전남대학교 법학연구소 공익인권법센터
인권법평론 제28호 (2022년 2월)
https://doi.org/10.38135/hrlr.2022.28.233

인공지능(AI) 기술의 윤리적·법적 관점에서의 談論

이 창 규*

목차

| 국문초록 |

인공지능 기술의 개발 단계에서는 윤리라는 사회규범(Social Norm)에 개발 방향의 판단을 의존하는 경향이 크고, 잘못된 목적으로 개발하지 않기 위해 윤리라는 규범(Norm)에 대한 의미 부여와 한계점을 명확히 할 필요가 있다. 이에, 합리적인 인공지능 기술을 활용하기 위해 설계부터 상용화까지 윤리적 기준이 적용되어야 할 것이다.

사실 인공지능 기술의 위험성에 대한 논의를 법률의 카테고리에서 논의되어야 하지만 과학기술의 발전과 규제의 균형을 맞추기 위해 규제가 필요한 영역과 윤리규범으로 최소한의 원칙을 요구하는 영역으로 구분하는 것이 필요하다. 이는 아직 위험성이 발현되지 않았으며, 영향 평가가 이루어지지 않았기 때문에, 선제적인 규제를 하는 것은 규제의 목적에도 맞지 않기 때문이다. 이에 인공지능 기술의 문제 되는 요소를 사전에 가늠하고 이

* 아주대학교 특임교수, 법학박사, 기술거래사
논문접수일 : 2022. 2. 6., 심사개시일: 2022. 2. 7., 게재확정일 : 2022. 2. 25.

에 대한 윤리규범을 제시하는 것이 필요하다.

인공지능 기술은 데이터 기계학습을 하므로 안전성 문제와 오남용에 대한 문제가 존재한다. 인공지능 기술은 대량의 데이터를 학습하여 성능을 향상하는 기계학습에 기반을 두고 있다. 이 때문에, 불확실성과 불투명성을 갖고 있으며 노이즈 데이터로 인한 오류를 발생시킬 가능성도 존재한다. 아울러 인공지능 기술이 특정 분야에서 인간의 능력을 넘어서고 있으며, 이러한 상황을 이용해 고의로 악용할 수 있다는 우려가 제기되고 있다. 특히, 인공지능에 의한 인간의 노동 대체, 사생활 침해, 양극화 심화 등의 사회 문제도 등장할 수 있다.

인공지능 윤리는 기술의 특수성 및 실제 적용 맥락, 다양한 이해관계자를 고려하는 윤리형태를 가져야 한다. 윤리가 기술의 기획, 형성, 배치 등에 필요한다는 것을 받아들이고, 인공지능 윤리를 기술의 개발에 있어서 적극적으로 수행되어야 할 것이다. 또한, 인공지능에 관한 윤리적 연구가 인공지능 기술에 매몰되지 않으면서 동시에 새로운 기술 및 변화를 이해하기 위한 윤리가 되려면, 가치, 방향성, 정당화를 고민해야 할 것이다.

주제어 : 인공지능, 윤리, 알고리즘, 사회규범, 윤리규범

Ⅰ. 서(序)

진화하는 인공지능(Artificial Intelligence, 이하 “AI”로 지칭함) 기술의 개발 단계에서는 윤리라는 사회규범(Social Norm)에 개발 방향의 판단을 의존하는 경향이 크고 잘못된 방향으로 가지 않기 위해 윤리라는 규범(Norm)에 대한 의미 부여와 한계점을 명확히 할 필요가 있다.[1] 그러므로 합리적인 AI 기술을 활용하기 위해 설계부

1) LEE RAINIE, JANNA ANDERSON AND EMILY A. VOGELS, *Worries about developments in AI*, PEW RESEARCH CENTERJUNE 16, 2021, 〈https://www.

터 상용화까지 윤리적 기준이 적용되어야 할 것이다. 가령, 자율주행 자동차에 탑재된 AI 기술이 적용된 프로그램은 어떤 판단을 할지에 관한 문제는 설계 단계에서 어떤 프로그램을 개발해야 할지와 연계된다.[2] 이에 AI 윤리는 이러한 논점에서 일종의 이정표를 제시할 수 있다.

사실 AI 기술의 위험성에 대한 논의를 법률의 카테고리에서 논의되어야 하지만 과학기술의 발전과 규제의 균형을 맞추기 위해 규제가 필요한 영역과 윤리규범으로 최소한의 원칙을 요구하는 영역으로 구분하는 것이 필요하다. 이와 같은 이유는 아직 위험성이 발현되지 않았으며, 이에 대한 영향 평가가 이루어지지 않았는데, 선제적인 규제를 한다면 법의 목적에도 맞지 않는다고 할 수 있다. 이에 AI 기술의 문제 되는 요소를 사전에 가늠하고 이에 대한 윤리규범을 제시하는 것이 필요하다 할 것이다.[3]

데이터 기계학습(machine learning)을 하는 인공지능 기술의 특징으로 인하여 AI의 안전성 결여와 오남용 등 역기능에 대한 우려가 존재한다. AI 기술은 대량의 데이터를 학습하여 성능을 향상하는 기계학습에 기반을 두고 있어 불확실성과 불투명성을 갖고 있으며, 노이즈 데이터(noise data)로 인한 오류를 발생시킬 가능성도 존재한다. 아울러 인공지능 기술이 특정 분야에서 인간의 능력을 넘어서면서, 이를 고의로 악용할 수 있다는 우려도 제기되고 있다. 이와 함께, 인공지능에 의한 인간의 노동 대체, 사생활 침해, 양극화 심

---

pewresearch.org/internet/2021/06/16/1-worries-about-developments-in-ai/〉(last visited Feb. 31, 2022).

2) 정석우, 심현철, "자율주행 자동차의 인공지능",「기계저널」제57권 제3호, 대한기계학회지, 2017, 43면.

3) European Parliamentary Research Service, *The ethics of artificial intelligence: Issues and initiativesm*, European Parliamentary, 2020, pp.13-14.

화 등의 사회 문제가 등장할 가능성도 있다.[4)]

현재 AI의 개발과 응용에 관한 법률과 윤리적 지침에 대해 정부, 교육기관, 공공기관, 민간기관 등에서 다양한 논의가 진행되고 있다. 이처럼 AI 기술에 대한 기대는 매우 크다고 할 수 있다. 그러나 AI에 대한 경계감도 크다. AI의 안전한 발전을 촉진하고 위험성을 적절하게 관리하는 구조와 사회에서 AI에 대한 지식과 이해를 촉진하는 구조가 요구되고 있다. 이러한 방어적 기능이 제대로 구축되지 않는다면, AI는 위험한 기술로서 분류되어 그 유익한 발전이 저해될 가능성도 있다. 기술의 발전과 법이나 제도 등의 사회 시스템, 가치관이나 윤리 등이 상충(相衝)되지 않는 대비책을 마련하는 것이 필요하다 할 것이다.

그리하여 이 글에서는 AI 기술의 윤리적, 법적 관점에서의 향후 발전 방안에 관한 논의를 진행하고 한다. 이를 위해 아래의 내용을 차례대로 전개한다.

첫째, AI 기술의 윤리적·법적 관점에서 AI 윤리의 필요성과, AI와 규범의 조화에 관한 내용을 법, 도덕, 윤리의 역할과 조화 방안을 논의한다.

둘째, AI 기술의 윤리적 발전을 위한 기본방향으로 AI 기술 윤리를 논의 하기 위해 AI 윤리에서의 특수한 문제와 가치중립적인 위치에서의 과학기술을 검토하고, AI 기술 윤리의 구성요소로써 5가지에 대해 차례로 검토한다.

셋째, AI 기술 윤리가 구체적으로 구현된 해외 주요국의 윤리지침과 국내지침을 차례로 검토하고 특징을 알아본다.

---

4) 신용우, “인공지능 관련 입법 현황 및 전망”, 「NARS 현황분석」, 제87호, 국회입법조사처, 2019, 1면.

Ⅱ. 인공지능(AI) 기술의 윤리적·법적 관점에서의 논의

1. 인공지능(AI) 윤리의 필요성

AI 윤리는 전문가로서의 사회적 책임을 부담하는 과학기술자에게 윤리적으로 행동할 수 있는 방향을 제시한다.[5] 이는 준법감시(Compliance) 업무와 겹치고 연구자로서 반드시 주의를 기울여야 하지만, 실무에서는 등한시되고 있다.[6] 그러나 AI를 활용하는 단계에서 잘못된 제어로 인한 위험성이 발생할 수 있다. 이에 AI 기술 개발자는 전문가로서의 행동 규범을 준수하는 것과 함께 연구개발의 단계에서 그 방향성이나 설계에 윤리적 문제가 발생하지 않도록 해야 하는 책임을 부담하게 된다. 이를 위해 AI 연구개발을 진행하는 각 단체에서 윤리지침의 제정이 필요하다.[7]

사실, 윤리지침이나 법정비는 개발을 저해하는 것이 아니라 장기적으로 보면 사회에 기술을 상용화하기 위해 필수적이라고 할 수 있다. 유사한 예로는 자동차와 도로교통법의 관계를 생각해볼 수 있다. 지금으로부터 20년 전의 우리나라에서는 안전띠 없이 자동차를 운행하였지만, 현재는 안전띠나 에어백과 같은 안전장치가 법적으로 요구되고 있다.[8] 이렇듯, 법률이 미정비된 상황에서는 많은

---

5) 송성수, 「과학기술자의 사회적 책임과 윤리」, 과학기술정책연구원, 2001, 7면.
6) 홍용희, "과학기술과 윤리의 상관성", 「윤리연구」 제72호, 한국윤리학회, 2009, 195면.
7) Mark Ryan, Bernd Carsten Stahl, Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications, *Journal of Information*, Communication and Ethics in Society, 2020, p.63.
8) 안전띠의 역사, 〈https://www.khan.co.kr/economy/auto/article/201011051115302〉 (최종방문일 2022년 1월 31일)

사상자가 발생하였지만, 반복된 피해가 발생하지 않도록 법정비가 진행되어 왔다.

