



인공지능의 규범구체화에 대한 연구*

이 희 옥**

인공지능의 혁신은 인류의 일상에 극적인 변화를 주고 있다. 머신러닝, 자연어의 이해, 음성·안면 인식, 클라우드 컴퓨팅, IoT, 빅데이터는 수십 년 동안 더디게 발전해 온 인공지능을 지능정보화시대를 향한 문턱에 올려놓았다. 오늘날 IT와 금융, 의료, 등 많은 산업이 기술로 연결된 디지털마켓으로 변모하고 있다. 인공지능의 의사결정시스템은 인간을 소모적인 일들로부터 자유롭게 하는 이점이 있지만, 기존 사회의 위험을 다른 차원에서 가중하기도 한다. 알고리즘의 공정성에 대한 의구심에도 불구하고, 인공지능 의사결정시스템은 정부의 행정행위에도 활용될 예정이다. 기술사회가 발전함에 따라 인공지능 의사결정시스템의 확대는 필연적이다. 그러나 인공지능 의사결정은 인간을 객체화하여 결과를 도출하는 시스템이라는 점에서 개인에게 중대한 영향을 미치는 결정까지도 인공지능에만 의존하게 된다면, 개인의 자기결정권은 위축될 수 있다. 이에 기술사회의 위험은 인공지능의 의사결정이 인간의 존엄과 자유에 근간한 '자기결정권'의 보장에 무게를 두고 접근되어야 한다.

한편 근래의 주요 국가에서는 인공지능에 대한 포괄적 규제입법을 하기보다는 윤리원칙을 설정하고 기술환경을 지켜보고 있다. 마찬가지로 국내에서도 인공지능에 대응한 규범의 문제를 구체적으로 다루지는 못하고 있다. 이는 법이 급변하는 기술에 앞서갈 수 없다는 현실을 반영하기도 하지만, 지금이라도 일상적인 경험들로 미래사회의 명암(明暗)을 진단하여 법적 대응을 모색하는 일은 가능하다.

이에 이 글은 인공지능에 대응한 윤리의 규범적 타당성을 살펴보고, 인공지능에 대한 법적 규율의 필요성을 논증하였다. 또한 입법 이전에 고려할 사전적 규제모델로서 '이원적 규제모델'과 개인의 자기결정의 전제조건을 돕는 '자기결정 보호모델'을 제시함으로써 인간의 자기결정권의 보장을 꾀하였다. 헌법적으로도 중요한 '절차적 적법절차'는 '데이터 처리'와 '알고리즘 의사결정의 처리'에서 투명성, 공정성, 그리고 합리성의 원리로서 당사자 간 이익조정을 꾀할 수 있다. 인공지능에 대한 규제는 인간존엄과 자기결정의 가치를 보장하면서도, 동시에 이를 제한하는 결과를 낳는다. 아울러 지금 논의되는 코드와 같은 인공지능의 규율 방안 등도 법적 규제로서 얼마만큼의 효용성이 있을지 예측하기 어렵다. 이 글은 인공지능에 대응해 헌법상 자기결정권의 개념을 구체화한 규범구체화의 시도를 통해서 남겨진 향후의 과제들에 대해서도 시론적으로 살펴보았다.

* 이 글은 필자의 박사학위 논문 “인공지능의 의사결정에 대응한 자기결정권의 보장에 관한 연구”의 일부 내용을 요약하여 발전시킨 것이다.

** 네이버 주식회사, 법학박사, moma02020@gmail.com

I. 들어가는 말

인공지능이 신기술 산업과 국가경쟁력의 핵심으로 여겨지면서 세계 각국의 이목을 끌고 있다.¹⁾ 구글을 위시한 글로벌 IT 기업들과 주요국가들에서는 인공지능이 가져올 특정할 수 없는 미래사회에 대하여 인간과 인공지능의 공존을 모색하는 방책으로 인공지능 윤리원칙을 하나, 둘씩 꺼내고 있다. 그러는 중에서도 각국은 인공지능기술의 선점에 정책의 주안을 두고 있으니, 향후에도 인공지능을 둘러싼 이해관계는 첨예하게 얽혀지게 될 것으로 전망된다.²⁾ 이러한 상황에서 인공지능에 대응한 규범론이 인공지능 윤리담론에 머물러 있는 점은 인공지능이 이제 겨우 시작된 기술이라는 점과 본래 법이 사실관계의 확정에 기초해 사후적으로 마련되는 규범이라는 점에서 수궁이 되는 바가 크다. 그러나 한편으로는 국가에서 법의 역할이 국민의 기본권과 인간존엄의 보장에 있음을 생각할 때, 인공지능에 대응한 규율을 외부로부터 제시된 윤리규범에만 기대어 무방비로 있다가 인간의 개입이 불가능한 기술의 특이점에 도달한 다음에서야 속수무책이 될까 우려스럽다. 특히 인공지능 윤리는 글로벌 사업자들이 기반한 미국의 논리가 될 수도 있다는 점에서 비판적 검토는 필요하다. 이에 기술의 특이점을 예측할 수 없는 시점에서 우리나라의 인공지능에 대한 법정정책적 관점은 ‘기술의 발전과 자국민의 보호’의 프레임에서 균형 있게 형성되어야 할 것으로 보인다.

이 글은 인공지능 윤리규범 못지않게 법학에서도 입법 이전에 사전적 규범에 대한 연구가 선행되어야 한다는 문제의식에서 출발해 인공지능에 요구되는 규범구체화에 이르는 논의를 해나가고자 한다. 특히 인공지능의 불투명성, 예측곤란성, 알고리즘의 개방성

1) 주요국가들의 인공지능과 로봇에 대한 관심은 2010년에 와서 뜨겁게 달구어지고 있다. 독일의 ‘Industry 4.0’은 사이버물리시스템(CPS)을 기반으로 스마트 공장을 구축하려는 독일 정부의 차세대 제조업 발전전략으로, 제조업과 ICT를 연결해 스마트 공장을 구축하는 것을 목표로 하였다. 미국의 오바마 행정부는 대통령 과학기술자문위원회의 권고로 2013년 ‘첨단제조 파트너십(AMP)’을 발표하고, 산·학·관 협력으로 산업 로봇(Co-Robot) 중심의 제조업 혁신전략을 실행해왔다. 또한 일본 정부도 로봇혁명을 경제성장전략의 핵심으로 10대 과제로 선정해 총리실 산하에 ‘로봇혁명회의’를 출범시키고 2015년에는 ‘로봇 신전략’ 계획을 발표해 로봇활용을 확대하려는 방침이다. 유럽연합(EU)에서도 제조, 농업, 헬스, 교통, 사회안전 등 다른 산업과의 융합을 통해 로봇시장을 선점하려는 전략으로 세계 최대 규모의 민간주도 로봇연구 프로그램(SPARC)을 운영하고 있다.

2) 차정미, “기술 패권의 미래”; 양화융합을 통한 중국의 혁신주도 성장, 그리고 핵심기술, 미래연구 포커스, Future Horizon(2018, 봄호), 15면.

및 확장가능성이 초래할 기술적 특성을 고려한 가치의 규범화와 다양한 이해관계에 따른 상충 문제와 환경에서 규범으로서 인공지능 윤리원칙에 대한 타당성을 살펴보고자 한다(II). 나아가 현재 고려해볼 수 있는 규범적 구체화의 시도로서 데이터와 알고리즘에 관한 규제모델과 자기결정 보호모델을 소개하고자 한다(III). 인공지능에 대한 규제는 인간존엄과 자기결정의 가치를 보장하면서도, 제한하는 결과를 낳기도 한다. 또한 사전적 규제를 행정기관이 주도하고, 강제화하는 경우에는 사전심의의 문제가 생길 수 있어 규제가 추구하는 공익에 대한 세심한 이익형량이 필요하다. 이에 마지막 단락에서는 결론을 대신하여 이같이 현재로서는 완전히 해결할 수 없는 인공지능 규제에 관한 문제점들을 시론적으로 살펴보고자 한다(IV).

II. 인공지능 규범론에 대한 논의

1. 인공지능 윤리론의 등장과 현황

1) 과학기술 가치중립성의 착시

과학이 일반적으로 자연 세계에 대한 진리나 보편적인 법칙의 발견을 목적으로 한다면, 기술은 과학적 지식을 활용하여 인간의 삶에 유용하도록 가공하는 수단을 말한다.³⁾ 과학기술은 과학과 기술의 합성어로 기술의 발전을 위해서는 과학이론과 그 발전이 필요하고, 기술에 종사하기 위해서는 과학지식이 필요하다. 과학과 기술은 본질적으로 동일하지 않고 역사적으로 볼 때 사회적 지위에도 차이가 있다고 볼 수 있다.⁴⁾ 오늘날 과학과 기술의 관계에 대한 평가는 일률적이지 않지만,⁵⁾ 20세기에 와서 양자는 국방,

3) 위키백과에서 '과학'의 검색결과를 참조,

<https://ko.wikipedia.org/wiki/%EA%B3%BC%ED%95%99>(2020. 11. 11. 최종방문).

