

# 인공지능 윤리 강화를 위한 제도적 개선방안 연구\*

차 건 상\*

## 요 약

인공지능 기술의 발전으로 우리의 삶은 혁신적으로 변화하고 있지만, 동시에 윤리적 문제들도 새롭게 대두되고 있다. 특히 알고리즘 및 데이터 편향성에 의한 차별문제, 딥페이크 및 개인정보 유출 문제 등은 인공지능 서비스 확대에 따라 사회적으로 해결해야 할 선결과제라 판단된다. 이를 위해 본 논문에서는 인공지능 윤리 측면에서 인공지능의 개념과 윤리적 이슈를 살펴보고 이를 예방하기 위한 각국의 윤리 가이드라인, 법률, 인공지능 영향평가 제도, 인공지능 인증제도와 인공지능 알고리즘 투명성 관련 기술 현황 등을 살펴보고 인공지능 윤리 강화를 위한 제도적 개선방안을 제시하고자 한다.

## Research on institutional improvement measures to strengthen artificial intelligence ethics

Gun-Sang Cha\*

## ABSTRACT

With the development of artificial intelligence technology, our lives are changing in innovative ways, but at the same time, new ethical issues are emerging. In particular, issues of discrimination due to algorithm and data bias, deep fakes, and personal information leakage issues are judged to be social priorities that must be resolved as artificial intelligence services expand. To this end, this paper examines the concept of artificial intelligence and ethical issues from the perspective of artificial intelligence ethics, and includes each country's ethical guidelines, laws, artificial intelligence impact assessment system, artificial intelligence certification system, and the current status of technologies related to artificial intelligence algorithm transparency to prevent this. We would like to examine and suggest institutional improvement measures to strengthen artificial intelligence ethics.

**Key words** : Artificial intelligence, Ethics, Trustworthy AI, AI ethics, AI ethics guideline

접수일(2024년 05월 10일), 수정일(1차: 2024년 05월 24일),  
게재확정일(2024년 05월 28일)

\* 건양대학교 스마트보안학과

★ 본 논문은 2023학년도 건양대학교 학술연구비 지원에 의하여 이루어진 것임.

## 1. 서 론

### 1.1 연구 배경

이제는 인공지능의 시대라 해도 과언이 아닐 정도로 의료, 금융, 교육, 채용, 예술 등 다양한 분야에서 인공지능이 광범위하게 이용되고 있다. 이러한 활용은 인공지능이 단순히 업무를 보조하는 수준을 넘어 사람을 대신하기도 하는 환경으로 진화하고 있다.[35]

그러나 인공지능이 완벽하지 않다는 점과 더 나아가 인공지능이 사람을 돕거나 지원하는 부분에서 도덕적 지위를 가지고 사람을 대신하여 윤리적 판단을 할 수 있는지에 대해서 다양한 우려의 목소리가 있었다. 이러한 우려는 챗봇 ‘이루다’ 사건(‘21), 영국 대학 입시 알고리즘 차별 논쟁(‘20), 미국 Amazon 채용 차별 논란(‘18), 마이크로소프트 AI 챗봇 차별 발언(‘16) 등에서 현실로 나타났다. 즉 인공지능 편향성과 개인정보 유출문제 등이 인공지능 활용에 앞서 선결해야 할 과제로 제시된 것이다. 본 논문에서는 인공지능 윤리와 관련하여 기존 연구와 국내외의 법제도 및 기술 등을 살펴보고 인공지능 윤리 강화를 위한 제도적 개선 방향을 제시하고자 한다.

### 1.2 관련 연구 및 연구 방법

인공지능 윤리와 관련하여 변순용(2020)은 데이터 관점에서 편향성 문제를 연구하였고 박남기(2020)와 박효용(2022)은 인공지능 윤리의 쟁점을 연구하였다. 조재규(2023)는 국방분야의 인공지능 윤리를 연구하였으며 김익현(2020)은 윤리적 인공지능 개발을 위한 시험평가검증확인(TEVV) 전략을 연구하였다. 또한 홍석한(2022)과 고학수(2021)는 유럽연합의 인공지능 법안 연구를 통해 인공지능 윤리 강화 방안을 제시하였다.

본 논문에서는 인공지능 윤리 강화방안 제시를 위해 2장에서 인공지능과 윤리에 대한 개념을 정의하고 3장에서 인공지능 윤리 관련 법·제도(가이드라인, 인증제도, 윤리영향평가, 윤리 관련 기술) 현황과 문제점을 분석하고 4장에서 인공지능 윤리 강화를 위한 제도적 개선방안을 제시하고 한다.