또한, 국경을 넘을 때마다 우측통행이 되거나 좌측통행이 되면 매우 불편하기 때문에, 국제 협조하에 일관된 체계가 설정되었지만 지금도 국가에 따라 도로교통법 그 자체가 미정비된 나라도 있다.[9] AI 기술의 적용에서도 비슷한 진행이 될 것으로 예상한다. 특히, 법정비에는 상당한 시간이 걸릴 것이며, 그러한 과정에 있어서 여러 가지 피해가 발생할 수 있다. 이에 발생될 문제에 대해 신중하게 검토하고 예측해 예방대책을 강구할 필요가 있다.

2. 인공지능(AI)과 규범의 조화

가. 법, 도덕, 윤리의 역할

윤리를 생각할 때 상위 개념인 규범(Norm)을 음미할 필요가 있다. 규범의 사전적 의미는 인간이 사회생활을 하는 데 있어, 구속되고 준거하도록 강요되는 일정한 행동 양식이다.[10] 이러한 규범은 사회생활을 영위하는 데 필요한 사회규범으로 이해할 수 있다. 사회규범 중에는 도덕, 법, 윤리, 관행, 습관이 있다. 사실, 사회규범은 사회구성원에게 일정한 구속을 요구하기 때문에, 집단의 조화를 유지하기 위해 지켜야 할 규범으로서 개인적인 욕망을 자제하는 강령으로 특징지어진다.[11] 사람은 사회적 동물이며, 자신의 욕망대로

9) 오른쪽에 핸들이 있는 자동차 이야기, 〈http://www.kama.or.kr/jsp/webzine/201803/pages/story_02.jsp〉(최종방문일 2022년 1월 31일)

10) 김윤명, "지능정보사회에 대한 규범적 논의와 법정책적 대응", 정보화정책 제23권 제4호, 한국지능정보사회진흥원. 2016년 겨울호, 26-27면.

11) 황경식, "도덕체계와 사회구조의 상관성",「철학사상」제32조, 서울대학교 철학사상연구소, 2009, 225면.

살 수 없고, 조화로운 사회생활을 영위하기 위해 다른 사람들과의 조화를 고려해, 자신이 원치 않는 형태로의 행동을 하기도 한다.

사회규범을 좀더 범주를 좁혀보자면 법(Law), 도덕(Moral), 윤리(Ethics)로 나눌 수 있다. 이 세 가지 특성의 차이점, 동일성, 기능 등에 대해서는 다양한 논의가 존재한다. 특히, 법학 쟁점 중에서 '법과 도덕'이라는 주제는 법철학에서의 중요한 논제이다. 법, 도덕, 윤리의 범주를 판단해보자면, 윤리는 좋음, 옳음, 쾌락 등 이상적 가치나 규범에 따라 행동해야 하는 당위이며, 도덕은 인간이 지켜야 할 도리 또는 바람직한 행동기준이다. 법은 국가권력에 의하여 강제되는 사회규범이다. 이러한 세 가지 범주의 차이점에 대한 강제력이라는 점에서 법이 윤리와 도덕과 다르다고 할 수 있다. 이는 법이 국가권력에 의한 물리적 강제력을 가진 점에서 윤리 및 도덕과 차별화된다고 할 수 있다.[12)]

법은 국가가 방패가 되어 실현하는 사회규범이며, 경찰권이나 법원과 같은 사법제도를 사용한 유형력으로 원하는 행동 규범을 실현할 수 있다. 한편, 도덕은 사람의 내심에서 작용하는 것이고, 원하는 행동 규범을 따르지 않았더라도 아무런 처벌을 받지 않는다.[13)] 양심의 가책을 느끼더라도 본인이 자발적으로 행동하지 않는 한 아무것도 변화는 일어나지 않는다.

강제력이 없다는 점에서 도덕과 윤리는 공통되지만, 법과는 그 성질이 다르다. 그렇다면 도덕과 윤리는 어떻게 다른가? 윤리와 도

12) Eva Hofmann, Barbara Hartl, Katharina Gangl, Martina Hartner-Tiefenthaler and Erich Kirchler, Authorities' Coercive and Legitimate Power: The Impact on Cognitions Underlying Cooperation, *Front. Psychol.*, 18 January 2017. 〈https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00005/full〉(last visited Feb. 31, 2022).

13) 설선혜·이승민, "도덕 판단에서 나타나는 도덕-인습 구분에 대한 논쟁과 함의", 「인지과학」, 제29권 제2호, 한국인지과학회, 2018, 137면.

덕에 대해서는 내심에 호소하는 규범이라는 점에서는 구별이 어렵다. 도덕은 개인이나 가족과 같은 작은 집단을 대상으로 매우 개인적인 내심에 크게 의존하는 반면, 윤리는 범용성을 가질 수 있다. 즉, 도덕은 지역성, 종교, 관습 등의 개인적인 요소를 잘 반영하고 있지만, 윤리는 특정 사회 집단에 공통적인 규범으로서 적용할 수 있다.[14]

예를 들어, 생명윤리학과 같은 응용윤리학(Applied ethics)의 한 분야에서는 생명과학에 종사하는 연구자 간의 연구 활동을 규율하는 규범으로써 작용한다. 해당 연구자는 종교, 국적, 성별을 불문하고 해당 집단에서 확립된 윤리는 그 집단 내에서 범용성을 가진다. 강제력을 가진 법은 내심의 자유까지 제어할 수 없고, 외형력에 의한 강제를 가할 뿐이며, 도덕처럼 개인 내심의 다양성을 인정하지 않다는 점에서 우위성을 가진다.[15]

이러한 점에서 법의 보편성을 언급하는 견해도 있지만, 실제로 법은 국가 단위로 기능하는 규범이며, 초월적 규범으로서는 윤리가 활용될 수 있다. 가령, 생명윤리 관점에서 유전자 조작에 관한 공통 윤리를 과학자들 사이에서 도출할 수 있더라도 법은 그 나라의 법감정, 법체계를 고려해야 한다. 모든 국가가 국제조약 비준을 하고, 국내법에 따른 입법규제로 실현한다는 것은 사실상 불가능에 가깝기 때문이다.

나. 윤리의 필요성

윤리가 도덕보다 뛰어난 점은 앞서 언급한 바와 같이 범용성이

14) Human-Centered Artificial Intelligence Institute(HAI), *Artificial Intelligence Index Report 2019 AI Index Report–Highlights, 2019*, p.34.

15) European Parliamentary Research Service, *Ibid.*, p.2.

크기 때문이다. 도덕이 개인이나 가족 등의 작은 집단에 사용되는 경우가 많지만, 윤리는 개개인의 관계로부터 사회에 이르기까지보다 광범위하게 사용된다. 이 때문에 도덕은 일상생활에 있어서 행동의 기준이 되더라도, 첨단 과학기술을 개발하는 현장에서의 판단 기준이 될 수 없다.[16] 따라서 다양한 응용윤리학이 성립되어 학문의 영역에서 크나큰 역할을 하고 있다. 이처럼 개인의 자발적 의사에 기초한 행동을 요청할 수 있는 것이 윤리라고 할 수 있다. 즉, AI 개발자에 대한 가장 바람직한 사회규범은 윤리라고 할 수 있겠다.[17]

무엇보다 윤리학에도 여러 가지 양상이 있으며, 그 논의의 역사는 오래되었다. 특히, 20세기에 들어서 첨단과학기술의 눈부신 발전으로 윤리학을 새로운 영역에 적용하려는 응용윤리학이 발전하였다. 즉, 윤리의 유용성으로 인해 입법기관에 의한 합의체 형성이 불필요하게 되었고, 그 조직 내에서 자유롭게 결정될 수 있는 유연성과 함께 개개의 차이를 전제로 하면서도, 공통의 사회규범을 실현할 수 있는 범용성이 높은 첨단과학의 규범으로서 응용윤리학이 활용된다고 할 수 있다.[18]

다. 인공지능과 윤리규범의 조화

AI는 그 개발 단계에서는 개발자의 의도대로 AI의 행동을 지배한다. 이는 AI 자체가 주어진 체계 내에서만 동작하기 때문이다. 개

16) Naomi Ellemers, Jojanneke van der Toorn, Yavor Paunov, The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017, *Pers Soc Psychol Rev,* Vol.23 No.4, 2019, p.335.

17) LEE RAINIE, JANNA ANDERSON AND EMILY A. VOGELS, *Ibid.*

18) Oriel FeldmanHall, Jae-Young Son, and Joseph Heffner, Norms and the Flexibility of Moral Action, *Personal Neurosci.* 2018, 〈https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7219684/〉(last visited Feb. 31, 2022).