4) 과학은 현상을 체계화하고 분석해 그에 대한 설명을 이론화한다. 과학자들에게는 실제적 효용이나 일반 대중의 반응보다 동료학자들의 인정이 중요하다면 기술은 일관성이나 합리성의 기준으로 기술 발전을 체계화하는 것보다 경제적 가치에 무게를 둔다. 제임스 매클렐란 3세·해럴드 도른(전대호 옮김), 『과학기술로 본 세계사 강의』 모티브(2006), 545면.

5) 첫째, 과학이론을 제품이나 공정에 응용하는 융합 관계이다. 둘째, 과학과 기술은 약한 정도의 시차를 두고 상호작용하는 연결 관계에 있다. 셋째, 기술이 과학과 분리되어 별도의 독자성·독립성을 유지하는 분리 관계에 있다. 제임스 매클렐란 3세·해럴드 도른(전대호 옮김), 앞의 책, 558-560면.

의료, 우주개발, 컴퓨터, 인공지능에 이르기까지 사실상 통합되었다고도 볼 수 있다. 우리 헌법 제127조는 과학기술 조항을 통하여 ‘과학기술’이 국민경제의 발전에 중대한 영향을 미치는 것으로 보고 있다. 더욱이 오늘날 과학기술이 자연을 포함한 인간생활에 미치는 영향력은 이전보다 커졌으며, 과학자-이용자의 관계가 독립적이지 않다.⁶⁾ 이제 과학기술은 어떠한 사회적 결과를 만들었는지에 따라 그 가치가 결정될 수 있으며,⁷⁾ 과학기술과 자율성과 독자성, 즉 중립성은 환상에 가깝다고 여겨진다. 이와 같은 맥락에서 인공지능은 인간 의사결정자가 아닌 인공지능의 제조자, 운영자의 통제권에 놓여 있다. 이와 같은 상황에서 인공지능 윤리론의 등장하였다.

2) 인공지능윤리 담론의 배경과 내용

가. 인공지능 윤리의 배경

본격적인 인공지능윤리 담론은 1942년 공상과학 소설작가인 Isaac Asimov(아이작 아시모프)의 단편소설 「Run around」에서 소개한 로봇 제3원칙에서 비롯되었다고 할 수 있다.⁸⁾ 이후 로봇에 관한 윤리 담론은 인공지능이 집중 조명된 2010년 전후로 본격적으로 시작되었다.⁹⁾ 실제로 로봇은 인간에게 해악을 초래할 여지가 있다. 그러나 그것은 사용자의 부주의나 기술적 결함에서 비롯된 물리적인 속성의 것이지, 영화 속에서 기계의 인간에 대한 공격과 같은 공포는 다분히 과장된 점이 있다.¹⁰⁾ 오히려 일본에서

6) 과거 Alfred Nobel(1833~1896)은 니트로글리세린을 보다 안전하게 사용하기 위해서 다이내마이트를 발명하였다. 그러나 다이내마이트는 본래 개발 의도와 달리, 전쟁 무기로 사용되었다. 원자폭탄을 제조하기 위한 과학자들의 맨해튼프로젝트(Manhattan Project)는 과학의 업적인가 아니면 과학연구의 불명예인가? 이처럼 과학기술은 연구자의 의도와는 달리 인간의 생명이나 가치를 훼손할 수 있는 위험성이 있다.

7) 김환석, 『과학사회학의 쟁점들』, 문학과지성사(2006), 30면.

8) 이 원칙에서 규범의 수범자는 로봇이고 이 원칙은, 사람을 다치게 하지 않고(1원칙), 인간의 명령에 복종하며(2원칙), 제1원칙이나 제2원칙에 상충되지 않는 한 자신을 보호한다(3원칙)는 내용이다. 이후에 로봇은 인류에게 해를 끼치거나 위협에 처한 인류를 방관해서는 안 된다(0원칙)는 것이 추가되었다.

9) Robin Murphy와 David Woods는 Asimov의 로봇 3원칙에 대한 수정안을 다음과 같이 제시한다. (1원칙) 인간은 안전과 윤리에 관한 최고 수준의 전문가적 기준을 충족시키는 인간-로봇 작업체계 없이 로봇을 작업에 배치해서는 안 된다. (2원칙) 로봇은 그것의 역할에 부합하는 방식으로 인간을 대해야 한다. (3원칙) 로봇에게는 스스로를 보호할 수 있는 충분한 자율이 주어져야 한다. 다만 그 보호는 앞의 두 원칙과 상충하지 않는 방식으로 자연스러운 통제권의 이전을 가능하게 하는 것이어야 한다. Robin Murphy & David Woods, "Beyond Asimov": The Three Laws of Responsible Robotics", IEEE Computer Society, IEEE Intelligent Systems(July/August, 2009), 19면.

는 고령화로 1인 가족의 건강과 일상생활을 도와줄 ‘휴머노이드(humanoid) 로봇’이 개발과 확대 방안이 모색되고 있다.¹¹⁾ 다만 Asimov의 로봇 3원칙은 그 상징성 때문인지 최근의 인공지능 윤리적 담론에서도 자주 거론되고 있다.¹²⁾

한편 오늘날 인공지능윤리 담론은 다음의 두 가지 배경에서 로봇윤리와 구분되어 형성되고 있다. 첫째, 빅데이터 발전에 힘입어 인공지능 머신러닝이 폭발적으로 발전하였고, 알고리즘의 성능과 예측모델의 수준도 그만큼 향상됨으로써 성숙했다고 보는 것이다. 즉 인공지능의 기술적 패러다임의 전환이 인공지능 윤리를 새로운 규범적 담론의 대상으로 등장시키고 있다.¹³⁾ 둘째, 미국을 중심으로 한 각국 정부와 기업의 이목이 인공지능기술에 집중되었고, 이에 인류에게 보편타당한 인공지능윤리 규범의 필요성이 강조되고 있다.

대표적인 인공지능 윤리는 2017년 1월에 인공지능 분야의 전문가로 구성된 민간단체인 ‘삶의 미래 연구소’가 발표한 아실로마 인공지능 원칙(Asilomar AI Principles)을 들 수 있다.¹⁴⁾ 총 23항목으로 구성된 이 원칙은 연구에서의 이슈(1~5), 윤리 및 가치(6~18), 그리고 장기적 이슈(19~23)에 대한 주제를 원칙으로 정하였다. 이 원칙은 인공지능 관련하여 연구의 목표를 ‘인간에게 유용한 지능을 개발하는 것’에 두고 있다. 또한 인공지능 윤리담론의 핵심 가치라고 할 수 있는 투명성, 개인정보 보호, 자유와 프라이

10) 로봇(robot)이란 용어는 체코어 ‘robota’에서 유래되었으며 Karel Capek의 1924년 연극에서 처음 등장한 이래로 기계들이 그들의 창조자를 파괴할 수 있다는 플롯은 공상과학 소설에서 보편화되면서 급속도로 확산되었다.

11) The dailypost, 2019. 2. 12.자 기사, “헬스케어의 새로운 패러다임, 돌봄로봇”, <https://www.thedailypost.kr/news/articleView.html?idxno=66473>(2020. 11. 5. 최종방문).

12) 이와 관련하여 일본은 2004년 후쿠오카에서 개최된 International Robot Fair에서 ‘세계로봇선언’을 발표하면서 인간과 로봇이 공존하는 윤리규범의 로드맵을 처음 제시하였다. 1. 차세대 로봇은 인간과 함께 공존하는 파트너가 될 것이다. 2. 차세대 로봇은 인간을 육체적, 정신적으로 보조할 것이다. 3. 차세대 로봇은 안전하고 평화로운 사회구현에 기여할 것이다.

13) 인공지능윤리는 로봇공학을 포함하는 부분이 있지만 인공지능기술은 로봇공학과 구분되는 독자성을 갖는다. 예를 들어 인공지능은 머신러닝을 통해 자연어를 이해하고 표현하는 기술이고, 데이터마닝과 알고리즘 훈련으로 데이터셋을 생산하여 새로운 지식체계를 구축할 수 있다. 또한 인공지능은 인간과 상호작용하며 온라인과 오프라인을 막론해 공간을 뛰어넘는 공유를 한다는 점에서도 로봇공학과 다르다. 한창수·류영선, 『지능형서비스로봇의 설계와 응용』, 청미디어(2008), 1-2면 참조.

14) 이 원칙은 인공지능의 위험을 사전에 통제할 수 있는 연구단계에서의 윤리문제가 활발해지고 있는 가운데, 각계 전문가들의 치열한 논의 끝에 발표되었다.

버시의 윤리와 가치가 언급되었으며, 미래의 초지능(super intelligence)은 윤리적 이상을 널리 공유하는 방식으로 모든 인류의 이익을 위해 개발되어야 한다는 점이 강조되었다.¹⁵⁾

나. 글로벌 기업의 인공지능 윤리원칙

미국의 Google은 인공지능 윤리와 사회적 문제를 연구하면서 ‘Google AI 이니셔티브’를 발표하였다. Google은 모두를 위한 AI를 활용할 것을 선언하고, “사회적으로 유익하며, 불공정한 편향을 만들어 내거나 강화하지 않고, 안전성을 우선으로 설계하고, 인간을 위해 책임을 다하며, 개인정보를 보호하고, 과학적 우수성에 대한 높은 기준을 유지하는, 이러한 원칙에 부합하는 용도에만 (인공지능을) 활용한다.”¹⁶⁾ 라고 하여 AI 개발원칙 7가지를 제시하였다.

