데이터 윤리 관점에서 데이터 편향에 미치는 요인 탐색

이광수*
대구가톨릭대학교 컴퓨터소프트웨어학부 교수

Exploring the factors affecting data bias from the perspective of data ethics

Gwang-Su Lee*
Professor, School of Computer Software, Daegu Catholic University

요 약 인공지능, 빅데이터, 사물인터넷 기술이 급속히 발전함에 따라 데이터 활용과 관련된 윤리적 문제가 대두되고 있으며, 특히 데이터 편향은 인공지능시스템의 공정성, 신뢰성, 투명성에 직접적인 영향을 미치는 핵심 이슈로 부상하고 있다. 본 연구는 데이터 윤리 관점에서 데이터 편향에 영향을 미치는 주요 요인을 규명하고, 윤리적인 인공지능시스템 설계를 위한 기반을 마련하고자 한다. 이를 위해 전문가 11명을 대상으로 심층 면접을 실시하여 데이터 편향에 영향을 미치는 요인을 도출하였고, 현장성 검증을 위해 인공지능 관련 분야 종사자를 대상으로 설문조사를 실시하였으며, 탐색적 요인분석, 신뢰도 분석, 회귀분석을 통해 요인의 통계적 유의성을 검증하였다. 본 연구에서 제시한 결과가 데이터 수집 및 처리과정에서 윤리 기준 수립, 인공지능 학습 데이터의 품질향상, 편향 방지를 위한 설계 원칙 제시에 기여하며, 인공지능 데이터 거버넌스 체계 강화 및 신뢰적 인공지능 형성에 기초 자료로 활용되기를 기대한다.

주제어 : 데이터 편향, 데이터 윤리, 인공지능 윤리, 데이터 공정성, 인공지능 데이터 거버넌스

Abstract With the rapid advancement of artificial intelligence (AI), big data, and the Internet of Things (IoT) technologies, ethical issues related to data utilization have become increasingly prominent. In particular, data bias has emerged as a critical issue that directly affects the fairness, reliability, and transparency of AI systems. This study aims to identify the key factors influencing data bias from the perspective of data ethics and to establish a foundation for the design of ethical AI systems. To achieve this, in-depth interviews were conducted with 11 experts to derive major factors contributing to data bias, followed by a survey of professionals in AI-related fields for empirical validation. Exploratory factor analysis, reliability testing, and regression analysis were employed to examine the statistical significance of the identified factors. The results of this study are expected to contribute to establishing ethical standards for data collection and processing, improving the quality of AI training data, and proposing design principles to prevent bias. Furthermore, the findings provide fundamental insights for strengthening AI data governance and fostering trustworthy artificial intelligence.

Key Words : Data Bias, Data Ethics, AI Ethics, Data Fairness, AI Data Governance

본 논문은 2025년도 대구가톨릭대학교 학술연구비 지원으로 수행되었음.
*교신저자 : 이광수(gslee@cu.ac.kr)
접수일 2025년 10월 29일 수정일 2025년 11월 27일 심사완료일 2025년 12월 17일

1. 서론

인공지능, 빅데이터, 사물인터넷 기술의 급속한 발전으로 인해 데이터의 생성과 활용은 전례 없이 증가하고 있고, 데이터 기반으로 한 의사결정의 과학화가 중요해지면서 한국 정부에서는 지능정보사회로 나아가는 일환으로 데이터 기반 의사결정을 위한 데이터기반행정법을 마련하여 다양한 산업과 사회 영역에서 데이터 기반 의사결정이 가속화되고 있다. 그러나 이와 함께 데이터 활용의 윤리적 문제에 대한 우려도 커지고 있으며, 그중 데이터 편향은 인공지능시스템의 공정성, 투명성, 신뢰성을 저해하는 핵심 문제로 부각되고 있다[1,3,11].

데이터 편향은 특정 집단이나 속성이 과도하게 반영되거나, 과거의 사회적 불평등이 데이터에 포함됨으로써 발생하며, 결과적으로 불공정하고 차별적인 의사결정을 유도할 수 있다. 예를 들어 대표성 부족, 편향된 출처의 데이터 수집, 주관적인 라벨링, 과거 차별의 반영 등은 알고리즘의 예측 결과에 심각한 왜곡을 초래할 수 있다. 이러한 문제는 단순한 기술적 이슈를 넘어 윤리적 사회적 이슈로 확장되고 있으며, 이를 체계적으로 분석하고 통제할 수 있는 기반 마련이 시급하다.

지금까지의 연구는 알고리즘 수준에서의 편향 완화 기술 개발이나 공정성 지표 설계에 초점이 맞추어져 있었고, 데이터 편향의 구조적 요인을 윤리적 관점에서 정량적으로 탐색한 연구는 상대적으로 부족한 실정이다[8]. 또한 정책 입안자, 개발자, 연구자 등이 실제로 활용할 수 있는 측정 가능하고 검증된 평가 지표의 개발 역시 미흡한 실정이다.

이에 본 연구는 데이터 윤리 관점에서 데이터 편향에 영향을 미치는 주요 요인을 다차원적으로 탐색하고, 데이터 공정성과의 관계를 실증적으로 분석하고자 한다. 이를 위해 OECD AI원칙, 한국지능정보사회진흥원의 데이터 품질 가이드라인 등 국내외 윤리지침 및 선행연구[1,2]를 토대로 데이터 자체 편향, 데이터 수집 편향, 데이터 처리 편향 등 다양한 편향 요인을 개념화하고, 전문가 심층면접 및 설문조사를 기반으로 탐색적 요인분석, 신뢰도 분석, 회귀분석을 수행하였다.

이를 통해 아직 연구가 미약한 분야인 데이터 윤리관점에서 데이터 편향에 초점을 맞추어 측정지표를 제안하였다.

본 연구의 학문적 기여는 다음과 같다. 첫째, 기존 연구들이 기술적 편향이나 특정 알고리즘에 국한된 분석에 집중했던 데 반해, 본 연구는 데이터 윤리라는 상위 개념을 기반으로 편향의 다차원적 원인을 체계적으로 분류하고 측정하였다. 이를 통해 윤리적 인공지능 연구의 이론적 기반을 공고히 하는 데 기여하였다.

