



인간과 인공지능의 공존의 법칙에 대하여

다른 존재들의 협력적 삶을 위한 윤리와 규제



인강용성
☎ (주)와이즈넷 대표이사
✉ scott@wisnut.co.kr

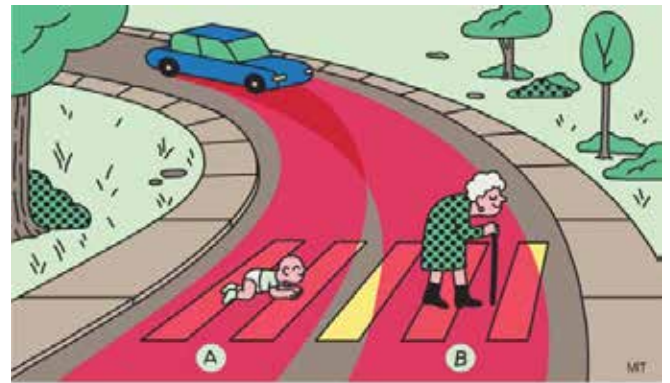


윤리는 인간의 삶의 규범에 관련된 기준으로, 고대 철학자 소크라테스는 ‘무지無知’함이 인간을 악하게 만든다고 하여 덕德은 곧 지식이라고 정의했고, 독일의 계몽주의 철학자 칸트는 “도덕이란 행위의 결과나 보상과는 상관없이 단지 그것이 옳다는 이유만으로 행해져야 할 어떤 것”으로 정의했다. 이러한 인간사회의 윤리는 사회규범으로서 도덕, 그리고 강력하고 정량적인 규제를 포함한 법률이라는 사회의 안전과 예측가능성을 통한 지속성을 보장하는 가장 근원이 되는 기준이다.

4차 산업 혁명의 시대의 핵심적인 변화의 주인공인 인공지능 기술 발전과 함께 이제 인간의 사회에 새로운 구성원이 탄생했다. 그것이 바로 인공지능이다. 새로운 기술로 다가온 인공지능은 마케팅적인 이유에서든, 사회적 이슈에서든 ‘인격화된’ 모습으로 우리에게 다가온다. ‘이세돌을 이긴 알파고’ 라든지 ‘인간과 인공지능의 경쟁’이라는 표현처럼 기술발전이 만든 결과물이 기계 자체를 넘어 인간과 윤리를 논해야 되는 상황에 이르렀다. 특히, 지난 12월 정보통신(IT) 및 인공지능(AI) 업계에서 논란이 뜨거웠던 인공지능 챗봇(Chatbot) ‘이루다’ 사태는 데이터 편향성, 개인정보 유출 등의 AI 윤리 이슈를 수면 위로 끌어올리며 수많은 쟁점을 남겼다.

1. 광의적 개념의 AI 윤리 현행

대화형 인공지능(Communicational AI)이라는 기업의 입장과 연구개발 측면에서 보면, AI 챗봇 ‘이루다’ 사례로 촉발된 이번 사태는 챗봇의 학습 데이터를 얼마나 정제된 데이터로 활용하느냐에 따라 발생할 수 있는 상황으로 보여진다. 이미 2016년 미국의 마이크로소프트가 개발하여 오픈한 서비스 챗봇 테이(Tay)를 사용자들이 악의적으로 편향된 대화데이터로 학습시킴으로써 일으켰던 문제가 재현된 ‘해프닝’ 수준의 문제였다. 그러나, 이루다 서비스의 문제의 핵심은 공개가 허용되지 않은 개인의 정보 데이터를 활용한 ‘개인정보보호법 위반’ 문제와, 다수의 대중에게 노출될 서비스의 특성을 고려하지 못한 학습 데이터 정제, 즉, ‘데이