발자가 선택하는 알고리즘(algorithm)이 AI의 행동이나 행동을 규정하는 것이기 때문이다. AI가 사회적 물의를 일으킨다면, 그러한 사회적 손해의 귀책성을 사후적으로 묻게 되는 상황에서도, 판단기준으로서 윤리를 적용할 수 있다. 이에 근거해 법적 귀책사유의 근거가 될 수 있다. 이 경우 AI 개발자에게 윤리적 판단을 맡긴다면 어떤 원리 및 원칙, 가치 판단에 근거해 윤리적 판단이라는 알고리즘을 구현시켰는지에 대한 설계사항이 이슈라고 할 수 있다.[19)]

그렇다면 윤리라는 규범을 AI 개발 단계에서 통합하는 것은 어떤 의미가 있는가? 여기서 말하고 싶은 것은 사회적 통제의 효율성이다. 사회에서 어떤 문제가 발생했을 때 그것을 어떻게 해결하는 것이 가장 효율적인가 하는 것이다. 이 경우의 효율성은 분쟁 해결에 걸리는 시간 등 사회적 비용을 절감할 수 있는 사항이 중요시 되어야 할 것이다. 그래서 가장 효율적인 것은 사회 구성원이 자율적으로 규범을 준수하고 문제를 발생시키지 않도록 하는 것이다.

그러나 현실에서는 사회분쟁이 발생한다면, 이에 대한 대처로서 법에 따른 통제가 이루어진다. 법이라는 사회규범에 의한 통제는 형사벌과 같은 법적 책임에 의한 강제력을 이용하여 인간을 정신적으로나 신체적으로 구속한다. 이러한 시스템의 유지에는 상당한 비용이 필요하다. 반면에 윤리는 사회에 대처하는 개인행동의 지표이며, 법과 같은 강제력을 가지지 않으며, 법에 따른 타율적인 규제가 아니라 개인에 의한 자율적인 행동 억제에 의한 사회의 행복과 발전을 실현하는 자발적인 행위를 가져온다고 할 수 있다. 자율·분산·협조를 중요시하는 소셜 네트워크(social network) 사회에 있어서는 더욱 바람직한 사회통제를 의미한다.[20)]

---

19) Felicitas Kraemer, Kees van Overveld & Martin Peterson, Is there an ethics of algorithms?, *Ethics and Information Technology* Vol. 13, 2011, p.255.

Ⅲ. 인공지능(AI) 기술의 윤리적 발전의 기본방향

1. 인공지능(AI) 기술 윤리의 향후 논의 방향

가. 인공지능(AI) 윤리의 특수한 문제

AI의 윤리적 문제는 가치(價値, value)와 관련된다. 그러나 이것은 경제학이나 공학이 다루는 가치로서 경제적 번영, 효율 등에 한정되지 않는다. 윤리학이 중심적으로 다루는 것은 보편성을 가진다. 가령, 공동체(community), 공평(fairness), 권리(rights), 도덕(moral), 자율성(autonomy), 정의(Justice), 존엄(dignity), 진실(truth,), 평등(equality), 행복(happiness) 등이 윤리학이 가지는 가치이다.[21]

따라서 인공지능의 윤리적 문제는 이러한 가치가 AI에 의해 침해되는 위험과 관련이 있다. 이 글에서 윤리지침이나 윤리원칙 등도 이런 것과 관련이 있다고 할 수 있다. 논점은 다양하지만, 일반화하고 요약하면 윤리적 과제란 인류 전체의 번영, 행복을 촉진하면서 희생을 가능한 적게 한다는 것이다. 그렇다면, 인공지능에서 특수한 논점은 무엇인가? 제리 카플란(Jerry Kaplan)은 "인공지능 : 모든 사람이 알아야 하는 것(Artificial Intelligence: What Everyone Needs to Know)"에 대해 AI가 무엇인가를 논할 때는 끊임없는 자동화의 진보를 고려해야 한다고 주장하고 있다.[22]

---

20) EKIM YURTSEVER, JACOB LAMBERT, ALEXANDER CARBALLO, KAZUYA TAKEDA, A Survey of Autonomous Driving: Common Practices and Emerging Technologies, *IEEE Journals & Magazines,* VOL. 8, 2019, pp.1-2.

21) 박종호, "인공지능 시대의 윤리와 법적 과제", 「과학기술법연구」 제24집 제3호, 한남대학교 과학기술법연구원, 2018, 185면.

22) Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to KnowR 1st Edition*, Jenson Books Inc., 2019, pp.23-25.

산업혁명은 단순한 물리적 노동을 자동화했지만, 컴퓨터 혁명은 지적 노동을 자동화했으며, 나날이 발전해 인공지능 혁명, 로봇 혁명, 사물인터넷(Internet of Things : IoT) 혁명에 있어서 점점 복잡한 물리적 노동, 지적 노동을 자동화하는 데 성공하였다고 한다. 그러나 이러한 새로운 기술혁명이 의미하는 것은 단순히 수행할 수 있는 업무의 복잡화·고도화만이 아니며, 개인 선택권의 자동화도 포함한다.[23)]

기계적인 작업은 정해진 법칙에 따르고 있는 것이며, 누가 해도 같은 결과가 나올 수 있다. 그러나 현재의 자동화가 진행되고 있는 것은 기계적인 작업만이 아니다. 인공지능의 네트워킹(networking)은 우리의 모든 행동에서 자료를 수집하여 사용자가 정의된 선택과 의사결정을 기계가 적용할 수 있도록 한다. 그러나 이러한 기술에는 의도적이든 의도적이지 않은 일정한 편향성(bias)이 발생한다. 이러한 편향성은 부당한 차별을 표출하는 등의 문제가 발생할 수 있고, 그렇지 않은 경우에도 편견의 존재 자체가 개인의 의사결정의 자율성이라는 가치를 위협하는 것일 수 있다는 것에는 주의가 필요하다.[24)]

나. 가치 중립적 위치에서의 과학기술

과학기술의 발전은 미지의 위험성이나 기존의 가치관과의 충돌을 가져오는 것은 확실하다. 우리는 AI에 어떤 위험이 있다고 예측해 말할 수 있지만, 미래의 이익을 포기하는 것은 잘못되었다고

---

23) Violeta Sima, Ileana Georgiana Gheorghe, Jonel Subi´c, Dumitru Nancu, Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review, *Sustainability* 2020, Vol. 12, No 10, pp.3-4.

24) 변순용, "데이터 윤리에서 인공지능 편향성 문제에 대한 연구",「윤리연구」제128호, 한국윤리학회, 2020, 148면.

할 수 있다. 과학과 기술은 가치 중립적인 지위를 가지며, 이는 선악(善惡)의 의미가 있는 것은 아니라고 할 수 있다. 가령, 전미 라이플 협회(National Rifle Association)의 표에서 "총이 사람을 죽이는 게 아니라, 사람이 사람을 죽인다(Guns don't kill people, people kill people)"라는 구호처럼, 과학기술의 자체만이 문제가 아니라 이를 활용하는 주체가 문제가 있다는 의미이다.[25]

구체적으로 과학기술의 가치 중립적 의미는 첫째, 과학과 기술 자체에 고유한 가치관이 있다는 점이다. 예를 들면, 과학에서는 일반성, 재현성, 정량성, 단순성 등 과학기술의 특유 기준이 존재한다. 기술에서는 효율성, 제어 가능성, 비용과 이득의 공제 등이 규범적 가치가 되고 있다. 과학이나 기술을 추진하는 것은 이러한 가치를 사회에 투영하는 한편, 그에 맞지 않는 개별성, 일회성, 非정량화, 非단순화, 非효율성, 非제어, 비용과 이익의 관점에서 취급할 수 없는 것 등을 병합한다. 둘째, 특정 기술의 산물은 악용에 기울기 쉬운 고유의 편향성을 가지고 있다. 이러한 개별 기술의 고유한 편견을 고려하는 것은 개발자, 사용자, 정책 결정자의 책임이다.[26]

2. 인공지능(AI) 기술 윤리의 구성요소

가. 응용윤리학에서의 인공지능(AI) 기술 윤리

AI 윤리에서 프로그램의 설계가 윤리적이어야 한다는 점은 향후

25) David Kyle Johnson, "Guns Don't Kill People, People Do?", 〈https://www.psychologytoday.com/us/blog/logical-take/201302/guns-don-t-kill-people-people-do〉(last visited Feb. 31, 2022).

