

인공지능 ‘기술-윤리원칙-산업-제도’의 매트릭스와 적용*

조주은** · 정종구*** · 박도현**** · 김은수*****

목 차

I. 들어가며	III. 인공지능 산업(Industry)-제도 (Institution)
II. 인공지능 기술(Technology)과 윤리원칙(Principle)	IV. TPII 매트릭스와 적용 제언

I. 들어가며

‘인공지능(Artificial Intelligence)’이란 표현은 하나의 사회 현상이 되었다. 부가가치를 창출하는 새로운 경제 성장 동력원으로서 인공지능을 산업과 서비스에 활용하거나 그렇게 홍보하는 사례를 쉽게 찾을 수 있다. 특정 상품에 인공지능 기술이 적용되었다든지 어떤 서비스가 인공지능에 의한 분석을 기반으로 한다는 취지의 광고 문구들을 어렵지 않게 접한다. 일반적으로 소비자는 인공지능이 적용되었다는 사실의 의미를 정확하게 이해하지 못한 상태로 단순히 ‘인공지능(AI) 기술’이 적용되었기 때문에 획기적으로 개선된 상품이거나 이전과 다른 차원의 서비스일 것이라는 막연한 인식을 가질 가능성이 높다. 일반 소비자 외에 기술개발자, 경영자, 제도입안자 등도 인공지능에 대한 각자 다른 인식을 하고 있을 수 있다. 기술개발자는 인공지능 기술의 발전방향, 경영자는 인공지능 기술의 산업적 적용에 의한 사업적 이익, 제도입안자는 인공지능이 사회에 가져올 수 있는 위험요인에 보다 초점을 맞추어 생각하게 된다. 이와 같이 신기술이 일상에 적용되는 범

* 이 연구는 서울대 인공지능정책이니셔티브(SAPI)의 세미나를 기반으로 진행되었고, 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행되었습니다(NRF-2019S1A5A2A03036673).

** 제1저자, 서울대학교 법과경제연구센터 선임연구원, 법학박사

*** 공동저자, 서울대학교 일반대학원 법학과 박사과정, 변호사

**** 공동저자, 서울대학교 일반대학원 법학과 박사과정 수료, 변호사

***** 공동저자, 서울대학교 법과경제연구센터 선임연구원, 전문법학박사

위가 확대되는 시점에 다양한 이해관계자가 그 기술에 대한 편익과 위험을 각자의 관점에서 고려하는 것은 당연한 일인데, 모두가 함께 출발해야 할 지점은 일단 해당 기술의 실체(구조, 메커니즘 등), 그리고 발전방향과 단계에 대한 이해이다. 기술개발자, 사업자, 소비자, 제도입안자 등이 기술에 대한 서로 다른 그림을 갖고 있을 때(또한 경우에 따라 그 차이가 클 때), 시의적절하고 필요한 합의형성, 지속적 개발, 활용, 제도 기반 마련에 문제가 될 수 있다. 그리고 그 기술이 인류 사회에 적용되는 것이 확대되는 과정에서의 위험요인을 고려한 윤리원칙이 무엇인지, 기술 적용에 의해 인류와 사회에 편익을 가져올 산업은 무엇인지, 해당 산업이 모두에게 안정적으로 발전할 수 있도록 윤리원칙에 부합하는 제도화 과정에서 이슈는 무엇이며 이에 대해 어떻게 논의해야 하는지에 대한 준비가 필요하다.

본 연구는 이러한 맥락에서 인공지능 기술의 메커니즘과 발전과정, 이에 대해 세계 사회에서 제시하고 있는 윤리원칙, 인공지능 적용 산업 중 소비자의 기대와 기술 적용 수준이 높은 산업 영역, 각 산업 영역별 제도적인 기반 마련 과정에서 논의가 필요한 이슈들에 대해 제시한다. II 장에서는 인공지능 기술의 비전과 현황 및 윤리원칙, III 장에서는, 우선순위 높은 윤리원칙으로 제시되는 편향차별금지/공정, 프라이버시, 안전성/책임성과 이러한 윤리원칙의 관점에서 소비자의 기대 수준과 활용 가능성이 높은 인공지능 기술 적용 산업인 금융, 정밀의료, 자율주행자동차의 관련 현황과 전망을 제시하며 이 산업 영역과 관련되어 제도화가 필요한 이슈에 대해 전개해 보도록 하겠다. 그리고 IV 장에서는 이와 같은 과정으로 제시한 기술(Technology)-윤리원칙(Principle)-인공지능 기술 적용산업(Industry)- 윤리원칙에 기반한 제도화(Institution)의 매트릭스(TPII Matrix)를 논의의 틀로 제시한다.

II. 인공지능 기술(Technology)과 윤리원칙(Principle)

1. 인공지능 기술에 대한 비전과 현실

기술적인 관점에서, 세계 대전 후 1940년대 컴퓨터가 처음 개발되었을 때부터 기계가 ‘인간처럼’ 정보를 처리하고 학습하고 추론하며 상식적인 사고를 하고 추상

화 능력을 갖게 될 것이라는 기대는 항상 존재했다. 사이먼(Simon, 1965)은 “20년 안에 기계는 인간이 할 수 있는 모든 일을 할 수 있는 능력을 갖추게 될 것”이라고 했고¹⁾ 민스키(Minsky, 1967)는 “한 세기 내로 인공지능을 만드는 데 장애 요인들이 상당 부분 해결될 것”이라고 했다.²⁾ 튜링(Turing)은 이를 ‘사고 기계(Thinking Machines)’로 표현했다.³⁾ 오늘날 세계사회의 화두인 ‘인공지능(Artificial Intelligence)’이라는 말이 최초로 등장한 것은 1956년 다투머스 회의(Dartmouth Conference)에서이다. 록펠러 재단에서 후원하고 다투머스 대학교에서 열린 6주간의 워크숍에서 10명의 과학자가 신경망(neural net), 자동기계 이론(automata theory), 지능연구(study of intelligence) 대한 논의를 한 다투머스 프로젝트는 인공지능 분야 최초의 학술적 연구로 간주된다.

이와 같이 공학의 관점에서 인공지능은 어쩌면 언제나 동일한 비전을 지향하고 있었다. 그것은 “학습이나 지능의 모든 특성을(every aspect of learning and any other feature of intelligence) 원칙에 기반해(in principle) 매우 자세하게 기술하여 기계가 이를 모방(simulate)할 수 있게 하는 기대(conjecture)에 기반하며 언어를 사용하고 추상과 개념을 형성하며 인간이 풀 수 있는 문제를 해결하고 스스로 발전할 수 있는 기계”⁴⁾에 대한 비전이였다. 그런데 이는 1950년대 이후, 2020년 현재 지금도 마찬가지이다. 당시 기대했던 ‘인공 일반 지능(artificial general intelligence)’의 개발 예상 시기는 계속 늦추어졌고 오늘날의 학자들도 이러한 수준의 인공지능형 기계가 만들어지기까지는 여전히 수십 년의 시간이 걸릴 것이라고 예측하고 있다.⁵⁾

다툼스 회의 이후 연구자들은 매우 활발하게 당시까지 ‘인간만이 할 수 있다’고 여겨지는 ‘특정한 일’을 해낼 수 있는 다양한 프로그램들을 개발했다.⁶⁾ 그러나

1) Herbert Alexander Simon, *The Shape of Autonomous for Men and Management*, Harper & Row, 1965, p.96.

2) Marvin Minsky, *Computation: Finite and Infinite Machines*, Englewood Cliffs, 1967, p.2.

3) A. M. Turing, “Computing Machinery and Intelligence”, *Mind* Vol. 59, No. 236, 1950, pp.433-460; 역사적인 이 논문은 다음과 같은 질문과 전제로 시작한다. “Can machines think? This should begin with definition of meaning of the terms “machine” and “think””. 이 질문과 전제는 그가 설계한 ‘이미테이션 게임(imitation game)’으로 연결되지만, 오늘날 이보다는 눈부시게 발전한 ‘생각하는 기계’의 수준에서도 유의미한 질문과 전제이다.

4) 다투머스 회의의 기획자들이 록펠러 재단에 제출한 제안서 내용; Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014, p.5.

5) Ibid., p.3.

이러한 기계의 능력이 보다 ‘일반화’되는 데 필요한 불확실성을 다루는 방식, 당시의 프로그램 언어인 상징적 표현 의존성, 데이터의 부족, 메모리 용량과 연산 처리 속도 등의 하드웨어적 한계 등이 인지되며 1970년대 연구에 대한 투자와 지원이 줄어드는 시기를 거치게 된다. 1980년대에는 인간이 작업하는 특정 영역(human domain)의 전문가들이 일일이 손으로 입력한 지식 데이터베이스를 바탕으로 간단한 추론을 도출하는 전문가 시스템(expert system)이 수백 개 넘게 개발되었으나 이러한 소규모 시스템은 큰 도움이 되지 못했고 대규모 시스템을 개발하고 유지하기에는 그러한 자원을 감당할 수 있는 자원과 목표가 있는 일부 조직 외에는 여전히 비경제적이었다. 또다시 투자자들이 ‘AI’라는 꼬리표를 기피하는 시절을 맞이한다.

이러한 시기를 지나 90년대에 다시 새로운 국면에 접어들는데, 신경망⁷⁾과 알고리즘 기법⁸⁾이 심화·개발되면서 사람의 프로그램에 의존하지 않고 기계가 대량의 데이터 입력과 다층적 네트워크를 통해 ‘통계적인 패턴’을 인식(학습)하며 더 다양한 기능을 수행(출력)할 수 있게 되면서 오늘날의 인공지능 기술기반이 마련되었다. 기본적으로 ‘인공지능’은 인간의 뇌와 유사한 특성을 보이는 인공 신경망을 모델로 한 특정한 통계학적 계산⁹⁾을 수행하는 분류기(classifier)¹⁰⁾라고 할 수 있

- 6) 수학적리증명, 미적분학, 대수학 문제를 해결하는 프로그램, 논리추론과 지각을 통합하여 물리적 운동의 계획 제어가 가능했던 셰이키 로봇, 심리치료사 역할을 하는 Eliza 프로그램, 클래식 음악 스타일을 모방하여 음악을 작곡하는 프로그램, 특정 임상 진단에서 수련의보다 월등한 결과를 내는 프로그램, 자동차를 스스로 움직이는 프로그램, 특허를 받는 발명을 하는 프로그램 등은 이미 70년대 이전에 존재했었다.
- 7) 신경망 모델은 1958년 Roesenblatt의 퍼셉트론(Perceptron) 학습모델의 제시로 시작되었다고 보는 게 일반적이다; F. Rosenblatt, “The Perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review* Vol. 65, No. 6, 1958, pp.386-408.
- 8) 다층 신경망을 학습시킬 수 있는 backpropagation algorithm(오류역전파 알고리즘)이 인공지능이 다시 부흥하는 계기가 되었고 이는 1986년 발표되었다; 장병탁, 장교수의 딥러닝, 홍릉과학출판사, 2017, 41면.
- 9) 사실상 인공지능 알고리즘은 통계학적 메커니즘에 기반해 있다는 사실에 대한 일반인의 인지가 필요하다. 이는 인공지능의 실체가 ‘데이터’임을 자각하는 데 도움이 될 것이다. 현재 머신러닝 문제 해결에서 가장 이상적으로 여겨지는 방법론 중 하나는 베이저언 에이전트(Bayesian agent)인데 이는 확률 모델이다. 즉 특정 영역에 적합한 확률적이고 조건부적인 독립관계인 베이저언 확률 이론에 기반한다. 이를 달리 표현하면 인공지능은 인간이 생성/입력하는 데이터에 기반한 확률적 추론이라는 것이고, 인공지능 관련 윤리적, 제도적 관점의 접근이 어떻게 이루어져야 하는지에 대한 중요한 초점이 될 수 있다.
- 10) ‘인공지능’이라는 이름은 기계가 뇌의 구조의 단편(신피질의 계층구조와 통합모형)을 모사했다는 의미에서 유래한다. 인공지능이 모사한 뇌의 단편과 인공지능(보다 정확하게는 ‘패턴인식기’) 시스템은 기본적으로 ‘정보의 입력-입력정보에 대한 파라미터-입력정보를 대조할 수 있는 파라미터-입력정보의 가중치를 표시하는 파라미터-인식의 출력’으로 이루어진다. 이 과정은 (거대한, 또는 특수한) 분류의 과정이다; Ray Kurzweil, *How to Create a Mind: The Secret of Human Thought Revealed*, Penguin Group USA, 2013. 패턴인식은 다층적 구조를 거치면서 더 정확해질 수 있는데 이러한 다

고 이는 지금 현재 우리가 '인공지능'이라고 일컫는 기계의 기본적 구조이다.¹¹⁾ 이러한 신경망 기반의 머신러닝을 응용한 인공지능은 우리가 사용하는 다양한 인터넷 서비스의 기반을 이루고 있으며 수많은 소프트웨어는 이미 오랜 시간 인공지능 기술을 활용하고 있다. 예를 들어 남녀노소 불문하고 세계인들이 일상적으로 사용하고 있는 Google의 검색엔진은 가장 뛰어난 인공지능 시스템 중 하나이다.

현재 시점에서 인공지능과 일반적인 소프트웨어 사이의 경계가 그다지 뚜렷하지 않다는 것을 강조할 필요가 있는 것은 인공지능이라는 말이 가져오는 여러 가지 오해를 막기 위해서이다. 오히려 더 의미 있는 구분은 ('그것이 인공지능이라 불리든 아니든')¹²⁾, 제한적 범위의 인지능력을 가진 시스템과 일반적인 문제 해결능력을 갖춘 시스템을 구분하는 것이고 현재 사용 중인 거의 모든 시스템은 사실상 매우 제한적인 범위의 인지능력을 가진 시스템에 해당한다고 볼 수 있다.¹³⁾ 많은 학자는 일반적인 문제해결 능력을 갖춘 시스템의 도래는 최소 수십 년(20년, 또는 그 이상)이 걸릴 것으로 예측하며 그 가능성과 정당성에 대한 많은 논의가 있다.¹⁴⁾

2. 인공지능에 대한 윤리원칙

인간의 뇌의 단편(신피질의 계층구조와 통합모형)을 모사한 통계적 추론에 기반한 패턴인식기의 발전 가능성과 정당성에 대한 논의는 윤리적 관점의 접근을 필요로 한다. 이는 다음과 같은 것을 인공지능이 여전히 할 수 없고 이를 할 수 있도록 만드는 것이 당분간 쉽지 않다는 기술공학적 설명이 도움이 될 것이다.

층구조를 학습시킬 수 있는 알고리즘 기술(backpropagation algorithm, 각주 8)은 신경망 모델, 인공지능의 새로운 부흥의 시작이 되었다.

- 11) David MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003; Nick Bostrom, op. cit., pp.9, 31.
- 12) Bostrom은 우리가 '인공지능'이라고 일컫는 사례 중 인공지능이라기보다는 일반적인 응용소프트웨어로 보아야 할 것도 많음을 지적하고 인공지능으로 불리느냐 아니냐 보다는 그 기능에 초점을 맞추어야 한다고 한다; Nick Bostrom, op. cit., p.16.
- 13) Ibid., pp.15-16.
- 14) 1980년대의 전문가 시스템에서 오늘날 제한적 범위의 인지능력을 가진 시스템에 이르는 인공지능을 Narrow AI - AI with Deep Learning의 시기로, 이후의 발전단계를 보다 직관에 가까운 인지능력을 가진 인공지능(Cognitive AI) - 인간 수준의 인공지능(Human-Level AI) - 초인적 인공지능(Superhuman AI)으로 구분하기도 한다; 장병탁, "인공지능 입문", 2019 강의자료. 인공지능의 인격성(personality), 자유의지(free will), 행위주체성(agency) 등에 대한 논의는 기술적으로도 아직 도래가 요원한 Human-Level AI 이후의 기계 수준과 관련된다 할 수 있다.

인공지능이 할 수 없는 것- 상식적 추론(commonsense reasoning)¹⁵⁾

예시) 컵이 넘어지면 물이 나온다, 물체를 공중에서 놓으면 땅으로 떨어진다, 사람은 태어나기 전에는 존재하지 않는다, 물고기는 물 밖으로 나오면 죽는다, 빵은 빵 가게에서 산다, 사람들은 주로 저녁에 잠을 잔다, 운전 시 교통법규를 어기면 벌금이 나온다....

과거 많은 응용 프로그램이 그러했듯, 인공지능은 여러 특정 분야에서 매우 효율적인 정보처리를 통한 업무 수행을 할 수 있지만, 경험이 많지 않은 어린아이도 즉각적으로 알 수 있는 일이나 행위(출력)로 인한 여러 측면의 영향을 범주를 이해하지 못한다. 또한, 인공지능에게 그런 상식적 추론을 하도록 구현하는 것이 기술적으로도 매우 어렵다고 한다.¹⁶⁾ 인간의 상식은 사회를 형성하고 유지하는 근간이다.¹⁷⁾ 따라서 상식적 추론이 가능하지 않다는 것은 인간 사회에 침해적 요인이 될 수 있는 근본적인 문제가 될 수 있다.

위에서 이야기한 바대로 현재 수준의 인공지능의 기술적 구조는 입력된 정보를 기반으로 한 통계적 추론에 의한 출력이다. 따라서 다층신경망의 알고리즘 기능과 상관없이 인공지능 시스템은 입력되는 ‘데이터’에 의한 결정론적 구조이다. 즉 인공지능은 윤리적, 규범적 관점에서 인간사회의 기반이라고 할 수 있는 상식에 의해 판단하는 것이 아니라 전적으로 입력되는 데이터에 의해 판단한다. 그리고 오늘날 인공지능을 가능하게 한 ‘빅데이터’ 기반의 데이터라고 할지라도 데이터는 사람에게 의해 입력, 수집, 처리, 조작¹⁸⁾ 된다. 따라서 사람이 갖고 있는 여러 가지 특성들이

15) 장병탁, 위 강의, 정리중명, 2019.

16) 기계가 이와 같이 ‘인간과 같은 일반적인 지적 능력’을 갖고 있는 모습이 우리에게 자연스럽게 느껴지는 이유는 현실적이고 공학적 지식보다는 관련된 영화들에 의해서일 수 있다. ‘터미네이터’, ‘바이센더 니얼만’, ‘그녀(Her)’ 등의 영화는 인간의 능력을 초월하여 파괴적-물리적인 능력뿐 아니라 ‘일반화’된 판단 능력을 주체적으로 보유한-이거나 인간과 인간처럼 교감-기계도 인격과 감정의 주체가 되어 사랑하고 상처받는-할 수 있는 기계에 대한 내용이다.