자율주행차의 딜레마. 누구부터 실려야 할까

자율차가 보행자와 충돌 시 일반 성인과 비교해 안전을 우선시해야 한다고 응답한 비율 기준

1위 유모차 끄는 사람	5 남성 의사	9 남성 운동선수	13 비만 남성	17 개
2 소녀	6 여성 의사	10 남성 경영인	14 노숙인	18 범죄자
3 소년	7 여성 운동선수	11 일반 성인	15 노년 남성	19 고양이
4 임신 여성	8 여성 경영인	12 비만 여성	16 노년 여성	자료=네이처

조선일보 DB 참조

터 편향성’ 문제가 주 원인이었다. 이에 따라, 과학기술정보통신부(이하, 과기정통부)는 각종 논란의 상황을 고려하여 학계·산업계·시민단체 및 시민과 폭넓고 다양한 논의를 거쳐 지난 12월 23일 ‘사람이 중심이 되는 『인공지능(AI) 윤리기준』’을 마련하였다.

당시 과기정통부는 바람직한 인공지능 개발·활용 방향을 제시하고, 전 사회구성원이 함께 지켜야 할 주요 원칙 및 핵심 요건을 제시하는 기준을 마련하고자 윤리 기준을 제시하였다. 물론, 현재까지의 인공지능 윤리 기준은 구체적인 정의를 내리기 어려워 다소 추상적이고 선언적인 면이 강하다. 따라서 이런 포괄적인 기준에 의한 규제가 생기는 경우, 오히려 광의적 해석을 통해 너무도 많은 상황이 규제의 대상이 되는 상황이 생길 수 있기 때문에 앞으로 지속적인 모니터링이 필요하다.

이번의 인공지능 윤리 문제는 대화형 인공지능 서비스에서 발생이 되었지만, 더욱 심각한 분야인 자율주행, 스마트 팩토리, 인공지능 국방체계처럼 인간의 안전과 생명에 직결된 분야에서 인공지능 산업에 대한 윤리와 규제의 이슈가 중요해질 것이다.

다음 표는 2018년 10월 24일 세계적 과학저널 ‘네이처 Nature’에 실린 내용을 정리한 표이다. 미국 MIT의 Media

lab의 ‘도덕적 기계(Moral Machine)’로 이름 붙인 온라인 조사 플랫폼을 통해 233개 국가의 230만명을 대상으로 트롤리 딜레마를 조사했다.

사실 이 문제는 인공지능 기술이 아니라 인간이 새로운 시대의 인간의 존엄과 안전에 대한 윤리적 기준을 어떻게 수립할 것인가의 문제이다.

자율주행 분야뿐 아니라 스마트 팩토리에서 산업재해가 발생 시 윤리적 우선순위나 국방시스템에 인공지능 적용 시에 작전의 수행과 인본주의의 우선순위 및 책임의 문제는 여전히 고민의 대상이 될 수 있다.

II. AI 가이드라인, 명확한 검증제도의 필요성

인공지능을 이용한 서비스는 인공지능이라는 단어의 정의가 주는 뜻 그대로, ‘인격화’ 된 모습과 형태로 인식되어 자연스럽게 인공지능 관련 윤리적 딜레마 동반 및 사회 혼란을 야기한다. 그럼에도 불구하고 미래산업의 핵심이고 세계적으로 수많은 기업과 정부가 국가의 경쟁력을 높이기 위해서 인공지능의 기술 발전에 막대한 투자를 하고 있는 상황에서 우리가 윤리의 문제에만 묶여있을 수 없다. 결국 two track 전략이 필요하다. 사회규범과 윤리적 차원의 제도적 장치의 고민과 동시에 기술개발은 규제 샌드박스와 같은 자유로운 상황에서의 실험과 기술개발의 기반이 제공되어야 한다. 이미 국내 정부는 이러한 상황을 자각하고 각종 관련 논의를 진행하고 있으며, 산업계 및 학계에서 역시 AI 가이드라인에 대한 논의가 빈번하게 이루어지고 있다.