Microsoft도 2018년 연례 개발자 컨퍼런스인 ‘MS 빌드 2018’에서 “마이크로소프트는 인공지능을 기반으로 클라우드와 그에 연결되는 모든 기기와 서비스 전반으로 영역을 확장할 것”이라고 밝히면서, 프라이버시, 사이버보안, 윤리적 인공지능을 제시하였다.¹⁷⁾ Microsoft는 2019년 ‘인공지능 개발자를 위한 AI 디자인 원칙’을 발표하였다.¹⁸⁾

15) <https://futureoflife.org/ai-principles/>(2020. 11. 5. 최종방문).

16) 한겨레, 2018. 6. 26자 기사, “Google, AI무기 개발 불참 원칙 정했다”, <http://www.hani.co.kr/arti/economy/it/850750.html>(2020. 11. 5. 최종방문).

17) Microsoft는 책임 있는 인공지능의 활용을 위해서 공정성과 신뢰성, 프라이버시와 보안, 포용성을 투명하고 책임성의 가치를 통해서 지켜나 가야 함을 제시하고 있다. Brad Smith·Carol Ann Brown, “The future Computed Artificial Intelligence and its role in society”, Microsoft Corporation (2018), 58-65면 참조.

18) 블로터, 2018. 5. 18 자 기사, “마이크로소프트의 미래는 창문 밖에 있다”, <http://www.bloter.net/archives/310467>(2020. 11. 5. 최종방문).

표 1 인공지능 개발자를 위한 AI 디자인 원칙

1	<p>“AI는 인류를 지원하고 인류의 자치권을 존중해야 한다.”</p> <ul style="list-style-type: none"> - 협업로봇(‘코봇’)은 위험한 노동을 해 인류를 보호할 역량 필요
2	<p>“AI는 투명성을 갖춰야 한다.”</p> <ul style="list-style-type: none"> - 인류는 인공지능 기술에 대해 정확히 이해해야 한다는 의미. 인공지능도 마찬가지로 인류를 이해해 공생관계를 구축하여야 함
3	<p>“AI는 효율성을 극대화하되 인류의 위엄을 훼손하면 안 된다.”</p> <ul style="list-style-type: none"> - 인공지능 기술이 발전에 투입될 필요. 인류의 역사와 문화유산이 보존되려면 인공지능 기술발전에 다양한 문화권의 연구원이 투입되어야 함
4	<p>“AI는 인텔리전트 프라이버시를 수행할 수 있도록 디자인이 되어야 한다.”</p> <ul style="list-style-type: none"> - 인류의 개인 및 조직 정보를 수집하되 인류 신뢰를 할 수 있도록 철저한 보호가 필요함
5	<p>“AI의 책임 체제를 구축할 수 있는 알고리즘이 있어야 한다.”</p> <ul style="list-style-type: none"> - 예상할 수 있거나 예상치 못한 상황들을 대비해 인공지능이 수행한 피해를 되돌릴 수 있는 대책이 필요함
6	<p>“AI는 편견이나 차별을 배우지 않도록 도덕 교육을 받아야 한다.”</p> <ul style="list-style-type: none"> - 도덕적인 태도에 대한 자료를 활용해 인공지능을 바르게 교육하여야 함

자료: Guidelines for human-AI Design Principle, MS Blog의 내용을 표로 재구성.

2. 인공지능 윤리론의 규범화의 한계

1) 윤리의 규범화 가능성의 전제

우리나라 표준국어대사전에서 발견되는 윤리(倫理)의 정의는, “사람으로서 마땅히 행하거나 지켜야 할 도리”로, 철학과 윤리학에서 인간 행위의 규범을 연구하는 학문의 의미로도 쓰인다.¹⁹⁾ 윤리(ethics)는 그리스어 ‘ethos’에서 파생된 ‘ethikos’에서 비롯하였으며,²⁰⁾ 철학적에서는 좋음과 옳음, 쾌락 등 이상적 가치나 규범에 따라 행동하는 당위성을 나타내기도 한다.

19) 네이버 국어사전, <표준국어대사전>에서 ‘윤리’의 검색결과(표준국어대사전의 정의), <https://ko.dict.naver.com/#/entry/koko/ed7f6ad4e2a84b228a6e48da4b55b698>(2020. 11. 3. 최종 방문).

20) Geoffrey Hazard, Jr, “Laws, Morals, and Ethics”, Southern Illinois University Law Journal Vol.19(1995), 453면.

한편 인공지능 규범화에 있어서 윤리와 도덕, 법은 어떻게 다른가?

첫째, Geoffrey Harzard는 인간 상호 간, 집단 상호 간의 상호인식(mutual intelligibility)을 할 수 있는 집단에서 공유하는 규범을 윤리라고 정의한다. 이에 윤리는 일정한 공동체에서 인간의 의사결정과정에서 수반되는 규범이 될 수 있다.²¹⁾ 윤리는 개인에 대하여 가정이나 사회에서의 역할 규범을 제시하는 암묵적인 기준이 되지만, 법적 판단의 대상이 되는 상호명료성과 주관성을 제시해주지는 못한다. 법과 달리, 윤리는 개방적(narrative)이어서²²⁾ 사회적 합의만 있다면 언제든지 수정될 수 있다.

둘째, 도덕은 일상생활에서 개인에게 개별적·주관적으로 형성되는 옳고 그름에 관한 관념이다. 도덕은 개인이 소속된 사회나 공동체의 문화를 반영하는 개념이지만,²³⁾ 개인적 신념에 따라서 형성되므로 윤리와 같이 “상호인식을 바탕으로 인격체간에 이행을 요구한다거나 그 내용을 완벽하게 전달하는” 기능을 하기 어렵다. 도덕은 주관적인 개념으로서 개인에 따라 계기와 그로 인해 형성된 관념이 다르다. Geoffrey Harzard의 ‘alienation’이라는 표현을 추론하면, 개인은 도덕이라는 가치 내에서 각자의 행성에서 사는 존재이다. 따라서 도덕의 가치를 인공지능 환경에서 규범화하기는 어려워 보인다.

셋째, 법은 국가 안에서 정치적·입법적인 권위에 의해 공식적으로 공포되는 규범이다. 그리고 재판이라는 법적 절차를 통해서 집행가능한 규범으로서 작동하게 된다. Geoffrey Harzard는 약속을 이행·강제하며, 집행가능성을 포함하는 법의 규범력을 인정한다. 법에서는 법관과 입법자, 변호인, 그리고 시민이 같은 의미에서 법을 이해한다는 전제가 중요하다. 이러한 법의 개념은 도덕이나 윤리와 달리 개념적으로 명확하므로, 안정성을 갖기에 규범으로서 잘 기능할 수 있다.²⁴⁾

21) Geoffrey Hazard, Jr, 앞의 논문, 454면.

22) “Ethics not only are mutually intelligible but are conveyed by action or narrative rather than by precept or other formulary.”, Geoffrey Hazard, Jr, 위의 논문, 454면.

23) The ethics center, “Ethics, morality, law—what’s the difference?”(Sep. 27, 2016), <https://ethics.org.au/ethics-morality-law-whats-the-difference/>(2020. 11. 5. 최종방문).

24) 법적 이행절차에서 각각의 행위자들은 상호 간의 인식(intersubjective intelligibility)을 통해서 사법 기능을 이행할 수 있게 된다. “Law” is predicated on the assumption that written rules are intersubjective intelligible. “intersubjective intelligibility” means, in simpler terms, that a rule written by one person can be substantially understood by another. Geoffrey Hazard, Jr, 앞의 논문, 450면.

위에서 살펴본 바에 따르면 윤리와 도덕, 법은 개념상 차이를 보인다. ‘강제성 여부’의 측면에서 윤리와 도덕을 같은 범주로 이해할 수 있는 것에 달리, ‘규범성’의 측면에서, 윤리는 오히려 법적 규범과 가까운 지점이 있어 보인다. 법이 규범의 한계로서 작동하고, 계약체결을 통해 그 이행을 요구한다면, 윤리는 공동체에 공유되고 상호 인식한 규범에 대한 구체적인 행위를 요구한다. 이렇게 보면 ‘규범성’의 측면에서 윤리는 법규범에 가깝지만, ‘강제성 여부’의 측면에서 법에 비할 수 없다. 법이 사회 안에서 개인과 단체에 대한 지위를 설정하여 그 지위에 따른 필요최소한의 행동기준을 부여한다면, 윤리는 그 이상의 범위에서 개인과 단체에 의무와 권한을 부여할 수 있을 것이다.²⁵⁾

인공지능의 의사결정 및 그 실행에는 기존에 존재하는 법을 위반하는 정도에 이르지 않더라도 인간의 가치를 위협하는 여러 위협에 마주할 수 있다. 존재 이때 윤리는 입법이 부재한 경우 및 법규범으로 해소되지 않는 위협에 대하여 규범적 역할을 할 수 있다. 다만 그 경우 인공지능 윤리규범은 ‘윤리’와 구분되는 강한 규범성을 갖춘 형태이어야지 그 실질적 기능을 다 할 수 있을 것이다.²⁶⁾ 그런데 우리는 강제성이 수반되는 규범을 윤리라고 할 수 있을 것인가? 혹은 윤리규범을 통해서 개인의 기본권과 권리를 청구할 수 있을 것인가? 라는 의문을 갖지 않을 수 없다.

