직관적으로 보여주는 시각화 기법으로, 단어의 크기는 해당 단어가 기사 내에서 등장한 빈도에 비례한다.

2022년 워드클라우드에서는 “기술”, “기업”, “산업”, “디지털”, “정부”, “데이터”, “서비스”, “반도체” 등이 크게 나타나며, 이는 당시 논의의 중심이 기술 발전과 산업적 응용, 정부의 정책적 개입, 데이터 활용과 서비스 혁신에 맞춰져 있었음을 시사한다. 즉, 2022년은 AI 윤리를 둘러싼 사회적 담론이 주로 산업 경쟁력 확보와 기술혁신의 맥락에서 전개되었음을 확인할 수 있다.

반면 2024년 워드클라우드에서는 여전히 “기업”과 “기술”이 핵심 키워드로 자리하고 있으나, 이와 더불어 “미국”, “규제”, “투자”, “지원”, “글로벌”, “인공지능”과 같은 키워드가 두드러지게 부상하였다. 특히 “미국”과 “글로벌”의 부각은 국제적 경쟁 구도 속에서 AI 윤리가 논의되고 있음을 보여주며, “규제”와 “투자”의 빈도 증가는 정책 · 제도적 대응과 자본 투입의 중요성이 한층 강조되고 있음을 반영한다.

따라서 두 시점의 비교를 통해, 2022년은 주로 기술·산업 중심의 발전적 담론이 강하게 나타난 반면, 2024년에는 국제 경쟁과 규제·투자 환경이 주요 담론으로 확장되었음을 확인할 수 있다. 이는 AI 윤리 담론이 단순히 기술 개발의 차원을 넘어, 글로벌 정책, 규제, 투자 전략 등 복합적이고 다층적인 차원으로 발전하고 있음을 시사하는 결과라 할 수 있다.

[그림 1] 연도별 키워드 출현 빈도 워드클라우드(상위 100 키워드)



(a) 2022년 (b) 2024년

2) N-그램(N-gram) 분석

N-그램 분석은 텍스트 내에서 연속적으로 등장하는 n개의 단어 묶음을 추출하여, 특정 단어가 다른 단어와 어떤 맥락 속에서 사용되는지를 보여주는 기법이다. 본 연구에서는 2-그램(bigram)을 활용하여 두 단어가 함께 출현하는 패턴을 분석하였다. 이를 통해 단어의 단순 빈도뿐만 아니라, 어떤 단어들이 의미적·맥락적으로 결합하여 담론을 형성하는지를 파악할 수 있다.

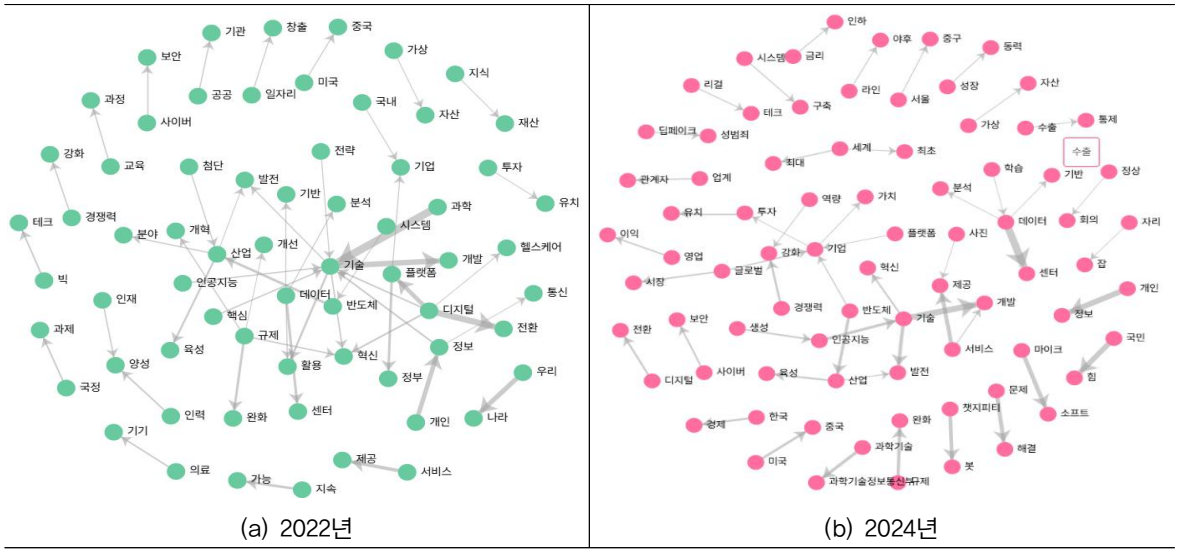
2022년 2-그램 네트워크([그림 2(a)])에서는 "기술-기업", "산업-발전", "정부-정책", "데이터-활용", "서비스-혁신" 등과 같이 산업과 정책 중심의 결합이 뚜렷하게 나타난다. 이는 당시 AI 윤리 담론이 주로 기술적 혁신을 통한 산업 성장, 정부 주도의 규제 및 정책 방향에 초점이 맞추어져 있었음을 보여준다. 특히 "반도체-산업", "디지털-혁신"과 같은 결합은 한국 사회의 특수한 산업적 맥락과 직결되어 있으며, AI 윤리가 기술 경쟁력 담론과 긴밀히 얽혀 있었음을 시사한다.

반면 2024년 2-그램 네트워크([그림 2(b)])에서는 "규제-정책", "데이터-보호", "프라이버시-보안", "윤리-책임", "글로벌-경쟁"과 같은 표현이 새롭게 부상하였다. 이는 2024년 들어 담론의 중심축이 단순한 기술·산업 발전에서 벗어나, 윤리적 책임성과 개인정보 보호, 그리고 국제 경쟁 구도로 확장되었음을 보여준다. 또한 "투자-지원", "산업-혁신" 등의 연결 역시 여전히 강하게 나타나고 있어, 경제적 성장과 사회적 책임

이 병렬적으로 논의되는 양상이 확인된다.

종합하면, 2022년 2-그램 네트워크는 산업 성장·기술 혁신 담론이 강하게 나타난 반면, 2024년에는 윤리·보호·규제·국제 경쟁이 중요한 연결축으로 추가되며 AI 윤리 담론의 스펙트럼이 다층적으로 확장되었음을 확인할 수 있다. 〈표 6〉은 두 연도의 바이그램 네트워크를 비교하여 정리한 것이다.