## 2. 인공지능 윤리 개념과 주요이슈

### 2.1 인공지능 개념 및 분류

인공지능(AI, Artificial Intelligence)은 인간의 학습능력, 추론능력, 지각능력을 인공적으로 모방하려는 컴퓨터 과학의 세부 분야 중 하나라고 정의하고 있다. [36] 이러한 인공지능은 약인공지능(Weak AI), 강인공지능(Strong AI), 초인공지능(Super AI)으로 분류된다. 약인공지능(Weak AI)은 특정 문제 해결을 목적으로 설계된 인공지능으로 대표적으로 구글의 알파고(AlphaGo), IBM의 왓슨(Watson), 아마존의 알렉사(Alexa), 애플의 시리(Siri) 등이 서비스되고 있다. 강인공지능(Strong AI)은 인간과 동일한 수준의 지성을 구현하고자 하는 인공지능으로, 인공일반지능(Artificial General Intelligence)이라고 부르기도 한다. 반면에 초인공지능(Super AI)은 인공지능이 기술적 특이점(Singularity)을 넘어 모든 면에서 인간의 능력을 훨씬 초월하는 인공지능이다. 특히 세계적인 물리학자 스티븐 호킹 박사는 2017년 강인공지능과 초인공지능이 인류에 심각한 위협이 될 것이라는 우려를 표명한 바 있다.[37]

### 2.2 인공지능 윤리 개념과 주요 이슈

윤리(ethics)의 어원은 ‘관습’을 의미하는 그리스어 ethos이다. ethos는 원래 동물들이 서식하는 장소(축사)를 의미하였으나 품습, 개인의 습관 또는 품성을 의미하게 되었다. 오늘날 윤리는 사회적 존재로서 나와 타인이 서로에게 마땅히 지켜야 할 사회적 규범을 의미한다.[38] 윤리의 발전은 윤리학의 출현으로 이어졌으며, 윤리학(Ethics)은 도덕의 원리, 기원, 발달, 본질과 같은 인간의 올바른 행동과 선한 삶을 사회 전반에 걸쳐 규명하는 철학의 주요 분야이다. 윤리학에서 윤리는 크게 규범윤리(Deontological Ethics), 상황윤리(Situation Ethics) 등으로 구분된다. 규범윤리는 행동의 본질적인 성격에 따라 윤리적 가치를 판단하고 특정한 행동이 윤리적으로 옳거나 그르다고 보는 것에 중점을 둔다. 칸트의 의무론이 규범윤리의 한 예로 설명될 수 있다. 반면에 상황윤리는 특정한 윤리적

규칙이나 원리에 따르지 않고, 상황에 따라 윤리적 판단을 내리는 것을 중요시하는 윤리적 접근 방법이다. 대표적으로 인공지능 자율주행 자동차의 고민 중 하나인 트롤리 딜레마의 해결 방안으로 인식하는 공리주의가 대표적이라 할 수 있다. 인공지능 윤리 관점에서 바라보면 칸트의 의무론은 행위의 결과와 상관없이 행위자체가 선이기 때문에 무조건 수행하는 도덕적 명령을 강조하는 규범윤리 보다 행위의 결과를 강조하는 공리주의에 가깝다고 할 수 있다. 인공지능 자율주행 자동차의 윤리 문제를 해결하고자 MIT의 라환교수는 '모럴 머신(Moral Machine)'이라는 플랫폼을 개발하였다. MIT 모럴 머신(Moral Machine)에서는 2,600만 개의 트롤리 딜레마에 관한 판단을 정량화하였으며 4년간 233개 국가에서 수백만명의 사람들이 4천 만건의 응답을 기록한 바 있다.[39]

최근 자율주행 자동차 분야 외에 다양한 곳에서 인공지능 측면에서 윤리적으로 문제가 야기되고 있다. <표 1>를 살펴보면, 인공지능 계획 단계에서 살인을 목적으로 인공지능을 기획하거나 설계단계에서 알고리즘의 공정성 기준이 모호한 부분 오남용사례로 나타나고 있다. 또한 학습단계에서 오염된 데이터를 사용하거나 활용단계에서 인공지능을 악의적으로 사용하는 경우, 할루시네이션, 저작권 침해, 딥페이크, 개인정보 유출 등이 인공지능 활용에 대한 윤리 문제로 제기되고 있다.