26) Nicol Turner Lee, Paul Resnick, and Genie Barton, "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms," Brookings Institution 〈https://www.brookings.edu/research/algorithmic-bias-detection-andmitigation-best-practices-and-policies-to-reduce-consumer-harms〉

AI의 개발에 있어서 방향성을 주는 근원적 체계이다. 앞으로 AI 윤리는 기술적 진보를 반영해야 하지만, 지금까지의 윤리학의 지식과 경륜을 무시할 수는 없다. 사실 윤리학의 논의는 매우 추상적이며, 도움이 되지 않는 것처럼 보일 수 있다. 그러나, 수학이나 물리학의 이론에서 정교한 논의가 기술개발을 지지하는 기반이 되는 것과 같이, 철학과 윤리학의 논의는 실제로 윤리적 판단에 관한 이론의 기초가 되고 있다.[27)]

직관에 기초한 도덕의 모델링은 인간의 능력 범위에서의 행위나 좋고 나쁜 것에 대해 추상적 이론과 크게 다를 수 있다. 그것은 인간이 지각 할 수 있는 크기의 물체에 대해 양자 역학에 기반을 둔 뉴턴 역학(Newtonian mechanics)의 이론과 크게 다르지 않다. 하지만 AI의 경우에는 센서가 기능하는 범위, 또 추론속도, 통신속도 등 다양한 부분에서 인간의 능력을 뛰어넘고 있어, 향후 더욱 확장해 나갈 것이다.[28)]

이 같이 사용자가 윤리적 측면을 이해할 필요는 없지만, 연구자와 개발자의 경우 이를 경시할 수 없다. 어떠한 이론이라도 개발을 거쳐 상용화된 물건이 된다면, 가치 중립적 의미를 가진다는 시념으로 개발은 진행되었다. 하지만, 회색 지대(Gray Zone)로 취급되고 있는 AI의 경우에는 논쟁이 계속될 것이다.[29)] AI의 윤리는 응용윤리학의 한 분야에서 이정표를 제시하게 된다. 이론적 기반과 실

27) Paul Ernest, Mathematics, ethics and purism: an application of MacIntyre's virtue theory, *Synthese*, Vol. 199, 2021, p.3138.

28) Fernando Martínez-Plumed, Emilia Gómeza, JoséHernández-Orallob, Futures of artificial intelligence through technology readiness levels, *Telematics and Informatics*, Vol. 58, May 2021, pp.1-2.

29) Jake Harrington, Riley McCabe, Detect and Understand: Modernizing Intelligence for the Gray Zone, 〈https://www.csis.org/analysis/detect-and-understand-modernizing-intelligence-gray-zone〉(last visited Feb. 31, 2022).

무적 응용의 인터페이스(interface)가 기반이 되는 것이 바로 응용윤리학이며, 특히, 정보윤리, 생명윤리, 동물윤리의 식견이 융합하면서 AI의 윤리를 구성하게 된다, 이때 다른 응용윤리학의 적용 분야와 마찬가지로 AI의 윤리에서는 메타 윤리학(meta-ethics)이라는 윤리학 이론도 검토를 해야 한다.[30]

나. 윤리적 판단 주체

윤리적 판단을 할 수 있는 주체는 인간이다. 그래서 도덕적으로 배려해야 할 대상도 인간뿐이었다. 그래서 만일 사람이 키우던 동물이 다른 사람을 물더라도 책임을 부담하는 것은 동물이 아니라 소유주였다. 이러한 논리에 따라 자동차 사고가 발생해 이에 대한 책임 소재가 운전자 과실, 도로 정비 상황에 대한 문제가 발생한다면, 각각의 책임 주체에 대해 법적 책임을 묻는다. 이는 종래의 생각대로 인간이 아닌 그 어떤 생물 등이 법적 책임을 부담할 수 없다고, 전제하기 때문에 AI는 책임을 부담하지 않게 된다.[31]

그렇다면 AI에 윤리 기준을 적용하는 것은 과연 어떤 체계에서 가능할지에 대한 판단이 필요하다. 사실 AI는 물건에 지나지 않기 때문에, 지금까지의 자동차와 같이, 사용자나 설계자·제조자가 책임을 부담해야 한다. 하지만 AI가 적용된 시스템의 경우에 이를 설계한 자와 사용자가 행동을 제어할 수 없을 가능성이 있으므로, 인간이 책임을 전부 부담하는 것이 타당할지가 의문이다.[32]

---

30) Sudhi Sinha, Metaethics, Meta-Intelligence And The Rise Of AI, 〈https://www.forbes.com/sites/forbestechcouncil/2021/01/21/metaethics-meta-intelligence-and-the-rise-of-ai/?sh=2a2992aa46e2〉(last visited Feb. 31, 2022)

31) 장재옥, 김현희, "인공지능의 법적 지위에 관한 논의", 「법학논문집」 제43집 제1호, 중앙대학교 법학연구원, 2019, 113면.

32) 최민수, "인공지능 로봇의 오작동에 의한 사고로 인한 불법행위책임", 「민사법의 이론과 실무」 제23권 제3호, 민사법의 이론과 실무학회, 2020, 5면.

AI 기술은 다양한 레벨로 나누어져 있으며, 레벨마다 윤리성을 묻는 방법, 책임 주체가 다르다. 자율주행주행차와 같은 자율 시스템(Autonomous System : AS)은 이하 6개의 자율성 수준으로 나눌 수 있으며, 이 수준에 따라 윤리성·책임의 취급이 다르다고 할 수 있다. 레벨 0~2에서는 인간이, 레벨 3~5에서는 시스템이 대응 주체가 된다. 이때 시스템의 책임을 잡는 방법에 대해서는 실무적인 수준과 개념적인 수준 모두에서 논의가 진행되고 있다.[33]

lv.0	lv.1	lv.2	lv.3	lv.4	lv.5
인간 모든 작업 수행	시스템 : 인간 작업 지원	시스템 : 인간 작업 부분 대체	시스템 : 제한된 영역에서 모든 작업 수행 비상시 인간 대응	시스템 : 제한된 영역에서 모든 작업 수행 비상시 제어 불가능	시스템 : 모든 영역에서 모든 작업 수행

위와 같은 구성에 따라 생각해야 할 몇 가지 요소에 관해 설명한다. 첫째, 기계는 인간과 같이 판단할 수 있는 것은 아니다. 법인과 같이 한정된 책임을 인간과 기계로 나누어 부담하도록 사회적으로 조정할 수 있다. 2016년에 EU 의회(European Parliament : EP)의 전자적 법인격(electronic personality) 검토 제안은 이러한 맥락에서 논의되었다. 실제로 EU 의회(EP)는 AI에 대한 제한적 전자인(e-person)으로 취급을 검토하는 것을 제언하고 있다.[34]

둘째, 인간과 같거나 그 이상의 기능을 가지는 AI 시스템을 만드

33) 村上祐子, 人工知能の倫理の現在—研究開発における技術哲学·倫理の意義—, IEICE Fundamentals Review Vol.11 No.3, 2018, 161面.

34) Directorate-General for Internal Policies of the Union (European Parliament), *ARTIFICIAL INTELLIGENCE AND CIVIL LIABILITY*, 2020, p.35.

는 것이 기술적으로 가능하며, 그러한 연구가 적어도 일부의 사회에서 용인되어야 한다. AI는 인간과 같거나 그 이상으로 사회적·윤리적 책임을 부담한다. 이러한 가능성을 고려해야 한다는 것이 후술하는 아실로마 AI 원칙(Asilomar AI Principle)의 제언이다. 또한, AI의 경우 인간과 기계가 융합된 시스템을 고려할 필요가 있다.[35]

다. 윤리와 시장 기능

AI 기술 윤리는 기술개발에 따른 경제 발전 요소를 고려해야 한다. AI의 기능이 문제 되는 것은 AI가 적용된 프로그램이 어떠한 제품에 탑재되어 제품으로서 시장에서 소비자에게 공급되는지와 관련이 있다. 즉, 일반 소비자가 AI의 가해행위의 피해자가 되는 상황을 생각할 필요가 있다.[36] 가령, 자율주행 주행차에 탑재된 모듈로서 甲 모델과 乙 모델이라는 상품이 있다고 한다면 甲 모델은 사람을 존중하고, 어떤 상황에서도 인신사고를 회피하는 행동을 취하는 것을 선택하는 윤리 모듈이다.

그리고 乙 모델은 인명보다 주행의 효율성을 우선시 하게 된다.[37] 목적지에까지 확실히 단시간에 도달하는 것이 중시되어 인신사고를 일으킬 가능성이 있는 경우에서 공리적인 고려를 하고, 피해 지수가 적은 경우를 선택해야 한다. 이를 자율주행 프로그램에 개발해 탑재할 때 자율주행 자동차를 사는 일반 소비자가 자동차 보험의 보장 내용을 선택할 수 있게 되는 윤리의 상품화가 될 수 있다. 이는 인공지능에 탑재되는 윤리 모듈을 상품화해 소비자가 선택할 수 있게 하는 문제가 발생할 수 있다.

---

35) 서형준, "4차 산업혁명시대 인공지능 정책 의사결정에 대한 탐색적 논의", Informatization Policy Vol. 26, No.3, 한국정보화진흥원, 2019, 7면.
36) 장재옥, 김현희, 앞의 논문, 106면.
37) 박종호, 앞의 논문, 190면.