[그림 2] N-gram 네트워크 비교(n = 2),(상위 50 단어)



(a) 2022년 (b) 2024년

〈표 6〉 2022년과 2024년 bigram 네트워크 비교

구분	2022년 bigram 특징	2024년 bigram 특징	변화 요인
핵심 연결어	기술-기업, 산업-발전, 정부-정책, 데이터-활용	규제-정책, 데이터-보호, 프라이버시-보안, 윤리-책임	담론의 중심이 산업 성장 → 윤리·규제 강화로 이동
산업 중심성	반도체-산업, 디지털-혁신, 서비스-혁신	투자-지원, 산업-혁신, 글로벌-경쟁	산업 담론이 유지되면서도 국제 경쟁 맥락 강화
정책·정부 역할	정부-정책, 규제-기준(제한적)	규제-정책, 글로벌-정책, 책임-정부	규제 및 글로벌 거버넌스 강조
윤리·사회적 가치	상대적으로 약함	윤리-책임, 프라이버시-보안, 데이터-보호	윤리·사회적 가치의 중요성 부각
국제적 맥락	중국, 미국 등 특정 국가 언급 제한적	미국-기술, 글로벌-경쟁, 국제-협력	국제 경쟁·협력 구도의 본격화
전반적 담론 성격	산업 성장과 기술 혁신 중심	산업 성장 + 윤리적 규제 + 국제 경쟁 병행	담론 스펙트럼의 확장

3) TF-IDF 분석

TF-IDF는 특정 단어가 한 문서 집합 내에서 얼마나 중요한지를 측정하기 위한 대표적인 가중치 기법이다(Salton & Buckley, 1988). 이는 단순히 출현 빈도가 높은 단어를 추출하는 방식과는 차별화된다. 빈도 분석은 한 연도 혹은 전체 코퍼스에서 많이 등장한 단어를 보여주지만, 본 연구에서와 같이 '기업', '기술'과 같은 단어처럼 거의 모든 문서에 공통적으로 나타나는 단어는 그 자체로는 특정 시기의 주제적 특징을 설명하기 어렵다. 반면 TF-IDF는 한 문서 안에서 상대적으로 많이 등장하면서도, 동시에 전체 문서 집합에서는 흔하지 않은 단어에 높은 가중치를 부여한다. 이를 통해 단순히 많이 사용된 단어가 아니라, 해당 연도 기사 집합을 특징짓는 차별화된 핵심어를 도출할 수 있다는 점에서 중요한 의미를 갖는다.

본 연구에서는 2022년과 2024년 각각의 기사 코퍼스를 대상으로 TF-IDF 값을 계산한 뒤, 상위 30개 단어를 추출하여 비교하였다(〈표 7〉). 그 결과 두 연도 모두에서 '반도체'가 가장 높은 TF-IDF 값을 기록하였다(2022년: 15,937.87; 2024년: 26,475.062). 이는 단순히 출현 빈도 상위라는 의미를 넘어, 두 시기 모두에서 인공지능 윤리 담론을 다룬 기사들 가운데 반도체가 시대별 차별성을 설명하는 핵심어로 기능했음을 보여준다. 즉, 반도체는 공통적으로 많은 기사에서 언급되었을 뿐 아니라, 특정 시점의 기사 맥락을 특징적으로 드러내는 역할을 했다.

연도별 상위 단어의 구성을 살펴보면 몇 가지 차이가 나타난다. 2022년 상위권에는 '디지털', '교육', '데이터', '금융'이 포함되어 있었으며, '정부', '후보', '대학'과 같은 항목도 상위 10위 내에 위치했다. 이는 당시 인공지능 윤리 논의가 정책적 차원과 사회제도, 특히 교육 및 공공 담론과 밀접히 연결되어 있었음을 시사한다. 반면 2024년에는 '중국'(23,429.083), '금융'(21,016.562), '데이터'(20,424.705), '교육'(19,266.684), '기업'(18,330.853) 등이 상위권을 차지하였다. 이 가운데 특히 '중국'이 두드러진 상승세를 보였는데, 2022년에는 6위(10,262.671)에 머물렀으나 2024년에는 2위로 상승하였다. 이는 인공지능과 윤리 담론이 국내적 차원을 넘어 국제적 경쟁 구도, 특히 미·중 기술 패권 논의와 밀접하게 연결되고 있음을 반영한다.

또한 '기업' 역시 2022년에는 13위(8,450.494)에 불과했으나, 2024년에는 8위(18,330.853)로 올라서며 기업 활동과 산업 적용이 인공지능 윤리 논의의 핵심 맥락으

로 부상했음을 보여준다. 반대로 2022년에 상위 10위권에 포함되었던 '후보', '대학', '정부' 등은 2024년에는 30위 이내에서도 상대적으로 낮은 순위를 차지하거나 상위권에서 이탈하였다. 이는 선거와 정책 담론 중심의 논의가 약화되고, 산업·투자·국제 경쟁이라는 경제적 프레임이 보다 강조되고 있음을 보여준다.

TF-IDF 상위 100개 단어를 바탕으로 시각화한 워드클라우드는 2022년과 2024년의 인공지능 윤리 담론에서 강조되는 주제적 차이를 보다 뚜렷하게 보여준다([그림 3]). 2022년 워드클라우드에서는 '반도체', '디지털', '데이터', '교육', '금융'과 같은 단어들이 중심에 크게 위치하고 있으며, '정부', '산업', '규제', '후보', '대학' 등이 함께 나타난다. 이는 2022년의 인공지능 윤리 담론이 정책적 · 제도적 논의와 교육적 필요성, 디지털 전환과 금융 산업의 변화에 초점을 맞추고 있었음을 시사한다. 특히 '후보'나 '대학'과 같은 단어의 존재는 선거 및 제도적 차원에서 AI 윤리 이슈가 공론화되었음을 의미한다.

반면 2024년 워드클라우드에서는 '반도체', '중국', '금융', '데이터', '교육', '기업', '산업', '투자', '시장'과 같은 단어들이 더 크게 부각되며, '글로벌', '플랫폼', '삼성전자', '미국' 등 국제적 · 산업적 키워드가 두드러진다. 이는 AI 윤리 담론의 중심축이 국내 정책적 · 제도적 논의에서 벗어나 글로벌 경쟁, 산업 적용, 투자와 시장 확산으로 이동했음을 잘 보여준다. 특히 '중국', '미국'과 같은 국가명이 크게 등장한 점은 국제적 기술 패권 구도가 인공지능 윤리 담론의 중요한 배경으로 자리잡았음을 시사하며, '삼성전자'와 같은 특정 기업명이 드러난 것은 산업 현장에서 AI 윤리의 실질적 적용과 기업 책임 논의가 부각되었음을 보여준다.

결론적으로, TF-IDF 분석은 단순 빈도 분석에서 드러나는 공통 단어의 단순 나열을 넘어, 각 연도별로 차별적인 의미를 가지는 단어를 추출하는 데 유효하다. 2022년은 '교육', '정부', '후보', '대학' 등이 상위에 위치하여 제도적 · 정책적 맥락이 뚜렷했으며, 2024년은 '중국', '기업', '투자', '시장'과 같은 항목이 부각되면서 글로벌 경쟁과 산업적 적용이라는 맥락이 더욱 강조되었다. 이는 인공지능 윤리 담론이 짧은 기간 동안 정책 · 제도 중심에서 산업 · 국제 경쟁 중심으로 무게 중심이 이동했음을 보여주는 실증적 근거로 해석할 수 있다.