<표 1> 인공지능 윤리 이슈

단계	구분	오남용 사례
계획	인공지능 살인	미국 국방부는 머신러닝으로 전투 지역 데이터를 분석해 표적을 포착하거나 전술을 제안하는 군용 AI 프로젝트인 '프로젝트 메이븐(Project Maven)'을 진행[36]
설계	알고리즘 공정성 기준 모호	2016년 재범 예측 인공지능 알고리즘 COMPAS(Northpointe社 개발)에서 흑인 재범 가능성이 높은 것으로 예측되었지만 실제로 재범을 한 사람의 비율은 흑인에 비해 백인 훨씬 높은 것으로 나타남(47.7% vs 28.0%)[40]
학습	오염된(	미국 Amazon은 2014년부터 채

	편향된) 학습데이터 사용	용 부분에 인공지능 이용하였으며 인공지능 학습시 기존 인사 데이터를 이용함. 그런데 남성과 달리 여성지원자는 점수를 깎아 내리는 문제점을 발견함[40]
활용	할루시네이션 (Hallucination)	챗GPT 초기, '조선왕조실록에 기록된 세종대왕의 맥북 프로 던짐 사건'에 대해 알려달라는 질문을 하면 실제로 세종대왕이 맥북 프로를 던진 것 같은 이야기를 지어서 답변함[41]
	저작권 침해	미국 작가들이 챗GPT의 학습 데이터로 자신들의 작품 속 글을 오용한 혐의로 오픈AI社를 고소함[42]
	딥페이크	2023년 펜타곤 옆 건물이 폭발하는 가짜 사진이 트위터를 통해 확산됨. 각종 언론이 사건을 보도하였고 주식시장도 일순간 80p 하락하는 사태가 발생함[43]
	개인정보 / 중요 정보 유출	프랑스의 개발자 리보는 챗GPT에 의해 자신의 개인정보가 유출된 사실을 인지하고 프랑스 개인정보보호 담당기관에 진정서를 제출함[44]

앞으로 인류가 인공지능에 더 많이 의존할수록, 인공지능에 더 높은 윤리적 책임성과 주체성을 요구할 것이다. 특히 <표 1>의 인공지능 윤리 이슈는 약인공지능(Weak AI)에서 발생하는 윤리문제로 강인공지능(Strong AI)과 초인공지능(Super AI) 등과 같이 인공지능에 인간과 유사한 윤리적 주체성을 요구할 경우, 지금까지 경험하지 못한 윤리의 문제가 발생할 수 있다. 다시말해 2.2에서 언급된 윤리의 범위가 '사회적 존재로서 나와 타인이 서로에게 마땅히 지켜야 할 사회적 규범'이 아닌 '사람과 인공지능이 서로에게 마땅히 지켜야 할 규범'으로 확대될 것이다.

이러한 문제점에 대한 대책으로 미시적인 측면에서 각국에서는 인공지능의 편향성을 방지하고 공정성, 투명성 등을 강조하는 가이드라인을 개발·배포하고 있다. 또한, 거시적인 측면에서 강인공지능(Strong AI) 윤리 이슈를 해결하기 위해 유럽위원회(EU)는 세계

에서 처음으로 ‘AI를 활용한 실시간 생체 정보 수집·식별 시스템이 사실상 금지하거나 개인의 특성·행동 데이터로 점수를 매기는 소셜 스코어링 금지’하는 내용을 포함하는 인공지능법안을 승인하여(2024년) 시행을 앞두고 있다.

### 3. 인공지능 윤리 관련 법제도 및 기술 현황

#### 3.1 인공지능 윤리 관련 국내외 가이드라인

국외 인공지능 윤리 관련 가이드라인을 살펴보면, 유럽위원회(EU: European Commission)는 2019년 ‘인공지능 고위급 전문가 그룹’ AI HLEG: High Level Expert Group on AI)을 통해 ‘신뢰할 수 있는 인공지능을 위한 가이드라인(Ethics guidelines for trustworthy AI)’을 발표하였다. 또한, WHO는 의료 관련 윤리 가이드라인을 발표하고 미국은 ‘AI 애플리케이션 규제 가이드라인’을 통해 인공지능 신뢰성 확보를 위한 10대 원칙을 제시한 바 있다. 국내에서는 2020년 관계부처 합동으로 ‘사람이 중심이 되는 인공지능 윤리기준’을 통해 인간성을 위한 인공지능 구현을 위해 개발, 활용 전단계에서 함께 지켜야 할 3대 기본원칙과 10대 핵심요건을 제시한 바 있다. 또한, 2021년에는 개인정보보호위원회에서 AI 개인정보보호 6대 원칙과 함께 법령상 준수해야 할 의무사항을 포함한 ‘AI 개인정보보호 자율점검표’를 발표하였으며 방송통신위원회, (사)한국인터넷자율정책기구, 서울특별시 등에서 ‘생성형 AI 윤리 가이드라인’을 발표한 바 있다. 그리고 국토교통부는 ‘자율주행차 윤리 가이드라인’을 통해 자율주행자동차 행위주체가 참고할 수 있도록 윤리 가이드라인을 발표하였으며 식품의약품안전처는 ‘인공지능 의료기기의 허가·심사 가이드라인’을 발표하였다. 또한, 2024년에는 과학기술정보통신부에서 의료분야, 일반분야 등을 포함한 ‘신뢰할 수 있는 인공지능 개발안내서’를 제작하여 배포하였다.