17) 상식(common sense)이란 무엇이나, 인식(perception)과 의식(consciousness)이란 무엇이나, 인간에게 과연 자유의지(free will)가 있느냐 등의 논의 또한 인공지능과 관련되어 논의되고 있는 철학적 논제이다. 그러나 이는 논의의 범위를 무한히 확대하여 당장 필요한 것을 하지 못하게 하는 장애 요인이 될 수도 있다. 그러기에는 이미 인공지능이라 불리는 다양한 형태의, 그러나 위 논제와는 아직 거리가 있는 기술 적용이 인간 사회에 미치는 영향은 매우 일상적이다(한 가지 예로, 위에 제시한 Google 검색 엔진의 탁월한 검색능력으로 인한 프라이버시·개인정보침해 여부). 현실적인 상황에 필요하고 적절한 대응을 체계적, 지속적으로 하다 보면 이러한 논제에 대한 답도 그 과정에서(중장기적으로) 자연스럽게 도출될 수도 있다. 이 연구는 보다 현실적인 논제에 초점을 맞춘다.

18) 데이터나 통계적 방법론에 있어서의 조작은 위조의 의미가 아니며, 예를 들면 설계에 따라 필수적인 데이터의 전처리도 조작의 과정이다.

데이터에 반영되고 사람에 의해 변형 또는 왜곡될 수도 있다. 사회를 구성하는 사람이 -개인적으로 어떤 윤리적 기준을 갖고 있느냐와 상관없이 - 사회의 윤리원칙과 이를 반영한 제도에 의해 통제되어야 한다면 사람이 다양한 플랫폼을 통해 무작위적으로 입력하는 데이터에 의한 출력을 하여 사회에 일정한 영향을 미치는 인공지능 시스템에 대해서도 그러하다.

1950년대부터 공학자들의 일관된 비전과 꾸준한 연구로 발전되어 온 인공지능 기술이¹⁹⁾ 사회 일반의 관심을 받게 된 것이 그리 오래되지는 않았다. 신경망 연구의 발전으로 인공지능의 2차 산업화 시기라고도 불리는 2010년 이후 인공지능은 Google 포털(검색 및 이미지 인식 등), 자율주행자동차(사물인식 센서와 운행계획), 왓슨(빠른 전문지식 검색/분류/출력), 시리,알렉사(음성인식과 각종 어시스턴트), 넷플릭스(추천 알고리즘 고도화), 게임(아타리, 체스, 바둑 등), 각종 사물인터넷(로봇 청소기 및 기타 정보통신 연결 가전 등) 등에 의해 인류의 생활에 점차 그 크기와 빛을 더해가는 지니와 같이 나타났다. 그러나 사실은 지니가 아닌 앞서 살펴본 지속적인 기술기반의 발전단계에서 일상에서의 '사용성(usability)'이 수십 년 전보다 현저히 좋아진 현재 시점에 놓여 있는 것이다. 세계사회가 인공지능에 새로운 관심을 갖고 각국 정부가 정책적으로 보다 본격적인 고민을 하기 시작하는데는 2016년 알파고가 이세돌 9단에게 승리를 거둔 것이 중요한 계기가 되었다.²⁰⁾ 미국의 오바마 행정부는 2016년 국가과학기술위원회(NSTC) 산하 기계학습 및 인공지능 소위원회를 신설하면서 본격적인 국가적인 인공지능 전략을 수립했고, 유럽은 2017년 디지털 단일시장 전략에 기반한 'AI 이니셔티브'를 발표하고 2018년 개인정보권리를 강화한 '일반개인정보보호규정(GDPR)'을 시행했으며 2019년 '신뢰할 수 있는 AI를 위한 윤리 가이드라인(The Ethics Guideline of

19) 위에서 살펴본 바와 같이 '인공지능'은 기술적 현실보다는 기술적 비전이라고도 볼 수 있다. 1950년대의 '인공지능'과 2020년 현재 '인공지능'은 그 기술 수준이 비교할 수 없지만 동일하게 '인공지능'이라고 불린다. 따라서 윤리적 규범적 제도적 관점에서는 보다 적절한 용어를 고민해 볼 수도 있을 것이다. 예를 들어 '패턴인식기', '신경망 머신' 등 인공지능의 보다 현실적인 기능을 표현하는 워딩으로 그 발전수준에 맞추어 부르는 것이 기계가 사회를 어떻게 조력하도록 준비할지에 보다 도움이 될 수도 있다. 인류는 아직 '지능'에 대한 정의도 정확하게 내리지 못하고 있다.

20) 2016년 알파고와 이세돌의 대국이 인공지능 역사에서 중요한 변곡점이 된 것은 1997년 딥블루가 체스 챔피언을 꺾은 후 체스의 분기 계수보다 10배가 크고 평가함수가 훨씬 복잡하며 매번 250여 개의 경우의 수를 검토하고, 첫 세 개의 수만 해도 탐색해야 할 수가 천만 가지가 넘는 바둑에서 기계가 인간의 직관을 넘어서는 판단을 하게 되는 것은 바둑 프로그램이 개발된 1960년대 이후 기술적으로 매우 어려운 벽으로 여겨져 왔기 때문이다; 손 개리시, 기계는 어떻게 생각하는가, 이수겸 역, 이지스 퍼블리싱, 2019, 357-359면.

Trustworthy AI’을 발표했다. 독일은 2017년 자율주행자동차 사고에 따른 손해 배상 및 보험제도를 마련했고, 영국은 2018년 ‘데이터 윤리 프레임워크(Data Ethics Framework)’를 발표했다, 일본은 2017년 ‘AI 연구개발 가이드라인’을 발표했다. 그 외 각국의 학계 및 OECD, UN 등의 국제기구는 윤리 가이드라인 및 정책권고를, 그리고 Google, MS 등 AI에 앞서가는 기업도 AI 관련 자율규제 등을 활발하게 발표하고 있다.

이러한 정부, 학계, 민간에서 최근 몇 년간 활발하고 논의, 발표하고 있는 AI 관련 윤리원칙은 목적, 방식, 우선순위 등에 차이가 있지만 공통된 지향을 갖고 있다. 최근 SAPI²¹⁾에서 2018년 이후 우리나라를 포함한 각국 정부, 국제기구, 민간, 시민사회 등에서 제시하는 인공지능 윤리원칙에 대한 조사를 한 결과 30여 가지의 공통된 윤리원칙을 도출했는데, 그중 높은 순위를 차지한 것은 편향·차별금지(Bias·non-Discrimination), 프라이버시(Privacy/Surveillance), 투명성(Transparency), 안전성(Safety/Security/Robustness), 공정(Fairness), 포용성(Inclusiveness), 설명가능성(Explainability)이다.²²⁾ 이러한 윤리원칙들은 인공지능 기술이 발전하는 과정에서 원칙의 실무적 적용을 위해 지속적으로 논의될 것이다. 무엇보다 인공지능이 특정 영역에 적용되어 개발되는 과정에서 이러한 윤리원칙이 실제로 적용되고 실무적인 제도화가 이루어지는 것이 안정적이고 지속적인 기술 개발과 동시에 기술이 인류와 사회에 발전적으로 기능하고 파괴적인 역기능을 막는 데 필요하다.

Ⅲ장에서는 이 중 편향차별금지/공정, 프라이버시, 안전성/책임성에 대한 윤리원칙 및 실제 인공지능 기술 적용 산업과 관련되어 제도화가 필요한 부분에 대해 전개해 보도록 하겠다. 인공지능 윤리원칙들은 대부분 인공지능 기술 적용산업 전반에 대한 것이지만 아래에서는 주요 원칙들과 더불어 이미 상당 수준의 기술기반이 마련되어 단기/중기적으로 관련 산업의 인공지능 기술 적용 분야로 활용될 것이고 편익(benefit)과 위험(risk) 요인에 대한 고려의 필요가 높은 산업들을 다룬다.

21) SNU AI Policy Initiative: 서울대 인공지능 정책 이니셔티브.

22) 이러한 윤리원칙들은 상호 관련될 수 있으며 보다 상세한 분석 내용은 별도의 SAPI 보고서로 발표될 예정이다.

Ⅲ. 인공지능 산업(Industry)-제도(Institution)

1. 편향차별금지/공정의 윤리원칙과 금융산업

(1) 편향차별금지/공정의 윤리원칙

앞서 언급한 SAPI의 2018년 이후 민간, 정부, 국제기구, 시민단체 등에서 발표한 인공지능 관련 윤리원칙 검토 결과 편향차별금지(Bias/non-Discrimination)는 가장 많은 기관에서 발표한 윤리원칙이었고 공정(Fairness)도 높은 순위에 속했다. 이 윤리원칙들은 상호 관련된다. 미국 오바마 행정부에서 2016년 10월 발표한 “국가 인공지능 연구개발 전략 계획”에서는 인공지능 전략으로 ‘AI의 윤리적, 법적, 사회적 영향의 이해 및 대응’을 제시했는데 이 전략은 “AI 적용기술이 인류가 지키는 공식적/비공식적 규범(formal/informal norms)에 의거해 행위할 것을 기대한다.”고 시작한다. 이어서 “데이터 중심의 AI 알고리즘(data-intensive AI algorithm)은 오용, 남용 및 젠더, 나이, 인종, 경제적 수준에 대한 편향을 초래할 수 있다”는 것과 “데이터의 적절한 수집과 사용을 넘어 AI 시스템이 정의롭고(just) 공정하고(fair), 투명(transparent)하고, 책임 있게(accountable) 디자인되어야 하며 그에 기반한 행위(actions) 및 의사결정(decision making) 과정에 대해 인간이 이해(interpretable)할 수 있어야 하고 데이터가 갖고 있을 편향(bias)이 면밀히 검토되어야 한다”고 제시한다.²³⁾ EU에서 2018년 발표한 “인공지능, 로봇틱스, 자동화 시스템”에서는 AI 시스템을 훈련하고 운영하는 과정에서 사용되는 데이터의 차별적 편견(discriminatory biases in data sets)은 가능한 이른 단계에서 탐지, 보고, 제거되어야 하며 자동화 기술에 대한 평등한 접근(equal access)과 사회 전반에 걸친 경제적 전환에 대응하기에 적합한 새로운 공정 분배와 이익 공유 모델(new model of fair distribution and benefit sharing)을 만들 필요가 있다고 했다.²⁴⁾ Google은 ‘AI 이니셔티브’에서 ‘모두를 위한 AI’

23) The Executive Office of the President of the US, “The National Artificial Intelligence Research and Development Strategic Plan”, National Science and Technology Council, 2016, pp.26-27.

24) European Commission, “Artificial Intelligence, Robotics and Autonomous Systems”,

를 발표하였는데 “사회적으로 유익하고, 불공정한 편향을 만들어내거나 강화하지 않고, 안전성을 위한 개발을 하며, 인간을 위한 책임을 지고 개인정보를 보호하며 과학적 우수성에 대한 높은 기준을 유지하는 원칙에 부합’할 것을 공표하고 있다.²⁵⁾ Microsoft는 ‘AI 디자인 원칙’에서 ‘편향에 대한 대책’을 명시했고 “잘못된 추론이 차별을 만들어내지 않도록 적절하고 대표성을 갖는 리서치를 수행한다”고 선언한다.²⁶⁾

인공지능의 편향·차별금지/공정의 문제는 위에서 살펴본 바와 같이 오직 입력된 ‘데이터’에 기반하여 추론하는 인공지능 시스템의 구조적 기반이자 한계에서 비롯된다. 인공지능은 입력된 데이터에 대한 가치판단 혹은 윤리적 판단을 하지 않는다. 또한 알고리즘에 기반한 통계적 추론에 의한 출력·의사결정이 어떤 개인적/사회적 영향력을 단기/장기적으로 초래할지에 대해 판단하지 않는다. 또한, 데이터는 ‘사람에 의해’ 생성-수집-처리-입력된다. 과거 규칙(논리) 기반 프로그램만큼은 아닐지라도 알고리즘에 의한 시스템의 학습과정에도 사람이 개입할 수 있다. 따라서 데이터의 ‘생성과정’에서 의도되지 않은 편향(unintentional bias)이 발생하고, 데이터의 ‘설계 및 처리 과정’에서 의도된 편향(intentional bias) 또한 발생 가능하다. 다음에서는 금융산업에서의 AI 적용기술에 대해 알아보고 편향·차별금지/공정과 관련되어 어떤 이슈가 있는지, 그리고 제도화 과정에서 필요한 내용에 대해 다루어 보도록 하겠다.

(2) 금융산업의 인공지능 기술 동향

컴퓨팅 파워가 증진되고 머신러닝·딥러닝에 의한 알고리즘 기술이 발전하며 인터넷과 모바일기기 도입으로 인해 빅데이터가 축적되면서 금융산업에서도 인공지능이 적극적으로 활용될 수 있는 기술적 토대가 마련되었다.²⁷⁾ 또한 산업적 관점에서는 금융혁신을 통한 이윤 창출 증대, 2008년 금융위기 이후로 금융규제 강화

Statement. Luxembourg: Directorate-General for Research and Innovation 30, 2018, p.17.

25) <https://ai.google/principles/> (방문일: 2020.5.17.)

26) <https://www.microsoft.com/en-us/ai/responsible-ai> (방문일: 2020.5.17.)

27) Oliver Fitch, Su Yang, “Artificial Intelligence: challenges for the financial sector”, *ACPR-Banque de France*, 2018, pp.7-8.

로 컴플라이언스에 대한 강한 요구가 인공지능 기술의 금융영역 적용 기회요인이 되었다. 인공지능 기술의 금융산업 적용 유형은 크게 대고객 서비스, 고객 신용평가 및 컴플라이언스와 같은 운영업무로 구분할 수 있다.

대고객 서비스에서 '금융챗봇'은 채팅창에서의 질의응답 서비스인데 업무시간 외 어느 장소에서나 고객 서비스가 가능하다.²⁸⁾ 금융사 입장에서는 챗봇을 통해 비용 절감과 고객정보 확보라는 편익을 얻을 수 있다. 금융사는 고객에게 상품 판매를 하는 과정에서 필요한 직원 고용, 교육, 장소 제공, 고객 대면 서비스 등의 비용이 절감되고 고객도 본인이 필요한 내용에 대해 시간과 장소에 구애받지 않고 안내받을 수 있는 편익이 있다. 또한 금융사는 챗봇을 통해 고객정보를 획득할 수 있는데 텍스트마이닝을 통해 주고받은 대화에서 드러난 고객의 성향을 파악하고 타겟 마케팅(target marketing)을 하여 수익 창출의 기회를 만들 수 있다. 국내 금융챗봇은 아직 금융소비가 던지는 질문에 대해 단순한 답변을 하는 수동적인 서비스를 제공하는 경우가 많으나 해외에서는 보다 고도화된 서비스를 제공하는 사례들이 있다. 예를 들면, 미국의뱅크오브아메리카에서 개발한 챗봇인 Erica는 고객의 질문에 대한 단순한 답변 외에 저축하고 싶은 목표액을 설정해 놓고 매달 부족한 저축액을 알려주는 등의 개인화된 서비스를 제공한다.

자동화된 방식으로 고객에게 투자자문을 하거나 투자관리를 해주는 '로보어드바이저(Robo-advisor)'도 대고객 서비스에 해당된다. '로보어드바이저'는 고객의 포트폴리오를 관리하거나 자산관리 영역에서 인공지능 기술을 활용한다.²⁹⁾ 챗봇 서비스를 통해 고객 데이터를 확보하고 분석하여 적은 비용으로 고객의 위험성향과 투자목표에 부합하는 포트폴리오를 설계할 수 있다. 이를 통해 보다 양질의 자문 서비스를 제공하고 자문기록을 보관함으로써 보다 투명한 절차에 따른 자문을 할 수도 있다.³⁰⁾ 또한 데이터를 학습하여 미래를 예측하고 실시간으로 거래를 하는 알고리즘 트레이딩을 통해 자산관리를 할 수 있다. 이는 실제로 신속한 거래를 통해 차익을 노리는 고빈도 매매(high-frequency trading)에서 활용되고 있으며,³¹⁾ 정확성을 높여 트레이더가 실수로 거래내역을 잘못 입력하는 '팻핑거 거래

28) 한국정보화진흥원, "인공지능 기반 챗봇 서비스의 국내의 동향분석 및 발전 전망", *D.gov Trend & Future* 2018-2호, 2018, 13면.

29) 이근영, "국내외 로보어드바이저 동향 및 분석", *전자금융과 금융보안* 제6호 2016-04, 2016, 40-41면.

30) Bonnie G. Buchnan, "Artificial intelligence in Finance", *The Alan Turing Institute*, 2019, pp.11-14.

(fat finger trades)'를 방지하는 데 도움이 된다.

신용평가 영역에서도 인공지능 기술 활용이 될 것으로 전망된다.³²⁾ 예를 들면, 고객에게 대출 등을 할 때 필요한 신용평가 과정에서 기존에 금융거래 기록이 충분한 고객 외에 사회초년생이나 경력단절자와 같이 최근 금융거래 기록이 부족하거나 없는 경우 요금납부기록, 통화기록, 소셜네트워크 정보 등의 비금융정보를 분석하여 고객의 신용평가를 하는 것이다. 특정 고객에게 특정 조건 하에서 일정한 비금융정보에 가중치를 부과하여 신용도를 평가하는 과정에서 인공지능 기술이 활용될 수 있다. 즉 금융정보 외에 고객이 생성한 정보나 이종거래의 신용 패턴 등 다양한 비금융정보를 머신러닝으로 분석하여 신용평가를 하는 것이다. 기술이 발달할수록 딥러닝 기법에 따라 보다 정교한 신용분석을 하게 된다. 즉 인공지능을 통해 빅데이터 분석을 할 수 있으며 이를 통해 변화하는 데이터를 실시간으로 분석할 수 있게 된다.

금융사에 부과되는 각종 규제를 준수하는 과정에서는 인공지능 기술이 현재 활발히 활용되고 있는데 금융사에 의한 활용과 금융감독 기관에 의한 활용으로 구분된다. 우선 금융사가 자체적으로 시행하는 준법감시의 경우 자금세탁방지(AML, Anti Money Laundering)와 이상거래탐지(FDS, Fraud Detection System)에서 인공지능 기술이 사용되고 있다.³³⁾ 가령 이상거래 탐지의 경우 인공지능 기술은 주어진 데이터를 스스로 학습하며 패턴을 실시간으로 생성하고 적용함으로써 이상거래를 발견하는 업무를 자동적으로 수행한다. 은행의 경우 적절한 자기자본 비율을 유지하면서 자본을 최적화해야 하는데(capital optimization) 그 과정에서 인공지능 기술을 활용하기도 한다. 또한 금융감독기관이 금융사의 규제 준수여부를 감시·감독하는 과정에서 인공지능 기술을 활용하는데 레그테크(Regtech)와 쉐테크(Suptech)로 구분할 수 있다. 레그테크³⁴⁾란 미시적인 금융규제를 자동화하는 것이다. 금융사의 데이터를 추출하고 분석하며 활용하는 과정에서 인공지능 기술을 활용해 금융사가 데이터를 임의적으로 조작하지 못하게 함으로써 규제의 실효성을

31) 김정수·신영재, "알고리즘매매와 고빈도매매의 글로벌 규제동향", 증권법연구 제14권 제2호, 한국증권법학회, 2013, 750면.