〈표 7〉 연도별 TF-IDF 순위(상위 30 단어)

순위	2022년		2024년	
	단어	TF-IDF	단어	TF-IDF
1	반도체	15937.87	반도체	26475.062
2	디지털	12023.111	중국	23429.083
3	교육	11946.014	금융	21016.562
4	데이터	11442.838	데이터	20424.705
5	금융	10385.589	교육	19266.684
6	중국	10262.671	투자	19028.284
7	산업	9801.156	미국	18833.277
8	정부	9511.355	기업	18330.853
9	후보	9453.164	딥페이크	17945.018
10	대학	9133.752	산업	17470.516
11	투자	8817.129	사업	17274.232
12	서비스	8722.26	서비스	17205.531
13	기업	8450.494	디지털	16842.024
14	플랫폼	8361.229	한국	16720.422
15	미국	8341.492	시장	16647.426
16	규제	8008	경제	15818.148
17	사업	7961.879	정부	15799.874
18	경제	7610.642	기술	15503.328
19	시장	7386.65	규제	15294.634
20	의료	7349.942	정보	15174.743
21	지원	7331.253	지원	15113.409
22	혁신	7281.922	국회	15002.87
23	기술	7238.639	혁신	14853.878
24	대통령	7215.255	플랫폼	14840.715
25	네이버	7213.48	삼성전자	14719.249
26	연구	7206.263	정책	14633.35
27	대표	7205.123	생성	14360.962
28	정보	7193.091	국가	14331.997
29	한국	7082.399	보안	14208.208
30	정책	6933.836	트럼프	14137.112

[그림 3] 연도별 TF-IDF 비교(상위 100 단어)

(a) 2022년 (b) 2024년

4) LDA 토픽모델링

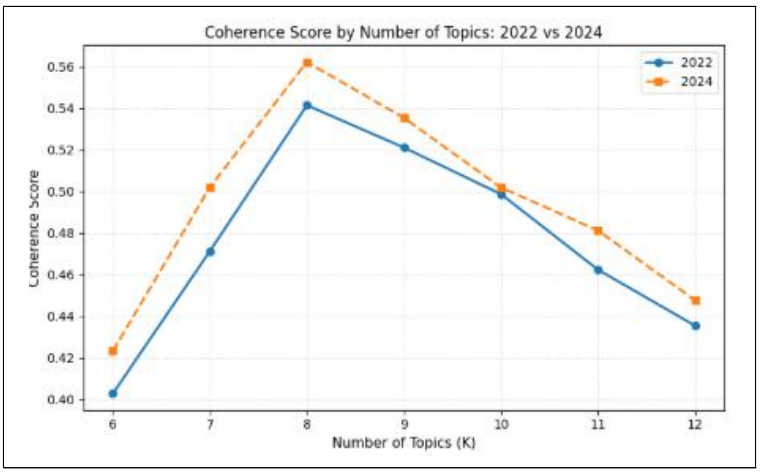
토픽모델링(Topic Modeling)은 대규모 문서 집합에서 잠재적으로 존재하는 주제(토픽)를 자동으로 추출하는 통계적 기법이다. 기본 가정은 각 문서가 여러 주제의 혼합으로 구성되어 있으며, 각 주제는 특정 단어들의 확률 분포로 표현될 수 있다는 것이다(Blei, Ng, & Jordan, 2003). 이를 통해 연구자는 문서 집합을 직접 읽지 않고도, 텍스트 데이터의 주제 구조와 담론의 패턴을 정량적으로 파악할 수 있다. 특히 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)은 가장 널리 사용되는 토픽모델링 기법으로, 문서-주제 분포와 주제-단어 분포를 동시에 추정하여 텍스트의 잠재적 의미 체계를 효과적으로 설명한다(Griffiths & Steyvers, 2004; Blei, 2012).

토픽 수의 결정은 토픽모델링 과정에서 중요한 단계이며, 연구 결과의 신뢰성과 해석 가능성에 직접적으로 영향을 미친다. 본 연구에서는 2022년과 2024년 기사 데이터를 대상으로 토픽 수를 탐색하기 위해 다양한 후보 값을 적용한 후 응집도(coherence)와 복잡도(perplexity) 지표를 종합적으로 고려하였다. 응집도는 특정 토픽 내 주요 단어들의 의미적 유사성과 일관성을 측정하는 지표로, 값이 높을수록 해당 토픽이 의미적으로 잘 응집되어 있음을 의미한다. 반면 복잡도는 모델이 새로운 데이터를 얼마나 잘 설명할 수 있는지를 나타내며, 값이 낮을수록 일반화 성능이 우수하다고 볼 수 있다(김지영·노동조, 2023; 박현정·이태민·임희석, 2025; 이대영·이현숙, 2021).

실제 분석 과정에서 토픽 수를 6개에서 12개까지 변화시키며 비교한 결과, 8개 토픽에서 응집도가 상대적으로 높은 패턴을 확인할 수 있다. 특히 7개 이하의 경우에는 개별 토픽의 범위가 지나치게 광범위하여 해석의 구체성이 떨어졌고, 9개 이상에서는 토픽 간 경계가 모호해지고 유사 토픽이 중복적으로 출현하는 문제가 발생하였다. 따라서 8개 토픽이 가장 해석 가능성이 높으면서도 데이터의 의미 구조를 충실히 반영할 수 있는 최적의 수로 판단되었다. 이와 같은 기준을 적용하여 2022년과 2024년 모두 8개의 토픽을 최종적으로 도출하였다([그림 4]).

도출된 각 토픽의 명명은 토픽별 상위 키워드와 해당 토픽에 높은 확률로 할당된 대표 문서를 종합적으로 검토하여 이루어졌다. 구체적으로, 각 토픽에서 확률값이 높은 상위 단어를 1차적으로 확인하고, 해당 토픽 비중이 높은 대표 언론기사를 추가적으로 분석하였다. 이를 바탕으로 연구자가 해당 토픽이 반영하는 주요 이슈와 의미를 해석하고 토픽명을 부여하였다. 이러한 절차를 통해 정량적 분석 결과와 질적 해석을 결합하여 토픽 해석의 타당성을 확보하였다.