그러나 이러한 소비자에 의한 선택의 제공은 실현되기 어렵다고 할 수 있다. 이는 단순하게 시장에서의 수요 공급의 균형이 이끄는 매매 가격으로 결정되는 문제가 아니다. 즉, 안전하지 않은 AI 윤리 모듈이 시장에서 표준(standard)이 되어 버린다면, 사용으로 인한 위험성이 커지게 된다. 무엇보다, AI 기술로 인해 발생된 사회적 책임을 전부 제조자가 부담하게 된다면, AI 개발자가 선택할 수 있는 선택사항은 법적책임을 회피할 수 있는 알고리즘이며, 제조자의 이윤을 극대화하는 알고리즘을 구현할 것으로 예상되며, 사회규범으로서의 윤리가 적용되기 어렵다고 할 것이다.[38]

지금까지 예로서 생각해 온 자율주행 자동차의 운행으로 발생할 문제를 해결하는 것은 법이라고 생각할 수 있다. 즉, 자율주행 자동차에 탑재되는 AI를 개발하는 것은 영리기업인 자동차 제조회사이며, 영리기업은 이윤을 추구하는 것이 설립 목적이다. 영리기업은 AI 기술에 윤리를 적용하는 것이 아니라 자율주행 자동차의 운행에 있어서 어떻게 법적 위험성을 회피할 수 있는지를 고민한다.

즉, AI의 행동으로써 사고가 일어나더라도 가장 법적 책임이 덜 부담하는 선택사항을 선택하는 알고리즘이 선택되어 프로그램을 사용할 것이다. 만일 사고가 일어나 AI로 발생된 손해에 대한 손해배상책임을 묻는다면 배상 책임 주체로서 기업 자체의 이윤이 감소하게 된다. 이를 회피하기 위해서는 가장 저렴한 손해를 상정하게 된다. 법적 책임을 가장 잘 피할 수 있는 행동이 AI 윤리라고 할 수 있다. 이러한 사고방식 자체가 AI 개발에 있어서 우려되어 피해야 하는 문제이며, 윤리가 통제해야 할 문제이다.[39]

---

38) Directorate-General for Internal Policies of the Union (European Parliament), *Ibid.*, p.15.

39) 정도범, 유화선, "인간과 인공지능(AI)의 공존을 위한 사회·윤리적 쟁점 : 신뢰할

라. 윤리 내용의 구성요소

AI 개발자가 예상할 수 있는 한 모든 위험을 회피하는 알고리즘을 개발하는 것은 당연하다. 특히, 인신 손해로 이어지는 위험을 피하는 알고리즘을 설계하게 된다. 그러나 예측할 수 없는 위험에 대해 프로그래밍할 수 없는 것과 위험으로써 프로그래머가 생각하지 않는 것에 대해 윤리적으로 비난하는 것이 가능한지에 대한 고민이 필요하다.[40]

대부분 회사에서 AI 프로그램을 개발하고, 이에 대한 문제가 발생하였을 때 개발자와 회사(법인)이 책임을 부담한다. 그러한 책임 소재를 검토하는 과정에서 법적 분쟁으로 이어지면 법원이 시시비비(是是非非)를 판단하게 되고, 향후 사회 구성원 간의 여론으로 판단하게 된다. 오히려 AI의 개발 단계에서 윤리감독을 AI에 이식하는 것은 현재까지 생각하기 어렵고, AI의 페일 세이프(fail safe) 기능으로서 소스 코드(Source code)가 인간의 사회규범인 윤리가 될 것이다.[41]

민주주의 국가에서 사회 구성원의 합의로 판단한다. 이후 이러한 올바른 정보를 얻은 후의 동의(informed consent)가 윤리적 의미를 가진다. 그러기 위해서도, 윤리에 대한 숙고와 논의가 필요하다. 또, 대중들의 의견을 무시하고, 일부의 개발자가 개발한 윤리가 인공지능의 행동을 제어하는 것은, 블랙박스(black box) 기능을 더욱 부각해, 민주주의의 이념을 등지는 결과를 발생하게 된다.[42]

수 있는 인공지능 실현 방안", KISTI ISSUE BRIEF 제35호, 2021.11, 1-2면.

40) 박기주, "인공지능 알고리즘을 활용한 전문(추천) 서비스 제공의 법적 성격에 관한 연구", 「법제논단」 2020년 3월호, 법제처, 105면.

41) Lukas Gloor, *Suffering-focused AI safety: In favor of "fail-safe" measures, Suffering-ocused AI safety－Center on Long-Term Risk*, 2016, p.5.

42) Karl M. Manheim, Lyric Kaplan, Artificial Intelligence: Risks to Privacy and

결국, 윤리를 구성해야 하는 요소는 인간이 그랜드 트루스(우리가 원하는 답, grand truth)이다. 그랜드 트루스는 현장에서 얻을 수 있는 정보이며, 메타데이터(metadata)가 아니고, 경험 지적인 데이터라고 할 수 있다. AI에게 윤리의 기본으로서 가르침에 있어서의 이 지식은 잘못되지 않았다고 가르쳐지는 것이다. 모든 분야에서 뛰어난 최적화된 표준(gold standard)이라고도 불리며, 과학 실험에서 틀림없는 기준이나 전문적 행위의 최적 형태나 순서·결과를 기초로 해야 한다. 그리고 이러한 그랜드 트루스를 올바르게 적용하는 의지가 필요하다.

마. 윤리교육의 필요성

국가의 강제력에 기반을 둔 법규제에 의한 관리통제를 벗어나 개인의 인식을 함양을 촉구하는 윤리교육이 중요하다. 이는 인공지능 개발자를 위한 그것뿐 아니라 고도의 정보사회에서는 사회 구성원들의 윤리의식 함양이 필요하며, 이를 실현할 수 있는 것이 바로 윤리교육이다. 지금까지 AI와 윤리라는 고찰에서는 인간인 AI 개발자의 윤리에 대해 생각해 왔지만, 기술적 특이점(Technical Singularity)이 발생한 이후에는 AI가 자율적으로 생각하고 행동하게 될 것으로 예측 된다.[43] 이때 어떤 윤리에 근거하여 인공지능은 행동할지는 윤리를 AI에게 프로그램밍해, 윤리적인 행동하게 하는 단계를 구축해야 한다.

윤리는 집단지성에 기초하므로, 윤리에 기초한 인공지능을 창출할 수 있다. 윤리를 구성하는 이념이 지식의 조합이라면, 모든 이념

Democracy, *21 Yale Journal of Law and Technology* 106, 2019, p.111.

43) Murray Shanahan, *The Technological Singularity*, The MIT Press Essential Knowledge series, 2015, pp.189-190.

을 코드화(code化)될 수 있다. 이에 코딩할 수 있다면, 무한하다고 생각되는 윤리 코드로 구성된 빅데이터 내에서 수의 조합 중에서 텍스트 마이닝(Text Mining) 기술을 사용하여 AI가 올바른 것을 선택하면 된다. 인류의 축적된 지식재산에 기반을 둬 AI를 사용하여 윤리를 찾아내는 것이다. 이것을 진행하면 가치 판단이 가능해진다. 가능한 한 많은 사회 구성원의 파라미터(parameter)를 사전에 등록해 두고 판단을 요구받을 때 이를 이용하여 순간적으로 판단할 수 있다.[44]

Ⅳ. 인공지능(AI) 기술 윤리의 구현

1. 2017년 이후의 인공지능(AI) 윤리지침

2017년 이후 AI 윤리지침은 국내, 국외에서 AI의 개발과 활용에 관한 윤리지침을 공표하였다.[45] 해외 주요 AI 윤리지침에서는 AI 기술이 발전이 빠르게 진행되고 있으므로, 이에 대한 규제를 강화한다면 기술적인 발전을 저해한다는 의견이 대부분이다. 이에 AI 윤리지침은 자율적 규제 방안을 제시하는 지침이라고 할 수 있다. AI 윤리지침의 실효성 갖게 방법으로는 표준으로 제정하는 방법도 고려할 수 있다. 주지하다시피 국제 표준은 각국의 개발자에 미치는 주는 영향이 크다고 할 수 있다. 대표적으로 ISO 표준, IEEE P7000 시리즈가 있으며, 이 표준은 IEEE EAD ver.2, 1e를 기초로 제정되었다.

44) 정지선, 김동성, 이홍주, "텍스트 마이닝 기법을 활용한 인공지능 기술개발 동향 분석 연구: 깃허브 상의 오픈 소스 소프트웨어 프로젝트를 대상으로", 「지능정보연구」 제25권 제1호, 2019, 3면.

45) 2017년 이후 미국과 유럽에서 공개된 주요한 AI 윤리지침을 선별하여 정리함.

〈표 1〉 해외 주요 AI 윤리지침

지침명	약어	제정 단체	활용대상	공표 시기
① 아실로마 AI 원칙 (Asilomar AI Principle)	아실로마 AI 원칙 (Asilomar AI Principle)	생명의 미래 연구소 (Future Life Institute)	개발자, 정책 입안자	2017
② 윤리적 내용 기획판 2: 자율 및 지능 시스템으로 인간의 복지를 최우선으로 하는 비전 (Ethically Aligned Design version 2: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems)	IEEE EAD ver. 2	자율 및 지능 시스템의 윤리에 관한 IEEE 글로벌 이니셔티브 (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems)	개발자, 정책 입안자	2017
③ 윤리적 내용 기획판(초판): 자율 및 지능 시스템으로 인간의 복지를 최우선으로 하는 비전 (Ethically Aligned Design (first edition): A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems)	IEEE EAD 1e	자율 및 지능 시스템의 윤리에 관한 IEEE 글로벌 이니셔티브 (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems)	개발자, 정책 입안자	2019
④ 신뢰할 수 있는 AI를 위한 윤리지침 (Ethics Guidelines for Trustworthy AI)	신뢰할 수 있는 AI (Trustworthy AI)	EU 집행위원회의 인공에 관한 고위급 전문가 그룹 (The European Commissions High-Level Expert Group on Artificial Intelligence)	개발자, 정책 입안자	2019

지침명	약어	제정 단체	활용대상	공표 시기
⑤ 인공지능 위원회의 권고 OECD/LEGAL/044914 (Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/044914)	OECD 권고 (OECD Recom-mendation)	OECD	정책 입안자	2019

2. 주요 내용

아실로마 AI 원칙(Asilomar AI Principle)[46]은 AI 연구 및 개발을 위한 23가지 지침이다. 이 지침은 유익한 AI 개발을 위한 AI 개발 문제, 윤리 및 지침을 설명하고 유익한 AI 개발을 위해 제정되었다. 이 지침은 2017년 캘리포니아 퍼시픽 그로브에서 열린 유익한 AI에 관한 Asilomar 회의에서 만들어졌다. Asilomar AI 원칙은 연구(Research), 윤리 및 가치(Ethics and Values), 미래 문제(Longer-Term Issues)의 3가지 범주로 세분화된다고 할 수 있다.

