연구보고서 2016-011

지능정보사회 대응을 위한 법제도 조사연구

A Proposal of the Ethical Guideline for Artificial Intelligence

김윤명/정필운/고인석/김대규

2017.04.

SPRi 소프트웨어정책연구소

이 보고서는 2016년도 미래창조과학부 정보통신진흥기금을 지원받아 수행한 연구결과로 보고서 내용은 연구자의 견해이며, 미래창조과학부의 공식입장과 다를 수 있습니다.

목 차

제 1 장 서론 ····· 1

제 1 절 연구의 목적 ····· 1

제 2 절 연구의 범위와 방법 ····· 3

제 2 장 국외 인공지능 관련 윤리 정책 및 법제 동향 ····· 4

제 1 절 미국의 윤리 정책 및 법제 동향 분석 ····· 5

1. 백악관 국가과학기술위원회 “인공지능 국가 연구 개발 전략 계획” ····· 5

2. 백악관 국가과학기술위원회 “인공지능의 미래를 위한 준비” ····· 15

제 2 절 유럽연합(EU)의 윤리 정책 및 법제 동향 분석 ····· 20

1. 유럽로봇연구네트워크 “로봇윤리 로드맵” ····· 20

2. 유럽에서의 기술 규제 : 로봇에 대한 법과 윤리의 당면과제 (약칭 : LoboLaw Project) ····· 21

3. 2014년 “로봇규제 가이드라인” ····· 22

제 3 절 중국의 윤리 정책 및 법제 동향 분석 ····· 25

1. 개요 ····· 25

2. 로봇산업발전규획 ····· 25

3. 인공지능과 인간의 역할에 대한 담론 ····· 30

제 4 절 일본의 윤리 정책 및 법제 동향 분석 ····· 32

1. 경제산업성 “로봇 신전략” ····· 32

2. 지식재산전략본부 “지식재산추진계획 2016” ····· 43

제 5 절 국제단체의 윤리 정책 및 법제 동향 분석 ····· 51
1. IEEE “윤리적으로 조화로운 설계” ····· 51
제 6 절 소결 ····· 60

제 3 장 국내 인공지능 관련 윤리 정책과 법제 동향 분석 ····· 63
제 1 절 인공지능 윤리 관련 정책 동향 분석 ····· 63
1. 로봇윤리헌장 제정을 위한 작업의 사례 ····· 63
2. 기타 최근의 정책적 시도 ····· 68
제 2 절 인공지능 법제 동향 분석 ····· 69
1. 기존 법제 ····· 69
2. 최근 논의 ····· 70
제 3 절 소결 ····· 72

제 4 장 윤리 정책과 법제의 대안 ····· 73
제 1 절 현실문제와 그에 대한 대응 방안 ····· 73
1. 인간행위의 규율수단 ····· 73
2. 변화하는 현실에 적응할 수 있는 법 정립을 위한 원칙 ····· 81
제 2 절 윤리적인 관점의 분석과 쟁점 연구 ····· 83
1. 인공지능의 존재론적 지위 ····· 83
2. 공학윤리의 활용 ····· 89
3. 인공지능 및 인공지능 로봇에 관한 윤리의 기본틀 ····· 93

제 3 절 가이드라인의 제안 ····· 106
1. 가이드라인의 필요성 ····· 106
2. 인공지능 윤리 가이드라인 제안 ····· 107
제 4 절 법제 정비방안 ····· 113
1. 가이드라인의 불완전성과 보충 필요성 ····· 113
2. (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 설치 ····· 113
3. 국제적인 인공지능 윤리 레짐의 창설 ····· 115
4. 그 밖의 법제정비방안 ····· 119
제 5 절 가이드라인과 법제정비방안이 SW산업분야에 미치는 영향 분석 ·· 121

제 5 장 결론 ····· 122

참 고 문 헌 ····· 126

표 목 차

〈표 4-1〉 레이어 모델에서 각 레이어의 특징 ····· 80

그 림 목 차

[그림 2-1] 창작 주체별 창작물과 현행 지적재산제도 ····· 50

[그림 4-1] 사회문제를 해결하는 규율수단 ····· 75

요 약 문

1. 제 목

지능정보사회 대응을 위한 법제도 조사연구

2. 연구 목적 및 필요성

인간을 대신하여 더 빨리, 더 정교하게 일을 해 줄 인공지능 덕분에 생산성이 향상되고 사람의 여가 시간을 더욱 늘어날 것으로 예상된다. 그러나 그 때문에 사람의 일자리가 감소하고 안전이 위협받을 수 있다고 예상되기도 한다. 따라서 인공지능 기술 발전을 위한 노력과는 별도로 그것이 가져올 역기능에 대한 국가・사회적 대처가 필요하다.