둘째, 실제 데이터를 다루는 전문가와 실무자들을 대상으로 한 설문 기반의 정량적 분석을 통해, 실질적인 문제 영역과 인식 격차를 확인하였다. 이는 향후 데이터 기반 AI 시스템의 데이터 거버넌스 설계 및 윤리 기준 수립에 있어 실천적 방향을 제시할 수 있는 중요한 단서가 된다.

셋째, 데이터 편향은 단순한 기술 오류가 아닌 사회적 맥락과 제도적 구조 속에서 재생산될 수 있음을 이론적으로 강조함으로써, AI 기술에 대한 신뢰 확보는 투명한 데이터 설계와 책임 있는 의사결정 구조를 통해 가능하다는 점을 부각시켰다. 이는 AI 기술에 대한 사회적 신뢰 회복과 규범 기반의 지속 가능한 AI 생태계 구축에 실질적 시사점을 제공한다.

따라서 본 연구는 윤리적 인공지능 구현을 위한 학문적 논의뿐 아니라, 공공 및 민간 부문에서 데이터 편향 완화와 AI 거버넌스 체계를 수립하고자 하는 실천적 시도에 있어 중요한 토대를 마련하고 신뢰적 인공지능 형성에 기초 자료로 활용할 수 있을 것이라 판단된다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련 선행연구를 고찰하고, 제3장에서는 전문가 심층 면접을 통해 측정요인 도출 및 연구 가설을 설정한다. 제4장에서는 실증분석 결과를 통해 데이터 편향의 측정지표를 제시하고, 마지막 제5장에서는 결론 및 향후 연구 방향을 제시한다.

2. 이론적 배경

2.1 윤리적 관점

인공지능 기술이 사회 전반으로 확산되면서, AI의 활용 과정에서 발생하는 윤리적 문제를 규범적으로 다루는 연구가 활발히 진행되어 왔다. OECD와 UNESCO 등 국제기구는 인간 존엄성, 기본권 보호, 공정성, 투명성, 설명 가능성을 포함하는 AI윤리 원칙을 제시하고 있으며[1,3], 이를 토대로 각 국가는 정책적·제도적으로 대응을 강화해오고 있다.

국내 연구에서도 So and Ahn(2021, 2022)은 국제적 윤리 가이드라인을 분석하여 투명성, 책임성, 공정성, 통

제성, 안전성의 다섯 가지 원칙으로 AI 윤리를 분류할 수 있는 모형을 제안하였고[4,5], Yoo(2024)은 대화형 AI가 생성하는 문장에 나타나는 사회적 편견을 진단하고, 윤리적 민감성 측정 모형을 개발하여 인공지능이 갖는 잠재적 위험을 정량화하려는 시도를 하였으며[6], Song(2024)은 인공지능 로봇이 사회에 도입될 때 발생할 수 있는 다양한 위험 요인을 측정 지표로 제시하여 윤리적 위험 관리의 필요성을 강조하였다[7]. 이처럼 기존 연구는 AI의 역기능을 예방하기 위해 윤리적 원칙 정립 및 측정 도구 개발에 중점을 두어왔으며, 이는 본 연구의 이론적 토대가 되고 있다.

또한 데이터 활용이 디지털 경제와 AI 혁신의 기반이 됨에 따라 데이터의 수집, 처리, 활용 과정에서 발생하는 윤리 문제가 중요하게 논의되고 있다. 특히 데이터의 공정성, 신뢰성, 프라이버시 보호 등은 데이터 윤리에서 가장 중요하게 다루고 있는 내용이다.

Byun(2020)은 인공지능 시스템의 편향성 문제를 데이터 윤리 관점에서 분석하고 편향성, 객관성, 공정성의 기준을 규정하고, 이를 최소화하기 위한 윤리 인증 기준을 제시하였다[8]. Jung(2020)은 알고리즘의 편향성 구조를 분석하고, 이를 보안하기 위한 사회적 대응 방안을 논의하였으며[9], Lee(2023)는 국내외 데이터 윤리 가이드라인을 종합적으로 분석하여 데이터 활용 환경에서 발생할 수 있는 윤리적 문제를 예방하기 위한 데이터 관리 원칙을 도출하였다[10].

국내·외적으로는 NIA의 AI학습 데이터 품질 관리 가이드라인과 EU의 신뢰할 수 있는 AI가이드라인 등이 발표되고[2,11], 데이터 품질과 윤리적 활용을 제도화하려는 노력이 이어지고 있다. 이러한 논의는 단순히 기술적 품질 관리에 머무르지 않고 사회적 불평등, 권리침해, 데이터 독점 문제로까지 확장되고 있다.

2.2 데이터 편향성

데이터 편향은 AI윤리와 데이터 윤리 연구의 교차 영역에서 중요한 과제로 다루어져 왔다. 편향은 주로 데이터 부족, 샘플링 오류, 편향된 데이터 출처, 라벨링 과정의 주관성, 사회적 불평등의 반영 등에서 발생한다.

Choi and Lee(2023)은 CNN 및 이미지 캡셔닝 모델을 활용하여 데이터 편향이 AI학습 및 예측 정확도에 미치는 영향을 정량적으로 분석하여 데이터 편향이 모델 성능뿐 아니라 예측 결과의 공정성에도 직접적 영향을 미친다는 것을 확인하였고[12], Kim(2021)은 자율지능 시스템에서 발생하는 편향 문제를 탐구하여 공정성 기준과 편향 완화 방법을 제시하였다[13].