[그림 4] 토픽 수(K)에 따른 coherence score 변화(2022년 vs. 2024년)



(1) 2022년 토픽 구성 및 주요 키워드

2022년 토픽 구조를 보면, 인공지능과 데이터 산업을 둘러싼 초기 담론의 방향성을 뚜렷하게 확인할 수 있다. 우선 토픽 1은 기술, 기업, 반도체, 산업, 정부, 투자가 핵심 키워드로 나타나 산업과 기업 중심의 기술 발전과 국가 차원의 전략적 투자를 강조

하는 흐름을 보여준다. 이는 당시 AI 기술이 경제성장과 산업경쟁력 강화의 핵심 동력으로 인식되었음을 반영한다. 토픽 2는 기술, 산업, 정부, 투자, 미국, 데이터, 경제가 주요 축을 이루며, 디지털 전환의 가속화와 글로벌 경쟁 환경 속에서 국내 산업의 적응 전략을 중심으로 논의되었다. 토픽 3은 기업, 산업, 디지털, 규제, 금융이 결합된 구조로, 플랫폼 산업을 어떻게 규제하고 기업 혁신을 동시에 촉진할 것인가에 대한 균형적 논쟁을 보여준다.

토픽 4에서는 정부, 산업, 인공지능, 교육, 개발이 주요 키워드로 등장해 정부의 정책적 지원과 공공서비스 내 AI 적용을 위한 초기 전략들이 부각되었다. 토픽 5는 기업, 개발, 서비스, 시장, 투자 등이 연결되며 신기술의 상용화와 산업 내 서비스 혁신을 통한 시장 확장에 대한 논의가 주를 이룬다. 이어서 토픽 6은 데이터, 인공지능, 규제라는 조합이 중심을 이루며 개인정보 보호, 데이터 관리 체계, 알고리즘의 공정성과 책임성 등 윤리적 · 규제적 의제들이 본격적으로 다루어졌다. 토픽 7은 기업, 산업, 데이터, 서비스, 글로벌, 성장 등이 키워드로, 글로벌 경쟁과 수출 중심의 성장 전략에 관한 담론을 나타냈다. 마지막으로 토픽 8은 기업, 정부, 산업, 서비스, 규제 등이 함께 등장하며, 미래 혁신 전략과 규제 정합성을 동시에 고려해야 한다는 문제의식을 보여주었다.

〈표 8〉 2022년 토픽별 주요 키워드

토픽	주요 키워드
토픽 1: 산업과 기업 중심의 기술 발전	기술, 기업, 반도체, 산업, 정부, 투자
토픽 2: 디지털 전환과 글로벌 경쟁	기술, 산업, 정부, 투자, 미국, 데이터, 경제
토픽 3: 산업 정책과 규제 환경	기업, 산업, 디지털, 플랫폼, 서비스, 규제, 금융
토픽 4: 정부 정책과 AI 활용 확대	정부, 산업, 인공지능, 경제, 교육, 혁신, 개발
토픽 5: 서비스 혁신과 플랫폼 경제	기업, 개발, 디지털, 서비스, 투자, 필요, 시장, 플랫폼
토픽 6: AI 윤리와 규제 담론	기술, 데이터, 인공지능, 규제, 한국, 분야, 혁신
토픽 7: 글로벌 시장과 디지털 성장	기업, 산업, 데이터, 서비스, 글로벌, 성장, 투자
토픽 8: 정책 · 규제와 미래 혁신 전략	기업, 정부, 산업, 사업, 디지털, 서비스, 경제, 규제

[그림 5(a)]는 2022년의 8개 토픽의 상대적인 근접도와 분리도를 LDA 토픽간 거리지도(Intertopic Distance Map)로 표현한 것이다(Blei et al., 2003; Sievert & Shirley, 2014). 원의 크기는 각 토픽의 상대적 비중을 나타내고, 원 중심 간 거리는 토픽-토픽

간 젠슨-섀넌(Jensen-Shannon) 거리를 다차원척도법(Multidimensional Scaling)으로 2차원에 투영한 값으로, 가까울수록 어휘·의미가 많이 겹치고, 멀수록 서로 다른 단어를 갖는다(Lin, 1991).

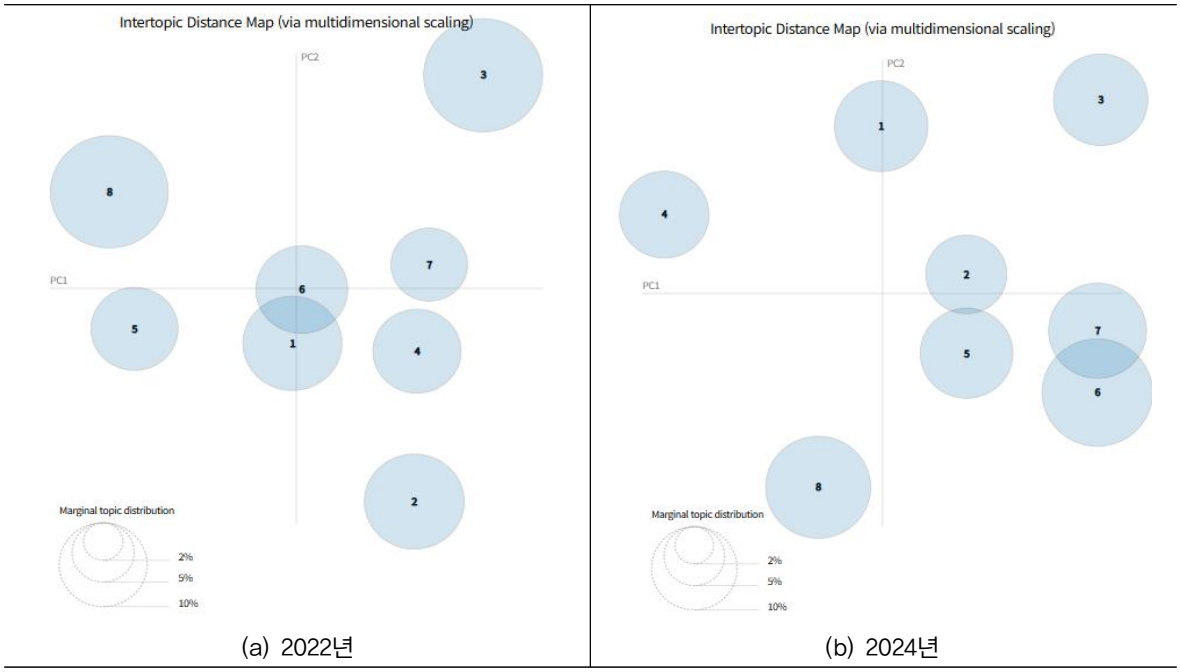
먼저 중앙부에서는 토픽 6(AI 윤리와 규제 담론)이 위치해 있다. 이 토픽은 주변의 토픽 1, 토픽 7, 토픽 4와 가장 짧은 거리를 보이는데, 이는 6번 토픽이 세 토픽과 공통 어휘를 많이 공유하는 연결 허브 역할을 한다는 뜻이다. 실제로 2022년 결과 해석에서 토픽 6을 AI 규제/윤리 담론으로 보았는데, 산업·기업 중심(토픽 1), 글로벌 성장 프레임(토픽 7), 정부 정책·공공 적용(토픽 4)과 모두 접점을 갖는 특성상 중앙에 놓이는 배치가 타당하다. 같은 맥락에서 토픽 1과 토픽 5 간의 거리도 비교적 짧아 산업/기업 중심 담론과 서비스 혁신/상용화 담론이 어휘적으로 맞물려 있음을 보여준다. 즉 2022년 담론의 핵심 축은 토픽 6을 매개로 토픽 1, 토픽 5, 토픽 7, 토픽 4가 서로 이어지는 중앙 클러스터에 형성되어 있다.