그와 같은 국가・사회적 노력은 정책과 이를 뒷받침할 법제의 형태로 구체화된다. 이 보고서는 지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시하는 것을 그 목적으로 한다.

3. 연구의 구성 및 범위

지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시를 위하여 우선 인공지능에 대비하여 미국, 유럽연합, 중국, 일본과 같은 선진국에서는 어떠한 준비를 하고 있는지 그 동향을 살펴보았다(제2장). 그리고 인공지능이 대비하여 한국에서 어떠한 준비를 하였는지 그 동향을 살펴보았다(제3장). 그리고 한국에서 준비가 어떤 문제가 있으며, 그 개선 방안은 무엇인지 제시하였다(제4장).

4. 연구 내용 및 결과

미국, 유럽연합, 일본, 중국에서는 지능정보사회에 적절히 대응하기 위하여 적극적인 정책을 추진하고 있었다. 미국은 2016년 10월 “인공지능 국가 연구 개발 전략 계획”, “인공지능의 미래를 위한 준비” 보고서를 통해 기술과 산업 진흥, 국가·사회적 대응을 위한 시사점을 제시하고 있었다. 유럽연합은 2012년 “유럽에서의 기술 규제 : 로봇에 대한 법과 윤리의 당면과제(RoboLaw Project)”와 2016년 “규칙 초안을 위한 보고서”(Draft Report)를 통하여 국가·사회적 대응을 위한 구체적인 윤리와 법제적인 대응 방안을 제시하고 있었다. 중국은 2016년 “로봇산업발전규획”을 통하여 2020년까지 로봇산업의 건강하고 지속가능한 발전을 추진할 것을 목표로, 제조업분야의 새로운 국면을 창출하고 공업수준을 한 단계 끌어올려 중국로봇기술이 세계 로봇 산업을 선점하기 위한 방법을 제시하고 있었다. 일본은 2015년 “로봇 신전략”을 통해 시계의 로봇 기술 혁신 거점, 세계 제일의 로봇활용 사회, 세계를 선도하는 로봇신시대에 대한 전략을 제시하고 있었다. 또한 “지식재산추진계획 2016”에서는 인공지능과 관련한 지식재산 정책 방향을 제시하였다. 이와 같이 유럽연합은 인공지능이 초래할 국가·사회적 영향을 검토하고 이에 대비하기 위한 윤리·법제적 대응에 중점이 있는 반면, 중국·일본은 인공지능 기술의 발전과 산업 진흥을 목표로 이를 달성하기 위한 법제적 대응에 중점을 두는 특징을 보였다. 미국은 기술의 발전과 산업 진흥, 국가·사회적 영향을 검토하고 이에 대비하기 위한 윤리·법제적 대응이 필요하다고 인정하면서도 구체적인 대응 방법을 제시하기보다는 지속적인 모니터링을 할 것을 권고하는 특징을 보였다.

우리나라는 지난 2007년부터 행정부 주도로 로봇윤리헌장을 제정을 위한 노력을 전개하였다. 그 목표 중 하나는 IEEE 국제 컨퍼런스에서 한국이 로봇윤리헌장 제정을 제안하는 것이었다. 그리고 그러한 정책 추진을 뒷받침하기 위하여 「지능형 로봇 개발 및 보급 촉진법」도 제정하였다. 그러나 행정부 주도 작업 경향, 전문가 경시 경향, 기술자와 사회과학자, 정책입안자 등의 협업 부족 등으로 목표를 달성하지 못하였다. 그 후 로봇 산업을 비롯한 인공지능 산업은 다른