Park et al.(2024)은 AI 모델의 편향성이 단지 알고리즘 설계나 학습 과정만의 문제가 아니라, 라벨링 오류, 학습 데이터 불균형 등이 주요 요인이라 지적하면서 편향 완화는 단순히 제거하는 것이 아니라 정확도 및 공정성 간의 균형을 고려해야 한다고 제시하였고[14], Kang et al.(2022)은 자연어 처리의 대규모 데이터가 언어, 문화, 집단 편향을 내포할 수 있음을 윤리적 관점에서 규명하고, 다양성 확보 및 데이터 설계 개선의 필요성을 제시하였으며[15], Jung et al.(2020)은 성 불평등, 교육 수준, 경제적 지표 등 사회적 요인이 편향이 미치는 영향을 분석하여 AI 시스템의 공정성 확보를 위해서 사회, 문화적 거버넌스 차원에서의 접근이 필요하다고 제시하였다[16].

Song et al.(2025)은 KoSBI 데이터셋을 활용해 성별, 연령, 지역 관련 편향을 측정하고 모델별 편향 정도를 계량적으로 분석하여 거대언어모델 개발 시 사회적 편향 완화를 위한 데이터 구성 및 평가 기준의 중요성을 제시하였고[17], Lee et al.(2023)은 ChatGPT의 응답에서 나타나는 편향성인 정치 성향, 지역감정, 외국인 혐어 등 다양한 주제를 대상으로 실증 분석하여 특정 주제에 따라 ChatGPT가 편향된 응답을 생성할 수 있음을 확인하였고, 편향성 완화를 위한 거대언어모델의 윤리적 함의와 공정성 검토의 필요성을 제시하였다[18].

Kim(2021)은 자연어 처리 기술이 내포한 연령 및 지역 편향 문제를 지적하며, 이를 해결하기 위한 대안으로 방언 자료의 활용을 제안하고 윤리적 데이터 구성과 처리 과정의 중요성을 제시하였다[19].

이처럼 기존 연구들은 주로 AI 윤리 원칙 및 측정지표 제시, 데이터 편향의 유형 분류 및 편향 완화 방안 제시에 집중해 왔으며, 데이터 편향이 데이터 공정성 인식에 어떤 영향을 미치는지를 실증적으로 검증한 연구는 상대적으로 부족하였다. 이에 본 연구는 기존 문헌에서 제시된 개념적 논의를 확장하여 데이터 편향 요인을 정량적으로 탐색하고, 데이터 편향 요인별 측정지표를 도출하여, 그 결과를 윤리적 데이터 설계와 거버넌스 체계 마련을 활용하는 것을 목표로 하였다.

〈Table 1〉은 연구 영역별 인공지능 및 데이터 편향 현황을 나타낸 것이다.

〈Table 1〉 Current Status of AI and Data Bias by Research Area

Category	Researcher	Main Content
AI Ethics	OECD[1]	Proposing AI Ethical Principles
	UNESCO[3]	Proposing AI Ethical Principles
	So and Ahn [4,5]	Presentation of AI Ethics Classification Model and Ethical Measurement Indicators
	Yoo[6]	Presentation of an Ethical Sensitivity Measurement Model for Conversational AI
	Song[7]	Presentation of Risk Factor Measurement Indicators for AI Robots
Data Ethics	Byun[8]	Proposal of Ethical Certification Standards and Guidelines for Data Bias, Objectivity, and Fairness
	Jung[9]	Proposal of Social Response Strategies for Mitigating Algorithmic Bias
	Lee[10]	Proposition of Data Management Principles for Preventing Ethical Issues
Data Bias	Choi and Lee [12]	Analysis of the Impact of Data Bias on AI Learning Outcomes
	Kim[13]	Proposed Fairness Standards and Bias Mitigation Methods
	Jung et al.[16]	Identified that AI Bias is Shaped by Social and Technical Factors
	Park et al.[14]	Presentation of Mitigation Strategies for AI Model Bias
	Song and Lee [17]	Presentation of Mitigation Strategies for Social Bias in Large Language Models
	Kim[19]	Presentation of Mitigation Strategies for Bias in Natural Language Processing
	Kang et al.[15]	Presentation of Improvement Directions for Social Bias
	Lee et al.[18]	Presentation of Mitigation Strategies for Social Bias in ChatGPT

3. 연구 방법

본 연구에서는 데이터 윤리 관점에서 데이터 편향에 미치는 측정지표를 도출하기 위해 전문가 심층 면접과 회귀분석을 통해 탐색적으로 접근하였다.

전문가 심층 면접은 2025년 8월 19일부터 8월 31일까지 면접을 3회 진행하여 총 43개의 측정지표를 도출하였다. 또한 도출된 측정지표의 현장성 검증을 위해 인공지능 시스템 운영 및 구축한 경험이 있거나, 데이터 분석 및 인공지능 관련 연구 경험이 있는 사용자를 대상으로 설문조사를 수행하였다.

설문기간은 2025년 9월 1일부터 10월 12일까지 42일 동안 이루어졌으며, 설문기간 동안 총 319부의 설문지가 회수되었고, 기재 내용이 부실한 13부를 제외한 306부가 통계분석의 연구자료로 활용되었다. 분석방법으로 SPSS 20을 사용하여 빈도분석, 탐색적 요인분석과 신뢰도 분석을 실시하였으며, 다중회귀 분석을 통해 데이터 편향 유형별 측정지표를 검증하였다.

3.1 전문가 심층 면접

데이터 편향 유형별 측정지표 도출을 위해 선행 연구를 토대로 인공지능을 이용해 정보시스템 구축 경험이 있는 프로젝트 관리자(3명), 데이터 분석 경험이 있는 데이터 분석자 및 과학자(3명), 인공지능 관련 연구 경험이 있는 교수(2명) 등 총 11명을 대상으로 전문가 심층 면접을 실시하였다. 〈Table 2〉은 전문가 심층 면접 정보를 나타낸 것이다.