<표 2> 인공지능 윤리 관련 국내외 가이드라인

구분	가이드라인명	연도
EU	신뢰할 수 있는 인공지능윤리 가이드라인	2019
	신뢰할 수 있는 인공지능 평가목록	2020
	인공지능 규제법률(안)	2021

구분	가이드라인명	연도
	Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for Educators	2022
UNESCO	인공지능 윤리권고	2021
로마교황청	Rome Call for AI Ethics	2020
WHO	Ethics and Governance of Artificial Intelligence for Health	2021
	Generating evidence for artificial intelligence based medical devices	2021
OECD	OECD 인공지능 권고안	2019
미국	AI 애플리케이션 규제 가이드라인	2020
	국방혁신위원회 인공지능 윤리원칙	2020
	인공지능과 알고리즘 이용에 관한 지침	2020
	AI BILL OF RIGHTS	2022
영국	데이터윤리혁신센터	2018
	Guidance on Artificial Intelligence and Data Protection	2020
일본	인간중심 AI의 사회 원칙	2019
중국	차세대 인공지능 관리 원칙	2019
한국	사람이 중심이되는 인공지능 윤리기준 (관계부처 합동)	2020
	인공지능 윤리기준(과학기술정보통신부)	2021
	스캐터랩 AI 챗봇 윤리점검표	2022
	금융분야AI가이드라인(금융위원회)	2021
	인공지능(AI) 개인정보보호 자율점검표 (개인정보보호위원회)	2021
	챗봇 윤리 가이드라인 ((사)한국인터넷자율정책기구)	2023
	생성형 AI 윤리 가이드북(방송통신위원회)	2023
	인공지능 기반 추천 서비스 이용자 보호를 위한 기본원칙(방송통신위원회)	2021
	로봇 윤리 기본 원칙(산업통상자원부)	2019
	서울시 생성형 AI 윤리 가이드라인 (서울디지털재단)	2023
	생성형 AI 저작권 안내서(문화체육관광부)	2023
	자율주행차 윤리 가이드라인(국토교통부)	2020
	인공지능 의료기기의 허가·심사 가이드라인(식품의약품안전처)	2022
	인공지능 학습용 데이터 품질관리 가이드라인(한국지능정보사회진흥원)	2021
	교육분야 인공지능 윤리원칙 (교육과학기술부)	2022
신뢰할 수 있는 인공지능 개발안내서 (과학기술정보통신부)	2024	

<표 2>와 같이 국내외에서는 인공지능의 신뢰성을 확보하고 편향성을 제거하기 위해 의료, 교통, 데이터, 금융, 방송, 생성형 AI, 교육 등 다양한 분야에서 인공지능 윤리 내용을 직간접적으로 포함한 가이드라인을 배포하고 있다. 이렇듯 다양한 여러 분야에서 가이드

라인이 수립되어 배포되고 있다는 점은 인공지능 서비스 확대에 따른 상대적으로 인공지능 윤리 이슈가 다양한 분야에서 발생하고 있거나 발생할 수 있다는 점을 인식하고 있기 때문이다.

### 3.2 인공지능 윤리 관련 국내의 법률 현황

국내 인공지능 관련 법률 현황을 살펴보면, 국내는 2023년 11월 시행된 ‘지능형 로봇 개발 및 보급 촉진법’에 일부 인공지능 관련 조문이 포함되어 있다. 2023년에는 국회에서 ‘인공지능 산업육성 및 신뢰기반 조성에 관한 법률’과 고위험 인공지능 이용사업자의 책무를 규정하는 ‘인공지능책임법’ 등을 발의하였다. 더불어 인공지능 윤리 이슈와 관련된 법률 개정(안)으로는 <표 3>과 같이 인공지능 회사의 알고리즘으로 개인정보가 유출된 경우 개인정보보호위원회가 알고리즘 제출을 요구할 수 있는 개인정보보호법 개정(안), 기업이 채용 분야에 인공지능을 활용할 경우 피 채용인에게 사전 고지의무를 부여하는 채용절차공정화법 개정(안) 등이 있다.