ICT 산업에 비하여 진흥되지 못하였으며, 관련 연구 성과도 축적되지 못하였다. 한동안 소강상태에 있었던 정책적 대응은 2016년 알파고 현상을 계기로 재점화되었다. 그리하여 「지능정보사회 중장기 종합대책」이 발표되고, 「지능정보사회기본법」 제정을 논의하고 있다. 윤리헌장을 만들려는 노력도 재개되고 있다. 그러나 행정부 주도의 산업 진흥 정책을 추진하기 위한 법적 근거는 이미 마련되어 있고, 현재 기술 수준을 고려하여 보았을 때 법령을 통한 규제 정책은 너무 이르다. 인공지능 연구와 그 응용을 위한 개별적인 법적 장애가 무엇인지는 현장에서 이미 제기되었고 그 해결책도 제시되고 있다. 한편, 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하기 위한 구체적인 정책이 필요하다.

이 보고서에서는 ICT 영역에서 사회문제를 해결하는데 법과 윤리 등 사회 규범의 기능과 시장, 기술 등과의 기능 분담을 고려하여야 하고, 인공지능 기술 발달 수준과 인공지능에 대한 존재론적 지위에 기초하여 인간의 존엄성과 안전을 보호하기 위하여 당장 필요한 것은 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 가이드라인 형식으로 제안하였다.

구체적으로 서술하면 현재 인공지능의 기술 수준을 발전기에 진입한 단계로 파악하고, 법적 강제력 없는 가이드라인(guideline) 형식을 통한 규율이 필요하다고 결론지었다. 그리고 인공지능을 판단과 행위 능력을 지닌 주체로 보거나 특별한 속성을 지닌 인공물로 보는 것을 넘어 '외화된 사회적 정신'(externalized social mind)으로 인식하는 것이 타당하다고 결론지었다. 이러한 인식에 따르면 현재 기술 수준의 인공지능을 별도의 권리와 의무의 주체로서 인정하기 곤란하지만, 앞으로 기술 발전이 진전된다면 별도의 권리와 의무의 주체로서 인정할 가능성도 열어둘 수 있다. 이러한 전제에 기초하여 현재 인공지능 윤리는 사람의 윤리이어야 하며, 인간의 존엄성과 안전을 보호하기 위하여 현재 시점에서 필요한 것은 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 총20여개 조문의 가이드라인 형식으로 제안하였다.

한편 강제력을 바탕으로 한 법적 효력이 없는 가이드라인을 보충하기 위하여 국내적으로는 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회를 설치하고, 국제적으로는 인공지능 윤리 레짐(regime)을 창설할 것을 제안하였다.

한편, 이 보고서는 지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시하는 것을 그 목적으로 하다보니, 인공지능이 고용, 교육 등 국가와 사회에 가져올 변화를 예측하고 이에 대한 정책의 대응에 관해서는 본격적으로 다루지 못하였다. 이 보고서의 한계이고 앞으로 연구 과제이다.

5. 정책적 활용 내용

인공지능 연구와 그 응용 상황을 고려하여 보았을 때 현재 상황에서, 인간의 존엄성과 안전을 보호하기 위하여 연구자・설계자・제작자・관리자・이용자가 윤리적으로 행위할 수 있도록 구체적인 행위 규칙이 필요하다. 이 보고서에서 제시한 가이드라인은 국가가 이와 같은 구체적인 행위 규칙을 제정하는데 초안으로서 기능할 수 있을 것이다. 한편, 가이드라인을 보충하기 위한 조직법적 제안, 즉 국내적으로는 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 설치하고, 국제적으로는 인공지능 윤리 레짐 창설 제안은 인공지능 연구와 그 응용을 규율하는 정책적 대안이 될 수 있을 것이다.

6. 기대효과

미국 전국공학인협회(NSPE)의 공학윤리헌장(Code of Ethics for Engineers)이 공학자에 대한 사회적 신뢰를 높여 공학자에게 자유를 부여하였듯, 인공지능 관련 가이드라인은 관련 연구와 그 응용에 사회적 신뢰를 높이고, 연구자・설계자・제작자・관리자 등에게 안정적인 자유 영역을 지키게 해 줄 것이다. 그리고 궁극적으로 인공지능 산업과 그 핵심인 소프트웨어 산업이 예측가능성을 가지고 연구와 개발, 그 투자를 할 수 있도록 하는 기반이 될 것이다.