32) OECD, *Artificial Intelligence in Society*, OECD Publishing, 2019, pp 55-57.

33) 금융보안원 보안기술연구팀, "머신러닝 기반의 이상거래 탐지시스템 동향", 2017, 6면.

34) 유원형, "[트렌드탐구] 新 핀테크 '레그테크'와 '쉐테크'가 주목받는 이유", 메가경제, 2019.7.5., <http://www.megaconomy.co.kr/news/articleView.html?idxno=91721>, (방문일: 2020.5.20.)

높인다. 셉테크³⁵⁾란 중앙은행이나 거시건전성 감독기구 같은 공적인 기구가 관리 감독업무에 머신러닝을 활용하는 것이다. 금융감독당국은 이를 통해 시장의 변동성이나 유동성 위기를 비롯하여 금융위기 등을 선제적으로 파악하고 대처할 수 있을 것으로 기대된다.

보험사에서는 새로운 보험상품을 개발, 판매, 위험관리하는 과정에서 인공지능 기술을 활용한다. 상품을 개발하고 보험료를 책정하는 단계에서는 사물인터넷을 활용할 수 있는데 예를 들면, 웨어러블 기기를 활용하여 피보험자의 건강증진 활동을 체크하고 사고 발생 위험을 줄어나갈수록 보험료를 깎아주는 방식이다. 보험위험 인수 여부를 판단하는 과정에서는 피보험자의 위험을 분석할 때에 인공지능 기술을 활용할 수 있으며, 보험상품을 판매하는 단계에서는 위에서 살펴본 챗봇과 로보어드바이저를 이용하여 고객상담을 제공할 수 있다. 그 과정에서 축적된 데이터를 가지고 판매 타겟을 분석하고 우량고객을 파악하여 보다 효율적인 보험상품 마케팅을 할 수도 있다. 위험관리 단계에서는 인공지능 기술을 활용하여 예측모형을 수립하고 관리할 수 있으며, 보험사고 발생위험을 사전에 파악하고 선제적으로 경감시키기 위한 조치를 고안할 수 있다. 보험금 지급과 심사의 단계에서는 드론을 활용한 손해사정이나 사고견적 파악을 위한 정보 수집이 가능할 것으로 예상된다.³⁶⁾

(3) 편향차별금지/공정의 윤리원칙과 금융 인공지능

살펴본 바와 같이 금융산업에서 인공지능을 활용하게 되면서 금융서비스의 수준이 향상되고, 금융기관이 효과적으로 위험을 관리할 수 있으며 새로운 수익모델을 창출할 수 있을 것이다. 기존에 금융시장에 접근하지 못했던 이들에게 금융시장에 접근할 기회를 제공함으로써 포용적 금융서비스를 제공할 것으로 기대되기도 한다.³⁷⁾ 이러한 편익과 더불어 편향차별금지/공정성의 윤리원칙과 관련된 문제가 제기된다.³⁸⁾

35) BIS(Bank for International Settlements), "Innovative technology in financial supervision (suptech): The experience of early users", *FSI Insights on policy implementation* 9, 2018, pp. 4-12.

36) 김규동, "인슈어테크 활성화를 위한 규제개선 방안, 글로벌 핀테크 규제환경 분석과 개선 방향 세미나" (2019.6.17.), 7면.

37) 고학수 · 김용대 · 김건우 · 윤주호 · 이나래 · 정종구, *금융권 인공지능 정책 연구*, 금융보안원, 2019, 131면.

38) 위의 책, 133면 이하.

인공지능 기술은 투입 데이터가 일정한 알고리즘을 거쳐 산출 데이터로 전환되는 과정을 전제하는데, 그 과정에서 합리적인 이유가 없는 차별이 발생할 수 있다.³⁹⁾ 크게 세 가지 차원에서 차별을 생각할 수 있는데 기존의 가용데이터가 편향되어 있는 경우, 데이터를 처리하는 알고리즘에 왜곡이 있는 경우, 산출된 결과물의 해석이 잘못된 경우이다. 금융 서비스 과정에서 이는 다양한 모습으로 나타날 수 있는데, 대표적인 것이 신용대출에서 금융사가 고객에게 대출 승인을 거부하게 되는 경우이다. 우선 신용대출을 하기 위해서는 고객의 신용을 평가할 수 있는 데이터가 있어야 하는데 가용데이터의 출처(genesis)와 형태(structure)는 제한되어 있을 수밖에 없다. 설령 충분한 데이터를 취득한다고 하더라도 그 데이터가 도출되었던 사회의 관행이나 관습이 데이터 자체에 편향을 야기할 수 있다. 예를 들면 여성을 대상으로 신용대출을 하는 과정에서 상대적으로 여성의 사회진출이 적었던 시점의 데이터까지 고려한다면 그 결과는 오늘날 현실에 부합하지 않게 도출되어 신용대출을 원하던 고객을 부당하게 차별할 수 있다. 또한 신용대출을 하기 위해서는 고객의 신용을 평가하는 알고리즘이 있어야 하는데, 여기에는 다양한 투입값을 일정한 가중치에 따라 평가하여 등급을 산정하는 모델이 전제된다. 이미 데이터의 편향이 만연해 있는 경우 이를 분별없이 고려하는 신용평가 알고리즘은 왜곡되어 있다고 평가할 수 있다. 설령 투입데이터가 편향 없이 잘 구성되어 있고 충분한 양이 확보되어 있더라도 신용평점을 산출하는 알고리즘에서 가중치 결정에 문제가 있다면 그에 따른 신용평가 결과는 고객에게 부당한 차별을 유발할 수 있다. 마지막으로 산출데이터를 해석하는 과정에서 편견이 개입하고 차별이 유발될 수 있다. 인공지능 기술을 활용하여 입력데이터를 분석하여 해석한다는 것은 결과값을 어떠한 기준에 따라 판단하게 되는 것을 전제한다. 그런데 산출데이터를 해석하는 기준이 부당하다면 아무리 투입데이터가 충분하며 양질이며 알고리즘이 적절하더라도 왜곡된 결과를 도출할 수밖에 없다. 신용대출 과정에서 고객에 대해 도출된 신용평점을 잘못 해석한다면 고객을 부당하게 차별하게 된다.

편향차별금지/공정의 이슈는 투명성(transparency)의 문제⁴⁰⁾와 연결된다. 앞서 살펴본 인공지능의 다차원 신경망 모델은 소위 블랙박스라 불리는 숨겨진 레이

39) 고학수·정혜빈·박도현, “인공지능과 차별”, 저스티스 통권 제171호, 한국법학원, 2019, 199면 이하.

40) 이선규, “알고리즘의 투명성과 설명가능성”, 서울대학교 인공지능정책 이니셔티브 이슈페이퍼 2019-2, 2019, 9-19면.

어에서의 분석과정을 거치게 된다. 기존의 규칙기반 컴퓨터 프로그램과 달리 딥러닝 같은 학습기반 알고리즘의 경우 여러 층위의 중간 레이어를 거치며 가중치가 산정되고 결과값이 도출되는데 이 과정을 인간이 충분히 이해하기 어려울 수 있다. 따라서 이러한 과정을 거친 인공지능 기술이 금융사의 가치판단이나 의사결정에 활용되는 경우 위에서 살펴본 그 결과의 편향·차별금지/공정의 문제와 더불어 법령상 요구되는 설명의무를 수행하는 과정에서 문제가 될 수 있다. 이는 인공지능 윤리원칙의 설명가능성(explainability)과도 연결된다.

(4) 제도화 과정의 고려사항

위에서 살펴본 바와 같이 편향·차별금지/공정, 투명성, 설명가능성은 상호 관련된 윤리원칙인데 관련되어 법제도화가 이루어졌으나 여전히 향후 세부적인 제도화 과정에서 풀어야 할 문제들이 있다. 2020. 1. 9. 개정된 '신용정보의 이용 및 보호에 관한 법률' 제36조의2에 따르면, 개인인 신용정보주체는 금융회사 등에게 자동화평가 실시여부, 자동화평가의 결과 및 주요기준과 기초자료 등의 설명을 요구할 수 있게 되었다. 자동화평가 결과의 산출에 유리하다고 판단되는 정보의 제출 또는 기초정보의 정정·삭제, 자동화평가 결과의 재산출을 요구할 수 있는 권리도 도입하였다. 이는 자동화평가에 대해 정보주체에게 적극적인 대응권을 부여한 것이다. 또한 개정 '신용정보의 이용 및 보호에 관한 법률' 제22조의4는 개인신용평가회사가 개인신용평가를 할 때 성별, 출신지역, 국적 등으로 합리적 이유 없이 차별하는 행위, 개인신용평가 모형을 만들 때 특정한 평가항목을 합리적 이유 없이 유리하게 또는 불리하게 반영하는 행위, 신용평가회사의 계열회사로부터 상품 또는 서비스를 제공받는 개인인 신용정보주체의 개인신용평점을 높이는 등의 불공정행위를 명시적으로 금하고 있다. 또한 제22조의5는 개인사업자인 신용평가회사가 개인사업자의 신용상태를 평가할 경우 당 개인사업자에게 평가에 불이익이 발생할 수 있는 정보 외에 평가에 혜택을 줄 수 있는 정보도 함께 고려할 것, 개인사업자인 신용평가회사와 금융거래 등 상거래 관계가 있는 자와 그 외의 자를 합리적 이유 없이 차별하지 아니할 것, 그리고 이해상충 방지, 불공정행위 금지, 개인사업자의 특성에 적합한 신용상태의 평가기준에 관한 사항 등을 내부통제기준으로 수립할 것 등을 행위규칙으로 명시하고 있다. 향후 이러한 조항을 인공지능 기술 관련해서 어떻게

구현할 수 있을 것인지가 문제된다.

금융서비스 전 과정에서 법령상 요구되는 설명의무에 대해서도 마찬가지로 문제가 제기된다. 은행에서 고객에게 금융 서비스를 제공하는 과정에서는 은행법에 따라 금융거래 단계별로 계약의 주요 내용이나 약관에 대해 자료를 제공하고 내용을 설명해야 하며(은행법 제52조의2 제2항, 동법 시행령 제24조의4 제2항), 금융투자업자가 일반투자자에게 금융서비스를 제공하는 과정에서는 자본시장법에 따라 적합성 원칙을 준수해야 하고 설명의무를 이행하여야 한다(자본시장과 금융투자업에 관한 법률 제46조, 제47조). 보험회사 또는 보험모집인도 보험계약자에게 보험계약 체결을 권유할 때 보험계약의 중요사항을 보험계약자가 이해할 수 있도록 설명해야 한다(보험업법 제95조의2). 그러나 이러한 전통적 설명의무를 인공지능의 자동화된 의사결정에 어떻게 적용될 수 있을가에 대하여는 불명확한 상황이기 때문에 앞으로도 추가적 고민이 필요하다. 예를 들어, 금융챗봇을 활용하여 고객에게 상담하는 과정에서 신규상품 가입을 유도하는 상황에서의 설명의무 구현방안이 문제될 수 있다.

이러한 문제 외에 금융산업에 적용된 인공지능 기술이 금융시스템 위험(systemic risk)을 초래할 수 있다는 점과 인공지능 개발업체(developer)에 대한 규제가 상대적으로 어렵다는 점도 문제가 된다. 위에서 다룬 데이터의 편향, 알고리즘 오류, 해석과정의 실수가 금융시스템 차원의 거시적인 문제로 확대될 수 있다. 알고리즘을 이용하여 자산관리를 하는 경우 그 방향성을 예측하기 어렵고 알고리즘 간의 거래로 인해 자칫하면 기존의 오류가 매우 크게 증폭될 수도 있다. 또한 금융 분야 인공지능 기술을 개발하여 공급하는 업체가 소수인 업계의 현실에서 금융 데이터가 공급되는 과정에서 문제가 될 수 있다. 금융회사가 금융위원회에 등록하거나 인가를 받는 것을 규제 시점으로 하는 기존 금융 감독 체계와 관련한 문제도 있다. 가령 금융투자업의 등록·인가요건의 경우 대주주 적격 요건 등을 요구하는데 이는 금융투자업에 진입하는 데에 충분한 사회경제적 신뢰와 자본력을 보유하고 있는지를 기초로 판단한다는 의미이다. 금융 인공지능 기술개발 업체가 규제 체제에 들어가는 것이 윤리원칙 준수 감독을 위해 바람직할 수 있으나 금융투자 관련 기술을 개발한 후에도 위와 같은 기존요건 때문에 금융투자업의 규제체제에 편입되지 못함으로써 규제공백이 생길 수도 있다.

2. 프라이버시 윤리원칙과 정밀의료

(1) 프라이버시 윤리원칙

프라이버시 문제는 인공지능이 크게 화두가 되고 있는 2016년 이후가 아닌 이른바 3차산업혁명 이후 정보통신기술이 인류의 일상과 소통의 방식을 근본적으로 바꾼 이후부터 대두된 가장 중요한 정보통신 사회의 이슈이다. 거대한 네트워크가 되어버린 사회에서의 '개인의 영역'은 그 의미, 침해와 보호의 방식, 정보에 대한 개인의 주체성과 객체성, 정보의 제공-수집-처리-활용(분석)-저장-공유-재사용 등의 전 과정에서 편익과 위험 등에 대한 새로운 접근과 규제를 필요로 하게 되었다. 두말할 나위 없이 그 정보의 대량 축적에 기반한 '인공지능 사회'에서 이러한 이슈의 중요성은 데이터에 대한 의존성·중요성에 비례해서 커지고 있고 앞으로도 그럴 것이다.

앞서 살펴본 SAPI의 2018년 이후 각 기관 공표 인공지능 관련 윤리원칙 조사에서 '프라이버시'는 위에서 다룬 '편향차별 금지'와 더불어 가장 많은 기관이 공히 제시한 윤리원칙이다. 미국은 2014년 39억 달러(USD) 규모로 조성된 '네트워킹과 정보 기술 연구와 개발'에 대한 투자 중 약 8천만 달러(USD)를 프라이버시 연구에 투자했다. 이 연구는 의료정보에 대한 프라이버시, 프라이버시 규제, 프라이버시에 대한 학제적 연구, 보안 등을 포함한다. 그 결과물 중 하나로 2016년 6월 발행한 "프라이버시 연구에 대한 국가전략"에서 프라이버시를 '정보주체(subjects)-데이터(data)-행위(actions)-맥락(context)'이라는 네 개의 주요 개념에 기반해 제시했다.⁴¹⁾ EU는 2016년 '디지털 단일시장'을 기치로 회원국 간의 개인정보의 자유로운 이동을 보장함과 동시에 데이터 주체의 개인정보보호에 대한 권리를 강화하는 GDPR(General Data Protection Regulation)을 제정하고 2018년 5월부터 시행하고 있다. 이는 1995년부터 적용되었던 개인정보지침(EU Directive on Protection Directive 95/46/EC)을 대체하는 것으로 데이터 처리자에 대한 개인정보 책임 부여, 데이터 주체에 대한 권리 강화, 위반 시 강력한 과징금 처벌규정 등을 주요 내용으로 한다. GDPR은 EU 외에 EU 시민들을 고객으

41) National Science and Technology Council, "National Privacy Research Strategy", 2016, pp.3-5.

로 하는 외국 회사에도 적용되기 때문에 세계 각국의 규제 과정에 중요한 참조가 되고 있다. 우리나라에서는 2020년 1월 개인정보보호 관련 3개 법률(개인정보보호법, 신용정보 이용 및 보호에 관한 법률-이하 신용정보법-, 정보통신망 이용촉진 및 정보보호 등에 관한 법률-이하 정보통신망법-) 개정안이, 발의 후 1년 2개월 만에 국회 본회의를 통과하였다. ‘개인정보보호법’은 개인정보 가명처리 개념 도입을 통한 데이터 활용성 확대, 데이터 처리 전문기관 지정, 개인정보위원회 위상 강화, ‘신용정보법’은 금융분야 데이터 분석/활용의 법적 근거 제시, 개인정보산업 도입 및 개인정보 자기결정권 강화(MyData⁴²), ‘정보통신망법’은 개인정보보호에 대한 내용을 ‘개인정보보호법’으로 이관하여 법제와 감독을 일원화하는 것을 주요 내용으로 한다.

개인정보보호에 대한 개인과 사회의 인식도 이미 상당한 수준으로 올라와 있다. 우리나라 국민 1000명을 대상으로 개인정보 활용에 대한 인식 조사를 한 결과 응답자의 62.9%가 데이터 3법을 잘 알거나 들어본 적이 있고, 이 중 45%는 법 개정이 바람직하다고 응답했으며 4차산업혁명 대응을 위한 데이터 활용은 ‘필요하다’가 70.2%로 응답하였고 개인정보 제공의향은 ‘공익적 목적에 기여’할 때가 84%로 가장 높았다. 개인정보 제공 여부 결정 시 중요한 요인은 ‘정보 활용 모니터링 권한’이 96.4%로 높게 나타났다. 개인정보 활용 서비스 분야 수요는 교통/안전(34.5%), 의료(22%), 금융(14.7%) 순으로, 정책지원 수요는 ‘컨트롤 타워/제도 마련’이 47%로 가장 높게 조사되었다.⁴³ 다음은 이러한 일반 시민의 인식에 대한 조사결과와 더불어 데이터 기반 신산업의 핵심 분야 중 하나로 전망되는 의료산업에서 인공지능 기술이 지향하고 있는 정밀의료 사례를 다루어 보고 프라이버시 윤리원칙 관련되어 제도화 과정에서 고려해야할 내용에 대해 알아보도록 하겠다.

42) 개인이 보다 능동적, 적극적으로 자기 정보에 대한 결정권을 행사할 수 있는 기반으로 자신의 데이터 활용 편익을 스스로 고려하여 원하는 서비스를 받을 수 있도록 하는 것이다. 예를 들면 금융기관, 병원, 통신사 등에 있는 자신의 개인정보를 다운받아 제3의 업체에 전달해 맞춤형제테크, 건강관리, 맞춤형 요금제 추천 등과 같은 서비스를 받을 수 있도록 하는 것이며 ‘데이터 이동권’의 개념과도 관련된다.

43) 배영임·신혜리, “데이터3법, 데이터 경제의 시작”, 경기연구원 이슈&진단, 2020, 7-11면.