<표 3> 인공지능 윤리 관련 법률 제정 및 개정 현황

법안명	발의일	주요내용
인공지능 산업 육성 및 신뢰기반 조성 등에 관한 법률안	2022.12	인공지능 윤리 규정, 고위험영역 인공지능에 고지의무 부과
인공지능 책임법안	2023.2	고위험인공지능에 대한 개발사업자, 이용사업자의 책무와 이용자의 권리 규정
개인정보보호법 개정안	2023.2	인공지능으로 개인정보 유출시 알고리즘 제출을 요구할 수 있도록 규정
채용절차공정화법 개정안	2023.3	채용과정에서 인공지능 활용시 편향성 사전 검증 및 고지의무 부과

국외에서는 세계적으로 인공지능 규제 내용을 담은 법안은 유럽위원회(EU)의 인공지능법안(EU AI Act) 최초이다. EU의 인공지능법안은 인공지능 시스템을 효과적으로 관리하기 위해 위험 기반 접근(risk-based approach)을 도입하였다. (그림 1)은 EU 인공지능법의 인공지능서비스 위험분류체계로 인공지능 서비스의 위험도에 따라 ‘수용 불가능한 위험(unacceptable risk)’, ‘고위험(high risk)’, ‘제한적 위험(limited ri

sk)’, ‘저위험(minimal risk)’ 등 총 4단계로 분류하고, 위험도에 따라 규제 수준을 차등적으로 적용하고 있다.



(그림 1) EU 인공지능법의 인공지능 서비스 위험분류체계[21]

인공지능 서비스 위험 분류체계에서 ‘인간의 잠재 의식을 조작하거나 기반하는 기술로 의사결정 능력을 손상시켜 특정 행동을 유도하는 시스템’, ‘연령, 장애, 특정한 사회·경제적 상황에서 비롯된 취약성을 악용해 개인이나 집단을 착취하고 행동을 조작하는 시스템’ 등을 ‘수용 불가능한 위험’으로 분류하고 해당 기술에 대해서는 기술의 개발과 시장 출시는 원천적으로 금지하고 있다. 또한 ‘고용이나 근로자 관리와 관련된 시스템’ 등 고위험 인공지능 서비스는 위험 관리 시스템 구축, 데이터와 데이터 거버넌스, 기술문서, 기록보존, 투명성 및 활용자에 대한 정보 제공, 인간에 의한 감독, 정확성, 견고성, 보안 등의 의무 사항으로 규제하고 있다. 반면 ‘제한적 위험’이 있는 인공지능 서비스는 투명성 의무(고지 또는 공개 의무)를 부여하고 있으며 ‘저위험’ 서비스 별도의 규제가 없는 것이 특징이다.

### 3.3 인공지능 윤리영향평가 및 인증제도

국의 인공지능의 윤리영향평가는 UNESCO, ‘AI Ethics Impact Assessment’와 캐나다의 2019년 ‘자동화된 의사결정에 대한 지침(Directive on Automated Decision-Making)’에 따른 공공부문의 자동화된 의사결정 시스템에 대한 알고리즘 영향평가 등이 있다. 또한 유럽위원회(EU)에서도 ‘인공지능 시스템에 대한 인권·민주주의·법치 영향평가’를 제공하고 있으며 미국에서는 인공지능 알고리즘에 대한 영향평가 등이 있다.

국내에서는 2022년 국가인권위에서 인공지능 개발 및 활용 과정에서 발생할 수 있는 인권침해와 차별을 방지하기 위하여 ‘인공지능 개발과 활용에 관한 인권

가이드라인’ 발표하였다. 해당 가이드라인에서는 국가가 인공지능의 개발과 활용에서 인권침해와 차별의 가능성 및 정도, 영향을 받는 당사자의 수, 사용된 데이터의 양 등을 고려하여 공공기관 및 민간기업을 대상으로 인권영향평가를 하도록 권고하고 있다.

국내 인공지능 관련 인증제도는 <표 4>과 같이 인공지능 신뢰성 측면에서 한국표준협회의 ‘인공지능 신뢰성 인증’ 등이 있으며 지능정보화산업협회의 ‘AI 신뢰성 인증’ 등이 있다.