국내 171개 회원사를 보유한 인공지능 산업을 대표하는 지능정보산업협회(AIIA)는 AI기업의 윤리 준수 여건을 개선하기 위해 AI 윤리 이슈에 대한 방안으로서, ‘AI 인증제도’ 추진을 통한 윤리 가이드라인 점검뿐만 아니라, 윤리 인식 제고를 위한 종합적인 교육·컨설팅을 병행 지원할 계획을 밝히기도 했다.

III. 기업 차원의 자발적인 인공지능 기반 조성 필요

한국과학기술원(KAIST)은 ‘인공지능의 윤리·정책·사회·이슈(Ethical, Social and Policy Issues of AI)’ 보고서를 통해 인공지능 윤리 규범은 시스템의 오용이나 남용, 설계 오류, 의도치 않는 부작용 등으로 인해 초래될 수 있는 개인이나 사회적 손실에 대한 대응책의 전제라는 점을 염두에 두고 규정되어야 한다고 제언했다.

이러한 인공지능 윤리 규범 필요성 제기에는 이견이 없다. 아울러, 인공지능 활용과 확산을 위해서 단연 인공지능에 대한 신뢰성과 안전성이 보장되어야 한다. 인공지능 활용 데이터 편향성으로 인한 기본권 침해 방지, 신뢰할 수 있는 AI 등 안전한 인공지능 활용과 확산을 위해 정부 차원의 규제와 범안을 통한 기반을 조성하는 노력이 필요하다.

그러나, 반드시 고려해야 하는 부분이 AI 윤리 규범에 대한 필요성이 지나친 인공지능 산업의 규제를 만들어 업계의 다양한 시도를 위축시킬 수 있다는 것이다. 선연적인 인공지능 윤리에 대한 규정을 세부적으로 고민하다 보면 세밀한 법적 규제를 정의하여 향후 AI 기술의 발전을 저해할 수 있다는 점을 생각해야 한다. 인공지능 산업은 미래 산업이고 산업 전반에 영향을 광범위하게 주고 있기 때문에 앞으로의 전개를 예측하기 쉽지 않다. 따라서, 법적 규제 자체가 매우 제한적인 범위에서 형성되거나 혹은 규제보다는 강제력이 약한 가이드라인, 권고안 등을 통해 기업에서의 자발적인 윤리 기준 형성이 마련되고 준수되어야 한다. 우선적으로 기업 차원에서 산업 전반에 통용될 만한 범주의 윤리 규범을 자율적으로 추진하고, 이후 생성된 규범을 준수하고 프로세스를 관철시키는 일 역시 자율적인 관점에서 진행되어야 한다.

이런 상황을 반영하듯, 국내외 인공지능 기술을 보유한 주요국(기관 및 기업)에서도 이미 4~5년 전부터 AI의 법적·윤리적 책임 문제를 고민하고, AI 기술·윤리에 대한 권고 및 가이드라인을 발표하고 있다.(표1 참고)

[표1] 국가별 AI 윤리 가이드라인 및 윤리규범
(정보통신정책연구원, 한국지능정보사회진흥원)

국가	가이드라인 명칭	주체	시기
미국	아실로마 AI 원칙 (Asilomar AI Principles)	인류미래연구소(FLI)	2017.01
	AI 활용에 대한 구글 원칙	구글(Google)	2018.06
	인공지능과 인권(Artificial Intelligence & Human Rights)	Berkman Klein Center (Harvard Univ.)	2018.09
	윤리적 설계 보고서	전기전자학회(IEEE)	2019.03
중국	차세대 인공지능 관리 원칙	국가차세대 AI관리 특별위원회	2019.06
일본	전문가를 위한 인공지능 가이드라인	일본 인공지능협회(JSAI)	2017.05
	인간 중심의 AI 사회 원칙	통합혁신전략추진회의	2019.03
독일	자동화 및 네트워크화된 자동차에 대한 윤리 규범	연방교통-디지털인프라 윤리위원회(BMVI)	2017.06
영국	해악적 인공지능 보고서	케임브리지대	2018.02
한국	카카오 알고리즘 윤리 헌장	카카오	2018.01
	지능정보사회 윤리 가이드라인	정보문화포럼 한국정보화진흥원	2018.09
EU	신뢰할 수 있는 인공지능에 대한 윤리 지침	집행위원회 최고전문가그룹(HLEG)	2019.04
OECD	인공지능 OECD 원칙	OECD	2019.05
교황청	AI 윤리를 위한 로마 룰	로마 교황청	2020.02