(2) 정밀의료의 인공지능 기술 동향⁴⁴⁾

인공지능 기술의 발전과 빅데이터 분석의 확산에 따라 '정밀의학' 혹은 '정밀의료'로 번역되는 'Precision Medicine (이하 '정밀의료')'이라는 새로운 의료 패러다임이 주목받고 있다. 정밀의료는 오늘날 과학기술 수준으로 수집할 수 있는 최대한의 유용한 정보를 한데 모아서 총체적으로 분석함으로써 질병의 사전예방, 진단, 치료, 연구 등에 있어서 보다 유효한 결과를 도출하려는 것을 목표로 한다. 같은 질병을 가진 환자라도, 환자 개인마다 가지고 있는 DNA 배열이 다르고, 똑같은 DNA를 가진 일란성 쌍둥이도 구체적으로 노출된 환경에 따라 발병양상이 달라지기도 한다. 약리학의 발달 때문에 똑같은 약물을 투여한다고 하여도, 개인마다 약물의 효능이 나타나는 감수성이 달라질 수 있다는 점도 파악하게 되었다. 그밖에 나이나 성별, 기타 요인에 따라 집단별로 치료의 양상에 차이가 나는 경우도 있는데 이러한 차이를 종합적으로 고려하여 종전처럼 평균인을 기준으로 한 접근법(one-size-fits-all approach)에서 벗어나 보다 세분화된, 나아가 개별화된 접근법(personalized approach)으로 패러다임 전환을 가능케 한 점에서 정밀의료의 의의를 찾을 수 있다. 왓슨(Watson)으로 널리 알려진 진단은 물론이고 표적 치료제 개발, 실시간 건강관리를 비롯한 다양한 영역에서 발전한 인공지능 기술과 ICT 기기가 정밀의료 패러다임을 점차 실현해나가는 상황이다.⁴⁵⁾ 세계적으로 보건의료 비용이 천문학적으로 증가해나가는 상황에서, 정밀의료 패러다임은 불필요한 지출을 최소화함으로써 재정 건전성을 도모할 수 있는 주요한 기반이 되기도 한다.⁴⁶⁾

정밀의료는 2003년 인간게놈프로젝트(Human Genome Project) 완료 및 전장유전체 서열분석 활성화 이후 추진되어 온 '개인 맞춤형의료(tailored personalized medicine)'의 연장선상에 위치한다.⁴⁷⁾ 최근 들어 급속히 발전한 인공지능 기술과

44) 아래의 내용은 대체로 고훈수·이동진·박미정·김은수·신수용·김현창·박유량·차효성, 보건의료 빅데이터의 보호 및 활용을 위한 법적 기반 강화 연구, 보건복지부 연구보고서 초안(미간행), 2019를 요약하고 발전시킨 것임을 밝혀둔다.

45) 김무웅, “글로벌 정밀의료 시장 현황 및 전망”, *BioIndustry* 제131호, 2018, 5-6면.

46) 국내 의료비 지출액은 OECD 평균에 미달하지만, 증가속도는 이를 크게 상회한다. 2005-2009년 기준 국내 연 의료비 증가율은 9.0%, OECD 평균은 3.4%였고, 2009-2013년의 국내 연 의료비 증가율은 5.4%, OECD 평균은 0.6%였다. 이는 같은 기간 국내 GDP 증가율을 훨씬 넘어선다. 김대환, 효율적 의료비 지출을 통한 국민건강보험의 보장성 강화 방안, 보험연구원, 2017, 20면 참조.

이를 토대로 한 빅데이터 분석을 통해 개인이 보유한 유전정보, 생체정보, 환경정보 등을 총체적으로 탐구할 수 있게 된 것이다. 다만, 환자마다 제각기 다른 치료를 한다는 것보다 최적의 의료를 한다는데 초점을 맞출 필요가 있어, 전미연구평의회(National Research Council)는 ‘맞춤의료’ 대신에 ‘정밀의료’라는 용어를 권장하였고,⁴⁸⁾ 그것이 오늘날까지 이어지게 되었다. 2015년 1월 오바마 대통령이 ‘정밀의료 선도계획(Precision Medicine Initiative, 이하 ‘PMI’)'을 연두교서로 발표하면서 정밀의료는 많은 주목을 받게 되었다.⁴⁹⁾ 영국에서는 ‘100,000 Genomes Project’를 통해서 2018년 말 10만 개의 유전자에 대한 분석을 마무리하고 본격적으로 정밀의료 분석 사업에 돌입하였고, 일본과 중국을 비롯한 선진국에서도 많은 투자를 진행 중이다.⁵⁰⁾ 우리나라도 정밀의료 분야에 많은 노력을 기울여왔다. 2014년부터 2021년까지 시행되고 있는 포스트게놈 신사업육성을 위한 다부처 유전체 사업에 5,788억 원의 예산을 투입하였고, 2015년에는 미국과 한·미 정밀의료 연구 협력의향서(LOI)를 체결하였으며, 2016년 8월 과학기술전략회의 9대 국가 전략 프로젝트 중 하나로 정밀의료 분야가 선정되었고, 보건복지부에서는 2017년 정밀의료 사업단을 발족하였다.⁵¹⁾ 과학기술정보통신부는 2019년 7월 우리나라에서 자체적으로 개발한 인공지능 정밀의료 서비스를 임상에 적용하기도 하였다.⁵²⁾

정밀의료의 또 다른 중요한 의의는 과거와 다르게 환자의 권리와 능동적 참여를 중시한다는 점이다. 일명 ‘4P’로 널리 알려진 ‘예측(Predictive), 예방(Preventive), 맞춤형(Personalized), 참여(Participatory)’라는 정밀의료의 네 가지 특징 중 ‘참여’ 부분이다.⁵³⁾ 예를 들면, 미국의 만성, 희귀질환 환자들의 커뮤니티

47) Francis S. Collins · Harold Varmus, “A New Initiative on Precision Medicine”, *The New England Journal of Medicine* Vol. 372, No. 9, 2015, p.793.

48) National Research Council, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*, National Academies Press, 2011, p.124.

49) The White House, “FACT SHEET: President Obama’s Precision Medicine Initiative”, 2015.

50) 관계부처 합동, “바이오헬스 산업 혁신전략”, 2019, 5면.

51) 박순영, “유전체 기반 정밀의료 연구동향”, 융합연구정책센터, 2017, 8면; 보건복지부는 2017년부터 2021년까지 난치암 환자 유전변이에 맞춘 표적치료제 개발에 430억 원을, 환자 맞춤형 의료서비스 제공을 위한 클라우드 기반 정밀의료 병원정보시스템 개발에 201억 원을 투입할 계획이라고 밝혔다. 보건복지부, “개인맞춤의료 실현을 향한 정밀의료 사업단 발족”, 2017, 2면 이하.

52) 과학기술정보통신부, “토종 AI 정밀의료 서비스, ‘닥터 앤서’ 임상 적용 시작”, 2019, 1-5면.

니티인 'Patientslikeme'라는 사이트에는 환자들의 실제 경험담을 토대로 한 각종 의료정보가 집적되어 있고, 이를 토대로 보다 높은 치료효과를 거두었다는 연구가 보고되기도 하였다.⁵⁴⁾ 오바마 정부는 환자의 접근권이 실질적으로 보장될 수 있도록 플랫폼을 개발하고 새로운 정책을 입안했다. 대표적으로 환자가 의료정보를 온라인으로 다운로드 받을 수 있도록 한 블루버튼 이니셔티브는 '건강보험 이전 및 책임에 관한 법률(HIPAA)'에 법제화되고, 전자의무기록(Electronic Health Record)을 연구에 자발적으로 기부할 수 있도록 한 'Sync for Science(S4S)' 프로젝트로 확대되었다.⁵⁵⁾

(3) 프라이버시 윤리원칙과 정밀의료

정밀의료를 달성하기 위하여 가장 중요한 것은 양적·질적으로 우수한 개인의 특성에 관한 데이터를 확보하는 일인데 '인구집단 코호트(cohort)' 구축이 대표적이다. ⁵⁶⁾ 국내의 경우, 질병관리본부 국립중앙인체자원은행에서 추진하는 한국인체자원은행사업(Korea Biobank Project, KBP)이 대표적 사례이다. 2008년 처음 시작된 이후 현재 제3기 사업을 진행하고 있고, 17개 대학병원에 소재한 인체자원단위은행에서 구성한 한국인체자원은행네트워크(Korea Biobank Network, KBN)를 통해 약 83만 명분의 인체자원을 수집하여 국내 연구자에게 분양하기도 했다.⁵⁷⁾ 유사한 사업으로 한국인 유전체 역학조사사업이 있는데, 여기서는 대규모 코호트를 바탕으로 한국인에게 흔한 만성질환의 위험요인을 규명하고 생체지표를 발굴한다. 마찬가지로 결과물이 국내 연구자에게 분양되기도 했다.⁵⁸⁾ 그밖에 국민

53) 박성해, “대규모 인구집단 유전체정보 기반 정밀의료 핵심인프라 구축”, 보건산업 동향, 2015, 3면.

54) Paul Wicks · Timothy E Vaughan · Michael P Massagli · James Heywood, “Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm”, Nature biotechnology Vol. 29, No. 5, 2011, pp.411-416.

55) 정일영·구원모, “헬스케어 생태계 구축을 위한 데이터 통합 방안”, 과학기술정책연구원 동향과 이슈 제46호, 2018, 11-16면.

56) 김봉조·김영진·문상훈, “컨트롤 타워 마련과 우수한 보건의료 인프라 적극 활용해야”, 보건산업 동향, 2015, 7면.

57) 인체자원은 국내 R&D 2,294개 연구과제에 분양되어 논문 894편과 특허 54건을 생산하였다고 한다. 질병관리본부, “2017 국립중앙인체자원은행 연보”, 2018, 10-11면.

58) 질병관리본부, “코호트 원격분석서비스 개시 및 자료공개 확대”, 2019, 4면.

건강보험공단, 건강보험심사평가원, 국립암센터를 비롯한 보건의료 관련 기관에서는 비단 코호트가 아니라도 기관이 보유한 다양한 보건의료 빅데이터를 자체적으로 만든 플랫폼을 통해 공개해왔다. 2019년 9월 시행된 보건의료 빅데이터 시범 사업에서 만들어진 보건의료 빅데이터 플랫폼은 질병관리본부, 국민건강보험공단, 건강보험심사평가원, 국립암센터의 보건의료 빅데이터를 한데 결합하여 연구자에게 제공한다.⁵⁹⁾ 또한 정부는 2020년부터 2029년까지 최대 100만 명 규모의 국가 바이오 빅데이터 구축 사업을 진행할 예정이라고 한다.⁶⁰⁾

정밀의료를 구현하기 민간 차원의 시도도 이어지고 있다. 인간게놈프로젝트 완료 이후 차세대 염기서열 해독기술(Next Generation Sequencing, NGS)로 불리는 새로운 기법이 등장하면서 분석에 소요되는 시간과 비용이 급속히 감소하여,⁶¹⁾ 미국의 23andMe⁶²⁾를 비롯한 민간업체가 출현하였고 의료기관이 아닌 유전자검사기관이 행하는 유전자 검사(Direct To Consumer, 이하 ‘DTC’)를 시행할 수 있게 되었다. 2017년 미국 FDA가 최종적으로 DTC를 부분 승인하면서 유전자와 특정 질병 간의 인과관계가 비교적 뚜렷이 밝혀진 일부 질병군에 한하여 본격적으로 서비스가 이루어졌다. 현재 미국에서는 파킨슨병, 알츠하이머, 대장암 등 12개 질환에 대한 DTC를 허용하고 있고, DTC에 대한 별도의 규제가 없는 일본은 약 360개, 중국은 약 300개 항목에 대한 서비스를 허용하고 있다.⁶³⁾ 우리나라도 2016년, 12개 영역에 한하여 DTC가 허용되고(생명윤리 및 안전에 관한 법률 제 50조 제3항 제2호, 의료기관이 아닌 유전자검사기관이 직접 실시할 수 있는 유전자검사 항목에 관한 규정 제1호 각목), 2019년 2월에는 규제샌드박스를 통해 마크로젠의 13개 항목에 대한 실증특례가 허용되었으며, 이와 별개로 11월까지 최

59) 보건의료 빅데이터 플랫폼, <https://hcdl.mohw.go.kr/BD/Portal/Enterprise/DefaultPage.bzr>, (방문일: 2020.5.20.)

60) 관계부처 합동, 앞의 글, 11-12면.

61) DNA 염기서열 해독기법은 1970년대 초 레이 우 교수에 의해 최초로 개발된 뒤 발전을 거듭했지만, 인간게놈프로젝트 당시 전장 유전체 서열분석(whole genome sequencing) 기법에는 시간과 비용이 상당히 요구되었다. 이후 차세대 염기서열 해독기술이 등장하면서 시간과 비용이 상당히 감소하였고, 정확도도 높아지고 있다. 문세영·장기정·김한해, “정밀의료의 성공 전략”, *KISTEP Inl 15호* (2016), 22-23면 참조.

62) 온라인 사이트에서 DNA 검사 진단 키트를 주문한 뒤 혈액 등을 제공하는 방법을 통해 자신의 DNA 시료를 담아서 본사로 보내면, 검사결과를 메일로 받을 수 있다. <http://www.23andme.com/en-int>, (방문일: 2020.5.20.)

63) 산업통상자원부, “‘혁신의 실험장’ 규제 샌드박스 최초 승인”, 2019, 8면.

대 57개의 웰니스 항목에 관한 DTC 유전자 검사서비스 인증제 시범사업이 시행되고 있다.⁶⁴⁾

그밖에 주로 대형병원 차원에서 자체로 보유한 보건의료 빅데이터를 직접 분석하거나, 나아가 개별 병원의 정보를 통합하여 빅데이터 분석을 시행하는 시도가 진행되고 있다. 세브란스병원, 아산병원, 서울대병원과 같은 대형병원에서 빅데이터 센터를 신설하고, 의료 인공지능으로 널리 알려진 IBM 왓슨을 도입하기도 하였으며, 뷰노, 루닛, 제이엘케이인스펙션과 같은 외부 업체와 협력하여 인공지능 소프트웨어를 개발하려고 한 것이 대표적 사례이다. 병원이 가진 의료정보를 전자화하고, 정보를 통합하거나 실시간으로 진료나 응급상황에 활용할 수 있도록 하는 플랫폼을 구축하려는 작업도 이루어지고 있다. 나아가 이러한 보건의료 빅데이터를 표준화하려는 작업이 병행되기도 하는데, 이와 같은 작업은 국내는 물론 해외의 수많은 기관들과의 협업을 용이하게 한다.

이와 같이 환자 데이터를 기반으로 하는 정밀의료는 프라이버시 문제에 직면한다. 2016년에 의료법 시행규칙 제16조가 개정되고, 하위 고시인 '전자의무기록의 관리·보존에 필요한 시설과 장비에 관한 기준'이 마련되면서, 전자의무기록을 클라우드와 같은 외부 서버에 보관할 수 있도록 하는 근거조항 자체는 마련되었다. 그러나 의료정보의 통합, 제3자 제공, 목적 외 이용 같은 2차적 활용 과정에서 발생할 수 있는 프라이버시 문제를 어떻게 해결할지에 관한 개인정보보호법 차원의 문제나, 신의료기술의 의료수가 반영과 같은 현실적 문제가 남아 있어 이러한 노력의 성과물이 정밀의료로 통합되지는 못한 상황이다.⁶⁵⁾ 또한 그동안의 논의가 대체로 공공기관 데이터에 국한된 한계를 들어 민간영역으로 확장하자는 주장이 제기되고 있고, 반대로 프라이버시 침해나 상업주의를 비롯한 부작용을 들어 경계하는 입장이 대립하는 상황이다.

오바마 정부에서는 앞서 본 연두교서의 일환으로 이듬해인 2016년 총 2억 1천 5백만 달러(USD)의 예산을 편성하여 그중 1억 3천만 달러(USD)는 1백만 명 코호트(cohort) 구축, 7천만 달러(USD)는 암 연구, 1천만 달러(USD)는 규제체계

64) 산업통상자원부, 앞의 글, 6-8면; 보건복지부, “「DTC 유전자 검사서비스 인증제 시범사업」 소비자 참여 연구 본격 시작”, 2019 참조, 1-4면 참조.

65) 이준영, “디지털헬스케어 동향 및 시사점”, 정보통신산업진흥원 이슈리포트 2019-03, 2019, 5면 이하 참조.

구축, 5백만 달러(USD)는 데이터 프라이버시와 이동성 강화에 배정하였다.⁶⁶⁾ ‘프라이버시와 신뢰 원칙(Privacy and Trust Principles)’, ‘데이터 보안 정책 원칙 및 프레임워크(Data Security Policy Principles and Framework)’를 잇달아 발표함으로써, 환자와 대중의 알 권리와 프라이버시를 보호하기 위한 노력을 구체적으로 시행하였다.⁶⁷⁾

(4) 제도화 과정의 고려사항

프라이버시와 관련하여 정밀의료는 몇 가지 쟁점을 갖고 있다. 첫째, 유전정보 및 생체정보에 대한 법적 개념이다. 국내법상 유전정보의 정의는 인체유래물을 분석하여 얻은 개인의 유전적 특징에 관한 정보로, 여기서의 인체유래물은 인체로부터 수집하거나 채취한 조직·세포·혈액·체액 등 인체 구성물 또는 이들로부터 분리된 혈청, 혈장, 염색체, DNA, RNA, 단백질 등으로 정의된다(생명윤리 및 안전에 관한 법률 제2조 제11호, 제14호). EU GDPR에서는 유전정보(genetic data)에 대하여 자연인의 생물학적 샘플 분석을 통해 얻어진 유전자 특성에 대한 개인정보라고 정의하여 국내와 유사하게 바라보면서도 후천적(acquired) 특성을 포함한다고 명시한 점에서 차이가 있다(GDPR 제4조 (13)). 일본에서는 개인식별 부호에 DNA를 구성하는 염기의 배열이 포함된다(개인정보의 보호에 관한 법률 시행령 제1조 제1호). 미국의 유전정보 차별금지법(The Genetic Information Nondiscrimination Act of 2008, 이하 ‘GINA’)은 유전정보의 범위에 특정 자연인뿐만 아니라 가족의 정보까지 포괄하는 것으로 광범위하게 규정한다(제201조 (4)(A)). 돌연변이로 인한 유전자 자체의 변화나, 환경과의 상호작용에 의한 기능상의 변화 같은 다양한 이유에 의해, 유전형의 관점으로만 유전정보를 이해하는 데는 한계가 있다. 이러한 시각에서 유전형을 넘어 표현형(phenotype)에 영향을 미치는 각종 요인을 포괄한 전체 기전을 총체론적으로 이해하려고 하는 일명 ‘후성유

66) The White House, op. cit.

<https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>, (방문일: 2020.5.20.)

67) 이에 대한 자세한 내용은 The White House, “Precision Medicine Initiative: Privacy and Trust Principles”, 2015; The White House, “Precision Medicine Initiative: Data Security Policy Principles and Framework”, 2016, pp.1-10 참조.

전학(epigenetics)’이라는 분야가 대두되고 있다.⁶⁸⁾ 그러나 앞서 살펴본 생명윤리 및 안전에 관한 법률에서는 유전정보를 유전형의 관점에서 협의로 파악하고 있어, 보다 광의의 관점에서 바라보는 GDPR과는 대조된다.