반면 상대적으로 분리된 토픽도 있다. 우상단에 위치한 토픽 3은 다른 토픽과의 거리가 가장 멀어 독자적 어휘장을 갖는다. 2022년 해석에서 토픽 3번은 "산업 정책과 규제 환경(플랫폼·금융 등)으로 보았는데, 이처럼 제도·시장 구조 중심의 어휘가 다른 토픽(산업·서비스·정책집행)과 겹침이 적어 가장 외곽에 위치한다. 우하단의 토픽 2 역시 중앙 클러스터와 충분한 거리를 두고 자리하며, 디지털 전환·글로벌 경쟁을 핵심으로 하는 전략·외연 프레임이 다른 토픽의 현장·운영 프레임과 어휘적으로 다름을 시사한다. 좌상단의 토픽 8도 중앙에서 떨어져 있으며, 정책·규제와 미래 혁신 전략이라는 메타 수준의 설계 담론이 다른 토픽들과의 어휘 중첩이 제한적임을 보여준다.

근접도만 놓고 보면, 가까운 쌍은 대체로 토픽 6과 토픽 1, 토픽 6과 토픽 7, 토픽 6과 토픽 4, 토픽 1과 토픽 5로 요약된다. 토픽 6과 토픽 1은 규제/윤리-산업/기업 접점, 토픽 6과 토픽 7은 규제/윤리-글로벌 성장 접점, 토픽 6과 토픽 4는 규제/윤리-정부정책/공공적용 접점, 토픽 1과 토픽 5는 산업/기업-서비스/상용화 접점이다. 이들 조합은 2022년 담론에서 전환의 실행 경로(정책—산업—서비스—규제)가 서로 유기적으로 연결되어 있었음을 시각적으로 확인시켜 준다. 반대로 거리가 먼 조합은 토픽 3과 다른 다수의 토픽들, 토픽 3과 좌측 토픽들, 토픽 8과 우측 토픽들로서 각각 제도·시장구조, 디지털 전환·대외전략, 미래 설계·거버넌스가 다른 실무·운영 중심 토픽과 주제 결이 다름을 의미한다.

원의 크기는 토픽 3과 토픽 2가 상대적으로 큰 편, 토픽 1, 토픽 6은 작은 편으로 관찰된다. 이는 2022년 담론에서 정책·제도/시장구조와 디지털 전환·대외전략이 기사 비중상 크게 나타났고, 반면 산업 현장/윤리·규제 집행은 비중은 다소 작지만 중앙 허브로서 역할을 수행했음을 시사한다. 요컨대 2022년의 토픽 공간은 중앙의 연결 허브(토픽 6)와 주변부의 특화된 담론 축(토픽 3, 토픽 2, 토픽 8), 그리고 산업·서비스 실행 축(토픽 1, 토픽 5)이 공존하는 별 모양의 구조를 이룬다.

[그림 5] LDA 시각화 – LDA intertopic distance map



(a) 2022년 (b) 2024년

(2) 2024년 토픽 구성 및 주요 키워드

2024년 토픽 구조는 2022년에 비해 훨씬 구체적이고 세분화된 논의가 나타난다. 우선 토픽 1은 기술, 기업, 서비스, 금융, 미국, 인공지능, 혁신, 데이터가 주요 키워드로, AI 기술의 산업 적용과 서비스 확산이 핵심적으로 다루어졌음을 보여준다. 토픽 2는 기업, 기술, 미국, 산업, 한국, 서비스, 데이터, 투자가 결합되어 산업 경쟁력과 기술 정책, 그리고 국가 간 기술 경쟁 구도가 함께 논의되었음을 나타낸다. 토픽 3은 기업, 시장, 미국, 데이터, 세계, 규제, 사용자, 중국이 주요 키워드로 등장하여 시장 중심의 글로벌 경쟁과 규제 이슈가 부각되었음을 보여준다. 토픽 4는 데이터, 정부, 산

업, 개발, 인공지능, 글로벌, 규제, 교육이 결합되어 정부 정책과 투자 주도의 AI 발전, 그리고 교육·규제 기반의 제도적 대응이 함께 논의되었음을 의미한다. 토픽 5는 기업, 투자, 미국, 서비스, 경제, 혁신, 교수, 문제가 주요 키워드로 나타나 AI 서비스 혁신과 사회적 활용에 대한 관심을 반영한다. 토픽 6은 기업, 기술, 한국, 글로벌, 반도체, 정부, 중국, 국가가 결합되어 글로벌 거버넌스와 규제 체계, 국가 차원의 전략 논의가 강화되었음을 보여준다. 토픽 7은 기업, 기술, 산업, 데이터, 중국, 글로벌, 서비스, 반도체가 주요 키워드로, 중국을 포함한 글로벌 경쟁 구도와 국가 전략이 부각되었음을 나타낸다. 마지막으로 토픽 8은 기업, 규제, 산업, 서비스, 인공지능, 반도체, 플랫폼이 함께 나타나 플랫폼과 산업 기반의 신산업 전략, 그리고 규제 정합성에 대한 논의가 형성되었음을 보여준다.

〈표 9〉 2024년 토픽별 주요 키워드

토픽	주요 키워드
토픽 1: AI 기술과 산업 적용 확산	기술, 기업, 서비스, 금융, 미국, 인공지능, 혁신, 데이터
토픽 2: 산업 경쟁력과 기술 정책	기업, 기술, 미국, 산업, 한국, 서비스, 데이터, 투자
토픽 3: 시장 중심의 글로벌 경쟁과 규제	기업, 시장, 미국, 데이터, 세계, 규제, 사용자, 중국
토픽 4: 정부 정책과 투자 주도 AI 발전	데이터, 정부, 산업, 개발, 인공지능, 글로벌, 규제, 교육
토픽 5: AI 서비스 혁신과 사회적 활용	기업, 투자, 미국, 서비스, 경제, 혁신, 교수, 문제
토픽 6: 글로벌 거버넌스와 규제 체계	기업, 기술, 한국, 글로벌, 반도체, 정부, 중국, 국가
토픽 7: 중국 · 글로벌 경쟁구도와 국가전략	기업, 기술, 산업, 데이터, 중국, 글로벌, 서비스, 반도체
토픽 8: 플랫폼 · 데이터 기반 신산업 전략	기업, 규제, 산업, 서비스, 인공지능, 반도체, 플랫폼

2024년의 토픽 구조는 2022년에 비해 보다 응집된 클러스터와 명확히 분리된 주변부 토픽이 공존하는 형태로 나타났다([그림 5(b)]). 우측 하단에 토픽 6, 토픽 7, 토픽 5가 서로 근접하여 하나의 밀집 클러스터를 형성하고 있는데, 이는 공통 어휘가 상당히 겹치는 주제 영역임을 시사한다. 실제 토픽 해석에서 이들은 AI 서비스와 사회적 적용, 글로벌 경쟁 구도, 투자와 규제 담론과 같은 주제들이며, 상호 간 밀접하게 교차하면서 산업 적용과 거버넌스, 국제 경쟁 전략이 얽힌 복합적인 논의 구조를 형성하고 있다. 특히 토픽 6과 토픽 7은 거의 겹쳐 나타나, AI 활용 정책과 글로벌 경쟁 전략이 2024년 담론에서 실질적으로 결합된 의제로 다뤄지고 있음을 보여준다.