〈Table 2〉 Expert Interview Participants' Information

Category		Area of Expertise			Frequency	
		IT	AI and Ethics	Public Sector	N	%
Gender	Male	2	1	6	9	81.8
	Female	–	1	1	2	18.2
	Subtotal	2	2	7	11	100
Age	30 ~ 39	1	–	–	1	9.1
	40 ~ 49	1	1	5	7	63.6
	Over 50 Years	–	1	2	3	27.3
	Subtotal	2	2	7	11	100
Occupation	Consultant	–	1	2	3	27.3
	Project Manager	1	–	2	3	27.3
	Data Analyst and Scientist	1	–	2	3	27.3
	Professor	–	1	1	2	18.1
	Subtotal	2	2	7	11	100
Experience	5 ~ 10	1	–	2	3	27.3
	11 ~ 20	1	1	4	6	54.5
	21 ~ 30	–	–	1	1	9.1
	Over 31 Years	–	1	–	1	9.1
	Subtotal	2	2	7	11	100

전문가 심층 면접 결과 6개 요인으로 데이터 자체 편향, 데이터 수집편향, 데이터 처리편향, 사회적 편향, 알고리즘 편향, 데이터 공정성이 도출되었고, 각 요인별 측정지표는 총 43개 도출되었다.

3.2 변수명 정리 및 연구가설 설정

전문가 심층 면접 결과를 통해 도출된 데이터 편향 유형별 6개 요인의 총 43개 측정지표는 다음과 같다.

데이터 자체 편향은 데이터 자체적으로 내재된 구조적 문제를 나타낸 것으로 8개 지표가 도출되었고, 데이터

수집 편향은 데이터를 수집하는 과정에서 발생하는 문제를 나타낸 것으로 7개 지표가 도출되었으며, 데이터 처리 편향은 데이터 전처리, 정제 과정에서 발생하는 편향을 나타낸 것으로 7개 지표가 도출되었다.

사회적 편향은 사회적 구조와 제도에 의해 반영되는 편향을 나타낸 것으로 7개 지표가 도출되었고, 알고리즘 편향은 데이터 분석 및 모델 학습 단계에서 발생하는 편향을 나타낸 것으로 7개 지표가 도출되었으며, 마지막으로 데이터 공정성은 데이터 및 알고리즘 결과가 특정 집단을 차별하지 않고 균형된 결과를 제공하는 정도를 나타낸 것으로 7개 지표가 도출되었다.

〈Table 3〉은 데이터 편향 유형별 요인 및 측정지표의 변수명을 나타낸 것이다.

〈Table 3〉 Measurement Variables by Factor

Factor	Measurement Variable	Variable Name
Intrinsic Data Bias	Lack of Representativeness	dabi1
	Source Bias	dabi2
	Data Accessibility Bias	dabi3
	Data Quality Bias	dabi4
	Data Generation Bias	dabi5
	Data Duplication and Distortion	dabi6
	Data Sparsity	dabi7
	Lack of Data Timeliness	dabi8
Data Collection Bias	Sampling Error	cobi1
	Sampling Frame Error	cobi2
	Temporal Bias	cobi3
	Collection Instrument Bias	cobi4
	Survey Response/Participation Bias	cobi5
	Data Collection Channel Bias	cobi6
	Data Filtering Bias	cobi7
Data Processing Bias	Labeling Error	pobi1
	Automated Labeling Bias	pobi2
	Missing Data Handling Bias	pobi3
	Outlier Handling Bias	pobi4
	Data Reduction/Amplification Bias	pobi5
	Data Selection Bias	pobi6
	Data Integration Bias	pobi7
Social Bias	Reflection of Social Inequality	sobi1
	Cultural Centrality Bias	sobi2
	Linguistic/Geographic Bias	sobi3
	Group Stereotyping Bias	sobi4
	Economic Bias	sobi5
	Education and Information Disparity Bias	sobi6
	Policy / Institutional Bias	sobi7
Algorithmic Bias	Feature Selection Bias	agbi1
	Class Imbalance	agbi2
	Algorithmic Amplification Bias	agbi3
	Model Overfitting Bias	agbi4
	Interpretation Bias	agbi5
	Hyperparameter Bias	agbi6
	Model Design Bias	agbi7
Data Fairness	Representativeness	fair1
	Balance	fair2
	Outcome Fairness	fair3
	Inclusiveness	fair4
	Decision-making Fairness	fair5
	Bias Control	fair6
	Ethical Standards	fair7

도출된 6개 요인 간의 관계를 파악하면 첫째, 데이터 자체 편향은 대표성 부족, 소스 편향, 접근성 한계 등 데이터 그 자체에 내재된 문제를 의미하고, 이러한 편향은 특정 집단이나 현상을 과소 및 과대 표현함으로써 결과적으로 데이터의 대표성과 결과 공정성을 저해할 것이다.

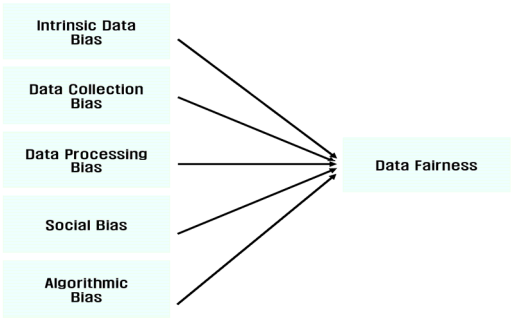
둘째, 데이터 수집 편향은 데이터 수집 과정에서 발생하는 샘플링 오류, 표집 프레임 문제, 수집 시점의 편향성 등으로 특정 집단의 정보를 과소 또는 과대 대표하게 만들고, 데이터의 균형성과 포용성을 저해하며 공정한 분석 결과 도출을 어렵게 만들 것이다.

셋째, 데이터 처리 편향은 데이터 전처리 과정에서의 레이블링 오류, 결측치 및 이상치 처리 기준의 자의성 등의 문제로 분석 결과를 왜곡시키고, 처리 기준이 일관되지 않거나 특정 특성을 의도적으로 제외하는 경우 의사결정 공정성과 결과 공정성이 저해될 것이다.

넷째, 사회적 편향은 데이터가 반영하고 있는 사회적 불평등, 지역 및 언어 차별, 고정 관념 등 사회적으로 누적된 불균형을 의미하고, 이러한 편향은 포용성과 대표성을 훼손하며 공정한 데이터 기반 의사결정을 방해 할 것이다.