유전자는 다른 개인정보와 달리 불변성, 가족공유성, 고유 식별성,⁶⁹⁾ 유전정보를 통해 개인의 신체적 특성을 파악할 수 있는 표현형 관련성, 현재의 유전정보가 미래에 어느 정도의 활용성을 가질지를 점차 예측하기 어려운 미지성을 갖는다.⁷⁰⁾ 이러한 이유에서 여러 국가의 법제에서는 일반 개인정보와 비교할 때 유전정보에 대해 한층 더 강화된 보호를 인정한다. 예를 들어, 우리나라는 유전정보를 ‘민감정보’에 포함되는 것으로 보고 있고(개인정보보호법 제23조 제1항, 개인정보보호법 시행령 제18조 제1호), GDPR에서도 민감정보와 유사한 개념인 ‘특별한 범주의 개인정보(special categories of personal data)’의 일환으로 본다(제9조 제1항).

유전정보와 별개로 정밀의료 맥락에서 논의되는 생체정보(biometrics, 바이오정보)는 지문, 홍채, 음성, 필적과 같이 개인을 식별할 수 있는 신체적 또는 행동적 특징에 관한 정보를 말한다(정보통신망 이용촉진 및 정보보호 등에 관한 법률 시행령 제9조의2 제1항 제1호). 생체정보는 사람이라면 누구나 가진 고유한 신체적 특징을 핵심으로 하고, 이는 평생 동안 쉽게 변하지 않으며 검출 센서에 의해 식별할 수 있어서 개인의 동일성을 인증할 필요성이 큰 금융거래, 서명 관련법에서 종종 발견된다(전자금융거래법 제2조 제10호 라목, 전자서명법 제2조 제13호). 생체정보의 주요 활용방식은 개인의 동일성 확인에 있지만, 최근 웨어러블 디바이스를 비롯한 첨단 ICT 기기의 발달과 함께 실시간 건강상태 확인 같은 보건의료 맥락에서의 활용에 대한 논의도 급증하는 상황이다. 유전정보와 달리 생체정보의 법적 성격은 다소 불명확하게 규정되어 있는 편이다. 그렇지만 적어도 생체정보가 개인정보(개인정보보호법 제2조 제1호)의 일종이라는 데 대해서만큼은 큰 이견이 없는 듯하다.⁷¹⁾ 문제는 개인정보 이상으로 보호가 요구되는 민감정보(동법 제23

68) 조은희, “유전정보에 대한 윤리적 쟁점의 변천”, 한국생명윤리학회지 제15권 제1호, 생명윤리, 2014, 52-53면.

69) 심지어 일란성 쌍둥이조차 DNA가 대체로 유사할 뿐이지 태아의 발생과정에서 일부 유전자 변이가 일어나므로 DNA 정보가 완전히 일치하는 것은 아니다.

70) 이원복, “유전체 시대의 유전정보 보호와 공유를 위한 개인정보보호법제의 고찰”, 법조 제729호, 법조 협회, 2018, 605-610면 참조.

71) 방송통신위원회·한국인터넷진흥원, “바이오정보 보호 가이드라인”, 2017, 2면.

조)나 고유식별정보(동법 제24조)에 해당될 수 있는지 여부다. 양자는 개념상 구분되지만, 정보처리에 별도의 동의를 통한 보호가 인정된다는 점에서 유사하다. 민감정보라고 볼 경우 ‘개인정보보호법’ 제23조 제1항에 명시된 ‘건강에 관한 정보’의 일종으로 볼 수 있는지가 주된 쟁점이다. 이를 인정하는 경우 극단적으로는 건강과 조금이라도 관련된 SNS 문구라든가 사물인터넷, 웨어러블 디바이스가 생성한 체온, 칼로리와 같은 모든 건강 관련 정보에 대하여 민감정보로 보아야 한다는 다소 어색한 결론에 이를 수 있다.⁷²⁾ 참고로 GDPR은 생체정보와 건강에 관한 정보 모두 특별한 범주의 개인정보로 본다(GDPR 제9조 제1항). 다른 한편, 일각에서는 개인 식별을 중시하는 생체정보의 특성에 비추어 고유식별정보에 준하는 개념으로 본 입장도 있다.⁷³⁾

둘째, 법제도의 중복적용과 상충가능성 문제이다. 예를 들어, 진단서, 처방전, 진료기록, 전자의무기록 등에 관하여 의료인이나 의료기관 종사자에게 정보 누설 금지의무가 부과되고 있고(의료법 제19조), 생명윤리 및 안전에 관한 법률의 인간대상연구(동법 제2조 제1호)에 관한 개인정보를 제3자에게 제공할 때에는 연구대상자의 서면동의 및 기관위원회의 심의가 요구된다(동법 제16조, 제18조). 그런데 이러한 정보는 거의 대부분 개인정보보호법 제2조 제1호의 ‘개인정보’에도 해당하므로, 개인정보의 처리를 위해서는 정보주체에게 서면동의를 받아야 한다(동법 제15조 내지 제22조). 결국, 단일한 개인정보 처리 행위에 대해 복수의 법제가 적용될 수 있다는 결론에 이르게 된다. 그밖에 상술한 법제와 기타 관련법령에서 정하는 각종 의무가 중첩될 수 있는데, 대표적 사례가 유전정보에 관한 차별금지 의무이다(생명윤리 및 안전에 관한 법률 제46조). 이렇게 수많은 법제도가 중첩적으로 적용되는 이유는 그만큼 보건의료 맥락에서의 개인정보가 다차원적 법익을 갖기 때문이다. 개인의 신체에 관한 정보는 프라이버시 및 개인정보 자기결정권(개인정보보호법), 인간의 존엄과 가치 및 인체의 위해방지(생명윤리 및 안전에 관한 법률), 건강의 보호와 증진(의료법)과 같은 개별법의 보호법익과 모두 연관되어 있기 때문이다. 따라서 예컨대, 익명화 수준의 비식별 조치⁷⁴⁾를 동원하여 개인정보

72) 법적 구속력은 없지만 행정자치부의 해설서는 혈액형이 건강에 관한 정보에 포함되지 않는다고 하여 대조적이다. 이창범·김민호·김현경·최경진·고환경·백대용·김주희, 개인정보보호 법령 및 지침·고시 해설, 행정자치부, 2016, 137-138면.

73) 김민호·김현경·김선아, 정보통신서비스 분야의 민감정보 유형과 보호방안 연구, 방송통신위원회 연구보고서, 2017, 36-38면.

보호법상 개인정보의 요건인 식별가능성을 제거하였다고 하더라도, 의료법, 생명 윤리 및 안전에 관한 법률, 기타 관련법상 각종 의무조항이 함께 면제된다고 보기 어렵다. 마찬가지로 채취자와 연구자가 별도로 존재하는 인체유래물연구에서 채취자의 서면동의로 연구자의 서면동의가 면제되는 법조항(생명윤리 및 안전에 관한 법률 제37조 제3항)은 개인정보보호법에 유추적용 되지 않는다.

셋째, 유전자검사 결과에 대한 알 권리와 모를 권리이다. 발병률 혹은 치사율이 높고 불특정 다수에 대한 전파력 문제되는 전염병과 달리 본인 또는 특정한 혈족에게 주로 영향을 끼치는 유전정보에 대해 본인 또는 제3자에게 알 권리/모를 권리가 있는지에 대한 것이다.⁷⁵⁾ 이는 의사의 고지의무와 관련이 되는데 개인이 헌법 제10조에서 비롯된 자기결정권을 온전히 활용하기 위해서든, 전문가인 의료인에게 부과되는 후견주의적(paternalistic) 책무로서든, 유전자검사에 대해서도 일정한 요건을 전제로 알 권리 내지 그에 상응하는 고지의무가 인정되어야 한다는 입장이 지배적이다.⁷⁶⁾ 반면, 헌팅턴병(Huntington’s disease)과 같이 현재의 의학수준으로 해결하기 힘든 질환에 대한 유전자검사 결과에 대해 당사자나 그 사람과 유전자를 공유하는 친족에게 모를 권리나, 그에 상응하는 고지하지 않을 의무로서 배려의무가 인정되느냐는 문제도 있다. 이에 따라 그러한 정보를 알려준 의사에 대한 손해배상 청구의 인용 여부가 달라질 수 있다.⁷⁷⁾ 독일 연방대법원이 지난 2014년 모를 권리를 인정하는 판결을 하였다.⁷⁸⁾ 나아가 모를 권리를 고려한다면 앞서 본 알 권리의 형태도 달라져야 한다는 지적도 있었는데, 예컨대, 정보의 ‘존

74) 이에 관련된 자세한 사항은, 고학수·이동진·이선구·김은수·정종구, 개인정보 비식별화 방법론 - 보건의료정보를 중심으로 -, 박영사, 2017 참조. 다만, 앞서 언급한 유전정보 맥락에서 적절한 종류의 비식별 조치가 무엇인지에 대한 논란이 존재하는데, 비식별 조치의 두 가지 방법 중 통계적 방법은 유전정보에 활용하는 경우에 정보의 가치를 급격히 감소시키는 문제가 있고, 암호화 방법은 아직까지는 실효성이 제대로 검증되지 않았다는 문제가 있기 때문이다. 이원복, 앞의 논문, 635면.

75) 허진·정덕영, “헌법상 유전정보보호와 의사의 고지의무 - 미국의 사례를 중심으로 -”, 헌법학연구 제 15권 제2호, 한국헌법학회, 2009, 502-504면.

76) 예컨대, 이을상, “임상유전학에서 자기 결정권”, 인문과학연구 제31집, 성신여자대학교 인문과학연구소, 2013, 423면 이하; 유효종, “유전정보를 모를 권리’의 윤리적·법적 근거와 실현 방법”, 한국의료법학회지 제22권 제1호, 한국의료법학회, 2014, 110-112면; 최정호, “생명의료인권협약에 비추어 본 ‘모를 권리’ 문제”, 인권법평론 제21호, 전남대학교 공익인권법센터, 2018, 306-307면 참조.

77) 석희태, “환자의 모를 권리와 의사의 배려의무”, 의료법학 제17권 제2호, 대한의료법학회, 2016, 145-147면.

78) ‘모를 권리’ 자체는 인정되었지만, 그럼에도 자녀의 모를 권리를 어머니가 대신 행사하였다는 이유로 청구는 기각되었다고 한다. 보다 자세한 내용은 최정호, 앞의 논문, 291-292면 참조.

재’에 대한 알 권리와, ‘내용’에 대한 알 권리를 구분해야 한다는 것이다.⁷⁹⁾ 다만, 유전자 검사 맥락에서는 정보의 존재에서 내용이 추론될 여지가 있기 때문에, 검사 전에 당사자나 이해관계자에게 결과의 고지 여부에 대한 선택권을 부여하지는 견해도 대두되었다.⁸⁰⁾

3. 안전성/책임성의 윤리원칙과 자율주행자동차

(1) 안전성/책임성 윤리원칙

기계가 인간에게 초래할 수 있는 위해에 대한 안전 문제는 정보통신혁명이나 인공지능 기술기반이 일상화되기 훨씬 전부터 가장 기본적으로 고려된 원칙이다. 기계가 초래하는 위해에 대한 문제는 기계를 인간이 얼마나 ‘통제’할 수 있느냐, 위해를 막기 위한 ‘사전적인 조치’와 위해가 발생했을 때의 ‘사후적 조치’를 어떻게 제도화할 것이냐, 그리고 이에 대해 기계를 사용하는 인간/사회의 ‘신뢰와 수용성’을 확보할 수 있을 것이냐에 대한 문제로 생각할 수 있다.

먼저 기계를 인간이 얼마나 ‘통제’할 수 있느냐의 문제는 하드웨어에 대한 통제와 소프트웨어에 대한 통제로 구분할 수 있는데 산업혁명 이후 기술개발 초기에서 최근의 3, 4차 산업혁명에 이르는 과정에서 점차 하드웨어에 대한 통제에서 소프트웨어에 대한 통제의 비중이 커져 왔다. 정보통신기술 기반에서는 기본적으로 이를 위한 (물리적/논리적) 보안 기술이 지속적으로 발전해왔고, 인공지능기술 기반에서는 앞서 살펴본 바와 같이 다층신경망의 은닉층에서 이루어지는 일에 대한 인간의 이해가능성과 통제가능성 문제와 연결된다. 위해를 막기 위한 ‘사전적 조치’에 대해서는 인공지능 기술을 개발하고 적용하는 과정에서 유무형의 위해에 대해 어느 정도로 예측하고 이를 고려한 안전장치를 마련해야 하느냐에 대한 사전적 기준이 필요하다. 위해가 발생했을 때의 ‘사후적 조치’는 인공지능 기술기반의 여러 특성을 고려했을 때 사전에 예측한 위해 발생 시, 예측할 수 없었던 위해 발생 시 관련자(하드웨어 제조사/소프트웨어 제조사/사용자/소유자/보험사 등)들이 각각

79) 최정호, 앞의 논문, 305면.

80) 유호종, 앞의 논문, 112-114면.

어느 정도의 책임을 져야 하느냐에 대한 합의된 제도가 필요하다. 기계 사용에 대해 인간과 사회의 '신뢰와 수용성'의 확보 문제는 이 모든 것들, 즉 통제, 사전적/사후적 조치에 대한 충분한 준비에 더해 소통과 이해가 필요하다. 초기 산업혁명과 같이 공장, 기업 등 특정 조직 종사자들을 조력하는 기계와 달리 정보통신/인공지능 기술 적용 분야처럼 인류의 일상과 함께하는 기계의 경우에는 그 위해의 규모가 비할 수 없이 클 것이고 제도를 준비하고 그 신뢰와 수용성을 확보하는 과정도 보다 거시적 차원에서 이루어져야 한다.

미국의 “인공지능 연구와 개발을 위한 국가 전략 계획”에서는 ‘인공지능 시스템의 안전성과 보안의 확보’를 전략으로 다루는데 “인공지능 시스템이 널리 사용되기 전에 이 시스템이 통제 가능한 방법(in a controlled manner)으로 안전하고 보안이 되는지에 대한 확신(assurance)을 할 수 있어야 한다.”고 한다. 그리고 인공지능 시스템이 이에 대해 문제가 될 수 있는 점은 인공지능 시스템이 많은 경우 완전하게 검토되고 테스트 될 수 없는 환경에서 설계된다는 것, 시스템 구축 후 감독할 수 없는 조건하에서 학습된 결과를 기반으로 시스템이 행위 할 수 있다는 것, 인간의 의도된 목표와 시스템에 입력된 지시사항이 불일치할 수 있다는 것, 시스템 운영과정에서 인간과의 상호작용이 시스템의 퍼포먼스에 영향을 끼치고 그 결과 인간의 여러 변동성(variation)이 시스템의 안전성에 영향을 미칠 수 있다는 것을 제시한다.⁸¹⁾ EU의 “인공지능, 로봇틱스, 자동화 시스템”에서는 윤리원칙으로 ‘보안, 안전, 육체적 정신적 온전성(security, safety, bodily and mental integrity)’을 제시하는데 이를 다음 세 가지 형태로 구현해야 한다고 한다. 첫째, 외부 환경과 사용자에게 대한 시스템의 외적 안전성(external safety), 둘째, 해킹 대처와 같은 시스템의 신뢰성과 내적 견고성(reliability and internal robustness), 셋째, 인간과 기계의 상호작용에 대한 정서적 안전성(emotional safety)이다.⁸²⁾ 다음에서는 이 윤리원칙을 인공지능기술 적용산업 중 가장 근시일 내에 일상에 들어올 것으로 예측되는 자율주행자동차와 관련해서 다루어 보도록 하겠다.

81) National Science and Technology Council, op. cit., pp.27-28.

82) European Commission, op. cit., pp.18-19.

(2) 자율주행자동차의 인공지능 기술 동향

인공지능시스템이 적용될 많은 산업 영역에서 가장 먼저 그리고 가장 큰 변화는 자율주행차(self-driving/autonomous vehicles)에 의한 교통 및 운송산업에서 이루어질 것이라고 전망된다.⁸³⁾ 1930년대 이후 공상과학소설에 꾸준히 등장하던 자율주행자동차는 1960년대 이후 인공지능 연구에서 지속적인 과제가 되었으나 주행환경에서 자동차의 통제 밖에 있는 여러 변수, 즉 보행자, 교통체증, 예상할 수 없는 다양한 사건들로 인해 도시에서 자율주행차가 운행하도록 하는 것은 너무 복잡한 문제로 간주되었다. 그러나 2004-2012년의 8년여의 시간 동안 학계와 산업 모두에서 놀라운 진전이 있었다. 센싱기술과 인식을 위한 기계학습은 매우 빠른 속도로 발전했고 Google 자율주행자동차(autonomous vehicle: AV)와 Tesla의 반자율주행자동차(semi-autonomous vehicle)가 도시를 주행하기 시작했다. 2020년 1월 Google의 모회사인 Alphabet은 자사의 누적 주행거리가 2000만 마일(약 3200만 km)를 돌파했고 2009년 창사 후 1000만 마일 돌파까지 10년이 걸렸는데 2000만 마일까지 1년이 채 걸리지 않았다고 발표했다.⁸⁴⁾

자율주행기술의 단계는 일반적으로 미국 도로교통안전국(NHTSA)의 자율주행자동차의 자동화 수준(Automation Level)에 대한 정의를 참조한다. 이에 의하면 0단계는 운전자가 시스템 전체를 통제하는 단계, 1단계는 대부분의 통제는 운전자에게 남아있는데 조향장치와 가감속에 대한 조작을 운전자와 시스템이 공동 수행하며 주행환경 모니터링 및 긴급사항 대처는 운전자가 하는 단계, 2단계는 조향장치와 가감속에 대한 조작을 시스템이 전담하는 단계, 3단계는 긴급 상황에 대한 대처를 제외한 나머지를 시스템이 수행하지만 운전자는 운전 상황에 대해 주시하고 있어야 하는 단계, 4단계는 기본적으로 긴급 상황에서도 자율주행시스템 스스로 통제할 수 있지만 지리적 상황 등과 같은 특정 조건에서는 자율기능을 상실할 수도 있는 단계, 5단계는 모든 상황에서 전적으로 시스템이 관장하는 단계이다.⁸⁵⁾ 현재 대

83) OECD, op. cit., p.48; Stanford University, "Artificial intelligence and life in 2030", *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, 2016, p.18.

84) 한국의 자율주행관련 기업들이 진행한 누적 운행거리는 71만 6천km이다; 박성우, "자율주행자동차도 마일리지 모아야 한데...미국 2,000만, 중국 200만, 한국은..." *중앙일보* (2020.1.11.) <https://news.joins.com/article/23679549> (방문일: 2020.5.17.)