먼저 연구(Research)는 인공지능 연구의 목표는 방향성이 없는 지능이 아니라 유익한 지능을 만드는 것이어야 하며, 연구 자금에 대해 AI에 대한 투자에는 유익한 사용을 보장하기 위한 연구 자금이 수반되어야 함을 의미한다. 그리고 윤리 및 가치(Ethics and Values)는 안전성 확보를 위해 AI 시스템은 운영 내내 안전해야 하며 적용이 가능하고, 검증이 가능해야 한다고 한다. 그리고 장애 투명성에 대해 AI 시스템에 문제가 발생한다면, 그 원인을 규명할 수 있어야 한다고 하였다. 개인정보 보호에 대해 AI 시스템이 해당 데

46) Future Life Institute: ASILOMAR AI PRINCIPLES, 〈https://futureoflife.org/ai-rinciples/〉(last visited Feb. 31, 2022).

이터를 분석하고 활용할 수 있는 경우, 개인이 생성한 데이터에 액세스, 관리 및 제어할 수 있는 권한이 있어야 한다고 하였다. 미래 문제(Longer-Term Issues)는 기능주의의 문제점에 대해 합의가 없는 한, AI 기능의 한계돌파에 대한 강한 가정을 피해야 함을 요청하고 있다. 그리고 AI 시스템에 의해 발생하는 위험, 특히, 재난 수준의 위험은 예상되는 영향에 상응하는 계획 및 완화 노력의 대상이 되어야 한다고 한다.

IEEE EAD ver. 2[47]는 IEEE EAD 1e[48]보다 자율 및 지능 시스템을 통한 인간의 웰빙, 일반인이 지침에 대한 피드백을 제공하거나 표준 작업 그룹에 참여하도록 권장하기 위해 행동하는 윤리 캠페인을 만들었다. 특히, IEEE EAD ver. 2에서는 자율 및 지능형 시스템의 윤리에 관한 IEEE 글로벌 이니셔티브(global initiative)의 13개 전문가 위원회의 의견을 반영하였다. 지침의 초판이 발표된 이후 5개의 새로운 위원회가 추가되었다. AI 시스템 분야의 250명 이상의 글로벌 사고 리더와 전문가의 의견을 모아 만들어졌다. IEEE 글로벌 이니셔티브(global initiative)와 여러 관련 연구 분야로서 기계 학습(Machine Learning), 인공지능, 지능형 시스템 엔지니어링 등을 더 잘 다룰수 있도록 설계에서 윤리적 고려 사항을 적용할 수 있도록 보완되었다.

IEEE EAD ver. 2의 I. 일반 원칙(General Principles), II. 자율 지능 시스템에 가치 내장(Embedding Values into Autonomous Intelligent

47) The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design version 2: A Vision for Prioritizing Human Well-being with Autonomous and with Autonomous and Intelligent Systems,* 2019. p.7.

48) The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems 2017), *ETHICALLY ALIGNED DESIGN A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems,* 2017, p.8.

Systems), III. 윤리적 연구 및 디자인을 안내하는 방법론(Methodologies to Guide Ethical Research and Design), IV. 인공 일반 지능(AGI) 및 인공 초지능(ASI)의 안전과 이점(Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)), V. 개인 데이터 및 개별 액세스 제어(Personal Data and Individual Access Control), VI. 자율 무기 시스템 재구성(Reframing Autonomous Weapons Systems), VII. 경제/인도적 문제(Economics/Humanitarian Issues), VIII. 법률(Law), IX. 감성 컴퓨팅(Affective Computing), X. 정책 목표(Policy Objectives), XI. A/IS의 고전 윤리(Classical Ethics in A/IS), XII. 정보 통신 기술(ICT)의 혼합 현실(Mixed Reality in Information and Communication Technology (ICT)), XIII. 웰빙(Well-being)의 내용을 갖고 있다.

IEEE EAD ver. 2는 IEEE EAD 1e 보다 자율 및 지능 시스템으로 인간의 복지를 최우선으로 하는 비전(Ethically Aligned Design (first edition): A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems)에서 없었던 IX. 감성 컴퓨팅(Affective Computing), X. 정책 목표(Policy Objectives), XI. A/IS의 고전 윤리(Classical Ethics in A/IS), XII. 정보 통신 기술(ICT)의 혼합 현실(Mixed Reality in Information and Communication Technology (ICT)), XIII. 웰빙(Well-being)이 추가되었다.

신뢰할 수 있는 AI를 위한 윤리지침(Ethics Guidelines for Trustworthy AI)[49]은 AI가 가져야 하는 세 가지 구성 요소를 제시하고 있다. 첫째, 모든 관련 법규를 준수하며, 적법해야 한다. 둘째, 윤리적

49) The European Commissions High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI, 〈https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai〉(last visited Feb. 31, 2022).

〈표 2〉 해외 주요 AI 윤리지침의 주요 내용

지침명	제정 목적	주요원칙	주요특징
① 아실로마 AI 원칙 (Asilomar AI Principle)	미국 보스톤의 비영리 연구단체인 삶의 미래 연구소 (Future of Life Institute) 주관으로 작성한 윤리원칙	인권보장, 개인정보보호, 해악금지, 공공성, 데이터 관리, 책임성, 통제성, 투명성, 무기경쟁 회피	· 스티븐 호킹·일론 머스크 등 다수의 AI 학자, 미래학자 및 산학연 관계자들이 서명 · AI 기술 연구자, 정책 입안자, 관련산업종사자에게 필요한 윤리원칙 제시
② IEEE EAD ver. 2 ③ IEEE EAD 1e	IEEE에서 Ethics in Action 캠페인과 함께 아울러 공개된 보고서	인권, 복지우선, 책무성, 투명성, 오용의 인식	· 각 원칙별로 이론적 배경, 참고 자료를 제시하고 윤리원칙 뿐만 아니라 관련 분야들에 대한 자료 수록
④ 신뢰할 수 있는 AI (Trustworthy AI)	EU 산하의 50여명으로 구성된 AI 전문가 그룹 주도	인간 권리·자율성 보장, 기술적 견실성, 사생활, 데이터 관리, 투명성, 다양성, 차별금지, 복지, 책무성	· 범국가 차원의 협업을 통해 신뢰할 수 있는 AI를 위한 윤리원칙 정립에 초점을 맞춤 · 각 원칙의 평가 리스트를 구체적으로 제시
⑤ OECD 권고 (OECD Recommendation)	OECD 디지털 경제 정책 위원회 주관하에 제작	포용적 성장, 지속가능 발전, 인간중심 가치, 공정성, 투명성, 설명가능성, 견고성, 보안 및 안전, 책무성	· 윤리원칙 뿐 아니라 정책 입안자들에 대한 제언 제시, 국가별 정책수립과 국제적 협력 도모

원칙과 가치 준수를 보장해야 한다. 셋째, AI 시스템이 좋은 의도를 가지고 있더라도 의도하지 않은 피해를 유발할 수 있으므로, 기술 및 사회적 관점에서 모두 강력해야 함을 언급하고 있다.

OECD 권고(OECD Recommendation)[50]는 AI에 대한 최초의 정부 간 표준인 인공 지능(AI)에 대한 권고로서 디지털 경제 정책 위원회(CDEP)의 제안에 따라 2019년 5월 22일에 OECD 이사회에서 채택되었다. 이 권고는 인권과 민주적 가치에 대한 존중을 보장하면서, 신뢰할 수 있는 AI의 책임 있는 관리를 독려함으로써 AI에 대한 혁신과 신뢰를 촉진하는 것을 목표로 한다. 개인정보 보호, 디지털 보안 위험 관리 및 책임 있는 비즈니스 행동과 같은 영역에서 기존 OECD 표준을 보완하는 이 권고안은 AI 관련 문제에 초점을 맞추고, 새로운 기술이 구현이 가능하도록 유연한 표준을 설정하였다. 또한 이 권고안은 신뢰할 수 있는 AI의 책임 있는 관리를 위한 5가지 상호보완적인 가치 기반 원칙을 식별하고, AI 행위자가 이를 촉진하고 구현하도록 촉구하였다. 구체적으로 포용적 성장, 지속 가능한 개발 및 웰빙, 인간 중심의 가치와 공정성, 투명성 및 설명 가능성, 견고성, 보안 및 안전성, 그리고 책임에 대하여 언급하고 있다.