다섯째, 알고리즘 편향은 알고리즘 설계 및 학습 과정에서의 특성 선택 왜곡, 클래스 불균형, 모델 과적합 등으로 인해 편향된 결과 도출로 나타날 것이고, 이는 데이터 기반 의사결정의 결과 공정성과 의사결정 공정성을 저해하는 주요 원인이 될 것이다.

이에, 본 연구는 데이터 편향 요인인 5개 요인을 독립변수로 설정하고, 데이터 공정성은 종속변수로 설정하여 데이터 편향 요인이 데이터 공정성에 미치는 영향을 분석하고자 [Fig. 1], 〈Table 4〉와 같이 연구모형과 연구가설을 설정하였다.



[Fig. 1] Research Model

〈Table 4〉 Research Hypotheses

Category	Hypothesis
H1	Intrinsic Data Bias is hypothesized to negatively affect data fairness
H2	Data collection bias is hypothesized to negatively affect data fairness
H3	Data processing bias is hypothesized to negatively affect data fairness
H4	Social bias is hypothesized to negatively affect data fairness
H5	Algorithmic bias is hypothesized to negatively affect data fairness

3.3 설문지 구성

본 연구에서는 가설 검증을 위해 설문 문항을 전문가 심층 면접 결과를 토대로 작성하였다.

데이터 편향 유형별 측정지표 도출을 위한 설문 문항으로 데이터 자체 편향 요인 8개 문항, 데이터 수집 편향 요인 7개 문항, 데이터 처리 편향 요인 7개 문항, 사회적 편향 요인 7개 문항, 알고리즘 편향 요인 7개 문항, 데이터 공정성 요인 7개 문항으로 구성하였고, 평가 척도는 리커드 5점 척도를 사용하였으며 설문 응답자의 일반적 배경을 측정하기 위해 4개 문항으로 구성하였다.

4. 연구분석 및 결과

4.1 인구 통계학적 특성

설문에 참여한 인원은 총 306명이고, 남성 230명, 75.2%이며, 여성 76명, 24.8%가 설문에 응답하였다. 연령대은 40대가 182명 59.5%로 높게 나타났고, 직업은 시스템 개발자가 117명 38.2%로 높게 나타났다. 〈Table 5〉는 인구 통계학적 특성을 나타낸 것이다.

〈Table 5〉 Demographic Characteristics of Survey Participants

Category		Frequency	Proportion (%)	Cumulative Percentage (%)
Gender	Male	230	75.2	75.2
	Female	76	24.8	100
	Subtotal	306	100	100
Age	30 ~ 39	56	18.3	18.3
	40 ~ 49	182	59.5	77.8
	Over 50 Years	68	22.2	100
	Subtotal	306	100	100
Occupation	Consultant	12	3.9	4.9
	Project Manager	53	17.3	21.2
	Data Analyst and Scientist	25	8.2	29.4
	System Developer	117	38.2	67.6
	Administrative Staff	49	16.0	83.7
	Research Staff	22	7.2	90.8
	Teacher	17	5.6	96.4
	Professor	11	3.6	100
	Subtotal	306	100	100
Experience	1 ~ 5	25	8.2	8.2
	6 ~ 10	57	18.6	26.8
	11 ~ 15	72	23.5	50.3
	16 ~ 20	95	31.0	81.4
	21 ~ 25	37	12.1	93.5
	26 ~ 30	16	5.2	98.7
	Over 50 Years	4	1.3	100.0
	Subtotal	306	100	100

4.2 탐색적 요인분석 및 신뢰도 분석

전문가 심층 면접을 통해 도출된 측정지표들 간의 타당성을 검증하기 위해 탐색적 요인분석을 실시하였고, 신뢰도를 분석하기 위해 크롬바하 알파계수(Cronbach α)를 구하였다.

전체 모형의 유의성을 검증하기 위해 Kaiser-Meyer-Olkin (KMO) 및 Bartlett 검정을 실시한 결과 〈Table 6〉과 같이 KMO 값은 0.962로 나타났고, Bartlett 값이 유의확률 0.000에서 유의한 것으로 나타나 요인분석의 적합성이 검정되었다.

〈Table 6〉 KMO and Bartlett's Test Results

KMO Measure of Sampling Adequacy		0.962
Bartlett's Test of Sphericity	Chi-Square	8621.497
	df	496
	p	0.000

데이터 편향 유형별 탐색적 요인분석 및 신뢰도 분석 결과는 〈Table 7〉과 같다.

〈Table 7〉 Exploratory Factor and Reliability Analysis Results by Type of Data Bias

Factor	Variable Name	Factor Loading	Communality	Eigen-value	Explained Variance	Reliability	
						D*	α**
Intrinsic Data Bias	dabi2	0.755	0.750	4.788	14.963	0.899	0.914
	dabi3	0.751	0.692			0.903	
	dabi4	0.729	0.694			0.902	
	dabi1	0.695	0.687			0.899	
	dabi5	0.648	0.668			0.902	
	dabi7	0.607	0.703			0.901	
	dabi6	0.580	0.694			0.902	
Data Collection Bias	cobi5	0.695	0.791	3.892	12.164	0.911	0.928
	cobi4	0.654	0.777			0.910	
	cobi2	0.639	0.778			0.910	
	cobi6	0.613	0.715			0.917	
	cobi3	0.574	0.750			0.919	
	cobi1	0.559	0.711			0.919	
Data Processing Bias	pobi2	0.816	0.844	4.144	12.951	0.900	0.920
	pobi1	0.789	0.810			0.903	
	pobi3	0.742	0.791			0.899	
	pobi7	0.570	0.681			0.907	
	pobi4	0.560	0.722			0.910	
	pobi6	0.551	0.719			0.911	
Social Bias	sobi2	0.817	0.821	4.059	12.686	0.869	0.909
	sobi3	0.759	0.813			0.873	
	sobi1	0.710	0.700			0.911	
	sobi4	0.704	0.772			0.877	
Algorithmic Bias	agbi4	0.689	0.772	2.880	9.000	0.821	0.867
	agbi2	0.676	0.745			0.820	
	agbi1	0.623	0.729			0.821	
	agbi5	0.582	0.626			0.860	
Data Fairness	fair3	-0.746	0.806	4.115	12.860	0.914	0.932
	fair4	-0.726	0.784			0.917	
	fair2	-0.725	0.791			0.920	
	fair1	-0.668	0.769			0.916	
	fair5	-0.660	0.776			0.913	