<표 4> 국내 인공지능 관련 인증제도

구분	인증기관	주안점
AI+인증	한국표준협회	성능, 품질
인공지능신뢰성 인증	한국표준협회	신뢰성
ISO/IEC 42001	한국표준협회	AI 경영
AI+X	한국인공지능협회	지능화
AI 신뢰성 인증	지능정보화산업협회	정확성
AI 시스템 품질인증	한국화학융합시험연구원	품질

### 3.4 인공지능 알고리즘 투명성 관련 기술 현황

인공지능 알고리즘 투명성 관련 기술은 IBM, 구글, 페이스북, 마이크로소프트 등 빅테크 기업들이 AI 모델의 편향성을 점검 및 완화하는 도구를 개발하여 무상으로 제공하고 있다. IBM은 2018년 ‘AI 공정성(Fairness) 360’이라는 오픈소스 툴킷을 제공하고 있으며 구글은 편향성 점검을 위한 Fairness Indicator & What-If Toolkit을 제공하고 있다. Microsoft는 2020년에 ‘Fairlearn’이라는 공정성 검증 도구를 발표하였으며 최근에는 미국의 스타트업 기업 중 하나인 트루에라에서 AI와 기계학습(ML) 개발 과정에서 프로그램의 편향성 등을 지속해서 검사하여 개발자에게 결과를 제품을 출시한 바 있다.

## 4. 인공지능 윤리 관련 법·제도 문제점 및 대응 방안

앞서 살펴본 바와 같이 인공지능 윤리 강화를 위해 가이드라인, 법률, 알고리즘 영향평가, 인공지능 신뢰성 인증 및 기술 측면에서 다양한 노력을 기울이고 있

다. 다만 <표 5>와 같이 현재의 문제점을 살펴보면 가이드라인 측면에서는 인공지능 서비스 제공자와 사용자에게 법적 구속력을 가지고 있지 않다는 점이며 의료, 교육, 생성형 AI 등 분야별 가이드라인은 디지털 융복합 환경에서 사업자들에게 혼선이 될 수 있다는 점이다. 또한 법률 측면에서 인공지능 윤리 원칙 등을 제시하는 기본법이 부재하다는 것이다. 특히 이 때문에 부처별 개별법에 인공지능 관련 조항을 추가하거나 개정하는 경우 방향성과 일관성이 문제가 발생하게 된다. 인공지능 영향평가 측면에서는 국가인권위에서 인권 중심의 인공지능 영향평가에 대한 방향을 국가에 권고하고 있으나 <표 1> 인공지능 윤리 이슈를 포괄하는 인공지능 영향평가 제도가 부재하다는 것이다. 또한 인공지능 영향평가 수행 여부도 의무가 아닌 권고 사항으로 되어 있는 것도 문제점이라 할 수 있다. 인공지능 인증제도와 관련하여 인공지능 서비스의 신뢰성을 확보하는 인증제도 활성화 정책이 미흡하며 특히 고위험영역 등의 인공지능 윤리 인증제도가 마련되어 있지 않다는 문제점으로 지적된다. 기술적 측면에서는 인공지능 편향성 관련 기술, 설명 가능한 AI 기술 등에 대한 중장기 로드맵 등을 통한 지속적인 기술개발이 미흡하다고 할 수 있다.

<표 5> 인공지능 윤리 관련 법·제도 문제점

구분	문제점
가이드라인	- 인공지능 서비스 개발자, 이용사업자, 이용자에게 법적 구속력을 가지고 있지 않음 - 디지털 융복합 환경에서 분야별로 가이드라인 혼재
법률	- 인공지능 윤리 등의 원칙을 제시하는 기본법 부재 - 인공지능 윤리 관련 개별법 개정시 일관성 및 방향성 부재
영향평가	- 국가인권위의 인공지능 인권영향평가의 경우 인권 이슈가 핵심으로 <표 1> 인공지능 윤리 이슈를 포괄하는 인공지능 영향평가 제도 부재 - 인공지능 영향평가의 경우 의무가 아닌 권고사항으로 되어 있음 - 인공지능 영향평가에 대해 사업자의 참여를 적극적으로 유도하는 정책 부재
인증	- 인공지능 인증제도 활성화 정책 미흡 - 특히 고위험영역 등의 인공지능 인증 제도 미흡
기술	- 국내 인공지능 편향성 관련 기술 개발 미흡

이러한 문제점 해결을 위해 무엇보다 중요한 부분은 첫째, 인공지능 윤리 관련 기본법 제정이 우선되어야 한다. 이는 사회 규범 대부분이 법률로 유지되고 있으며 이러한 규범적 측면에서 인공지능의 역할이 확대되고 있기 때문이다. 따라서 구속력이 결여된 가이드라인 보다 다소 포괄적이라 할지라도 인공지능 윤리 관련 법률 제정이 무엇보다 필요하다고 할 것이다.