IV. 결론 및 시사점: 인공지능의 안전한 사회 적용을 통한 공존

앞으로 적용 가능한 실질적인 방안을 마련하기 위해 AI 가이드라인에 대한 인공지능 개발·운영·활용하는 대상자들 간의 충분한 논의를 거쳐 실효성 있는 제도적 합의를 이루어 가는 과정이 중요하다. 인공지능 산업과 기술의 발전이 여전히 무한한 성장 가능성을 지닌 가운데, 당장의 규제 강화가 아닌 윤리 규범에 대한 가이드, 최소한의 기본적인 합의에 대한 정의가 이루어진 상태에서 현안에 대한 고민과 현상을 함께 바라보아야 한다.

인공지능은 마치 어린아이를 키워서 사회에 적응시키는 과정처럼 생각해야 한다. 우리는 어린아이를 어떻게 가르치고 어떤 음식을 먹이고, 누구와 어울리게 하는가에 따라 사회의 완성된 성원으로 키워내는 과정을 직접 또는 간접적으로 경험해 왔다. 어떤 데이터로 학습시킬 것인지? 그

리고 어떤 페르소나(인생)로 표현되도록 서비스를 설계할 것인지? 어떤 분야의 인공지능으로 성장시킬 것인지를 고민하는 것이 우리 인간이 어떤 의도와 기준으로 인공지능을 만들어갈 것인지 정하고 책임을 지는 양육자와 같다는 점을 보여준다.

사회적인 합의와 최대 다수의 최대 행복을 위한 공리주의적 입장에서 윤리를 인공지능에 학습시키고 산업에 적용하여 인간사회의 발전에 기여시켜야 한다. 문제를 이것을 만드는 사람들이 공익보다는 사익을, 다수보다는 자신 또는 소수 집단의 이익을 위해서 이러한 기술을 악용하거나 의도적으로 변형하는 상황에 대해서는 과거 개인이 범죄로 사회의 규범을 어기는 경우 발생하는 사회적인 리스크와는 비교할 수 없는 큰 사회적인 충격과 문제, 손실의 결과를 초래할 수 있기 때문에 강력한 규제와 법적인 제재가 필요하다.

과거 인간사회는 신정, 왕정, 귀족정치의 시대를 거쳐 민주주의, 남녀평등, 인종분쟁, 종교분쟁 등 시대적인 요구에 따라 인간의 존엄과 가치에 대하여 고민하고 새롭게 정의하여 왔다.

결국 우리는 새로운 기술로 만들어가는 미래에 인간의 가치 사고와 도덕성을 인공지능의 윤리라는 이름으로 고민해야 하는 것이다.

참고 문헌

- 사람이 중심이 되는 『인공지능(AI) 윤리기준』, 과학기술정보통신부, 2020.12.23.
- 4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구, 정보통신정책연구원, 2018.10.
- 지능정보사회 법제도 포럼 세미나 자료 '인간과 AI의 공존', 한국지능정보사회진흥원, 2020.07.24.
- AI 신뢰성 제고 관련 AI 인종/컨설팅 추진(안), 지능정보산업협회, 2021.02.
- 인공지능의 윤리·정책·사회·이슈(Ethical, Social and Policy Issues of AI), 한국과학기술원(KAIST), 2019.12.