*D. : Alpha if Item Deleted, **α: Cronbach α

〈Table 7〉의 데이터 편향 유형의 탐색적 요인분석 결과 총 43개 측정지표 중 11개 측정지표가 이론 구조에 맞지 않게 적재되어 최종적으로 데이터 자체 편향 7개 지표, 데이터 수집 편향 6개 지표, 데이터 처리 편향 6개 지표, 사회적 편향 4개 지표, 알고리즘 편향 4개 지표, 데이터 공정성 5개 지표, 총 32개 측정지표가 분석에 이용되어 6개 그룹으로 묶이는 것을 알 수 있었다.

4.3 다중회귀 분석

데이터 자체 편향요인, 데이터 수집 편향요인, 데이터 처리 편향요인, 사회적 편향요인, 알고리즘 편향요인이 데이터 공정성에 부(-)의 영향을 미칠 것이라는 가설을 확인하고자 다중 회귀분석을 실시하였다. 분석 결과는 〈Table 8〉과 같이 나타났다.

〈Table 8〉 Multiple Regression Analysis Results by Type of Data Bias

Dependent Variable	Independent Variable	SE	β	t	p	Tolerance
Data Fairness	Constant	0.099	-	57.769	0.000**	
	Intrinsic Data Bias	0.053	-0.108	-2.04	0.042*	0.387
	Data Collection Bias	0.062	-0.371	-5.809	0.000**	0.267
	Data Processing Bias	0.060	0.023	0.404	0.687	0.337
	Social Bias	0.049	-0.326	-6.933	0.000**	0.494
	Algorithmic Bias	0.058	-0.146	-2.551	0.011*	0.333
	$R = 0.821$, $R^2 = 0.673$, Adjusted $R^2 = 0.668$ $F = 123.735$, p=.000, Durbin-Watson=2.094					

*.p 〈 0.05 **.p 〈 0.01

데이터 자체 편향요인이 데이터 공정성에 미치는 영향은 t값이 -2.04로 유의수준 p〈0.05에서 통계적으로 유의하게 나타나 가설 H1는 채택되었고, 데이터 수집 편향요인이 데이터 공정성에 미치는 영향은 t값이 -5.809로 유의수준 p〈0.01에서 통계적으로 유의하게 나타나 가설 H2는 채택되었다.

데이터 처리 편향요인이 데이터 공정성에 미치는 영향은 t값이 0.404로 나타나 가설 H3는 기각되었고, 사회적 편향요인이 데이터 공정성에 미치는 영향은 t값이 -6.933으로 유의수준 p〈0.01에서 통계적으로 유의하게 나타나 가설 H4는 채택되었으며, 알고리즘 편향요인이 데이터 공정성에 미치는 영향은 t값이 -2.551로 유의수준 p〈0.05에서 통계적으로 유의하게 나타나 가설 H5는 채택되었다. 즉 데이터 자체 편향요인, 데이터 수집 편향요인, 사회적 편향요인, 알고리즘 편향요인이 통계적 유의수준 하에서 데이터 공정성에 부(-)의 영향을 미치는 것으로 나타났다.

회귀모형은 F값이 p=0.000에서 123.735의 수치를 보이고 있고, 회귀식에 대한 R^2 값이 0.673으로 67.3%의 설명력을 보이고 있으며, Durbin-Watson 값은 2.094로 2에 가까워 잔차들 간에 상관관계가 없어 회귀모형은 적합한 것으로 나타났다. 또한 공차한계는 모두

0.1 이상의 수치를 보이기 때문에 독립변수 간 다중공선성에는 문제가 없는 것으로 나타났다.

4.4 연구결과

데이터 편향 유형별 측정지표 개발을 위해 연구 가설을 검증한 결과, 데이터 자체 편향요인, 데이터 수집 편향요인, 사회적 편향요인, 알고리즘 편향요인은 채택되었고, 데이터 처리 편향요인은 기각되었다. 〈Table 9〉은 연구 가설을 검증한 결과를 나타난 것이다.

〈Table 9〉 Hypothesis Testing Results

Category	Hypothesis	Result
H1	Intrinsic Data Bias is hypothesized to negatively affect data fairness	Accepted
H2	Data collection bias is hypothesized to negatively affect data fairness	Accepted
H3	Data processing bias is hypothesized to negatively affect data fairness	Rejected
H4	Social bias is hypothesized to negatively affect data fairness	Accepted
H5	Algorithmic bias is hypothesized to negatively affect data fairness	Accepted

또한 본 연구에서는 타당도 분석시 공통성의 값이 0.4 미만이면 측정지표의 설명력이 부족한 것으로 판단하여 이를 제거하여 내용적 타당도를 확보하고자 하였고[20], 탐색적 요인분석 결과 이론 구조에 맞지 않게 적재된 특정지표는 제거하였으며, 연구가설 검증 결과 가설이 기각된 측정지표도 제거하였다. 즉, 총 43개 측정지표 중 17개 측정지표가 제거되고 26개 측정지표가 도출되었다. 〈Table 10〉은 회귀분석 결과 요인별 측정지표를 나타낸 것이다.