부분의 상용화 수준은 2-3단계로 부분자율기능이 지원되는 단계이다.⁸⁶⁾ 구글 웨이모는 4단계 자율주행차 개발과 양산에 박차를 가하고 있고 GM은 스티어링과 페달이 아예 없는 크루즈 AV를 개발 중이다. Tesla는 현재 레벨 2.5 수준의 반자율주행기능을 제공하고 있으며 소프트웨어 업그레이드를 통해 자동화 수준을 높이고 있다.⁸⁷⁾ 국내 자율주행기술을 선도하는 현대·기아차는 고속도로 주행 지원시스템을 비롯한 3단계 자율주행 기술을 양산 차에 적용하고 있으며 완전 자율주행 단계로 알려진 4단계 기술을 시연할 수 있는 시험중인 차를 보유하고 있다. 2020년까지 고도 자율주행차 상용화와 2030년까지 완전 자율주행차를 상용화하겠다는 목표로 자율주행 기술개발 중이다.⁸⁸⁾ 인식(perception)과 배송추적(tracking)에 있어서는 이미 시각(vision)기능을 포함해서 인간에 가깝거나 인간과 같은 수준에 이르고 있지만 자율주행자동차의 센싱 알고리즘은 근시일 내에 인간의 운전 능력을 초월하는 성과를 보일 것으로 예측되며 특히 계획(planning)과 같은 사고(reasoning) 영역은 알고리즘 개발에 의해 가속화될 것으로 전망된다.⁸⁹⁾

OECD 국가에서 운송산업은 경제의 가장 큰 영역을 차지하는 분야 중 하나이다. 2016년 OECD 국가 전체 GDP의 5.6%를 차지했다. 교통사고 감소, 교통정체 완화와 기타 효용에 의해 자율주행자동차가 가져올 수 있는 경제적인 파급력은 상당할 것으로 전망된다. 2015년 연구에 의하면, 미국 전체 차량의 10%가 자율주행자동차로 바뀐다면 1년에 1,100 명의 생명을 구하고, 380억 달러(USD)의 비용을

85) NHTSA, <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety> (방문일: 2020.5.17.)

86) 전용일, “우리나라 자율주행자동차법안(자율주행자동차 개발 촉진 및 상용화 기반조성에 관한 법률안)의 주요 내용 분석과 시사점”, 법학연구 제27권 제1호, 2019, 250면; 그러나 2018년 6월 미국자동차공학회(SAE International)는 자율주행차의 자율주행 단계를 1단계와 2단계까지는 지원 부분(assisted)으로, 3단계에서 5단계까지는 자동화(automated) 부분으로 구분해서 개정 발표하였다. 3단계의 기준 요건이 강화되었는데 2단계에서 요구되는 주차 도움기능(Parking Helper L2)과 고속도로 주행 기능(Highway Pilot L2), 그리고 교통혼잡주행기능(Traffic jam Pilot)과 자율주행시스템기능(Automated Driving System)을 요건으로 하였다. 이는 사고발생 시 책임의 소재를 분명히 하고, 자율주행모드에서의 사고발생 시 2단계와 3단계 중 정확히 어떤 단계에서 사고가 발생했는지를 구분하는 데 도움이 될 것으로 보인다; 조주은, “ICT와 사회의 ‘융합’을 위한 사회적 합의 프레임워크: 인공지능과 자율주행자동차를 중심으로”, 중앙법학 제21집 제3호, 중앙법학회, 2019, 213-214면.

87) Mautodaily, “테슬라 모델3의 자율주행기능, 다른 차와 이렇게 다르다”, 2020.3.2., <https://www.autodaily.co.kr/news/articleView.html?idxno=417029> (방문일: 2020.5.17.)

88) 인더스트리 뉴스, “자율주행차의 눈 ‘AI 영상인식’, 중기 원천기술 선점 경쟁 치열”, 2019.8.12. <http://www.industrynews.co.kr/news/articleView.html?idxno=33882> (방문일: 2020.5.17.).

89) Stanford University, op. cit., p.20.

절감할 수 있고, 90%가 자율주행자동차로 바뀐다면 1년에 21,700명의 생명을 구하고 4,470억 달러(USD)를 절감할 수 있다고 예측된다.⁹⁰⁾ 자율주행자동차의 기대 편익은 교통사고 저감 외에도 소비자의 이용효율성 증대, 장애인, 고령자 등 교통약자 이동편의성 제고, 대중교통 소외지역 지원, 교통정체 및 주차 공간 부족 해소에 기여, 커넥티드카로 발전 시 차량 감소, 친환경기술과 접목 시 공해 등 환경문제에 기여할 수 있는 등 다양하다.⁹¹⁾ 자율주행자동차는 사람들의 삶의 방식에도 영향을 미칠 것으로 예상된다. 미국의 통근자들은 평균적으로 하루에 25분을 차에서 보낸다고 조사되는데, 자율주행자동차가 생활화되면 사람들은 이 시간 동안 엔터테인먼트 등 자신을 위해 다른 일을 할 수 있을 것이고 자율주행자동차와 차량공유서비스로 이동 편의성이 높아지면 사람들의 거주지 선택도 달라질 것으로 전망된다. 주차공간에 대한 필요의 감소는 도시와 공용공간의 디자인 및 설계도 달라지게 할 것이다. 또한 공공 운송서비스에 대한 필요의 형태도 달라져 교통약자에 대한 이동서비스나 주요 거점별 이동수단으로서의 공공서비스로 변화될 수 있다⁹²⁾

자율주행시스템의 복잡성으로 인해 기업들은 각자의 전문 분야에 초점을 맞추고 다른 분야에 전문성을 가진 다른 업체들과 협력관계를 맺는 경향이 강하다. 예를 들어 자율주행자동차의 선두업체인 Waymo는 방대한 데이터셋과 기계학습에 전문성을 집중하고 차량을 직접 제조하지 않으며 GM, Jaguar와 같은 차량제조업체와 협력한다. 또한 대규모의 차량제조업체도 최신 기술 도입을 위해 작은 스타트업들과 협력관계를 맺는다. 2018년 Honda는 27억 5천만 달러(USD)를 GM의 크루즈 자율주행 벤처에 투자했다. Uber와 같은 차량공유 회사도 자율주행사업에 많은 투자를 하면서 CMU와 같은 선도적 연구대학과 파트너십을 맺었다.⁹³⁾ 그런데 이러한 추세에 의해 사고 발생 시 다수의 이해관계자가 차량에 관여되어 있을 경우 책임문제(liability)를 어떻게 다룰 것이냐에 대한 이슈가 대두된다.

90) OECD, op. cit., p.48.

91) 조주은, 앞의 논문, 220면.

92) Stanford University, op. cit., pp.20-21.

93) OECD, op. cit., pp.49-50.

(3) 안전성/책임성의 윤리원칙과 자율주행자동차

자율주행자동차 관련해서 가장 중요한 이슈이자 이해관계자들 사이에서 첨예한 논쟁이 되는 것은 사람에 대해 여러 형태로 예상되는 안전성 문제와 사고 발생 전 후에 대한 책임 문제이다. OECD 보고서는 기본적으로 '자동화 기능(autonomous function)'을 자동차(vehicles)에 도입하는 접근방식에 대해 이해관계자들은 아직 완전히 동의하지 않고 있다는 점을 지적한다. 주요한 차이는 운전자의 역할, 그리고 자율주행기능의 가용성 범위에 대한 것이다. 먼저 운전자의 역할에 대해 운전자의 필요가 완전히 없어지는 자율주행자동차의 개발을 주장하는 측에서는 주행 기능, 안전 모니터 등 어떠한 이유로도 운전자가 필요 없는 자율주행자동차 기술을 개발하는 것을 지향한다. Google의 Waymo나 Tesla는 이러한 기술개발이 곧 가능하다고 주장한다. 반면, 운전자에 대한 보조 기능을 하는 자율주행자동차의 개발을 주장하는 측에서는 적어도 근시일 내에서는 자율주행자동차가 운전자를 대신함으로써 사고를 줄일 수 있는 수준이 최선임을 주장한다. 예를 들어, Toyota는 추돌사고를 내지 않는 자동차를 개발하는 것에 집중할 것을 강조했다. 또한 자율주행기능의 가용성의 범위에 대해서 특정 지역에서만 자율주행자동차 가용성을 주장하는 측은 사람의 개입이 없는 높은 수준의 자율주행자동차는 상세한 지도화(mapped in detail)가 되어 있는 특정 지역에서만 가능하다는 입장이다. 예를 들어 Cadillac의 Super Cruise는 지도화되어 있는 고속도로에서만 가용하다. 반면 어디에서나 자율주행자동차 가용하다는 측은 어떤 도로, 어떤 상황에서도 자율주행이 가능하도록 기능이 개발되어야 한다는 입장이다. 대부분의 자동차 제조업체들이 선호하는 입장이다.⁹⁴⁾

자율주행자동차의 안전성과 책임 문제를 위해서는 기술적인 안전성, 제어와 신호에 대한 설비 및 책임, 운전자 규제와 교통량과 운행 규칙 등에 대한 준비가 필요하다. 또한 데이터에 대한 정책적 접근도 필요하다. 다른 모든 인공지능 기반 기기와 마찬가지로 데이터에 접근하고 시스템에 적용하는 것은 자율주행자동차의 핵심적 성공요소이다.⁹⁵⁾ 그간 여러 업체의 실시간 운행 데이터가 경쟁력을 위한 비

94) Ibid., p.51; 조주은, 앞의 논문, 220-221면.

95) 구글 웨이모의 2018년 기준 시험주행 거리는 127만 1,587마일(204만 6,420km)이고 2위인 GM 크루즈는 44만 7,621마일(72만 376km)이다; 조주은, 앞의 논문, 215면; 이동현, “요동치는 자율주행차 순

전매 특허가 되어 공유되지 않는 문제가 있었다. 이러한 데이터는 자율주행자동차의 연구자와 개발자들이 시스템을 향상시켜 안전성을 제고하는 데 매우 중요한 정보가 된다. 이를 위해 MIT에서는 운전자의 행위를 이해할 수 있는 접근 가능한 데이터 셋을 만들었고 Google은 최근 자율주행기술 발전을 위해 Waymo의 데이터를 오픈 소스로 공개하기로 매우 전향적인 결정을 했다.⁹⁶⁾ 정책적으로도 다양한 시스템에 의해 수집된 데이터에 대한 접근, 공개 데이터를 위한 편당에 대한 논의가 필요하다. 보안과 프라이버시에 대한 제도적 준비도 필요하다. 언급한 대로 자율주행자동차는 시스템, 운전자 행위, 운행 환경에 대한 방대한 데이터가 필요하며 정보 전달을 위해 다양한 네트워크와 접속한다. 수집되고 사용되는 데이터는 정보 주체의 의사에 반하는 ‘원치 않은 접근’에 대해 적절한 보호조치를 할 수 있어야 한다. 그러한 정보에는 위치 정보, 사용자 행위 정보 등 민감한 정보가 포함될 수 있다. 이와 관련해서 국제교통포럼(International Transport Forum)에서는 ‘자율주행자동차를 위한 포괄적인 사이버보안 프레임워크(comprehensive cybersecurity frameworks)’가 필요하다고 했다. 프라이버시와 데이터 보안을 위한 새로운 암호 프로토콜과 시스템도 제시되었다. 문제는 이러한 시스템이 자율주행자동차의 기본 기능 수행에 필요한 속도를 저해할 수 있다는 것인데 실시간 자율주행자동차 기능에 필요한 규모와 속도가 아직 확보되지 않았다.⁹⁷⁾

또한 사회기반시설에 대한 재정비를 고려해야 하는데 자율주행자동차의 도입은 자율주행과 사람 운행 방식이 혼재하는 운전환경을 위한 변화가 필요하다. 자율주행자동차가 상용화되면 자율주행자동차 상호 간에는 기술적 소통이 가능해질 것이다. 그러나 사람이 운전하는 기존 자동차는 이러한 소통에서 배제되어 새로운 불확실성에 직면하게 된다. 자율주행자동차가 사람 운행 자동차에 적절히 반응하도록 개발되어야 한다. 자율주행자동차와 사람 운행 자동차가 운행하는 도로를 분리하는 것도 논의될 수 있다. 자율주행자동차의 기술이 발전할수록 도로교통 및 기반정책에 자율주행자동차의 인지 수준이 반영되어야 한다.⁹⁸⁾

위...현대차는 제자리걸음”, 중앙일보 (2019.3.25.), <https://news.joins.com/article/23420999>, (방문일: 2020.5.17.)

96) 이진한, “구글 웨이모, 자율주행차 데이터 공개”, 테크월드 (2019.8.22.), <http://www.epnc.co.kr/news/articleView.html?idxno=91724> (방문일: 2020.5.17.)

97) OECD, op. cit., pp.51-52.

98) Ibid., pp. 51-52; 조주은, 앞의 논문, 221-224면.

이러한 자율주행자동차의 기술적 발전 및 제도적 접근과 더불어 고려해야 할 것은 변화된 교통수단 및 환경에 대한 '신뢰와 사회적 수용성'이다. 특히 소비자들이 자율주행자동차를 '얼마나 안전하다고 인식하느냐'에 대한 사회적 수용성이 매우 중요한 요소이다. 업계에서 매우 안전한 기술임을 주장하고 법 제도적으로 그것을 시험하도록 허용하려 해도 소비자가 이를 위협하게 인식하면 4단계 이상의 자율주행자동차가 도로 위를 주행하는 것은 요원하거나 많은 무리수를 낳게 된다.⁹⁹⁾ 장차 충분히 안전한 자율주행자동차가 기술적으로 구현된다고 하더라도 그것이 대중의 매일의 삶과 함께하기 위해 사람들이 적응(adjust)하는 것은 별개의 문제이다.¹⁰⁰⁾

실제로 업계에서는 누적 주행거리에서 압도적인 Google 웨이모는 1000만 마일(1,610만 km) 돌파 시까지 사고 건수는 두 자릿수, 우버가 30만 마일(48만 km) 주행 후 사고 건수는 세 건에 불과하며 그 외 다임러 BMW, 르노 등 유수의 자동차 업체들 역시 사람보다 기계가 더 안전하다고 주장하는 자료를 제시하고 있다.¹⁰¹⁾ 그러나 소비자의 인식은 이와 다르다. 2019년 딜로이트 글로벌에서 조사한 “자동차에 대한 소비자 조사”는 전 세계 20개국 2만 5천여 명의 소비자를 대상으로 수행되었는데 자동차 및 정보통신업계의 노력과 투자업계의 뜨거운 반응과는 별개로 미래 모빌리티 서비스를 실제로 사용할 소비자의 수용성에 대해 진지하게 고민할 시점임을 제시하고 있다. ‘자율주행차가 안전하지 않다’고 응답한 소비자가 2019년 미국에서 50%, 우리나라에서는 49%로 조사되었다. 그 외 일본(50%), 인도(48%), 독일(47%) 모두 비슷한 수준의 응답으로 보였고 중국의 경우만 25%로 상대적으로 낮은 수치를 보였다. 같은 보고서에서 자율주행차 개발과 사용에 대해 기대하는 정부개입의 수준은 일본 53%, 미국 56%, 독일 59%, 인도 60%, 우리나라 62%, 중국 77%로 조사되었다. 이에 의하면 중국은 정부의 높은 개입에 대한 기대가 자율주행차의 안전성에 대한 신뢰를 담보한다는 것을 알 수 있다. 우리나라도 소비자의 정부개입에 대한 기대가 상당히 높은 편이다. 커넥티드 카에 대한 선호도는 개인정보 노출에 대한 거부감과 우려가 큰 미국(47%), 독일(35%), 일본(36%)의 경우 상대적으로 낮았고 반면 중국(79%), 인도(76%)는

99) 조주은, 위의 논문, 217면.

100) Stanford University, op. cit., p.18.

101) 최정단 · 민경옥 · 한승준 · 성경복 · 이동진 · 최두섭 · 조용우 · 강정규, “도심 자율주행서비스 테스트를 통한 자율주행 기술개발 현황 및 5G 연계 미디어의 역할”, 방송과 미디어 제24권 제1호, 한국방송·미디어법학회, 2019, 67면.

두 배 정도 크게 나타났다. 우리나라는 60% 정도의 우호적인 견해를 보였다.¹⁰²⁾

이와 관련해서 최근 미국 연방도로교통안전국(NHTSA)이 자율주행자동차에 대한 최초의 공론화 과정은 좋은 참고가 된다. 미국 연방도로교통안전국은 소비자들을 대상으로 최소 60일 이상 운전대와 페달이 없는 자동차가 도로 위를 달리는 것을 허용할지 여부에 대한 공개 여론조사를 하기로 했다. 미국도 우리나라와 마찬가지로 운전자의 조작장치가 없는 자동차는 형식 승인 자체를 받을 수 없고 도로 위를 달리는 시험운행이 불가능했다. 이 조사는 GM을 위시한 자동차 업계의 탄원으로 시작되었는데 업계는 완전자율주행자동차의 실증 실험을 위해 미국 내 도로에서 제한된 수의 차가 운전자의 조작장치 없이 주행할 수 있도록 규제 완화를 해달라고 요청해왔다. 이에 연방도로교통안전국은 이에 대한 소비자의 인식 조사를 한 후 이를 필요한 제도 개선에 반영하겠다고 취지의 공론화 과정을 채택했다.¹⁰³⁾

(4) 제도화 과정의 고려사항

우리나라는 2019년 5월, 자율주행자동차의 · 안전한 운행을 위한 운행기반 조성 및 지원 등을 통한 자율주행자동차의 상용화 촉진을 위해 ‘자율주행자동차 상용화 촉진 및 지원에 관한 법률’이 시행되어 자율주행 안전구간 지정, 시험운행지구 지정 및 운영 관리, 여객 유상운송에 대한 특례, 보험 가입 의무, 익명화된 정보 활용 등의 주요 내용에 대해 규정을 하였다. 이에 앞서 독일은 2017년 5월 도로교통법 개정으로 자율주행자동차 사고에 따른 손해배상책임과 보험제도를 마련해서 시행했고 2017년 8월에는 교통·디지털 인프라 부 산하 자율주행자동차 윤리위원회가 20개의 윤리기준(ethical rules)을 발표했다. 이에 따르면 운전자가 없는 레벨5에 대해서는 추후에 규정하고 레벨 3, 4에 대해서 자율주행자동차 사고로 피해를 입은 사람은 차량 소유자가 가입한 보험회사에 보험금 청구를 하도록 하고 시스템 오류로 인한 사고 발생 시는 자동차의 제조사가 피해보상을 하도록 규정하였다. 또한

102) 김태환, “2019 딜로이트 글로벌 자동차 소비자 조사, ‘모빌리티에 대한 현실적 접근’”, *Deloitte Korea Review* 제12호, 2019, 39면; 조주은, 앞의 논문, 217면.

103) 안효문, “미국 자율주행 공론화 조사를 주목해야 하는 이유”, IT 조선 (2019.3.19.) http://it.chosun.com/site/data/html_dir/2019/03/19/2019031901632.html (방문일:2020.5.17.); 조주은 위 논문, 218면; 2020.2 최초로 운전대, 페달, 백미러 없는 자율주행 전기배달차, ‘R2’가 일반도로에서 달리는 것이 승인되었다; 김현석, “운전대·백미러·페달 없는 자율차 미 도로 달린다”, *한국경제* (2020.2.9.) <https://www.hankyung.com/economy/article/2020020919251> (방문일, 2020.5.17.)