3. 인공지능 윤리의 규범 적합성에 대한 비판

Ray Kurzweil이 가정한 미래의 인공지능은 스스로 알고리즘을 업로드하고, 프로그램에 개입하려는 인간의 통제를 거부하고 극복해 버리기까지 한다.²⁷⁾ 그럼에도 그는 인류가 안전한 사회를 만들어냄으로써 인공지능의 위협은 극복될 수 있다고 한다. 이러한

25) 예를 들어 기업이 소비자 관련법을 준수하게 되면 법에서 정해놓은 기본적인 소비자의 권리를 보호하게 되지만, 법에 강제되지 않는 의무를 이행하는 경우 윤리적 규범까지 준수한 것이라고 할 수 있다. 이때 기업이 윤리적 규범을 이행하게 되면 법적 권리 이상의 권리를 보호할 수 있다.

26) Nick Bostrom·Eliezer Yudkowsky, “The Ethics of Artificial Intelligence”, *Machine intelligence research institute*(2011), 12-13면.

27) Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking(2005), 198-203면.

관점에서 인공지능 윤리규범은 최상의 규제방안이자 이상적인 도구이기까지 하다. 그러나 인공지능 윤리규범이 실질적인 기능을 다 하려면, 로봇공학의 Asimov'의 로봇 3원칙이나 먼저 소개한 인공지능 윤리원칙의 수준에서 설계되어서는 안 된다. 오히려 인공지능의 예측곤란성에 대응해 법이 해결하지 못하는 위험에 대비하는 것이 필요하다.²⁸⁾ 지금까지 Google이나 Microsoft와 같은 글로벌 기업들이 인공지능 윤리원칙을 선제적으로 제시한 것은 신뢰있는 인공지능 활용을 도모하는 일로서 나름의 의미가 있다. 하지만 개별 국가들에서는 그들에게 윤리원칙을 강제할 장치가 없다. 특히 글로벌 기업들의 윤리원칙은 인공지능의 의사결정으로부터 보호되어야 할 핵심적 가치를 포함하지만, 대개는 지나치게 선언적인 내용으로 이행사항이 추상적인 수준에 머물러 있어 우려스럽다.

기업의 인공지능 윤리원칙이 실질적인 규범으로 작동하려면, 인공지능을 실행하는 주체와 향유하는 주체를 구분하여 규범적 형식으로 설계되는 것이 필요하다. 아울러 윤리원칙이 보호하려는 가치를 중심으로 규율대상과 내용이 명확하게 제시됨으로써,²⁹⁾ 이행과 책임의 내용이 세심하게 배분되는 것도 필요하다.³⁰⁾ 또한 인공지능 윤리원칙의 규범성을 유지하려면 해당 규범의 유효성이 검증되어야 하고, 규범에 약속된 사항이 일관되게 이행되는 것에 대해서 객관적인 검증절차가 필요한 것이다. 그런데 안타깝게도 이와 같은 제도적 장치는 아직 찾아볼 수 없다. 앞으로 인공지능에 대한 윤리원칙이 실질적 규범성을 가지려면, 최소한의 기술적, 규범적 제어장치나 견제가 수반되어야 할 것이다.³¹⁾

-
- 28) 아시모프의 로봇 3원칙을 윤리적 AI 모델로 적용하는 것에 대한 비판에 대해서는, Nick Bostrom · Eliezer Yudkowsky, 위의 논문, 15면 참조.
 - 29) Anders Hanssen과 Stefano Nichele는 인공지능 윤리규범에는 지금의 (약한)인공지능과 AGI를 구분할 필요가 있다고 주장한다. Anders Hanssen·Stefano Nichele, "Ethics of Artificial Intelligence Demarcations", *AI lab*, Oslo Metropolitan University(2019), 133-142면 참조.
 - 30) 약한 인공지능과 AGI는 세부적인 능력에 있어서 차이점을 갖게 되므로, 각자의 능력에 따른 윤리적 권리와 의무가 주어져야 한다. 예를 들어 약한 인공지능의 경우에는 AGI보다 취약한 시스템을 갖고 있고, 불투명성의 정도도 인공신경망이 복잡한 AGI가 훨씬 크다고 예상할 수 있다. 현재 우려되고 있는 인공지능의 불투명성과 예측불가능한 잠재된 문제들에 대한 우려는 대부분 AGI를 염두에 두고 제기되는 것이다. 따라서 인공지능윤리 규범은 구현된 인공지능의 수준에 따른 구분에서 비롯된 기준을 제시하는 것이 요구된다. Anders Hanssen·Stefano Nichele, 위의 논문, 8면.
 - 31) 이를 현실화하려면 외형적으로 법률과 유사한 형태를 취하거나 소스코드의 형태로 인공지능에 탑재할

III. 인공지능 환경에서의 규범구체화에 대한 다양한 모색

1. 인공지능에 대한 법적(사전적) 규율의 필요성

인공지능의 의사결정에 대응한 규범칙으로 윤리는 규범적 특성에서의 한계가 있다. 인공지능의 기술의 특징과 그로부터 발생하는 위험은 법적 침해를 낳기도 하고 개인에 대하여 법적 의무를 발생시키기도 하는 등 개인 스스로 전혀 예상하지 못하는 상황을 초래하기도 한다. 즉 인공지능 기술은 그 위험의 양상이 질적으로 다른 것임에도 불구하고, 기업 등이 내세우는 윤리원칙이나 관련 기술을 규제하는 정도의 수준에서 규율하는 것으로는 부족해 보인다. 이와 같이 인공지능의 위험에 대응한 국가 차원에서의 실정법적 규율의 필요성이 있음에도, 국내에서는 인공지능의 윤리적 쟁점과 법적 쟁점이 뚜렷하게 구분되지 않고 혼재된 채로 논의되고 있다. 대표적으로 로봇윤리는 인공지능과 로봇기술이 인간사회에 수용되기 이전에 필요한 담론이지만, 윤리규범만으로는 인공지능과 로봇이 인간사회에 미칠 파급력을 통제할 수는 없을 것이다.³²⁾ 인공지능이 자동화(automated)에서 자율화(autonomous)로 옮겨가는 길목에서 결국 법규범의 체제가 모색될 수밖에 없는 실정이다. 이에 인공지능 환경에 대응하는 윤리적, 법적 규범의 구조화를 모색하는 논의가 본격적으로 시작되어야 한다.³³⁾ 다만, 인공지능의 규율을 좁은

수 있는 '기계윤리(machine ethics)', '윤리적 알고리즘(ethical algorithms)'의 개발도 수반될 수 있을 것이다. Nick Bostrom·Eliezer Yudkowsky, 앞의 논문, 6면.

32) Ryan Calo는 로봇기술의 발전으로 기존 규범체계에 혼란을 줄 것이라며 Cyberlaw 규범이 적용되는 새로운 입법정책을 요구하고 있다. 그의 주장은 다음으로 요약된다. 첫째, 인간과 로봇의 경계가 없어지면서, 물리적인 상해를 입힐 가능성이 있으므로 법적 개입이 요구됨. 둘째, 인공지능의 사회적 영향을 고려해 독립된 기구인 연방로봇위원회(Federal Robotics Commission)와 같은 거버넌스가 요구됨. Ryan Calo, "Robotics and the Lessons of Cyberlaw", *California Law Review* Vol. 103/3 (2015), 556면 및 562면.

33) 우리나라의 대학 및 연구기관의 연구자, 법조계 실무가 총 53명에 대해 "인공지능 시대 윤리적·법적 쟁점"을 주제로 윤리적 접근을 우선할지 혹은 윤리적 접근과 법제도의 동시적 접근이 필요한지에 관한 설문 조사결과, 인공지능 시대의 윤리적·법적 쟁점에 대응하기 위해서는 윤리적 접근과 법적 접근이 모두 필요하다는 의견이 지배적으로 나타났다. 특히 인공지능 분야의 전문가들이 이러한 입장을 표시하였으며, 윤리적 접근이 우선시되고, 이후에 법제도적 접근이 필요하다는 의견(17.0%)에 비해, 윤리적 접근과 법제도적 접근을 동시에 하는 것이 필요하다는 의견(81.7%)이 상대적으로 높게 나왔다. 이러한 집계는 급변하는 기술적 환경을 고려하더라도, 윤리적 접근만으로는 인공지능의 위험에 효과적으로 대

법규범으로 미리 준비하기에는 미래사회의 위험과 기술발전의 예측범위가 넓지 않은 편이다. 따라서 인공지능에 대한 법적 규율은 특정한 입법론을 논하기 앞서서 기본적인 규율원칙과 대상을 구체화하는 규제모델을 착안해 두는 것으로부터 시작되는 것이 바람직하다.