3. 국내 '인공지능(AI) 윤리기준'의 주요 내용

우리나라의 '인공지능(AI) 윤리기준'은 윤리적 AI의 구현을 위해 정부·공공기관, 기업, 이용자 등 모든 사회 구성원이 함께 지켜야 할 주요 원칙과 핵심 요건을 제시하고 있다. 사실 AI 기술의 발전·확산과 함께 AI 기술의 윤리적 개발·활용 역시 세계 각국과 주요

50) OECD(2021), *Recommendation of the Council on OECD Legal Instruments Artificial Intelligence,* 2021, p.7.

국제기구의 관심 대상이 되어 왔으며, 2019년에 우리나라가 주도적으로 참여한 경제협력개발기구(OECD) 인공지능 권고안('19.5)을 비롯하여 OECD, 유럽연합(EU) 등 세계 각국과 국제기구, 기업, 연구기관 등 여러 주체로부터 다양한 인공지능 윤리 원칙이 발표되었다.[51]

이에 우리나라는 글로벌 추세에 발맞추어 '인공지능(AI) 윤리기준' 마련을 추진해왔다. '인공지능(AI) 윤리기준'은 AI·윤리 전문가로 구성된 인공지능 윤리연구반을 통해 국내외 주요 AI 윤리원칙을 분석하고, 그 결과를 윤리 철학의 이론적 논의와 연계하여 '인간성을 위한 인공지능(AI for Humanity)'를 목표로 하는 윤리기준 초안을 마련하였으며, 3개월에 걸쳐 학계·기업·시민단체 등 각계의 다양한 전문가로부터 의견을 수렴하였다.[52]

이러한 과정을 거쳐 마련된 '인공지능(AI) 윤리기준(안)'은 '사람중심의 인공지능'을 위한 최고 가치인 '인간성(Humanity)'을 위한 3대 기본원칙과 10대 핵심 요건을 제시하고 있으며, 주요 내용은 다음과 같다.[53]

첫째, 목표 및 지향점은 모든 사회 구성원이 모든 분야에서 자율적으로 준수하며 지속 발전하는 윤리기준을 지향해야 한다. AI 개발에서 활용에 이르는 전 단계에서 정부·공공기관, 기업, 이용자 등 모든 사회 구성원이 참조하는 기준이다. 또한, 특정 분야에 제한되지 않는 범용성을 가진 일반 원칙으로, 이후 영역별 세부 규범이 유연하게 발전해나갈 수 있는 기반 조성하였다. 구속력 있는 '법'이나

51) 관계부처 합동, 사람이 중심이 되는 인공지능을 위한 신뢰할 수 있는 인공지능 실현 전략(안), 2021.5, 3면.
52) 관계부처 합동, 사람이 중심이 되는 「인공지능(AI) 윤리기준」, 2020.12, 5면.
53) 관계부처 합동, 앞의 자료(주 51), 5면.

‘지침’이 아닌 도덕적 규범이자 자율규범으로, 기업 자율성을 존중하고 AI 기술발전을 장려하며 기술과 사회변화에 유연하게 대처할 수 있는 윤리 담론을 형성하였다. 사회경제, 기술 변화에 따라 새롭게 제기되는 인공지능 윤리 이슈를 논의하고 구체적으로 발전시킬 수 있는 플랫폼으로 기능을 하고 있다.

둘째, 최고 가치는 윤리기준이 지향하는 최고 가치는 ‘인간성(Humanity)’으로 설정하고, ‘인간성을 위한 인공지능(AI for Humanity)’을 위한 3대 원칙·10대 요건 제시하였다. 3대 기본원칙은 ‘인간성(Humanity)’을 구현하기 위해 인공지능의 개발 및 활용 과정에서 인간의 존엄성 원칙, 사회의 공공선 원칙, 기술의 합목적성 원칙을 지켜야 한다. 그리고 10대 핵심 요건은 3대 기본원칙을 실천하고 이행할 수 있도록 AI 개발~활용 전 과정에서 인권 보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성의 요건을 충족해야 한다.

‘인공지능(AI) 윤리기준’에서 AI의 지위는 ‘인간성을 위한 인공지능(AI for Humanity)’을 지향하며, AI가 인간을 위한 수단임을 명시적으로 표현하지만, 인간중심주의(human species-centrism) 또는 인간 이기주의를 표방하지는 않음을 언급하고 있다. 또한 AI는 지각력이 있고 스스로를 인식하며 실제로 사고하고 행동할 수 있는 수준의 강인공지능(strong AI)을 전제하지 않으며 하나의 독립된 인격으로서의 인공지능을 의미하지도 않음을 규정하고 있다.[54]

적용 범위와 대상은 AI의 개발부터 활용에 이르는 전 단계에 참여하는 모든 사회 구성원을 대상으로 하며, 이는 정부·공공기관, 기업, 이용자 등을 포함하며, AI 윤리기준의 실현방안으로 AI 윤리기

54) 관계부처 합동, 앞의 자료(주 51), 15면.

준을 기본 플랫폼으로 하여 다양한 이해관계자 참여하에 인공지능 윤리 쟁점을 논의하고, 지속적 토론과 숙의 과정을 거쳐야 함을 요청하고 있다.

V. 결(結)

AI 기술 윤리는 미래를 향한 규범이 될 것이다. 미래와 현재의 차이는 시간적 차이이지만, 그 시간의 양에 대해서는 알 수 없다. 이는 미래의 발전 정도를 정량화할 수 없다는 의미이다. 과학기술이 발전하는 이상, 윤리도 시대와 함께 변화한다. 보편성을 기대할 수 없는 윤리라는 사회규범을 적용할 수 없고, 인공지능이 해야 할 행동을 바람직한 형태로 제어할 수 있어야 한다. 이에 이를 윤리라고 명하지 않고, 법으로 지칭할 수 있다. 그러나 AI를 비롯한 첨단과학에서 요구되고 있는 것은 인류가 전통적으로 구분해 온 사회규범으로서의 윤리이다. 이 전통적인 윤리라는 사회규범에 새로운 요소를 통합함으로써 보다 나은 윤리가 태어날 것으로 기대하고 싶다. 이를 위해서도 이 분야의 연구는 인공지능뿐만 아니라 철학이나 인지과학 등 학제 간 연구가 요구된다.

AI와 관련한 윤리적 쟁점은 매우 넓다. 아직 우리가 예측하지 못한 사례들이 있다. 이는 어쩌면 새로운 사회 구조를 만들고 있기 때문이며, 앞으로 만들어진 사회가 AI를 새로운 주체로 가늠해야 한다. 윤리적 판단이 인간이 갖는 공감이나 비공감과 같은 의사결정에 영향을 준다면, AI의 윤리학습은 감정 상태에 대한 모의실험이 필요하게 된다. 아직 인공지능 윤리가 기술적으로 깊이 있는 연구 단계가 도달하지 못했지만, 오히려 인간의 이기심으로 제한적 지능

을 가진 시스템으로도 여러 가지 이슈들이 발생할 것이다.

다만, 아직은 그 결과가 사회에 큰 피해를 주거나 파국을 가져오는 것은 아니고 일부 사람들에게 차별을 가하거나, 반사회적 행동을 유발하고, 오해와 확대 해석으로 인한 오용의 문제가 발생할 것으로 본다. 초기의 많은 문제는 AI를 사용하는 또는 같이 공존해야 하는 인간이 갖는 특성 때문에 발생할 것이다. 이는 인류 초기부터 우리가 갖는 특성인 의인화와 감정 애착 또는 이입이라는 특징 때문에 발생한다.

AI 윤리는 기술의 특수성 및 실제 적용 맥락, 다양한 이해관계자를 고려하는 윤리형태를 가져야 한다. 윤리가 기술의 기획, 형성, 배치 등에 이미 녹아있는 것임을 받아들인다면, AI 윤리는 기술의 개발 및 사회 도입에 앞서 더욱 선제적이고 적극적으로, 그리고 기술의 전체단계와 함께 수행되어야 할 것이다. 또한, 인공지능에 관한 윤리적 연구가 AI 기술에 매몰되지 않으면서 동시에 새로운 기술 및 변화를 이해하는 윤리 논의가 되려면 가치, 방향성, 정당화를 고민해야 할 것이다.

참 고 문 헌

1. 국내문헌

김윤명, “지능정보사회에 대한 규범적 논의와 법정책적 대응”, 정보화정책 제23권 제4호, 한국지능정보사회진흥원. 2016년 겨울호.

박기주, “인공지능 알고리즘을 활용한 전문(추천) 서비스 제공의 법적 성격에 관한 연구”, 「법제논단」 2020년 3월호, 법제처.

박종호, “인공지능 시대의 윤리와 법적 과제”, 「과학기술법연구」 제24집 제3호, 한남대학교 과학기술법연구원, 2018.

변순용, “데이터 윤리에서 인공지능 편향성 문제에 대한 연구”, 「윤리연구」 제128호, 한국윤리학회, 2020.

서형준, “4차 산업혁명시대 인공지능 정책의사결정에 대한 탐색적 논의”, Informatization Policy Vol. 26, No.3, 한국정보화진흥원, 2019.