최근 미국에서도 애나 에슈(Anna Eshoo), 돈 베이어(Don Beyer) 하원의원이 ‘AI 과운데이션 모델 투명성 법안(AI Foundation Model Transparency Act)’을 발의하여 추진하고 있다. 다만 국내에서도 인공지능 윤리 관련 법안 추진시 고려해야 할 사항으로 투명성 확보를 위해 인공지능 알고리즘 제출 등과 같은 조항이나 설명 가능한 AI의 범위를 정하는 데 있어서 기업의 영업기밀 침해나 지적재산권 등을 침해하지 않는 범위가 명확히 논의되어야 할 것이다.

둘째, 인공지능 윤리 관련 기본법을 근간으로 인공지능 서비스의 자율성을 높이고 공정성, 투명성, 신뢰성과 같은 윤리적 규범을 준수할 수 있는 평가 및 인증 제도의 확립이 필요하다. 특히 인공지능 영향평가, 인증제도에 대한 기업 참여 유도를 위한 인센티브(세제혜택, 연구개발 지원금 확대, 정부주도 참여 프로젝트 혜택, 기술이전, 수출 및 해외진출 지원 등)에 대한 정책적 지원이 필요하다.

셋째, 인공지능 윤리 기본법을 충족하고 인공지능 서비스 기업의 자율성을 유도하는 실효성 있는 가이드라인 마련이 필요하다. 특히 가이드라인은 범의무사항을 준수하고 인공지능 윤리원칙을 지향하는 일관된 방향성이 중요한 부분이다.

넷째, 인공지능 윤리 기본법과, 인증제도 및 가이드라인을 충족할 수 있는 지속적인 기술 개발이 필요하다. 특히 편향성 제거를 위한 기술 이외에 인공지능 학습에 필요한 양질의 데이터 관리방안에 대한 대책 수립도 병행되어야 할 것이다.

앞서 살펴본 바와 같이 인공지능 윤리를 강화하기 위한 제도적 개선방안으로 법률제정, 영향평가 및 인증 활성화, 기본법을 충족한 실효성 있는 가이드라인과 인공지능 편향성 제거 관련 기술(알고리즘 및 데이터)에 대한 지속적인 개발이 체계적으로 추진되어야 한다.

## 5. 결 론

정부에서는 디지털플랫폼 정부를 핵심 국정과제로 추진하고 있다. 디지털플랫폼 정부는 AI·데이터 기반의 정부서비스 혁신모델로 이중 AI 가상비서 서비스를 챗봇단계에서 지능비서서비스, 감성인지서비스 까지 추진하겠다는 전략이다. 교육부에서도 AI 디지털 교과서를 통해 학생에 대해 AI 기반으로 진단·분석을 추진하여 1:1 맞춤형 교육을 추진하겠다는 계획이다. 이렇듯 행정, 공공, 기업 등 사회 전 분야에 걸쳐 인공지능 서비스가 확대되고 있으며 사용자의 요구도 늘어나고 있다. 앞으로 인공지능 기술의 발전과 보급으로 우리 사회와 삶에 더 많은 영향을 미치게 되면서, 윤리적 문제들이 더욱 뚜렷해질 것으로 예상된다. 더불어 우리 앞에 다가올 강인공지능(Strong AI)과 초인공지능(Super AI) 시대에 대한 윤리적 대비도 필요한 부분이라 할 것이다. 이러한 측면에서 인공지능의 발전을 도모하면서 인공지능 서비스의 부작용을 최소화할 수 있는 법제도, 기술 등의 생태계(거버넌스) 마련을 위한 노력이 시급하다 할 것이다.

## 참고문헌

- [1] 변순용, ‘데이터 윤리에서 인공지능 편향성 문제에 대한 연구’, 2020, 윤리연구
- [2] 박남기, ‘인공지능과 윤리적 이슈’, 2020, 언론정보연구.
- [3] 신소현, AI 알고리즘의 편향성 및 윤리적 이슈에 대한 고찰.
- [4] 박휴용, ‘인공지능 윤리의 주요 쟁점과 필수 고려사항’, Vol.30. No.6, 2022, 열린교육연구.
- [5] 이해령, ‘초등학생 인공지능 윤리 인식 척도 개발’
- [6] 김경주, ‘인공지능윤리교육의 국내 연구 동향 분석’, 논문지 제19권 제4호-2023년 12월, 디지털산업정보학회.
- [7] 조재규, ‘국방 인공지능 윤리에 대한 연구’, 2023년 봄 (39-1) 통권 제139호 pp. 213-240, 국방정책연구.
- [8] 김정남, 방은찬, 장희영, 송선영, ‘인공지능(AI)의 학습용 데이터 윤리 프레임워크 개발에 관한 연구’, 2023, vol.1, no.140, pp. 283-309, 윤리연구.