2. 이원적 규제모델

인공지능은 빅데이터와 머신러닝 기술이 융합된 시스템의 총체로 인공지능 규제모델은 (i) 데이터 처리와 (ii) 인공지능 알고리즘의 처리를 각 단계로 보는 기술적 처리과정으로 접근되는 것이 필요하다. 기존에 ICT 분야에 대한 규제는 주로 정보법과 입법에서 개인정보를 중심으로 ‘이용자 보호’라는 프레임 안에서 모색되어 왔으며, 이는 개인의 자기결정권을 직접적으로 보장하는 체계는 아니었다고 할 수 있다. 인공지능은 개인, 사업자, 정부가 각각의 고유한 이익을 추구하는 가운데 그들의 이익이 충돌하는 구조로서 당사자 간의 이익을 균형 있게 취급하기 위해서는 몇 가지의 원칙이 요구된다. 첫째, 투명성의 원칙이다. 이는 알고리즘의 배후에 있는 영향력으로부터 적절한 통제를 하기 위한 기준이 된다.³⁴⁾ 투명성의 원칙을 인공지능 의사결정 시스템에 구체화하는 기준을 적용하는 방안을 고려할 수 있다. 둘째, 공정성의 원칙이다. 인공지능의 복잡성으로 인한 불투명성을 배제하기 위해서는 불공정한 요소들을 걸러낼 수 있어야 한다. 예

응하기는 어렵다는 인식에서 비롯된 것으로 보이며, 답변의 이유는 다음으로 요약된다. 1. 인공지능 관련 규범에는 윤리적, 법제도적 접근이 혼용되고, 이에 융합방법론의 필요성이 있음. 2. 법은 최소한의 도덕이므로 사람과 공존하기 위한 윤리규범이 필요하고 관련 교육의 필요와 법제도적 접근이 필요함. 3. 법제도와 윤리는 분리하기 어려운 속성으로 윤리적인 고려에 따라 법제도를 수정해갈 수 있음. 4. 양자의 경계가 모호하고 별개로 취할 수 없음. 5. 윤리적 접근을 통한 사회적 합의는 법제도의 뒷받침 없이는 공허한 메아리일 수 있음. 6. 로봇의 윤리적 행위성이 법적 책임귀속 문제로 연결되듯이 인공지능의 윤리적 접근은 법제도의 측면의 접근에 맞닿아 있음. 7. 기본적으로 윤리적 접근이 우선하지만 궁극적으로 권리·의무에 관한 내용으로 귀결될 것임. 장민선, 「인공지능(AI) 시대의 법적 쟁점에 관한 연구」 연구보고 18-10 한국법제연구원(2018), 130-131면 참조.

- 34) 독일에서는 특히 기업의 데이터 처리 및 (인공지능)의사결정시스템은 조사되어야 하고, 개인의 참여(정보에의 접근)가 보장되는 것이 필요하다는 논의가 있다. Gerd Gigerenzer·Felix G. Rebitschek·Gert G. Wagner, Eine vermessene Gesellschaft braucht Transparenz, ZBW - Leibniz - Informationszentrum, Wirtschaft(2018), 867-868면.

를 들어 기업이나 정부의 의도적인 은폐, 사회의 감시 체계의 부재, 미래의 기술문맹과 같은 기술적 요인들이다.³⁵⁾ 셋째, 합리성의 원칙이다. 합리성의 원칙은 헌법상 적법절차에서 요구되는 것으로 데이터와 알고리즘의 검증 및 관련 절차에서도 적용될 수 있을 것이다. 이와 같은 원칙들은 인공지능의 의사결정으로 인한 부정적 요소³⁶⁾를 걸러내고, 잘못된 상관관계 또는 불공정한 취급에 대한 개인의 이익·거부·선택권을 확보함으로써 자기결정권을 보장하는데 나름의 기준이 될 것이다.

1) 데이터 처리과정

기술이 복잡해질수록 전통적인 개인정보 보호체계의 기능은 축소될 것으로 전망되는데, 이는 빅데이터 성능이 기존 규제를 회피하게 하기 때문이다. 맞춤형추천서비스와 통합의료·금융 서비스에서는 민감정보를 포함한 개인에 관한 중요정보가 수집·처리되고, 예측모델은 2차 처리된 정보를 통해서 정교한 결과를 제공할 수도 있을 것이다. 개인정보의 처리에 관한 문제는 차치하고라도 개인의 인지와 동의를 확인할 장치가 없으므로, 인공지능의 데이터 처리에는 제동장치가 필요하다.³⁷⁾ 인지와 동의제도는 개인의 자기결정권의 보장에서의 핵심이라고 할 수 있다. 데이터 처리는 알고리즘의 의사결정의 예측모델의 기반이 되므로 데이터 처리에 대한 타당성 및 공정성에 대한 항목을 구비해 평가하는 절차가 요구된다. 아울러 개인의 자유와 재산상 이익에 큰 영향을 미치는 의사결정(복지수혜, 채용, 의료보험)에 사용되는 데이터에 대해서는 데이터의 정확성과 일련의 메커니즘의 적정성이 개인에게도 검증(inspect)될 기회가 제공될 수 있어야 하는바, 이는 ‘공적 감사(audit)의 결과’가 이해하기 쉽게 통지되어야 한다는 뜻이다.³⁸⁾

35) 버클리 대학의 정보학 부교수인 Jenna Burrell은 투명성을 방해하는 세 가지 요소를 제시하고 있다. Jenna Burrell, “How the machine “thinks”: Understanding opacity in machine learning algorithms,” SAGE Publishing, *Big Data & Society* Vol.3(2016), 2-3면.

36) 예를 들어 프라이버시와 언론의 자유를 연구하는 Boston University Law School 교수 Danielle Citron은 알고리즘의 의사결정으로부터 개인이 평가되는 사회를 ‘점수화된 사회(scored society)’라고 하였다. Danielle Keats Citron·Frank Pasquale, “THE SCORED SOCIETY: Due Process for Automated Predictions”, *Washington Law Review* Vol 89/1(2014), 23면.

37) 빅데이터 연구자인 Kate Crawford는 데이터의 수집과 사용, 빅데이터의 분류, 범주화 등의 과정을 규제함으로써 개인의 이익과 기술의 효용을 함께 확보할 수 있다고 한다. Crawford·Jason Shultz, “Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms”, *Boston Law Review* Vol.55(2014), 109면.

특히 이러한 개인에게 이와 같은 통지권이 형태로 부여되는 것이 중요한데, 정부 기관의 행정처분과 같은 의사결정 시스템에서 개인의 적정한 통지권이 보장되는 않은 경우에는 청문(hearing)에 관한 개인의 권리가 침해될 수 있기 때문이다. 물론 사인 간의 구도에서도 객관적인 감독절차와 통지가 없다면, 이는 미래의 '예측된 (프라이버시의)결정 해악'을 초래한다.³⁹⁾ 이와 같은 데이터 처리에 관한 규제모델의 필수 요소들은 궁극적으로는 개인이 데이터 처리과정과 관련해서 이의를 제기할 수 있는 것을 전제한다. 인공지능의 의사결정에 주요한 상관관계에 기여한 데이터에 기록된 내용에 대한 수정·거절에 대한 권리가 실체화되는 것이 요구된다. 그리고 인공지능 알고리즘에 훈련에 제공된 데이터의 항목을 검토하는 기관은 여타의 이해관계로부터 독립된 제3자이어야 하며, 이들은 개인과 이해관계인들의 이익의 문제에서 중재자의 역할을 해낼 수 있어야 한다.⁴⁰⁾

2) 알고리즘 의사결정

알고리즘 의사결정 과정에 대한 규제는 데이터 처리에서와 마찬가지로 공정성과 투명성의 관점에서 실행되는 것들이 필요하다. 공인된 제3자의 검증절차와 감사추적(audit trail)을 기록하는 등의 방안으로 고려될 수 있으며, 처리 결과에 대한 개인의 이의·거부·선택권이 인정되기 위해서는 적절한 절차적 장치가 전제되어야 한다.⁴¹⁾ 여기서는 추가적으로 알고리즘의 책임성에 관한 부분만을 간략히 언급하고자 한다. 알고리즘의 책임성(Accountability)을 확보하는 것을 인공지능의 의사결정에 관한 모델링은 Nicholas

38) Kate Crawford·Jason Shultz, 앞의 논문, 126면. Danielle Keats Citron, "Technological Due Process", *Washington University Law Review* Vol. 85(2008), 1305-1306면.

39) Kate Crawford는 이 해악을 예측된 프라이버시 해악(predictive privacy harm)라고 한다. 그러나 미국의 프라이버시권의 인정 과정을 보면 이는 사적영역에서의 자기결정에 관한 해악이라고도 볼 수 있을 것이다. Kate Crawford·Jason Shultz, 앞의 논문, 95면.

40) Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press(2015), 141면.

41) 알고리즘의 예측모델에 대한 개인의 권리를 특히 유럽연합의 일반개인정보보호규정(GDPR)의 제22조 제1항의 '프로파일링에 대한 이의권'에서 근거를 찾는 주장들이 많은데, 이러한 권리는 다른 나라의 개인정보 보호법률에서 굳이 찾지 않아도 되는 것이다. 인공지능의 의사결정에 대응한 개인의 권리는 인간의 자유와 존엄, 그리고 우리 헌법상 인정되는 자기결정권이 보장하는 기본권으로 보다 정교하게 확인되고(적극적 자유와 선택과 통제의 측면에서), 입법 차원에서 구체화되어야 할 것이다.