이와 달리 지도 상단의 토픽 1은 중앙 클러스터에서 일정한 거리를 두고 위치해 있

으며, AI 기술 확산과 산업 적용이라는 주제를 상대적으로 독립적으로 강조한다. 토픽 1이 클러스터와 일정 부분 연결성을 가지면서도 독립된 축을 유지하고 있다는 점은, 기술 확산 담론이 다른 주제들과 상호 연관되면서도 자율적 흐름을 형성하고 있음을 의미한다. 비슷한 맥락에서 우상단에 위치한 토픽 3도 다른 토픽들과 거리를 두고 있으며, 국제 규범·표준화 및 규제 프레임을 중심으로 하는 독자적 어휘장을 갖는다. 이는 글로벌 규제 담론이 2024년 논의에서 상대적으로 독립적이고 차별적인 축으로 자리했음을 보여준다.

좌측의 토픽 4 역시 주변부에 위치하며, 데이터와 규제를 강조하는 어휘로 구성되어 있다. 이는 산업·투자 중심 클러스터와 거리를 두고, 기술 확산과는 별개의 윤리·사회적 권리 보호 담론으로 자리매김하고 있음을 드러낸다. 또한 하단에 위치한 토픽 8은 크기가 상대적으로 크고 다른 토픽과 상당한 거리를 유지하고 있는데, 이는 플랫폼·데이터 기반의 신산업 전략이 독자적으로 중요한 비중을 차지하면서도 기존 산업·규제 중심 담론과 차별화된 위치에 있음을 보여준다. 즉 토픽 8은 2024년 담론에서 별도의 성장 전략 축으로 부상한 것으로 해석된다.

전체적으로 보면 2022년이 중앙 허브(윤리·규제 토픽)를 중심으로 여러 토픽이 방사형으로 연결된 구조였다면, 2024년은 투자·서비스·경쟁 전략 토픽들이 밀집하여 핵심 클러스터를 형성하고, 나머지 기술·산업 적용 확산(토픽 1), 시장 중심의 글로벌 경쟁과 규제(토픽 3), 정부 정책과 투자 주도 AI 발전(토픽 4), 플랫폼 기반 신산업 전략(토픽 8)이 주변부에서 독립적 축을 이루는 구조로 변화했다. 이는 2022년이 담론의 교차와 연계에 방점이 찍혀 있었다면, 2024년은 주제별 세분화와 전문화가 뚜렷해졌음을 시각적으로 보여주는 결과라고 평가할 수 있다.

이처럼 2022년과 2024년의 토픽 구조를 비교해 보면, 2022년 담론은 산업 발전과 규제 정비 같은 거시적이고 준비적인 담론이 중심이었던 반면, 2024년에는 AI 기술의 실제 적용, 글로벌 규제 협력, 사회적 통합, 책임 있는 AI 등 보다 구체적이고 심화된 논의로 전환된 것을 확인할 수 있다. 다시 말해, 초기에는 가능성과 잠재력에 대한 담론이 강했다면, 시간이 지나면서 기술 적용의 현실성과 그에 따른 개인정보 보호, 알고리즘 편향, 책임성, 글로벌 규범 정합성과 같은 사회적·윤리적 과제가 본격적으로 부상한 것이다.

(3) 토픽 구조 비교

토픽간 거리 지도로부터 2022년과 2024년의 토픽 구조를 비교해볼 수 있다. 2022년의 토픽 구조는 비교적 중앙에 여러 토픽이 분포하며 상호 간 거리가 가깝게 나타나는 허브형 구조를 보였다. 특히 토픽 6, 토픽 1, 토픽 4 등은 중앙부에 위치하여 다른 주제들과 교차하며 연결되는 양상을 보였는데, 이는 당시 인공지능 윤리 담론이 아직 서로 구분되지 않은 채 혼합적으로 제시되었음을 시사한다. 다시 말해, 2022년의 논의는 기술 발전, 정부 정책, 산업 적용, 윤리·규제 담론이 서로 뚜렷한 경계를 갖기보다는 중첩되어 논의되는 융합적 담론 구조에 가까웠다. 반면 토픽 3이나 토픽 8처럼 지도 외곽에 위치한 일부 토픽들은 독립성을 갖기는 했지만, 전체적으로는 토픽 간 거리가 크지 않아 주제의 상호 연관성이 강한 상태로 해석된다.

이에 비해 2024년의 지도는 밀집 클러스터와 주변부 토픽의 뚜렷한 분화라는 특징을 보였다. 지도 우측 하단의 토픽 5, 토픽 6, 토픽 7은 서로 겹칠 정도로 근접해 있어, AI 서비스 적용, 글로벌 경쟁 구도, 투자 및 규제 전략이 강하게 결합된 하나의 핵심 클러스터를 형성하였다. 이는 2024년 담론이 산업화·시장 확산과 직접적으로 연결된 의제를 중심으로 수렴하고 있음을 보여준다. 반면 토픽 1은 기술 확산 담론을, 토픽 3은 국제 규범·표준화 담론을, 토픽 4는 데이터 보호 담론을, 토픽 8은 플랫폼 기반 신산업 전략을 각각 독립적인 축으로 형성하며 주변부에 자리했다. 이와 같은 주변부 토픽의 분화는 2022년에 비해 주제별 전문화와 세분화가 진전되었음을 의미한다.