- [9] 김경동, '새로운 노멸, 인공지능(AD)과 윤리규범에 관한 소고', 제23권 제2호(통권 90호), 2023년, 법학연구.
- [10] 김익현, '윤리적 인공지능 개발을 위한 시험평가검증 확인(TEVV) 전략', 2020, 한국국방기술학회 논문지.
- [11] 심지원, 이은재, 김문정, '인간의 윤리로서 인공지능 윤리', 2022, 철학·사상·문화.
- [12] 홍석한, '유럽연합 '인공지능법안'의 주요 내용과 시사점', 2022. 제38호, 유럽헌법연구.
- [13] 고학수, 임용, 박상철, '유럽연합 인공지능법안의 개요 및 대응방안', 2021, DAIG.
- [14] 배영임, 김유나, '인공지능의 윤리적 쟁점에 관한 탐색적 연구', 2022, 경기연구원 기본연구.
- [15] 김병주(2017). '인공지능 윤리의 필요성과 국내외 동향', 2017, 한국통신학회지(정보와통신).
- [16] 교육부, '교육분야 인공지능 윤리원칙(시안)', 2022.
- [17] 한국교육학술정보원, '윤리적 요소를 통합한 고등학생용 인공지능 교육 프로그램 설계 연구', 2023.
- [18] 한국교육학술정보원, '인공지능(AI)의 학습용 데이터 윤리 가이드라인에 대한 연구', 2023.
- [19] 한국데이터산업진흥원, '2023 데이터산업 백서', 2023.
- [20] 한국교육학술정보원, '인공지능 윤리영향평가의 교육분야 적용 가능성 탐색', 2023.
- [21] 한국언론진흥재단, 'EU AI법(AI act)의 주요 내용과 미디어 업계 영향, 시사점', 2024년 2호.
- [22] 국가인권위원회, '인공지능 개발과 활용에 관한 인권 가이드라인', 2022.
- [23] 관계부처 합동(2020. 12. 23.). "사람이 중심이 되는 '인공지능(AI) 윤리기준'".
- [24] 국회입법조사처, '인공지능 시대 도래에 따른 AI 입법수요 및 과제 연구', 2023.
- [25] 방송통신위원회, '생성형 AI 윤리 가이드북', 2023.
- [26] 한국인터넷자율정책기구, '챗봇 윤리 가이드라인', 2023.
- [27] 개인정보보호위원회, '인공지능(AI) 개인정보보호 자율점검표', 2021.
- [28] 과학기술정보통신부, '신뢰할 수 있는 인공지능 개발안내서(안), 2024.
- [29] 금융위원회 보도자료, '금융분야 인공지능(AI) 가이드라인', 2021.
- [30] 국토교통부, '자율주행자동차 윤리가이드라인', 2020.
- [31] 브레인리포트, '인공지능 기술보다 중요한 인공지능 윤리' 2023.
- [32] 한국인공지능윤리학회, '생성AI 신뢰성 평가 체계의 마련과 신뢰성 인증(TTA 발표자료)', 2024.
- [33] <https://horizon.kias.re.kr/17815/>.
- [34] <https://www.joongang.co.kr/article/25235203#home>.
- [35] [https://h21.hani.co.kr/arti/economy/economy\\_general/5110.html](https://h21.hani.co.kr/arti/economy/economy_general/5110.html).
- [36] <https://www.techtube.co.kr/news/articleView.html?idxno=4372>.
- [37] [https://www.hani.co.kr/arti/international/international\\_general/667299.html](https://www.hani.co.kr/arti/international/international_general/667299.html).
- [38] <https://www.edulabkorea.com/reference/general.php?ptype=view&idx=365&page=25&code=general>.
- [39] <https://mobislive.com/reportcontent/?content=1133>.
- [40] <https://www.hankookilbo.com/News/Read/A2021101409500001667>.
- [41] <https://www.mk.co.kr/news/economy/10766683>.
- [42] <https://www.yna.co.kr/view/AKR20240110165200009>.
- [43] <https://www.nocutnews.co.kr/news/6044322>.
- [44] <https://news.mt.co.kr/mtview.php?no=2023040616044055477>.

## — [ 저 자 소 개 ] —



차 건 상 (Gun-Sang Cha)  
 2012년 숭실대학교 컴퓨터학 박사  
 2006년 행정안전부 정보보안 전문위원  
 2015년 건양대 스마트보안학과 교수  
 2019년 건양대 정보통신원장  
 2021년 교육부 주관(제2권역)  
 정보보호영재교육원장  
 email : chagunsang@konyang.ac.kr