Diakopoulos가 알고리즘의 의사결정 과정을 분류한 것에 착안할 수 있다. 그에 따르면 알고리즘의 의사결정은 우선순위화(prioritizing), 분류(classfication), 연관지음(association), 필터링(filtering)으로 구분할 수 있다.⁴²⁾ 이 각각의 과정은 알고리즘에 대한 검증에 좋은 구조를 제공해준다. 첫째, 우선순위화(prioritizing) 과정에서 알고리즘이 우선순위로 처리한 결정의 근거는 추출할 때 그 요소가 정의로운지 혹은 차별적인지에 대한 판단이 가능할 것이다. 둘째, 분류(classfication) 및 연관지음(association) 과정에서 적용된 기준을 확인하고, 편향된 요소(예를 들어 정치적 가치)가 들어갔는지를 식별해 낼 수 있을 것이다. 셋째, 개별 단계에서 알고리즘 자체의 오류를 확인함으로써, 알고리즘의 데이터세트 처리에서 발생한 오류에 대한 기록을 남기고 보완조치 유무 및 그 적정성을 판단할 수 있다. 넷째, 알고리즘에 적용된 함수의 정의와 연관지음에 하는 알고리즘에 적용된 함수에 대한 분석을 검토할 수 있다. 다섯째, 알고리즘 처리에서 특별히 가중치가 적용되거나 아주 배제되는 정보의 적정성을 재확인함으로써 알고리즘의 공정성을 진단할 수 있을 것이다.⁴³⁾

3. 인공지능에 대응한 자기결정 보호모델

1) 자기결정 주체의 보호

인공지능의 의사결정에 대응해 개인의 자기결정권이 보장되려면 개인이 인공지능의 의사결정의 과정을 이해하고 그것의 실제적인 효과를 이해해야 하므로, 일반적인 인간의 사고능력보다 더 높은 수준에서의 판단력이 요구될 수 있다. 이러한 점은 자기결정권이 기반하는 자율 패러다임의 한계라고도 보여진다. 인공지능 환경에서 이러한 능력은 개인 스스로가 갖추어야 할 사항이라기보다는 인공지능을 통제하는 운영자의 자율적 협력구조와 공적 기구 등의 지원으로 보완되어야 하는 부분이라고 봄이 타당하다.

42) 알고리즘의 책임에 관한 조사 기준의 예시를 정리한 것이다. Nicholas Diakopoulos, "Accountability in algorithmic decision making", *Communications of ACM* Vol.59/2(2016), 9-10면.

43) 알고리즘의 책임의 문제에 대한 조사 기준의 예시를 정리한 것이다. Nicholas Diakopoulos, 위의 논문, 9-10면.

2) 인공지능 통제자의 선(善)의무

첫째, 스스로 이행이 의무화되는 자율적 원칙을 마련하는 것이다.

법국가적으로 확산되는 인공지능 윤리규범은 인공지능기술을 통해 서비스를 제공하는 기업들을 통해서 선제적으로 제시되기도 한 만큼 인공지능기술을 안전하게 활용하려는 목적에서 점차 더 필수적이라고 할 수 있다. 우리나라에서는 「자율주행자동차 상용화 촉진 및 지원에 관한 법률」⁴⁴⁾이 입법화되기도 하지만, 아직까지는 인공지능을 규율하는 입법은 제한적인 범위에서 이루어질 필요가 있다. 즉 기술발전의 유용성과 예측근란성을 충분히 고려하여야 하는 이유로 입법이 있기 이전에는 사업자의 자발적 사회적 참여로 인한 자율규제가 필요하다. 이 같은 맥락에서 인공지능 통제자인 사업자들에게 자체적인 윤리규범의 마련이 요구되지만, 윤리규범의 특성상 법적 강제력이 없기 때문에 이행 후 규범력을 승인하는 것은 사회의 몫이다. 이와 달리, 사업자에게는 세부항목에 대한 필수적 요소를 확인하고 평가하는 컴플라이언스(Compliance)제도의 도입도 고려될 수 있는데, 개별 인공지능서비스의 경우에는 일종의 서비스 약관에 의무규정을 넣게 되면, 계약적 성격을 가질 수 있어 권리분쟁으로 인한 혼란을 줄일 수 있다.

둘째, 설명된 대응권을 제공하는 것이다. 인공지능의 의사결정에 수반된 정보와 데이터 처리 및 알고리즘 처리의 객관성에 대한 평가가 담긴 정보는 인간 의사결정자가 인공지능의 의사결정에 대응(이의·거부·선택)하는데 중요한 요소가 된다. 앞서 데이터 처리의 규제모델의 요소이지만, 사업자의 자율규제로 채택되는 경우 비교적 세부적으로 접근될 수 있게 된다. 또한 사업자는 인공지능 의사결정의 알고리즘 처리과정에 대한 정보를 제공할 수 있는데,⁴⁵⁾ 개인에게 중요한 의사결정에 있어서 특정요소와 결과 간의 상관관계(로직) 처리의 적정성이 요구될 것이다. 특히 자기결정권의 인격적 측면과 관련된 자기표현에서 인공지능의 알고리즘의 개별 단계에서 적용된 기준이 공평하게 적용되었는지에 대한 것의 평가를 공개할 수 있어야 할 것이고, 알고리즘의 성격과 용처에 따라 사업자 혹은 타인의 가치와의 조정이 필요하다 할 것이다. 하지만 모든 의사

44) 2020. 5. 1. 시행, 법률 제16421호(2019. 4. 30. 제정)

45) 이하의 평가요소는 이 장의 제1절 IV. Nicholas Diakopoulos가 인공지능의 의사결정과정을 우선순위화(prioritizing), 분류(classfication), 연관지음(association), 필터링(filtering)으로 분석하고 이를 그 규율방식으로 수용한 것에 이은 전개이다.

결정에 대해서 후순위로 처리된 것들에 대해서 사업자에게 알릴 의무가 지워지는 것은 곤란하다. 다만 개인에게 중대한 영향을 미치는 의사결정의 결정에 개인이 거부와 설명 요구를 제기한 경우 인공지능 통제자는 배제된 정보나 알고리즘에 대해서도 기본적인 설명을 제공할 수 있을 것이다.

셋째, 선택과 표현에 대한 보호의무에 대한 배려가 필요하다. 인공지능의 의사결정은 알고리즘의 훈련을 통해서 무엇을 버리고 취할지에 대한 선택의 연속인바, 개인에게 투사된 인공지능의 의사결정은 개인의 자기결정권리의 선택과 표현에 대한 이익을 제한하는 측면이 있다. 그리고 개인의 결정은 타인 및 공동체와의 관계에도 영향을 끼치므로 개인은 인공지능의 의사결정을 거부하거나 다른 의사결정을 선택할 수도 있어야 한다. 인간은 자신의 의사결정이 타인이나 사회에서 거부당할 때 대안을 선택하는데, 그 대안으로 자기의 입장과 상충하는 타인과의 이익을 조율하면서 자기에게 최적이라고 판단되는 의사결정들을 해나가기 때문이다.⁴⁶⁾ 따라서 사업자는 개인의 그의 선택과 표현을 존중하는 방식으로 데이터와 알고리즘에 개입할 수 있는 옵션을 다각화할 필요가 있다.

넷째, 개인의 권리 보장측면에서 모색되기는 하였으나, 사업자의 선(善)의무⁴⁷⁾가 기술적, 경제적으로 가능한 수준에서 자율적으로 준수된다면 인공지능 통제자에게도 유리하게 작용될 것이다. 이는 개인에게 인공지능의 의사결정을 검증하고, 거부하는 기회를 줌으로써 인공지능의 의사결정에 대한 책임을 개인과 배분하는 결과를 낳기 때문이다. 한편 인공지능 의사결정의 신뢰성 측면에서 인공지능 의사결정시스템의 기술적 측면의 수준을 현실화할 필요가 있다. 특히 기술적 조치의무로 인한 문제 발생시, 사후적인 책임 소지를 분명히 하기 위해서 구체적인 기술적 의무사항을 정기적으로 업데이트하고,

46) 정규원, 「피해자의 승낙에 대한 연구 -의료행위를 중심으로-」, 서울대학교 법학박사 학위논문(1999), 의료행위에 있어서 피해자의 승낙에 대한 검토에서 제시된 개인의 의사결정과 잠재능력에 관련하여 153-155면 참조.

47) 자기결정 관점에서 인간 의사결정자는 결코 합리적이지만은 않은 존재이다. 하지만 인공지능 자동화된 의사결정시스템이 인간에게 투사하는 자기결정을 보호하려면 인공지능이 합리적인 의사결정을 할 것이라는 나름의 이상향을 가지고 접근하는 것이 필요하다. 이에 따라 인공지능을 통제하는 입장에 대해서는 자발적으로, Immanuel Kant가 바라보는 이상적 인간향의 선의지의 의미에서 규범적 의무를 설정하는 것이 필요하다. 이에 인공지능 통제자의 '선(善)의무'라는 용어로 표기하였다.

이에 대한 검증 및 평가결과를 남길 필요가 있다. 사업자의 경우 기업체의 규모와 인공지능 의사결정이 이용자에게 미치는 중요도, 영향력 등을 기준하여 기술적 조치의무에 관한 책임의 수준도 달리 판단될 수 있으므로 기술환경에 부합한 수준을 스스로 설정, 준수하는 것이 필요하다.