〈Table 10〉 Measurement Variables by Factor Based on Regression Analysis Results

Factor	Measurement Variable	Variable Name
Intrinsic Data Bias	Lack of Representativeness	dabi1
	Source Bias	dabi2
	Data Accessibility Bias	dabi3
	Data Quality Bias	dabi4
	Data Generation Bias	dabi5
	Data Duplication and Distortion	dabi6
	Data Sparsity	dabi7
Data Collection Bias	Sampling Error	cobi1
	Sampling Frame Error	cobi2
	Temporal Bias	cobi3
	Collection Instrument Bias	cobi4
	Survey Response/Participation Bias	cobi5
	Data Collection Channel Bias	cobi6
Social Bias	Reflection of Social Inequality	sobi1
	Cultural Centrality Bias	sobi2
	Linguistic/Geographic Bias	sobi3
	Group Stereotyping Bias	sobi4
Algorithmic Bias	Feature Selection Bias	agbi1
	Class Imbalance	agbi2
	Model Overfitting Bias	agbi4
	Interpretation Bias	agbi5
Data Fairness	Representativeness	fair1
	Balance	fair2
	Outcome Fairness	fair3
	Inclusiveness	fair4
	Decision-making Fairness	fair5

5. 결론

본 연구는 데이터 편향 문제를 데이터 윤리 관점에서 구조화하고, 데이터 편향에 영향을 미치는 주요 요인을 실증적으로 탐색하고자 하였다. 이를 위해 데이터 편향을 여섯 가지 요인인 데이터 자체 편향, 데이터 수집 편향, 데이터 처리 편향, 사회적 편향, 알고리즘 편향, 데이터 공정성으로 구분하고, 각 편향 요인에 대한 총 43개의 세부 측정 지표를 설정하여 설문조사를 실시하였다. 또한, 데이터 공정성을 종속변수로 설정하고, 다중 회귀분석을 통해 요인 간 인과관계를 분석하였다.

분석 결과, 데이터 자체 편향, 데이터 수집 편향, 사회적 편향 그리고 알고리즘 편향이 데이터 공정성에 유의미한 영향을 미치는 것으로 나타났다. 이는 데이터 편향이 특정 단계나 단일 요인에 의해 발생하는 문제가 아니라, 데이터의 생성·수집·활용에 이르는 전 주기적 과정에서 복합적으로 작용한다는 점을 실증적으로 보여준다. 특히 **데이터 수집 편향과 사회적 편향이 상대적으로 높은 영향력**을 보였다는 점은, 데이터 공정성 확보를 위해 기술적 요소뿐 아니라 데이터가 형성되는 사회적·제도적 맥락을 함께 고려할 필요가 있음을 시사한다.

반면 데이터 처리 편향 요인은 데이터 공정성에 통계적으로 유의미한 영향을 미치지 않는 것으로 나타나 가설이 기각되었다. 이러한 결과는 데이터 처리 단계의 중요성이 낮다는 의미라기보다는, 몇 가지 구조적 요인에 기인한 결과로 해석될 수 있다. 학문적 관점에서 볼 때, 데이터 처리 과정에서 발생하는 결측치 제거, 정규화, 이

상치 처리 등의 절차는 비교적 표준화되어 있으며, 이로 인해 응답자들이 인식하는 편향 수준의 변별력이 상대적으로 낮았을 가능성이 있다. 또한 데이터 처리 편향은 데이터 수집 편향이나 알고리즘 편향과 밀접하게 연관되어 있어, 회귀분석 과정에서 다른 요인들에 의해 설명력이 흡수되었을 가능성도 존재한다. 실무적 관점에서는 데이터 처리 과정이 자동화 도구나 기존 프레임워크에 의해 수행되는 경우가 많아, 실무자들이 해당 단계에서의 윤리적 위험을 명시적으로 인식하지 못하는 구조적 한계가 반영되었을 가능성도 고려할 수 있다. 이러한 점에서 데이터 처리 편향은 통계적으로 기각되었으나, 실제 현장에서는 다른 편향 요인과 결합된 잠재적 위험 요인으로 지속적인 관리와 추가 연구가 필요한 영역이라 할 수 있다.

본 연구 결과는 데이터 공정성 확보를 위한 실천적 방안에 대해서도 중요한 시사점을 제공한다. 데이터 자체 편향과 데이터 수집 편향의 영향이 유의미하게 나타난 점을 고려할 때, 데이터 기획 및 수집 단계에서부터 대표성, 접근성, 표집의 적절성을 체계적으로 점검하는 기준이 필요하다. 또한 사회적 편향과 알고리즘 편향의 영향은 데이터 및 알고리즘이 사회적 불평등이나 기존의 편향된 가치 판단을 재생산할 수 있음을 의미하므로, 데이터 활용 및 모델 적용 단계에서 집단 간 결과 차이를 점검하고 설명 가능성을 확보하는 절차가 요구된다.

정책적 측면에서 본 연구는 데이터 윤리 관점에서 데이터 편향을 점검할 수 있는 기초 틀을 제공한다는 점에서 의의를 지닌다. 본 연구에서 도출한 데이터 편향 측정 지표와 분석 결과는 공공 및 민간 영역에서 인공지능 및 데이터 기반 시스템을 도입·운영하는 과정에서 데이터 공정성을 점검하기 위한 가이드라인이나 평가 지표로 활용될 수 있다. 특히 데이터 수집 단계와 사회적 편향 관리에 대한 기준을 제도적으로 명확히 설정함으로써, 데이터 윤리를 고려한 데이터 거버넌스 체계 구축에 기여할 수 있을 것이다.

실무적 측면에서는 본 연구 결과가 조직 내부의 데이터 관리 및 활용 프로세스에 적용될 수 있다. 데이터 기획·수집 단계에서 편향 점검 항목을 포함한 체크리스트를 운영하고, 알고리즘 개발 및 활용 단계에서 공정성 관련 검증 결과를 문서화함으로써, 데이터 윤리 기준을 선언적 수준이 아닌 실제 운영 체계로 전환할 수 있다. 이는 데이터 기반 의사결정의 신뢰성을 높이고, 장기적으로는 인공지능 시스템에 대한 사회적 수용성을 제고하는 데에도 기여할 수 있을 것이다.