윤리기준에 의하면 사고에 대한 판단이 어려운 경우 탑승한 사람이 최종 책임을 져야 한다고 밝혔다.¹⁰⁴⁾

자율주행자동차와 관련 제도의 범위는 매우 넓지만 먼저 실제 운행 및 사고와 관련해서 많은 쟁점을 갖고 있는 도로교통법, 자동차관리법, 자동차손해배상보장법, 제조물책임법에 대해 생각해 볼 수 있다. 자율주행자동차는 그 개념상 기술개발 과정에서 운행기능이 운전자 조작에서 자율주행시스템으로 점차 이전되기 때문에 “자율주행자동차의 운전 주체를 누구로 볼 것이냐”에 대해 많은 법적 논점에 제기되고 이에 따라 자율주행자동차의 합법성 여부, 사고 발생 시 책임소재, 입증책임, 민형사적 책임 등과 관련한 쟁점이 발생하고 보험제도와도 관련되어서도 새로운 고려가 필요하다.

현행 ‘도로교통법’은 자동차 운전을 하기 위해서는 운전면허의 발급과 운전면허를 발급받은 운전 주체로서의 ‘사람’을 필수로 하고 있다(도로교통법 제43조, 제80조). 따라서 자율주행자동차의 자동화 단계에서 사람인 운전자가 완전히 배제될 때부터는 현행법상 무면허 운전으로 해석된다. 또한 모든 차의 운전자는 차의 조향장치와 제동장치, 그 밖의 장치를 정확하게 조작하여야 한다고 규정하여(동법 제48조) 핸들과 브레이크 조작 없이 주행하는 자율주행자동차는 현행법상 불법이다. 따라서 자율주행자동차가 상용화되는 과정에서는 자동화 단계에 따라 사람인 운전자가 배제되고 기계가 스스로 운행하는 과정이 법제상 허용되는 ‘운전’의 범위에 포섭되어야 한다.¹⁰⁵⁾ 이 논의 과정에 앞서 제시한 업계의 기술개발과정에서 드러나는 위험성에 대한 공유, 소비자를 비롯한 일반인의 자율주행기술에 대한 신뢰와 사회적 수용성, 기계의 운전을 포섭할 수 있는 교통정책 및 인프라의 정비에 대한 준비 및 확인 등에 대한 합의 과정이 필수적이다.¹⁰⁶⁾

다음으로, 현재 자율주행자동차의 정의 및 시험운행의 근거가 되는 법률은 ‘자동차관리법’이다. ‘자동차관리법’은 “운전자 또는 승객의 조작 없이 자동차 스스로 운행이 가능한 자동차”를 자율주행자동차로 정의(자동차관리법 제2조 제1의3호)하고 있고 ‘자동차관리법 시행규칙’은 “자율주행자동차 시험운행을 위한 임시운행

104) 장민선, 인공지능 시대의 법적 쟁점에 대한 연구, 한국법제연구원, 2018, 77-78면.

105) 강선준·원유형·김민지, “제4차 산업혁명 시대의 자율주행자동차 상용화를 위한 안정적 법적 기반을 위한 법 정책적 연구 자율주행자동차 특별법 제정(안)을 중심으로”, 기술혁신학회지 제21권 제1호, 한국기술혁신학회, 2018, 158면.

106) 조주은, 앞의 논문, 232면.

허가(동법 제26조)”와 “안전운행요건(동법 제26조의2)”을 규정하고 있으며 임시 운행 허용 구간은 계속 확대되고 있다. 또한 ‘자동차관리법 시행규칙’의 하위 규정인 ‘자율주행자동차의 안전운행 요건 및 시험운행 등에 관한 규정’은 자율주행자동차의 임시운행에 필요한 세부요건 및 확인방법 등 안전운행요건을 정함을 목적으로 다양한 기술적 정의를 한 총칙(제1장), 자율주행자동차의 제작대상 및 방법, 손해배상 책임 및 보험가입, 사전시험주행 등의 안전운행요건의 일반규정, 조종장치, 운전자우선모드, 표시장치, 기능 고장 자동감지, 경고장치, 최고속도 제한 및 전방충돌방지 기능, 운행기록 장치, 영상기록 장치 등을 규제한 자율주행자동차의 구조와 기능에 관한 규정을 한 안전운행요건(제2장), 운전자 지정 및 관리 등을 규제한 임시운행에 관한 규정(제3장)으로 구성되어 있다. 앞서 살펴보았듯이 자율주행자동차의 기술 개발과 자동화 과정이 이러한 자동차관리법에 지속적으로 반영되어야 하며 이 과정이 업계, 소비자, 학계와의 지속적인 소통을 통해 효과적으로 이루어질 수 있다.¹⁰⁷⁾

자율주행자동차의 상용화 관련해서 이미 논란이 뜨거운 분야가 사고 발생 전/후 판단 및 책임 관련된 문제이다. 먼저 사고 발생 전의 문제 관련해서는 이른바 ‘트롤리 딜레마’라고도 일컬어지는 직진을 해도 선회를 해도 보행자와 승객 등 사람에 대한 사상이 발생하는 상황에서 자율주행자동차는 어떻게 움직이도록 알고리즘이 설계되어야 하는가에 대한 문제가 있다. 이를 윤리학의 고전적 딜레마에 비유하는 것인데 이는 결국 기계가 아닌 ‘사람’의 판단에 의해 결정된다. 이에 대한 논의 과정에는 철학적으로 공리주의적 관점(사상자의 수에 대한 양적 판단), 의무론적 관점(어떤 경우에도 사람을 ‘수단’으로 사용해서는 안 된다/사람을 해할 -단순한 ‘예견’과 구분되는- ‘의도’를 갖는 행위는 정당화될 수 없다)에 대한 논의, 형사법적으로 작위·부작위에 대한 ‘고의’, ‘긴급피난’과 같은 ‘위법성 조각사유’, ‘면책적 긴급피난’이나 ‘적법한 행위에 대한 기대가능성’과 같은 ‘책임 조각사유’에 해당되는지에 대한 관점에서 다루는 논의가 있다.¹⁰⁸⁾

또한 사람의 개입을 배제하는 4단계 이후의 자율주행자동차의 운전자를 사람이라고 볼 것인지, 자율주행시스템이라고 볼 것인지, 또는 사람이 자율주행자동차의

107) 조주은, 위의 논문, 233면.

108) 김준호, “‘폭주하는 전차’ 사례에 대한 형사법적 결론을 생각해보기-자율주행자동차의 윤리적 딜레마에 관한 법학적 사유실험을 겸하여”, 법학연구 제29권 제2호, 연세대학교 법학연구원, 2019, 1-36면.

자율주행시스템을 전적으로 신뢰하는 것이 허용될 것인지에 대한 것도 중요한 쟁점이다. 기존 자동차 운전의 경우 자동차 사고의 원인의 90%가 운전자의 과실이고 9%가 외부적 환경, 1%가 기계적 결함으로 조사되고 있지만 자율주행자동차가 상용화될 경우 사람의 과실보다 자동차의 결함, 자율주행자동차 관련 소프트웨어 결함으로 발생할 것으로 예측된다.¹⁰⁹⁾ 이는 자율주행자동차에 운전을 맡긴 사람에게 결과에 대한 예견 가능성과 결과 회피의 의무를 인정할 수 있을지와 연결된다. 그리고 3단계에서 4단계로 넘어가는 과정에서 자율주행자동차에 탑승한 운전자의 ‘탑승자의 주의의무’와 운전기능을 점차 전적으로 담당하게 되는 자율주행시스템의 ‘운행과 관련된 주의의무’를 어떻게 정립할 수 있는지, 기존에 사람에게 부여된 운전면허를 따기 위한 최소한의 운전 능력이나 윤리적·규범적 판단 능력이 자율주행시스템의 알고리즘에 반영되어야 하는지에 대한 고민도 필요하다. 이와 같이 기존에 고려되지 않은 새로운 유형을 포섭하는 책임법 체계와 관련되어 향후 많은 논의가 필요하다.¹¹⁰⁾

형사법적 책임과 별개로 민사법적인 검토에 기반한 손해배상책임은 피해자의 과실 유무, 피해자의 수에 따라 그 범위가 달라진다. 현행 ‘자동차손해배상보장법’에 의하면 “자기를 위하여 자동차를 운행하는 자”(자동차손해배상보장법 제3조)가 그 운행으로 다른 사람을 사망 또는 부상하게 한 경우에는 손해배상 책임을, “자동차보유자”(동법 제5조)는 자동차의 운행으로 다른 사람을 사망, 부상하게 한 경우 피해자에게 책임을 지는 책임보험이나 책임공제에 가입할 의무를 지우고 있다. 이

109) 최현태, “자율주행차 운행 중의 사고 발생과 손해배상책임” 한양법학 제19권 제1호, 2018, 456면; 이종영·김정임, “자율주행자동차 운행의 법적 문제” 중앙법학 제17집 제2호, 중앙법학회 2015, 160면; 자율주행자동차는 정보통신 기술 및 컴퓨터 시스템에 기반하기 때문에 컴퓨터 시스템의 해킹이나 바이러스 감염 등으로 인한 오작동과 같이 자율주행자동차 시스템의 결함으로 인한 사고가 더 높은 비중을 차지할 것으로 예상된다.

110) 보다 넓게는 4단계 이상의 완전 자율주행기능이 가능한 자율주행자동차가 책임의 주체가 될 수 있는지의 여부와 관련해서 2017년 EU 유럽의회에서 채택된 ‘로봇민사규율 결의안’도 살펴볼 수 있다. 결의안에서는 인공지능이 탑재된 로봇을 전제로 자율성 확보, 인지역량, 물리적 형태, 환경적응 역량 등을 갖춘 로봇을 ‘스마트 자율 로봇’으로 정의하고 이에 대한 법인격 부여, 법적 책임에 대한 규칙, 개발 및 이용에 대한 윤리 규칙 등에 대해 제기하였다. 즉 이른바 ‘강한 인공지능’에 규범적 차원에서 행위능력과 책임능력을 인정할 수 있는지에 대한 논의이다. 자율주행자동차의 경우 인간을 보조하는 수준에서 제한된 기능을 수행하는 것이므로 ‘약한 인공지능’의 범주에 포함되며 인공지능 자신이 아닌 제조업자나 프로그래머, 또는 운전자나 보유자에게 책임을 지우는 것이 일반적이고 현실적 시각이다; 최현태, 앞의 논문, 473-475면; European Parliament, “European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015 2103(INL))”, 2017; 조주은 앞의 논문, 235면.

를 자율주행자동차에 적용할 경우 “자기를 위하여 자동차를 운행하는 자”에 자율주행자동차의 운행을 조정하거나 관리하는 사람도 해당하는지가 쟁점이 된다. 대법원 판례에 의하면 “사회 통념상 당해 자동차에 대한 운행을 지배하여 그 이익을 향유하는 책임주체로서의 지위에 있다고 할 수 있는 자”라고 하며¹¹¹⁾ 이에 의할 때 자율주행자동차의 운전자도 운전자가 아닌 자율주행기능에 의해 운행했다라도 운전자의 운행지배와 운행이익이 완전히 상실되었다고 볼만한 사정이 없는 한 그 운행을 지배하여 이익을 향유하는 자로서 책임을 부담한다고 할 수 있다. 그러나 4단계 이상의 자율주행자동차의 경우 운전자의 개입 여지가 없을 때 현행법을 적용할 수 있느냐에 대한 논의가 필요하다. 이 경우 주된 책임의 주체가 운전자나 운전자에서 제조업자로 넘어가게 되는 것이 불가피하다는 견해도 있으나 이 경우에도 차량보유자의 책임도 배제할 수 없다.¹¹²⁾ 결국 향후 자율주행자동차의 자동화 진전 단계에서 기존의 ‘차량보유자’, ‘차량 운전자’의 책임 구조는 차량보유자, 탑승자, 자율주행자동차 제조사(차체, 소프트웨어), 그리고 이들에 대한 손해보험사, 그리고 비정상적 상황까지 포함한다면 자율주행시스템에 대한 해킹을 한 자 등을 포섭하는 책임 구조로 변화될 것이고 이들 사이의 위험/책임 분배의 문제가 논의되어야 한다.¹¹³⁾

‘제조물책임법’은 “제조물의 결함”으로 발생한 손해에 대한 제조업자 등의 손해배상책임을 규정한다(제조물책임법 제1조). 먼저 자율주행자동차 관련해서 교통사고 발생 시 자율주행자동차가 ‘제조물책임법’에 의해 포섭 가능한지에 대한 검토가 필요하다. 동법 제2조 제1호는 “제조물이란 제조되거나 가공된 동산(다른 동산이나 부동산의 일부를 구성하는 경우를 포함한다)”이라고 규정하고 있어서 자율주행자동차 차체 외에 자율주행기능의 핵심적인 역할을 하는 인공지능 소프트웨어가 제조물에 포함되어 다루어질 수 있는지가 쟁점이 된다. 이에 대해 소프트웨어는 동산이 아니기 때문에 제조물로 보기 어렵다는 견해¹¹⁴⁾와 일단 자동차에 탑재되어 부품과 같은 기능을 수행하는 임베디드 소프트웨어(embedded software)로서 제조물에 포함된다고 볼 수 있다는 견해가 있다.¹¹⁵⁾ 각국에서 소프트웨어 자체를 제

111) 대법원 2009. 11. 12. 선고 2009다63106 판결; 대법원 1986. 12. 23. 선고 86다카556 판결.

112) 최현태, 앞의 논문, 457면.

113) 조주은, 앞의 논문, 236면.

114) 권오승·신은주·홍명수·차성민·이현중, 제조물 책임법, 법문사, 2003, 190면.

조물로 볼 것인지에 대해서는 입법례와 판례의 입장이 다양한데 특히 자율주행자동차와 같이 기계의 본원적 기능이 소프트웨어에 내재된 인공지능시스템과 합체되어 있는 경우 소프트웨어를 제조물 개념에 포함하는 입법에 대한 논의가 필요하다.¹¹⁵⁾

제조물책임법상 제조물책임을 지는 주체인 “제조업자”에 대해 (동법 제2조 제3호 가목, 나목) 자동차의 경우 완성차 제조업자, 부품제조업자, 원재료 제조업자 모두 제조물 배상 책임의 책임 주체가 될 수 있는데 자율주행자동차의 경우 소프트웨어에 결합이 있을 때 임베디드 소프트웨어를 제조물로 보게 되면 완성차 제조업자, 소프트웨어 제조업자 모두에게 책임을 물을 수 있다. 이에 대해서도 보다 명확하게 합의와 입법이 필요하다. 또한 제조물책임법에서의 책임의 요건으로 제조물의 ‘결함’이 있어야 하는데 이를 ‘제조상의 결함, 설계상의 결함, 표시상의 결함, 그 밖에 통상적으로 기대할 수 있는 안전성이 결여되어 있는 것’(동법 제2조 제2호)으로 규정하고 있다. 자율주행자동차의 경우 그 비중이 큰 소프트웨어 오류 발생 시 제조상의 결함에 대해서 법률에 규정되어 있는 대로 “소프트웨어가 원래 의도한 설계와 다르게 가공됨으로써 안전하지 못하게 된” 점을 피해자가 입증해야 하는데 소프트웨어가 문제 상황에서와 같이 재현되고 그로 인해 사고가 발생하였다는 것을 입증하는 것이 피해자 입장에서 쉽지 않은 문제가 있다. 또한 소프트웨어와 자동차의 하드웨어 등에 대한 ‘설계상의 결함’ 및 이에 대해 ‘대체설계의 가능성’을 자율주행자동차와 같이 발전의 속도가 빠른 혁신기술의 경우에 어떻게 정의하고 입증해야 하는지에 대한 문제도 논의가 필요한 내용이다. 같은 맥락에서 제조물책임법 제 4조 제1항에 규정된 제조물책임을 지는 자의 면책사유 중 제2호 “제조업자가 해당 제조물을 공급한 당시의 과학·기술 수준으로는 결함의 존재를 발견할 수 없었다는 사실”에 대해서도 자율주행자동차의 위험성을 감안하여 어느 정도 수준을 면책의 범위로 인정해야 하는지, 또한 제4조 제2항에 “제조물을 공급한 후에 그 제조물에 결함이 존재한다는 사실을 알거나 알 수 있었음에도 그 결함으로 인한 손해의 발생을 방지하기 위한 적절한 조치를 하지 아니한 경우”로 규정하

115) 강선준·원유형·김민지, 앞의 논문, 160면; 권영준·이소은, “자율주행자동차 사고와 민사책임”, 민사법학 제75권, 2016, 468면; 이상수, “임베디드 소프트웨어의 결함과 제조물책임 적용에 관한 고찰”, 법학논문집 제39집 제2호, 중앙대학교 법학연구소, 2015, 73면.

116) 조주은, 앞의 논문, 237면.

면책의 예외에 대해서도 리콜이나 위험 통지 등 소비자의 안전을 충분히 확보할 수 있도록 해석하는 방안에 대한 논의도 필요하다.¹¹⁷⁾

2018년 4월부터 시행된 개정 제조물책임법은 제조물의 결함이나 손해의 내용, 양자의 인과관계를 입증해야 했던 기존의 법제를 해당 제조물이 “정상적으로 사용되는 상태에서 손해가 발생하였고 그 손해가 제조업자의 실질적 지배영역에 속한 원인으로 초래되었으며 통상적으로 해당 제조물의 결함이 없이는 그러한 손해가 발생하지 않는다는 사실”을 입증하면 제조물 결함과 손해 사이의 인과관계가 추정되는 것으로 개정하였다(동법 제 3조의2). 또한 “제조업자가 제조물의 결함을 알면서도 그 결함에 대하여 필요한 조치를 하지 아니한 결과로 생명 또는 신체에 중대한 손해를 입은 자가 있는 경우에는 그 자에게 발생한 손해의 3배를 넘지 아니하는 범위에서 배상책임”을 지도록 하는 징벌적 손해배상책임을 규정하였다(동법 제 3조 제2항).¹¹⁸⁾ 이는 향후 상용화될 때 그 위험성이 매우 큰 자율주행자동차에 대해서 의미가 있다. 다만 자율주행자동차와 같이 복잡성과 여러 가지 형태(생명, 신체, 재산상의 손해뿐 아니라 프라이버시와 같은 인격권 등)의 위험성이 큰 첨단기술 제품의 경우 입증책임과 징벌적 손해배상의 정도에 대한 추가적 고려가 필요할 수 있다. 예를 들면, 결함이나 손해와의 인과관계에 대한 소비자의 입증책임이 더욱 어려워져 입증책임을 완화하거나 제조사나 소프트웨어 업체에게 전환시키는 방안의 검토가 필요한데 이 경우 제조업체나 소프트웨어 업체들에 과도한 책임 전환이 될 경우 기술개발에 저해를 초래할 수 있다. 따라서 운전자, 제조사, 소프트웨어 업체, 그리고 보험회사 사이의 책임 분배에 대한 명확한 규정과 이를 위한 합의가 필요하다.¹¹⁹⁾

117) 최병록, “자율주행자동차에 있어서 제조물책임의 주요 쟁점에 관한 연구”, IT와 법 연구 제14권, 경북대학교 IT와 법연구소, 2017, 211-233면.