3) 국가의 개인에 대한 자기결정 보호의무

인간은 누구나 추상적 의사결정 능력만 있다면 자기와 관련한 사항을 결정할 권리인 자기결정권이 인정된다. 따라서 인공지능의 의사결정에 대해서도 이를 수용하거나 거부할 수 있게 검증할 권리가 인정되어야 한다. 인공지능기술 환경에서 개인의 자기결정권의 행사에는 개인적 인식능력이 요구된다. 그러나 현실적으로 인공지능의 의사결정에 대한 개인의 인식능력은 연령, 빈부, 지식, 그 밖에 개인적 배경에 따라서 상당한 차이를 갖게 될 것이다. 특히 연소자와 노약자, 기술에 대한 이해도가 낮은 부류는 전반적으로 낮은 인식능력을 갖게 되리라고 쉽게 예상할 수 있다. 인공지능서비스에서 개인의 의사결정이 인공지능에 위임된 상황이라면, 이에 대한 고려는 일차적으로 기업 등의 의무사항으로 포섭시키는 것이 타당하다. 그런데 인공지능서비스의 성격과 용처는 천차만별일 것이어서 합리적인 규제책을 마련하기에는 관망이 필요한 상황이다. 이와 같은 상황에서 사인 간의 관계에서 관련 의무를 도출하기는 쉽지가 않다. 이에 사적 영역에서 해결되지 않는 부분에 있어서 국가 등의 공적 지원이 요청된다.

개인의 자율성과 관련하여 Martha Nussbaum은 역량이론을 살펴보면, 개인의 자율성이 필수적인데도 개인으로서는 확보되지 못하는 경우 공적 개입이 필요하다고 할 수 있다. 역량이론에 따르면, 개인의 자율성은 내적인 적합성의 측면에서 내부역량을 충족한 뒤 외적인 조건들과 결합해서 최종적인 실현 단계로 나아갈 수 있다.⁴⁸⁾ 그는 역량 개념을 ‘발전가능성’이라는 것과 연관시키는데, 개인의 자율성을 발휘할 수 있는 상태

48) 그가 제시한 역량에 대한 접근이론은 오늘날 인권의 구체적인 내용을 구성하며, ① 생명(life)과 ② 신체의 건강(bodily health), ③ 신체의 완전성(bodily integrity), ④ 감각(senses), ⑤ 감정(imagination), 상상(imagination)과 사고(thought), ⑥ 실천이성(practical reason), ⑦ 친교(affiliation), ⑧ 다른 종과의 유대(other species), ⑨ 놀이(play), ⑩ 환경적 통제(control over's environment)를 요소로 포함한다. Martha Nussbaum, *Creating Capabilities: The human Development Approach*, The Belknap Press of Harvard university Press(2011), 33-34면.

로 발전하게 하는 것이 민주주의 국가에서 사회정의를 위한 규범적 요청임을 설명하고 있다.⁴⁹⁾ 이는 인공지능 의사결정에 대응한 개인의 자기결정 보완책에 시사하는 바가 크다. 자유민주주의 국가체제에서 국민은 그들이 선택한 법과 원칙에 따라 인간답게 살 권리를 갖고 있다. 그런데 인공지능 환경에서는 법적 책임이 없는 글로벌 기업들이 개별 국가와 그 국민의 자기결정을 위협하게 된다. 특히 글로벌 기업들은 다양한 분야에서 인공지능 통제자로 등장하게 되므로 개별 국가는 이들의 협력을 이끄는 유인정책을 마련할 의무가 있는 것이다.

한편 국가가 공적 프로그램을 통해서 개인을 지원하는 경우 개인의 자기결정능력에 대한 ‘보충의 필요성’이 전제되어야 한다. 예를 들어 법적 효과가 있는 인공지능의 의사결정에 대한 개인의 자기결정능력에는 ‘신체적, 정신적 의사능력’이 전제되어야 하며, 지원대상이 세분화될 수 있을 것이다. 기본적인 개인의 인식능력은 충분한 능력을 보유한 경우와 일반적인 능력을 보유한 경우, 일반적인 중간 이하의 능력을 보유한 경우, 그와 같은 능력을 아주 결여한 경우로 나눌 것이다. 마지막 경우에는 일반인의 수준까지 끌어올리는 공적개입이 요구될 수 있어 보이며, 공적개입의 목표는 최소한의 독립적인 자기결정능력을 갖도록 하는 것에 있다. 이를 위해서는 두 가지 방향에서 지원방안을 마련할 수 있다. 첫째, 개인의 기본역량인 내적 역량이 결핍되는 경우 이를 보완하는 측면이다. 둘째, 인식능력의 외적 역량의 차원에서 인공지능 환경과의 상호작용 등을 지원하는 측면이다. 후자와 관련해서 Martha Nussbaum은 타인과의 친교, 다른 종과의 상호작용 등 외적 요소와 연결하는데, 이에 따르면 외부적인 요인이 부족한 경우 개인의 의사결정은 인공지능의 의사결정에 대해서 자율성을 갖추지 되므로 공적개입이 요구된다. 외부적인 요인으로 인공지능과 인간의 상호작용을 상정해볼 수 있다. 그러나 인공지능과 인간의 상호작용과 관련해서는 아직 개념조차도 불분명한 수준에 머무를 있을 뿐이다. 이에 본 논문에서는 추상적인 수준에서 방향성을 제시하는 것에 그치며, 다음의 연구에 맡기도록 한다.

49) Martha Nussbaum, 위의 책, 113-122면 참조.

IV. 남겨진 문제점들 - 결론을 대신하여-

1. 자기결정과 인간존엄의 접점

자기결정권은 자신의 스스로 자기의 행위를 지배하는 규칙(rule)을 선택하고 그것을 실행할 수 있는 자유에 권리성을 부여한 것이다. 따라서 자기결정권의 핵심요소는 ‘자기결정’을 보장하는 것이라고 할 수 있다. 헌법 제10조에서 보장되는 인간존엄은 인간이 자기답게 자유로운 결정을 할 수 있는 전제조건이 된다. 인간존엄은 상황에 따라 다양한 윤리적, 정치적인 표상들을 모두 포함해 규범화를 논하기에는 다소 공허한 지침이 되기도 한다. 그러나 인간존엄은 우리 헌법상의 가치이며, 독일 기본법 제1조에서도 “인간의 존엄은 불가침이다. 이를 존중하고 보호하는 것이 모든 국가의 의미이다”라고 규정하고 할 만큼, 모든 사회의 윤리적, 법적, 정치적 논의에서 논증의 만능패(All-zweckwaffe)로 작용한다.⁵⁰⁾ 한편 인간존엄은 인공지능에 대응한 자기결정권의 근거이면서도, 자기결정권의 개념을 구체화함으로써 중첩되기도 하고, 자기결정권과 갈등하는 가치 상호 간 양보해야만 하는 접점이 생기기도 한다. 자기결정권은 인간존엄의 하위개념이지만, 자기결정 측면을 보장하게 되면서 양보해야 하는 자유와 인격권이라는 인간존엄의 영역이 나타나기 때문이다. 그럼에도 불구하고 Niklas Luhmann은 이른바 “인간은 자기에게 고유한 자기결정적인 행위의 근거로서 존엄을 지닌다”라고 하듯이,⁵¹⁾ 인간은 자기결정을 통해서 자신의 정체성을 찾아가는 존재로서 존엄을 가진다고도 볼 수 있다. 오히려 인공지능에 대응한 인간존엄의 핵심으로서의 자기결정권의 내용을 확정하는 것으로부터 인간존엄은 시작될 수 있다. 자기결정권은 스스로 외부의 무엇을 향해서(to) ‘선택’ 할 수 있고, 간섭하는 무엇으로부터(from)의 ‘통제’ 가능한 권리어야 한다. 본래 자기결정권은 보충적 개념을 갖는 기본권이기에 향후 맞닥뜨린 현실 세계에서 그 의미를 명확화시키는 것이 법의 역할이요 과제라 생각한다.

50) 김영환, “인간의 존엄에 대한 논의의 재구성: 형이상학 없는 인간의 존엄”, 「법철학연구」 제23권 제1호, 한국법철학회, 12면.

51) Niklas. Luhmann, *Grundrechte als Institution*, Duncker & Humblot(2019), S. 53 ff.