본 연구는 데이터 편향에 영향을 미치는 요인을 데이터 윤리 관점에서 탐색하고, 이를 기반으로 데이터 공정성과의 인과관계를 분석하였다는 점에서 이론적·실천적 의의를 지닌다. 그러나 다음과 같은 한계점이 존재한다. 첫째, 설문조사는 데이터 실무 경험이 있는 전문가를 대상으로 이루어졌지만, 응답자 수가 상대적으로 제한적이며 특정 산업군인 시스템 개발자에 편중되어 있다. 이로 인해 연구 결과를 일반화하는데 한계가 있다. 둘째, 데이터 편향은 공공기관, 기업, 연구기관 등 데이터 생성 주체에 따라 다양한 방식으로 나타날 수 있으나, 본 연구는 이를 명확히 구분하지 못하고 통합적으로 접근하였다. 셋째, 데이터 편향이 데이터 공정성에 미치는 영향을 실증적으로 분석하였으나, 종속변수를 데이터 공정성에 한정하여 분석하였다는 한계가 있다. 데이터 편향은 공정성뿐 아니라 투명성, 신뢰성 등 다양한 윤리적 가치와 상호 연관되어 나타날 수 있음에도, 본 연구에서는 이러한 다차원적 효과를 포괄적으로 검증하지 못하였다.

이에, 본 연구에서 지적한 한계점을 극복한다면, 데이터 윤리 관점에서 데이터 편향 유형별 측정지표가 세부적으로 도출될 것이다.

본 연구는 데이터 편향 측정 지표 도출을 위해 회귀분석을 활용하여 실증적으로 분석한 최초의 연구로서, 향후 데이터 윤리 기준 수립, 인공지능 학습데이터 품질 향상 및 데이터 편향 방지를 위한 설계 원칙 제시 등에 기초 자료로 활용되기를 기대한다.

REFERENCES

[1] OECD, "OECD Principles on Artificial Intelligence," 2019.

[2] NIA, "Guidelines for Data Quality Management for AI Learning V1.0," 2021.

[3] UNESCO, "Recommendation on the Ethics of Artificial Intelligence(41 C/73)," SHS, General Conference, 41st, 2021.

[4] S. J. So and S. J. Ahn, "A Study on the Classification Model and Components of Artificial Intelligence Ethical Principles," The Journal of Korean Association of Computer Education, Vol.24, No.6, pp.119-132, 2021.

[5] S. J. So and S. J. Ahn, "A Study on the Artificial Intelligence Ethics Measurement indicators for the Protection of Personal Rights and Property Based on the Principles of Artificial Intelligence Ethics," Journal of Internet Computing and Services, Vol.23, No.3, pp. 111-123, 2022.

[6] K. S. Yoo, "Diagnosing the Ethical Development Stages and Developing an Ethical Sensitivity Measurement Model in Conversational Artificial Intelligence," Doctoral dissertation, Sungkyunkwan University, Seoul, 2024.

[7] H. K. Song, "Development of Measurement Indicators by Type of Risk of AI Robots," Journal of Internet Computing and Services, Vol.25, No.4, pp.97-108, 2024.

[8] S. Y. Byun, "A Study on the Problem of AI Bias in Data Ethics," Journal of Ethics, Vol.1, No.128, pp.143-158, 2020.

[9] W. S. Jung, "Discrimination and Bias of Artificial Intelligence," Human Beings, Environment and Their Future, No.25, pp.55-73, 2020.

[10] C. H. Lee, "Analyzing Data Ethics Guidelines," Journal of Ethics Education Studies, No.70, pp.323-366, 2023.

[11] EU Commission, "High-level expert group on artificial intelligence(AI HLEG)," Ethics Guidelines for Trustworthy AI, 2019.

[12] I. H. Choi and S. W. Lee, "A Quantitative Analysis for Effects of Data Biases on AI Model Outcomes," Journal of the Ergonomics Society of Korea, Vol.42, No.6, pp. 661-673, 2023.

[13] H. E. Kim, "Fairness Criteria and Mitigation of AI Bias," Korean Journal of Psychology: General, Vol.40, No.4, pp.459-485, 2021.

[14] M. J. Park, Y. J. Son and S. M. Chae, "A Study on Mitigation Strategies Based on Identifying Deepening Factors of AI Bias," The Korea Society of Management information Systems Conference, pp. 622-623, 2024.

[15] H. I. Kang, Y. J. Jang, S. Y. Park and H. S. Kim, "Ethical Issues in Natural Language Processing arising from Data," Annual Conference on Human and Language Technology, pp.26-31, 2022.

[16] H. K. Jung, M. Y. Cha and Y. Y. Ahn, "Social Factors Affecting Bias Research in AI Field," Proceedings of the Korean Information Science Society Conference, Vol.2020, No.12, pp.1460-1462, 2020.

[17] S. M. Song and S. B. Lee, "Evaluation of Social Bias Classification Performance of LLMs Using the KoSBI Dataset: A Comparative Analysis of GPT-3.5 Turbo and GPT-4o," Informatization Policy, Vol.32, No.3, pp. 43-72, 2025.

[18] S. Y. Lee, C. J. Park, G. M. Kim and H. S. Lim, "Analysis of Toxicity and Bias of ChatGPT within Korean Social Context," Annual Conference on Human and Language Technology, pp.539-545, 2023.

[19] J. U. Kim, "Ethical Problems and Solutions in Natural Language Processing: Collection of Dialect Data as a Starting Point for Overcoming Age and Regional Biases," Journal of Research Methodology, Vol.6, No.1, pp.157-180, 2021.

[20] J. J. Song, "SPSS/AMOS Statistical Analysis Method," Gyeonggi-do : 21st Century History, 2014.

이 광 수(Gwang-Su Lee) [정회원]

- 2005년 8월 : 한국외국어대학교 전자계산교육전공 (교육학석사)
- 2012년 2월 : 성균관대학교 컴퓨터교육전공 (교육학박사)
- 2005년 9월 ~ 2014년 7월 : 한국사학진흥재단 정보화팀장
- 2014년 7월 ~ 2024년 1월 : 중소벤처기업연구원 전문위원
- 2024년 9월 ~ 현재 : 대구가톨릭대학교 컴퓨터소프트웨어학부 교수

〈관심분야〉
데이터분석, 데이터 윤리, 정보기술아키텍처, 정보교육