설선혜·이승민, “도덕 판단에서 나타나는 도덕-인습 구분에 대한 논쟁과 함의”, 「인지과학」, 제29권 제2호, 한국인지과학회, 2018.

송성수, 「과학기술자의 사회적 책임과 윤리」, 과학기술정책연구원, 2001.

신용우, “인공지능 관련 입법 현황 및 전망”, 「NARS 현황분석」 제87호, 국회입법조사처, 2019.

장재옥, 김현희, “인공지능의 법적 지위에 관한 논의”, 「법학논문집」 제43집 제1호, 중앙대학교 법학연구원, 2019.

정도범, 유화선, “인간과 인공지능(AI)의 공존을 위한 사회·윤리적 쟁점 : 신뢰할 수 있는 인공지능 실현 방안”, KISTI ISSUE BRIEF 제35호, 2021.11.

정석우, 심현철, “자율주행 자동차의 인공지능”, 「기계저널」 제57권 제3호, 대한기계학회지, 2017.

정지선, 김동성, 이홍주, “텍스트 마이닝 기법을 활용한 인공지능 기술개발 동향 분석 연구: 깃허브 상의 오픈 소스 소프트웨어 프로젝트를 대상으로”, 「지능정보연구」 제25권 제1호, 2019.

최민수, “인공지능 로봇의 오작동에 의한 사고로 인한 불법행위책임”, 「민

사법의 이론과 실무」 제23권 제3호, 민사법의 이론과 실무학회, 2020.
홍용희, “과학기술과 윤리의 상관성”, 「윤리연구」 제72호, 한국윤리학회, 2009.
황경식, “도덕체계와 사회구조의 상관성”, 「철학사상」 제32조, 서울대학교 철학사상연구소, 2009.
관계부처 합동, 사람이 중심이 되는 인공지능을 위한 신뢰할 수 있는 인공지능 실현 전략(안), 2021.5.
관계부처 합동, 사람이 중심이 되는 「인공지능(AI) 윤리기준」, 2020.12.

2. 외국문헌

RAINIE, LEE., ANDERSON, JANNA AND VOGELS EMILY A., Worries about developments in AI, PEW RESEARCH CENTERJUNE 16, 2021, <https://www.pewresearch.org/internet/2021/06/16/1-worries-about-developments-in-ai/>(last visited Feb. 31, 2022).

Johnson, David Kyle, "GunsDon’t Kill People, People Do?", <https://www.psychologytoday.com/us/blog/logical-take/201302/guns-don-t-kill-people-people-do>(last visited Feb. 31, 2022).

Directorate-General for Internal Policies of the Union (European Parliament), *ARTIFICIAL INTELLIGENCE AND CIVIL LIABILITY*, 2020.

YURTSEVER, EKIM., LAMBERT, JACOB., CARBALLO, ALEXANDER., TAKEDA, KAZUYA., A Survey of Autonomous Driving: Common Practices and Emerging Technologies, *IEEE Journals & Magazines*, VOL. 8, 2019.

European Parliamentary Research Service, *The ethics of artificial intelligence: Issues and initiativesm*, European Parliamentary, 2020.

Hofmann, Eva., Hartl, Barbara., Katharina, Gangl., Hartner-Tiefenthaler, Martina and Kirchler, Erich., Authorities' Coercive and Legitimate Power: The Impact on Cognitions Underlying Cooperation, Front. Psychol., 18 January 2017. <https://www.frontiersin.org/articles/10.3389/fpsyg. 2017. 00005/full>(last visited Feb. 31, 2022).

Kraemer, Felicitas., Kees van Overveld & Peterson, Martin., Is there an ethics of algorithms?, *Ethics and Information Technology* Vol. 13, 2011.

Martínez-Plumed, Fernando., Gómeza, Emilia., José Hernández-Orallob, Futures of artificial intelligence through technology readiness levels, *Telematics and Informatics*, Vol. 58, May 2021.

Future Life Institute: ASILOMAR AI PRINCIPLES, <https://futureoflife.org/ai-principles/>(last visited Feb. 31, 2022).

Human-Centered Artificial Intelligence Institute(HAI), A*rtificial Intelligence Index Report 2019 AI Index Report–Highlights*, 2019.

Harrington, Jake., McCabe, Riley., Detect and Understand: Modernizing Intelligence for the Gray Zone, <https://www.csis.org/analysis/detect-and-understand-modernizing-intelligence-gray-zone>(last visited Feb. 31, 2022)

Kaplan. Jerry., *Artificial Intelligence: What Everyone Needs to Know*R *1st Edition*, Jenson Books Inc., 2019.

Manheim, Karl M., Kaplan, Lyric., Artificial Intelligence: Risks to Privacy and Democracy, *21 Yale Journal of Law and Technology* 106, 2019.

Gloor, Lukas., *Suffering-focused AI safety: In favor of "fail-safe" measures, Suffering-focused AI safety—Center on Long-Term Risk,* 2016.

Ryan, Mark., Stahl. Bernd Carsten., Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications, *Journal of Information, Communication and Ethics in Society*, 2020.

Shanahan, Murray., *The Technological Singularity*, The MIT Press Essential Knowledge series, 2015.

Ellemers. Naomi., Jojanneke van der Toorn, Paunov, Yavor., The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017, *Pers Soc Psychol Rev*, Vol. 23 No.4, 2019.

Lee, Nicol Turner., Resnick, Paul., and Barton, Genie., "Algorithmic bias detection and mitigation: Best practices and policies to reduce

consumer harms", Brookings Institution <https://www.brookings.edu/research/algorithmic-bias-detection-andmitigation-best-practices-and-policies-to-reduce-consumer-harms>

OECD, *Recommendation of the Council on OECD Legal Instruments Artificial Intelligence*, 2021.

FeldmanHall, Oriel., Son, Jae-Young., and Heffner, Joseph., Norms and the Flexibility of Moral Action, Personal Neurosci. 2018, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7219684/>(last visited Feb. 31, 2022).

Ernest, Paul., Mathematics, ethics and purism: an application of MacIntyre's virtue theory, *Synthese*, Vol. 199, 2021.

Sinha, Sudhi., Metaethics, Meta-Intelligence And The Rise Of AI, <https://www.forbes.com/sites/forbestechcouncil/2021/01/21/metaethics-meta-intelligence-and-the-rise-of-ai/?sh=2a2992aa46e2>(last visited Feb. 31, 2022)

The European Commissions High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>(last visited Feb. 31, 2022).

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *ETHICALLY ALIGNED DESIGN A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, 2017.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design version 2: A Vision for Prioritizing Human Well-being with Autonomous and with Autonomous and Intelligent Systems*, 2019.

Sima, Violeta., Gheorghe, Ileana Georgiana., Subi´c, Jonel, Nancu, Dumitru., Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review, *Sustainability* 2020, Vol. 12, No 10.

村上祐子, 人工知能の倫理の現在—研究開発における技術哲学‧倫理の意義—, IEICE Fundamentals Review Vol.11 No.3, 2018.

3. 기타자료

안전벨트의 역사, <https://www.khan.co.kr/economy/auto/article/201011051115302> (최종방문일 2022년 1월 31일)

오른쪽에 핸들이 있는 자동차 이야기, <http://www.kama.or.kr/jsp/webzine/201803/pages/story_02.jsp>(최종방문일2022년 1월 31일)

<Abstract>

A Study of artificial intelligence technology from an ethical and legal perspective

Lee, Chang-Kyu*

At the development stage of artificial intelligence technology, there is a strong tendency to rely on the social norm of ethics to determine the direction of development. And in order not to go in the wrong direction, it is necessary to give meaning to the norm of ethics and clarify the limits. Ethical standards must be applied from design to commercialization in order to utilize rational artificial intelligence technology.

In fact, discussions about the dangers of artificial intelligence technology must be discussed in the legal category, but require minimal principles in areas and ethical norms that require regulation to balance the development of science and technology with regulation. It is necessary to divide into areas to be treated. right. Although the danger has not yet emerged and its impact has not been assessed, it can be said that preemptive regulation does not meet the purpose of the law. Therefore, it is necessary to estimate in advance the problematic factors of artificial intelligence technology and present an ethical code for them.

Due to the characteristics of artificial intelligence technology for data machine learning, there are also concerns about adverse functions such as lack of safety and misuse of artificial intelligence. Artificial intelligence technology is based on machine learning, which learns large amounts of data to improve performance, has uncertainty and opacity, and can cause errors in noise data. In addition, there are concerns that artificial intelligence

* Professor for Special Affairs at Ajou Univ., S.J.D., technology transfer agent

technology can be deliberately abused while surpassing human capabilities in certain areas. At the same time, social problems such as human labor substitution by artificial intelligence, privacy invasion, and deepening of polarization can appear.

Artificial intelligence ethics must have a form of ethics that takes into account the peculiarities of the technology, the context of its actual application, and various stakeholders. If we accept that ethics is already integrated into the planning, formation, placement, etc. of technology, artificial intelligence ethics will be carried out more preemptively, positively and with all stages of technology prior to the development of technology and the introduction of society. Will have to. Also, in order for ethical research on artificial intelligence to become an ethical debate that understands new technologies and changes without being buried in artificial intelligence technology, we will have to worry about value, direction, and justification.

Key Words : artificial intelligence, ethics, algorithms, social norms, ethical norms