종합하면, 2022년이 윤리와 기술, 산업과 규제가 서로 얽혀 교차적이고 중첩된 담론 지형을 보였다면, 2024년은 일부 주제들이 하나의 밀집 핵심 클러스터로 응집되는 동시에, 나머지 주제들은 명확히 분리된 독립 축을 형성하는 이중 구조로 전환되었다고 평가할 수 있다. 이러한 변화는 인공지능 윤리 담론이 단순히 원칙 수준의 논의에서 벗어나, 산업 적용·사회적 영향·국제 경쟁 전략 등 구체적 의제 중심으로 재편되고 있음을 보여준다. 특히 이러한 구조적 변화는 인공지능 윤리 담론이 기술 수용에 대한 일반적 논의 단계에서 벗어나, 개인정보 보호, 알고리즘 편향, 책임성, 글로벌 규범 정합성 등 구체적 정책 이슈 중심으로 전환되고 있음을 시사한다. 또한 언론 담론에서 반복적으로 등장하는 이러한 쟁점들은 단순한 사회적 우려에 그치지 않고, 점차 제도화 가능한 정책 의제로 연결되고 있음을 보여준다. 이는 인공지능 윤리 논의가 공론장의 담론 수준에서 정책 설계 및 제도화 단계로 이동하고 있음을 의미한다.

5. 결 론

본 연구는 언론기사 데이터를 기반으로 토픽모델링을 활용하여 인공지능 윤리 담론의 구조와 시기별 변화를 분석하였다. 2022년과 2024년 두 시점의 국내 주요 언론 보도를 대상으로, 키워드·N-그램·TF-IDF·LDA 토픽모델링을 결합해 인공지능 윤리 담론의 구조와 변화를 정량적으로 추적하였다. 분석 설계를 통해 빈도적 특성(무엇이 얼마나 언급되는가), 결합 패턴(어떤 용어가 함께 언급되는가), 맥락적 중요성(시점별로 무엇이 '차별적' 핵심어인가), 주제 구조(담론이 어떤 토픽으로 조직되는가)를 다층적으로 파악하였고, 이를 바탕으로 한국 사회의 AI 윤리 담론이 초기 의제화 국면에서 제도화·집행 중심 국면으로 이동하고 있음을 확인하였다. 이는 인공지능 윤리 문제가 추상적 규범 논의에 머무르는 것이 아니라, 사회적 우려를 제도적 장치와 정책 수단으로 연결해야 하는 단계로 진입하고 있음을 의미한다.

이러한 결과는 정책 담론과 사회 담론이 분리된 두 영역이라기보다 상호작용 속에서 재구성되고 있음을 보여준다. 언론 담론에서는 개인정보 보호, 편향, 책임성, 딥페이크와 같은 사회적 우려가 반복적으로 제기되었고, 이러한 쟁점은 시간이 흐르면서 법·제도 정비, 규제 집행, 국제 규범 대응과 같은 정책적 언어와 결합하는 양상을 보였다. 다시 말해, 사회적 우려는 단순한 여론 차원에 머무르지 않고 점차 제도화 가능한 정책 의제로 전환되고 있으며, 이는 인공지능 윤리 담론이 공론장의 논의에서 정책 설계의 논의로 이동하고 있음을 의미한다.

먼저, 데이터 규모의 증가 자체가 담론 확장의 정황증거로 나타났다. 2022년 6,337건(12.4MB) 대비 2024년 14,612건(26.5MB)으로 기사 수와 용량이 모두 두 배 내외 증가했으며, 이는 AI 기술 확산과 더불어 윤리·규제 이슈가 공론장에서 빠르게 확대·심화되었음을 시사한다. 이러한 차이는 단순 양적 증가를 넘어 사회적 담론의 성숙도 지표로 해석할 수 있다.

둘째, 키워드 구조의 중심축이 이동했다. 2022년 상위권은 '기술-기업-산업-디지털-정부-데이터'가 결을 이루었으나, 2024년에는 '기업-기술' 축을 유지하면서 '미국·시장·개발·규제·투자·한국'이 두드러졌다. 특히 '미국'의 급상승(22위→3위)과 '시장·규제'의 상향 배치는 지정학·시장·집행 프레임의 강화를 가리킨다. 반대로 2022년에 상위였던

‘디지털·플랫폼·교육’ 등 전환·준비 성격의 항목은 하락·이탈하였다. 요컨대 2024년 담론은 정책 일반보다 규제·집행과 국제 경쟁 환경을 더 빈번히 나타나는 양상으로 변화하였다. 이와 같은 중심축의 이동은 인공지능 윤리 담론의 초점이 기술 일반에 대한 기대와 우려에서, 실제 적용 과정에서 발생하는 규제·집행·시장 질서의 문제로 이동하고 있음을 보여준다. 특히 2024년 담론에서 기업, 시장, 규제, 글로벌, 중국 등의 키워드가 상대적으로 부각된 것은 인공지능 윤리 문제가 더 이상 선언적 원칙의 차원에 머무르지 않고 산업 경쟁과 제도 운영의 맥락 속에서 구체화되고 있음을 시사한다.

셋째, N-그램(빅그램) 결합 패턴은 담론의 전환을 더욱 명확히 보여준다. 2022년에는 ‘기술-기업’, ‘산업-발전’, ‘정부-정책’, ‘데이터-활용’ 등 산업·정책 중심 결합이 뚜렷했으나, 2024년에는 ‘규제-정책’, ‘데이터-보호’, ‘프라이버시-보안’, ‘윤리-책임’, ‘글로벌-경쟁’이 새롭게 부상하며 윤리·보호·국제 경쟁 축이 강화되었다. 이는 산업 성장 담론이 윤리·보호·거버넌스로 확장되는 과정을 구조적으로 입증한다.

넷째, TF-IDF 분석은 각 연도의 ‘차별적’ 핵심어를 부각시켰다. 두 연도 모두 ‘반도체’가 최상위를 기록해 산업 기반과 AI 윤리 논의의 연결을 공통적으로 드러냈고, 2024년에는 ‘중국·금융·기업·시장’ 등이 상위권으로 이동하여 국제 경쟁·자본·시장 맥락이 2022년에 비해 더 강하게 부각되었다. 이는 단순 빈도 상위어와 달리, 해당 시기를 특징짓는 맥락 특이적 신호가 무엇인지 보여준다.

다섯째, 토픽모델링 결과 두 시점 모두 응집도(coherence)와 복잡도(perplexity)를 종합해 8개 토픽이 최적임을 확인했다. 7개 이하는 과도한 포괄로 해석력이 저하되고, 9개 이상은 유사 토픽의 중복·경계 모호성이 커졌다. 동일한 토픽 수를 적용함으로써 두 시점의 구조적 비교 가능성을 확보하였다.

여섯째, 토픽간 거리 지도는 구조 변화의 결절을 시각적으로 제시한다. 2022년에는 ‘AI 윤리·규제’ 토픽이 중심 허브로서 산업·정부·성장 관련 토픽들과 가까운 거리를 유지하며 교차·중첩된 담론 지형을 이루었고, 2024년에는 산업화·시장 확산·규제 전략이 밀집된 핵심 클러스터가 형성되는 한편, ‘기술 확산’, ‘국제 규범·표준화’, ‘데이터 보호’, ‘플랫폼 기반 신산업’ 등은 독립 축으로 분화했다. 이는 담론이 원칙 수준을 넘어 집행·시장·국제 전략 중심으로 재편되는 이중 구조로 전환되었음을 뜻한다.