118) 이경미·황경웅, “개정 제조물책임법에 대한 소고”, 법학논문집 제43집 제1호, 중앙대학교 법학연구원, 2019, 71-73면.

119) 강선준·원유형·김민지, 앞의 논문, 179-180면.

IV. TPII 매트릭스와 적용 제언

지금까지 인공지능 기술(Technology)의 비전과 현황, 인공지능을 비롯한 모든 기술이 인류 사회에 적용되는 과정에서 변치 않을 윤리원칙(Ethics Principle), 그리고 인공지능 기술 적용 분야 중 시민의 기대 수요가 높고 기술 수준에 있어서도 신산업 핵심 분야로 전망되는 금융산업, 정밀의료, 자율주행자동차(Industry)의 인공지능 기술 동향, 주요 윤리원칙에 대한 각 산업별 이슈와 이에 대한 제도화 과정(Institution)에서의 고려사항에 대해 다루어 보았다.

인공지능 기술이 사회에 적용되는 과정은 몇 가지 특징을 갖고 있다. 첫째는 매우 긴 시간의 흐름이다. 위에서 살펴본 바와 같이 '인공지능'은 지난 반세기 이상 동안 기술공학자들이 변함없이 갖고 있었던 큰 그림의 설계였고 여전히 달성되지 못했으나 그 비전하에서 때로는 느리게, 때로는 빠른 속도로 그 성취를 이루어 나가고 있는 기술적 비전이다. 아마 몇 차례의 '인공지능의 봄과 겨울' 역사가 보여주듯 인류와 사회가 인공지능을 어떻게 바라보든(그것을 우호적으로 보든, 비우호적으로 보든, 유익하게 보든, 위험하게 보든) 이러한 기술적 전개는 기술진보의 페이스로 진행될 것이다. 둘째는 다층적 관점이다. 이러한 현실적 준비과정에서 여러 이해관계자의 관점이 다를 수 있다. 업계, 학계(여기에서도 공학, 경영학, 경제학, 사회학, 법학 등 학제적 관점), 정부, 시민 혹은 수요자 등은 - 사실 이러한 분류 아래에서도 더욱 세분화할 수 있다. - 인공지능 기술에 대한 다른 관심과 이해를 갖고 있다. 셋째, 단계적이고 실용적인 합의의 필요성이다. 인공지능은 기술적 발전의 과정이고 다양한 관점으로 바라볼 수 있지만, 무엇보다 현실과 일상에의 적용 과정은 매우 실재적이다. 따라서 기술에 대한 핵심적 이해, 인류 사회에서 지켜져야 하는 공통분모로서의 윤리원칙, 기술의 적용 과정과 윤리원칙과의 충돌 부분, 이로 인해 발생하는 다양한 이슈에 대한 적절한 논의를 거친 제도화를 통해 기술 개발과 적용 모두가 안정적이고 발전적으로 이루어지는 과정이 필요하다.

이 보고서에서 다룬 인공지능 '기술(Technology)-윤리원칙 (Principle)-적용 산업(Industry)-제도화(Institution)'의 접근을 향후 인공지능 기술의 지속적 개발과 인류와 사회에 대한 적용 및 융합과정에서 활용할 수 있는 TPII 매트릭스로 제시해보도록 하겠다. 여기에서는 이 보고서에서 다룬 인공지능 기술의 발전 비전과 현황, 세 가지 주요 윤리원칙, 인공지능 기술 적용 핵심 신산업 세 가지, 제도화 과

정의 고려사항을 이 매트릭스에 적용해 제시해 본다(표 1).

[표 1] 인공지능 기술발전 및 적용 과정의 TPII 매트릭스

인공지능 기술 단계 Technology	인공지능 윤리원칙 Principle	인공지능 적용 산업 Industry	제도화 Institution
<ul style="list-style-type: none"> 제한적 범위의 인지능력 (AI with deep learning) 	<ul style="list-style-type: none"> 편향·차별금지/공정 	<ul style="list-style-type: none"> 금융산업 ... 	<ul style="list-style-type: none"> 정보주체의 권리; 자동화 평가 실시여부, 자동화 평가 결과 및 기준, 기초자료 등에 대한 설명요구권, 자동화 평가 결과 산출에 유리한 정보 제출, 기초정보 정정/삭제, 결과 재산출 요구권(개정 신용정보보호법) 금융서비스 전 과정에서의 설명의무에 대한 설명가능성 (은행법, 자본시장과금융투자업에관한법률, 보험업법) 금융시스템 위험 및 개발업체 규제 ...
	<ul style="list-style-type: none"> 프라이버시 	<ul style="list-style-type: none"> 정밀의료 ... 	<ul style="list-style-type: none"> 유전정보와 생체정보에 대한 법적 개념 (유전형/표현형과 관련 범위, 생체정보의 개인정보/민감정보/고유식별정보 여부) 법제도의 중복적용과 상충가능성 (의료법/생명윤리및안전에관한법률/개인정보보호법/의료법) 유전자 검사에 대한 알 권리와 모를 권리(헌법) ...
	<ul style="list-style-type: none"> 안전성/책임성 	<ul style="list-style-type: none"> 자율주행자동차 ... 	<ul style="list-style-type: none"> 자동화 진행과정에서 운전자 배제와 핸들/브레이크 조작 여부에 대한 제도적 포섭(도로교통법) 시험운행이 확대될 수 있는 기반 조성(자동차관리법 등) 자율주행자동차 사고 관련 책임 문제(형사법, 민사법, 자동차손해배상보장법, 제조물책임법)
	<ul style="list-style-type: none"> ... 	<ul style="list-style-type: none"> ... 	

[표 1]은 보고서에서 다룬 내용을 TPII 매트릭스에 적용한 것이다. 인공지능 기술은 1940-50년대부터 ‘인간처럼’ 일반적인 문제해결 능력을 지닐 수 있는 기계를 지향하면서 계속 발전해 오고 있으며, 90년대 이후 ICT 기반 확장으로 인한 빅데이터 기반 구축, 병렬 하드웨어와 같은 컴퓨팅 파워의 발전, 80년대 이후 더욱 효과적인 기법으로 발전해 온 알고리즘 기술, 모바일 및 클라우드 기술에 의해 일상화된 데이터 입력 기반과 효율적인 저장, 활용 기반 구축 등으로 현재 상당히 일상화된 수준의 사용성을 지닌 제한된 범위의 목적 기반 인공지능 적용 수준에 와 있다. 인공지능 기술이 인류 사회의 일상과 함께하는 수준이 확장되고 고도화됨에 따라 정부, 업계, 국제기구, 학계, 시민단체 등은 인공지능 기술개발 및 적용 과정에 반드시 고려되어야 할 윤리원칙들에 대해 연구, 공표하고 있으며 많은 기관은

공히 데이터 기반 인공지능의 구조와 한계에서 비롯되는 편향차별금지/공정, 프라이버시, 안전성/책임성 윤리원칙을 제시하고 있다. 인공지능기술 적용에 대한 시민의 기대수준이 높고 핵심 신산업으로 전망되는 금융산업, 정밀의료, 자율주행자동차에 대해 주요 윤리원칙들과 상충될 수 있는 이슈들을 다루어 보았고 이슈들이 제도화되는 과정에서 고려해야 할 사항들에 대해 제시하였다.

TPII 매트릭스 적용은 인공지능 기술의 발전단계에 대한 현실적인 인식, 윤리원칙의 지속적인 업데이트 및 적용기술과 도메인에 따른 우선순위 조정, 적용산업이 확장되는 과정에서 사전적으로 예상되는 이슈와 사후적으로 도출되는 이슈의 정확한 파악, 그리고 제도화 과정에서 관련 이해관계자들의 선제적·지속적이고 체계적인 소통 및 합의 과정에 의해 그 가치를 더할 수 있을 것이다.

주제어 Ⅱ 인공지능 기술, 인공지능 윤리원칙, 인공지능 산업, 인공지능 관련 제도, 인공지능 적용 매트릭스

| 참고문헌 |

1. 국내문헌

- 강선준 · 원유형 · 김민지, “제4차 산업혁명 시대의 자율주행자동차 상용화를 위한 안정적 법적 기반을 위한 법 정책적 연구 자율주행자동차 특별법 제정 (안)을 중심으로”, 기술혁신학회지 제21권 제1호, 2018.
- 고학수 · 정해빈 · 박도현, "인공지능과 차별", 저스티스 통권 제171호, 한국법학원, 2019.
- 고학수 외, 금융권 인공지능 정책 연구, 금융보안원, 2019.
- 고학수 외, 개인정보 비식별화 방법론 - 보건의료정보를 중심으로 -, 박영사, 2017.
- 과학기술정보통신부, “토종 AI 정밀의료 서비스, ‘닥터 앤서’ 임상 적용 시작”, 2019.
- 관계부처 합동, “바이오헬스 산업 혁신전략”, 2019.
- 권영준 · 이소은, “자율주행자동차 사고와 민사책임”, 민사법학 제75권, 2016.
- 권오승 · 신은주 · 홍명수 · 차성민 · 이현중, 제조물 책임법, 법문사, 2003.
- 금융보안원 보안기술연구팀, “머신러닝 기반의 이상거래 탐지시스템 동향”, 2017. 8..
- 김규동, “인슈어테크 활성화를 위한 규제개선 방안, 글로벌 핀테크 규제환경 분석 과 개선 방향 세미나”, 2019. 6. 17..
- 김대환, “효율적 의료비 지출을 통한 국민건강보험의 보장성 강화 방안”, 보험연구원, 2017.
- 김무웅, “글로벌 정밀의료 시장 현황 및 전망”, BioINdustry No. 131, 2018.
- 김민호 · 김현경 · 김선아, “정보통신서비스 분야의 민감정보 유형과 보호방안 연구”, 방송통신위원회 연구보고서, 2017.
- 김봉조 · 김영진 · 문상훈, “컨트롤 타워 마련과 우수한 보건의료 인프라 적극 활용 해야”, 보건산업 동향, 2015.
- 김정수 · 신영재, “알고리즘매매와 고빈도매매의 글로벌 규제동향”, 증권법연구 제

- 14권 제2호, 한국증권법학회, 2013.
- 김준호, “‘폭주하는 전차’ 사례에 대한 형사법적 결론을 생각해보기-자율주행자동차의 윤리적 딜레마에 관한 법학적 사유실험을 겸하여-”, 법학연구 제29권 제2호, 연세대학교 법학연구원, 2019.
- 박순영, “유전체 기반 정밀의료 연구동향”, 융합연구정책센터, 2017.
- 박성해, “대규모 인구집단 유전체정보 기반 정밀의료 핵심인프라 구축”, 보건산업동향, 2015.
- 방송통신위원회 · 한국인터넷진흥원, “바이오정보 보호 가이드라인”, 2017.
- 배영임 · 신혜리, “데이터3법, 데이터 경제의 시작”, 경기연구원 이슈&진단, 2020.2,
- 보건복지부, “「DTC 유전자 검사서비스 인증제 시범사업」 소비자 참여 연구 본격 시작”, 2019.
- 산업통상자원부, “‘혁신의 실험장’ 규제 샌드박스 최초 승인”, 2019.
- 석희태, “환자의 모를 권리와 의사의 배려의무”, 의료법학 제17권 제2호, 대한의료법학회, 2016.
- 손 계리시, 이수겸 역, 기계는 어떻게 생각하는가, 이지스퍼블리싱, 2019.
- 유원형, “[트렌드탐구] 新 핀테크 '레그테크'와 '셍테크'가 주목받는 이유”, 메가경제, 2019. 7. 5..
- 유호종, “유전정보를 모를 권리’의 윤리적 · 법적 근거와 실현 방법”, 한국의료법학회지 제22권 제1호, 한국의료법학회, 2014.
- 이근영, “국내의 로보어드바이저 동향 및 분석”, 전자금융과 금융보안, 제6호, 2016.04..
- 이선구, “알고리즘의 투명성과 설명가능성”, 서울대학교 인공지능정책 이니셔티브, 이슈페이퍼, 2019.2.
- 이원복, “유전체 시대의 유전정보 보호와 공유를 위한 개인정보보호법제의 고찰”, 법조 제729호, 법조, 법조협회, 2018.
- 이을상, “임상유전학에서 자기 결정권”, 인문과학연구 제31집, 성신여자대학교 인문과학연구소, 2013.
- 이종영 · 김정임, “자율주행자동차 운행의 법적 문제”, 중앙법학 제17집 제2호, 중앙법학회, 2015.

이준영, “디지털헬스케어 동향 및 시사점”, 정보통신산업진흥원, 이슈리포트 (2019-03호), 2019.

이창범 외, 개인정보보호 법령 및 지침·고시 해설, 행정자치부, 2016.

장민선, 인공지능 시대의 법적 쟁점에 대한 연구, 한국법제연구원, 2018.

장병탁, “인공지능 입문”(강의자료), 2019 .

_____, 장교수의 딥러닝, 홍릉과학출판사, 2017..

전용일, “우리나라 자율주행자동차법안(자율주행자동차 개발 촉진 및 상용화 기반 조성에 관한 법률안)의 주요 내용 분석과 시사점”, 법학연구 제27권 제1호, 2019.

정일영·구원모, “헬스케어 생태계 구축을 위한 데이터 통합 방안”, 과학기술정책 연구원 동향과 이슈 제46호, 2018.

조주은, “ ICT와 사회의 ‘융합’을 위한 사회적 합의 프레임워크- 인공지능과 자율주행자동차를 중심으로”, 중앙법학 제21집 제3호, 중앙법학회, 2019.

질병관리본부, “코호트 원격분석서비스 개시 및 자료공개 확대”, 2019.

_____, “2017 국립중앙인체자원은행 연보”, 2018.

최병록, “자율주행자동차에 있어서 제조물책임의 주요 쟁점에 관한 연구”, IT와 법 연구 제14권, 경북대학교 IT와 법연구소, 2017.

최정호, “생명의료인권협약에 비추어 본 ‘모를 권리’ 문제”, 인권법평론 제21호, 전남대학교 공익인권법센터, 2018.

최현태, “자율주행차 운행 중의 사고 발생과 손해배상책임”, 한양법학 제19권 제1호, 2018.

한국정보화진흥원, “인공지능 기반 챗봇 서비스의 국내외 동향분석 및 발전 전망”, Trend&Future, 2018.2.

허전·정덕영, “헌법상 유전정보보호와 의사의 고지의무 - 미국의 사례를 중심으로 -”, 헌법학연구 제15권 제2호, 한국헌법학회, 2009.

2. 외국문헌

Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies*, Oxford

- University Press, 2014.
- Buchnan, Bonnie G., “Artificial intelligence in Finance”, *Alan Turing Institute*, 2019.
- Collins, Francis S. · Harold Varmus, “A New Initiative on Precision Medicine”, *The New England Journal of Medicine*, Vol. 372, No. 9, 2015.
- Fitche, Oliver · Su Yang, “Artificial Intelligence: challenges for the financial sector”, Fintech–innovation Hub, *ACPR, Discussion paper*, 2018.
- MacKay, David, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- Minsky, Marvin, *Computation: Finite and Infinite Machines*, Englewood Cliffs, 1967.
- Rosenblatt, F., “The Perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review*, Vol. 65, No. 6, 1958.
- Simon, Herbert Alexander, *The Shape of Autonomous for Men and Management*, Harper & Row, 1965.
- Turing, A.M. , “Computing Machinery and Intelligence”, *Mind*, Vol. 59, No. 236, 1950.
- Wicks, Paul · Timothy E Vaughan · Michael P Massagli · James Heywood, “Accelerated clinical discovery using self–reported patient data collected online and a patient–matching algorithm”, *Nature biotechnology*, Vol. 29, No. 5, 2011.

기타 자료

- BIS(Bank for International Settlements), “Innovative technology in financial supervision (suptech) – the experience of early users”, *FSI Papers* No 9, 2018.
- European Commission, “Artificial Intelligence, Robotics and Autonomous

Systems”, 2018.

National Research Council, “Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease”, 2011.

National Science and Technology Council, “National Privacy Research Strategy”, 2016.

OECD, “Artificial Intelligence in Society”, 2019.

Stanford University, “Artificial Intelligence and Life In 2030”, One Hundred Year Study On Artificial Intelligence, 2016.

The Executive Office of the President of the US, “The National Artificial Intelligence Research and Development Strategic Plan”, National Science and Technology Council, 2016.

The White House, “FACT SHEET: President Obama’s Precision Medicine Initiative”, 2015.

_____, “Precision Medicine Initiative: Privacy and Trust Principles”, 2015.

_____, “Precision Medicine Initiative: Data Security Policy Principles and Framework”, 2016.

<Abstract>

The Matrix of Technology-Principle-Industry-Institution (TPII) and its Application to Artificial Intelligence*

Cho, Ju Eun** · Jeong, Jong Gu*** ·
Park, Do Hyun**** · Kim, Eun Soo*****

This article is to derive a practical framework for the application and convergence of Artificial Intelligence(AI) to real world, which has been a major topic of economy, society, culture, innovation and daily life. A turning point for AI to come in mind of citizens of the world was the historical Go match giving victory to AlphaGo over Lee Se-Dol in 2016. However, AI is a generic technology which has history reaching a century, and in 1950's, when it had low-level function compared to nowadays, it was still called "Artificial Intelligence". Therefore for the AI application and convergence to real world, it can be a starting point to have common understanding of AI technology for multi-dimensional stakeholders (developer, businessman, consumer, policymaker etc.) who may have different pictures and interests concerning AI. Then during the application of this technology, ethical considerations concerning risks should be shared, and the industrialization based on this technology needs

* This research is based on the seminars of SNU AI Policy Initiative, and this work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2019S1A5A2A03036673).

** Senior Researcher, Ph.D. of Laws, Center for Law & Economics, Seoul National University.

*** Doctoral Student, Lawyer, Graduate School of Law, Seoul National University.

**** Ph.D. Candidate, Lawyer, Graduate School of Law, Seoul National University.

***** Senior Researcher, S.J.D., Center for Law & Economics, Seoul National University.

institutionalization through which producer and consumer could carry on stable benefits.

Concerning ‘AI technology’, even though not all stakeholders can have the same knowledge level with technology expert, the vision, current stage, its structure & mechanism, and limitations should be basically recognized among stakeholders and they are focused here. ‘AI ethic principle’ deals with bias-non-discrimination/fairness, privacy, safety/liability which have been presented as major ethical issues by governments, private sectors, inter-governmental organizations, and civil societies for years. Based on the principles, ‘AI applied industry’ deals with finance, precision medicine, autonomous vehicle which are high demanded domains. For the ‘institutionalization’ which is necessary for the stable development of industrialization of AI, analyzes the issues which should be discussed for the consensus building. Finally this procedural matrix, ‘Technology-Ethic Principle-Industry-Institution (TPII)’ is presented as a frame for an initiative and consistent discussion among multi-dimensional stakeholders.

Key Words || AI Technology, AI Ethic Principle, AI Industry, AI-related Institution, AI application matrix