2. 인공지능에 대한 법의 통제 가능성

천년의 전환기에 Lawrence Lessig은 “Code is law”라고 운을 떼며 시스템의 아키텍처 구조와 코드, 알고리즘을 가르쳤다.⁵²⁾ 그는 인터넷의 상용화와 디지털화로 열린 인터넷 공간은⁵³⁾ 정부와 상거래를 지배하는 세력들이 규제할 것이고, 시민들이 선택한 자유주의는 위협받을 것이라 예측하였다.⁵⁴⁾ 그는 기술의 위협에 ‘인터넷 공간의 헌법’을 논하면서, 법적 도구로써 코드(Code)를 현실 사회에서 규제도구로 제시한다. 그의 주장에 따르면 인터넷 공간의 질서를 구조화하는 일에서도 코드는 도구로써 기능할 수 있으며 인터넷 공간을 만들어내는 소프트웨어와 하드웨어를 만들어 구조화해내는 코드는 그 공간에서 법으로써 작동하게 된다.⁵⁵⁾

코드는 알고리즘 내에 존재하므로 결국 코드 규범화의 논의는 알고리즘 규제와 밀접하다. 알고리즘은 우리 주변에서 존재하지만, 인공지능의 알고리즘이 주목받으면서 디지털과 컴퓨팅 기술로 실행되는 알고리즘이 규제대상으로 부각되고 있다.⁵⁶⁾ 알고리즘 규제가 현재로서 가능하다고 보는 것은 특정 단계의 알고리즘 결정이 다음 단계에서 이루어지는 의사결정에 기초가 된다는 점에 있다. 즉 알고리즘 코드화는 규칙으로 표현되기 때문에 성취된 과제를 역산한다거나, 방대한 양의 데이터에도 추적이 가능한 부분이 있기 때문이다. 예를 들어 머신러닝의 학습 알고리즘은 다른 알고리즘이 데이터에서 학습된 패턴을 기반으로 적절한 결정을 내릴 수 있도록 한다. 이러한 인공지능의 단계적 의사결정 과정은 인간 운영자가 의사결정 프로세스의 최종 단계에 있는 경우에도

52) Lawrence Lessig은 헌법학자로서, 헌법 또한 법률 텍스트가 아닌 삶의 방식이고 구조(architecture)로서 건설되는 것이라고 본다. 사이버공간은 현실적인 세계이며 이를 방치할 때 정부와 상거래에 의해서 지탱되는 세력으로 통제되고 있기에 인간의 본질적인 자유를 보호하기 위해서도 이 공간에서의 코드를 이해할 때 그 질서에 개입할 수 있다고 주장한다. 로렌스 레식(김정오 옮김), 「코드 2.0」, 나남(2009), 40-41면.

53) 원문에는 ‘Cyberspace’라고 표기되었다.

54) Lawrence Lessig이 Cyberspace에서 위협으로 생각하는 규제의 본질은 정부나 이익을 추구하는 자들로부터 개인의 자유를 제한하는 것을 말하는 것이다. 그의 저서에서는 대개는 공산주의 국가들의 조직된 질서체계에서 개인의 자유가 크게 제한되지만, 인공지능 알고리즘이 통제하는 미래세계에서는 예외를 인정하지 않고 이익을 제거할 수조차 없는 완벽해 보이는 질서의 코드화야말로 규제가 된다고 한다.

55) 로렌스 레식(김정오 옮김), 앞의 책, 44-45면 참조.

56) Nicholas Diakopoulos, 앞의 논문, 3면.

결정적인 영향을 끼친다. 따라서 알고리즘의 체계는 알고리즘 소스코드에 대한 규범적 탑재, 기술적 적법절차와 같은 규제모델을 적용하기 적절한 측면이 있다.⁵⁷⁾ 다만 알고리즘은 빅데이터에 기반한 동적이고 활발한 프로그래밍을 수행하므로 규칙성을 찾거나 예측이 어려운 점, 그리고 데이터를 특정 방향으로 유도함으로써 의사결정에 편향을 줄 수 있다는 점에서 주의가 필요하다.⁵⁸⁾ 즉 알고리즘의 규제에 관한 논의는 머신러닝 알고리즘의 확장가능성에 따라 복잡한 양상으로 전개될 수 있으므로 아직은 그 효용성을 논하기는 이른 단계라고 생각한다.

3. 당사자 간 이익조정 문제

인공지능의 의사결정에 대한 위험을 사회공동체의 위험으로 인식한다는 전제에서 국가는 헌법에 따라 일차적인 책임을 지게 된다. 그 위험이 국가로부터 비롯되면 국가가, 사인으로부터 비롯되면 사인이, 사인으로부터 유발되더라도 국가가 책임지는 경우가 있을 것이다. 각자의 역할과 책임은 입법과 정책의 결정에 따라 정해지겠지만, 모든 법적 책임은 인간 의사결정자의 자발적이고 독립적인 의사결정의 실현을 목적으로 한다. 인공지능의 의사결정권에 대한 위험은 개인과 국가, 그리고 기업이 함께 분배하여야 한다. 이렇게 역할과 책임을 분배하는 구조에서는 국가, 기업, 개인이 각자의 자율성과 책임 내에서 그 기능에 충실할 수 있는 장점이 있다. 따라서 규제모델을 적용한 규제책을 도입하는 경우 검증·조사 영역을 행정기관이 주도해 강제하는 것보다 제3의 기구에 위임하고 그 결과를 각자가 평가하는 것이 사전심의의 우려를 최소화할 수 있을 것이다. 다만 현실적으로는 기업이 알고리즘의 설계와 자체 감독 및 평가에 대한 1차적 권한을 갖게 되므로, 각자의 책임은 권한에 따라서 차등적인 배분이 필요하다고 생각한다.

정리하면, 개인은 자기책임에 입각한 최종결정자로서 자기결정을 행사할 수 있어야 하며, 기업은 스스로 설계한 알고리즘의 구조화에서 개인의 자기결정을 존중하여야 한

57) Danielle Keats Citron·Frank Pasquale, 앞의 논문 43면.

58) 우선순위화(prioritizing), 분류(classification), 연관지음(association), 필터링(filtering)은 인공지능 알고리즘의 데이터학습에 따른 과정으로, 데이터의 원자적 결정을 생각함으로써 알고리즘의 힘을 평가할 수 있다. 때때로 이러한 결정은 더 높은 수준의 의사결정과 정보변혁을 형성하기 위해 강한 사슬에 묶인다. 각 과정에 대한 자세한 설명은 Nicholas Diakopoulos, 앞의 논문, 4-8면 참조.

다. 마지막으로 국가는 자기결정의 보호가 사회적 공조로 모색되어갈 수 있도록 하는 것에 그 역할이 있다. 이러한 삼각 구도에서 각각 영향력에 따른 책임을 다하는 규율 체계가 전제될 때에 인공지능과 인간의 안전한 공존을 기대할 수 있을 것이다.

참 고 문 헌

[국내문헌]

- 김영환(2020), “인간의 존엄에 대한 논의의 재구성: 형이상학 없는 인간의 존엄”, 「법철학연구」 제23권 제1호, 한국법철학회
- 김환석(2006), 『과학사회학의 쟁점들』, 문학과지성사.
- 로렌스 레식(김정오 옮김)(2009), 『코드 2.0』, 나남.
- 장민선(2018), 「인공지능(AI) 시대의 법적 쟁점에 관한 연구」 연구보고 18-10 한국법제연구원.
- 정규원(1999), 「피해자의 승낙에 대한 연구 -의료행위를 중심으로-」, 서울대학교 법학박사 학위 논문.
- 제임스 매클렐란 3세·해럴드 도른(전대호 옮김)(2006), 『과학기술로 본 세계사 강의』모티브.
- 차정미(2018), “기술 패권의 미래”; 양화융합을 통한 중국의 혁신주도 성장, 그리고 핵심기술, 미래연구 포커스, Future Horizon.
- 한창수·류영선(2008), 『지능형서비스로봇의 설계와 응용』, 청미디어

[국외문헌]

- Anders Hanssen·Stefano Nichele(2019), “Ethics of Artificial Intelligence Demarcations”, AI lab, *Oslo Metropolitan University*.
- Brad Smith·Carol Ann Brown(2018), “The future Computed Artificial Intelligence and its role in society”, *Microsoft Corporation*.
- Crawford·Jason Shultz(2014), “Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms”, *Boston Law Review* Vol.55.
- Danielle Keats Citron(2008), “Technological Due Process”, *Washington University Law Review* Vol. 85.
- Danielle Keats Citron·Frank Pasquale(2014), “THE SCORED SOCIETY: Due Process for Automated Predictions”, *Washington Law Review* Vol 89/1.

- Frank Pasquale(2015), *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge, MA: Harvard University Press.
- Geoffrey Hazard, Jr(1995), “Laws, Morals, and Ethics”, *Southern Illinois University Law Journal* Vol.19.
- Gerd Gigerenzer·Felix G. Rebitschek·Gert G. Wagner(2018), “Eine vermessene Gesellschaft braucht Transparenz”, ZBW – Leibniz – Informationszentrum, Wirtschaft.
- Jenna Burrell(2016), “How the machine “thinks”: Understanding opacity in machine learning algorithms,” SAGE Publishing, *Big Data & Society* Vol.3.
- Martha Nussbaum(2011), *Creating Capabilities: The human Development Approach*, The Belknap Press of Harvard university Press.
- Nicholas Diakopoulos(2016), “Accountability in algorithmic decision making”, *Communications of ACM* Vol.59/2.
- Nick Bostrom·Eliezer Yudkowsky(2011), “The Ethics of Artificial Intelligence”, *Machine intelligence research institute*.
- Niklas. Luhmann(2019), *Grundrechte als Institution*, Duncker & Humblot.
- Robin Murphy & David Woods(2009), “Beyond Asimov’: The Three Laws of Responsible Robotics”, IEEE Computer Society, *IEEE Intelligent Systems*.
- Ryan Calo(2015), “Robotics and the Lessons of Cyberlaw”, *California Law Review* Vol. 103/3.