종합하면, 2022년의 담론은 기술·산업 중심의 발전 프레임과 정부·정책 주도 준비 국면의 색채가 강했고, 2024년에는 규제·보호·책임과 국제 경쟁·시장 집행이 전면화

되었다. 데이터 규모의 증가, 키워드·결합 패턴의 변화, TF-IDF의 맥락 특이 신호, 그리고 토픽 간 거리 구조의 재편이 같은 방향으로 수렴하는 점에서, 한국 사회의 AI 윤리 담론은 의제화에서 제도화·집행으로 이행하고 있음을 보여준다. 이러한 결과는 향후 정책 설계에서 규제 명료성·책임성·데이터 보호의 구체화와 더불어, 국제 표준·산업 전략과의 정합성을 동시 달성하는 이중 과제가 중요함을 시사한다. 또한 학술적 측면에서는, 언론 담론을 매개로 한 사회적 윤리 인식의 변화를 계량적으로 추적하는 접근이 정책-사회 간 간극을 가시화하는 유효한 방법임을 확인하였다.

다만 본 연구는 몇 가지 한계를 지닌다. 우선, 분석 범위를 2022년과 2024년 두 시점으로 한정함으로써 연속적 추세의 세밀한 변화를 추적하기에는 제약이 있었다. 또한 데이터 출처가 언론 기사에 국한되었기 때문에, 실제 시민 담론이나 학술·정책 현장의 논의가 충분히 반영되지 못했을 가능성이 있다. 또한 본 연구는 언론기사 데이터를 기반으로 토픽모델링을 중심으로 한 담론 구조 분석에 초점을 두었으며, 정서적 반응에 대한 정량적 분석은 연구 범위에 포함하지 않았다. 향후 연구에서는 감성분석 기법을 추가적으로 적용하여, 담론의 구조뿐 아니라 정서적 특성까지 함께 분석하는 확장이 가능할 것이다.

향후 연구에서는 본 연구의 방법론적 틀(키워드·N-그램·TF-IDF·토픽모델링)을 유지하면서, 감성분석 기법(KoBERT 기반 분류기 등)을 결합하여 담론의 정서적 흐름을 계량적으로 분석하는 시도가 필요하다. 아울러 데이터 출처를 학술 논문, 정책 보고서, 온라인 커뮤니티 등으로 확장하면, 인공지능 윤리에 대한 사회적 인식 지형을 보다 종합적이고 정밀하게 파악할 수 있을 것이다.

끝으로, 본 연구의 정책적 시사점은 다음과 같다. 첫째, 언론 담론에서 반복적으로 등장하는 개인정보 보호, 알고리즘 편향, 책임성, 딥페이크 문제는 향후 인공지능 정책 설계에서 우선적으로 고려해야 할 핵심 쟁점임을 보여준다. 둘째, 2024년 담론에서 산업 활용과 글로벌 경쟁 관련 이슈가 강화된 것은 규제와 혁신의 균형을 고려한 정책 접근이 필요함을 시사한다. 셋째, 국제 규범과 국내 제도 논의가 동시에 부상하고 있다는 점은 한국형 인공지능 윤리 거버넌스가 국내 현실과 국제 정합성을 함께 고려하는 방향으로 설계될 필요가 있음을 보여준다. 넷째, 사회적 우려가 제도화 가능한 정책 의제로 전환되고 있다는 점에서, 향후 정책 수립 과정에서는 언론 담론과 시민 수용성에 대한 지속적 모니터링이 병행될 필요가 있다.

참고문헌

김지영, 노동조 (2023), 토픽모델링을 활용한 4차 산업혁명 분야의 국내 연구 동향 분석, 〈한국비블리아학회지〉, 34권 4호, 207-234.

박미정 (2024). 인공지능 시스템과 데이터 프라이버시를 위한 입법 고찰. 〈한국의료법학회지〉, 32권 2호, 33-52.

박현정, 이태민, 임희석 (2025), 토픽 표현의 N-그램 변화에 따른 토픽 모델 평가: 응집도와 다양성을 중심으로, 〈인터넷정보학회논문지〉, 26권 1호, 19-33.

박휴용 (2022). 인공지능 윤리의 주요 쟁점과 필수 고려사항. 〈열린교육연구〉, 30권 6호, 173-201.

이대영, 이현숙 (2021). LDA 토픽모델링의 적정 토픽 수 결정 방법 탐색: 혼잡도와 조화평균법 활용을 중심으로, 〈교육평가연구〉, 34권 1호, 1-30.

임미가 (2021). 인공지능 윤리 연구에 관한 체계적 문헌고찰. 〈윤리연구〉, 1권 135호, 47-66.

홍종현 (2025). 인공지능기본법의 제정에 따른 개인정보보호법과의 관계에 대한 공법적 고찰. 〈법학연구〉, 33권 2호, 167-206.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671-732.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of NAACL-HLT 2019,* 4171-4186.

European Commission. (2019). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on Artificial Intelligence. European Union.

Fast, E., & Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence.*

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, *1*(1), 2-15.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(Suppl. 1), 5228-5235.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99-120.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory, 37*(1), 145-151.

McCombs, M., & Shaw, D. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176-187.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501-507.

Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

OECD. (2019, updated 2024). *OECD AI principles*. OECD.AI. https://oecd.ai/en/ai-principles

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management, 24*(5), 513-523.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63-70.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. Paris: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Zuboff, S. (2020). *The age of surveillance capitalism*. PublicAffairs.

Abstract

Structural Changes in AI Ethics Discourse: A Topic Modeling Analysis of Korean News Articles

Sungyeol Yu

Professor, Department of Computer and Information Engineering, Catholic University of Pusan

This study analyzes the structural characteristics and temporal changes of artificial intelligence (AI) ethics discourse using large-scale media texts. Focusing on 2022 and 2024, it examines how AI ethics-related discourse has been socially constructed and transformed over time. News articles are treated as data reflecting public discourse rather than individual ethical awareness.

The dataset consists of 6,337 articles from 2022 and 14,612 from 2024 collected from Naver News. The analysis employs keyword frequency, n-gram, TF-IDF, and Latent Dirichlet Allocation (LDA) topic modeling, with the number of topics determined based on coherence scores.

The results show that the 2022 discourse exhibited an integrated and overlapping structure centered on regulatory concerns, while the 2024 discourse became more differentiated and specialized, focusing on industrial applications, governance, and global alignment. This indicates a shift from broad normative discussions to issue-specific and policy-oriented discourse.

The study provides empirical evidence of the structural evolution of AI ethics discourse and highlights its implications for policy development and AI governance.

Keywords: AI Ethics, Topic Modeling, Discourse Analysis, News Articles

논문 투고일: 2025. 9. 15.
논문 수정일: 2026. 4. 8.
게재 확정일: 2026. 4. 30.