SUMMARY

1. Title

A Proposal of the Ethical Guideline for Artificial Intelligence

2. Purpose and Necessity

Thanks to artificial intelligence that would replace human's works, it is expected that productivity will improve and leisure time will increase. On the other hand, it is also expected that artificial intelligence will reduce the number of human's jobs and even threaten our safety. To prevent these side effects of artificial intelligence, appropriate national and social measures need to be established.

These sorts of national and social measures can be specified as policies or laws supporting those policies. This report aims to diagnose the current state of laws for artificial intelligence society and their problems, and to present alternative solutions.

3. Composition and Scope

To diagnose the current state of laws for artificial intelligence society and their problems, and to provide alternative solutions, this report has examined the trend of developed countries including the U.S., European Union, China and Japan on preparations for the future artificial intelligence society. (Chapter 2) Moreover, this paper covers how Korea has been preparing for the future artificial intelligence society. (Chapter3) Furthermore, the report points out some problems about preparations that Korea has been set up for artificial intelligence, and provides right solutions for them. (Chapter 4)

4. Findings and Result

The U.S., EU, Japan and China have been enforcing aggressive policies to respond appropriately to artificial intelligence society. The U.S. published the reports named “National Artificial Intelligence Research and Development” and “Preparing for The Future of Artificial Intelligence” in October 2016. EU also published “Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics – also known as RoboLaw Project–” in 2012 as well as “Draft Report” in 2016. China announced “Robot industry development plan.“ Also, Japan has publicized “Japan’s Robot Strategy” in 2015, and “Intellectual Property Promotion Plan 2016” in 2016.

Overall, EU focuses more on scrutinizing national and social effects of artificial intelligence and establishing ethical and legal solutions for them. On the other hand, China and Japan put more importance on setting goals of improving artificial intelligence technology and its industry and preparing legal steps in order to achieve those goals. In case of the U.S., it acknowledges the necessity of reviewing technological development, industrial advancement and national and social effects. Instead of providing specific measures, however, it only suggests more and continuous monitoring on artificial intelligence issues.

Since 2007, Korea has been making efforts to enact “The Chapter of Robotics Ethics” led by Ministry of Industry and Energy (present. Ministry of Trade, Industry and Energy). One of their goals was that Korea proposes to enact of “The Chapter of Robotics Ethics” at IEEE international conference. To support those drive policies, Korea has enacted 「Intelligent Robots Development and Distribution Promotion Act」. The policy response that came to a lull got a chance to reignite after AlphaGo syndrome in 2016. As a result, 「Intelligence information society Mid-to Long-Term Comprehensive Plan」 was announced, and whether to enact 「FRAMEWORK ACT on Intelligence

information society」 is on discussion table. Plus, the effort has recently been remade to create the Chapter of Ethics.

This report argues for the division of functions among social norms such as laws and ethics, markets, and technologies to deal with social issues related to ICT field. Therefore, this report presents ethics in forms of guidelines of 20 provisions that is believed necessary for ethical and social cognitive improvement for designer, manufacturer, maintenance and administration, in order to protect human dignity and safety based on the level of artificial intelligence technology development and ontological status of artificial intelligence.

Meanwhile, it has suggested to establish National Artificial Intelligence Ethics Committee and Artificial Intelligence Ethics Commission within the country and to put regime of artificial intelligence ethics internationally to complement the guidelines lack in legal validity.

5. Practical Use of Policy

The guidelines that this report provides can work as a governmental draft for specifying ‘The Behavior Regulation’. Also, legal organization suggestions to complement guidelines such founding establish National Artificial Intelligence Ethics Committee and Artificial Intelligence Ethics Commission within the country and putting regime of artificial intelligence ethics internationally would work as alternative policies on artificial intelligence studies and regulations on its applications.

6. Expections

As Code of Ethics for Engineers of the NSPE, the U.S, has given the

freedom to engineers by raising the social trust toward engineers, the guideline of artificial intelligence also can be kept a stable independence area to designers, manufacturer, maintenance and administration by raising social trust on the related research and applicant. Finally, it is guaranteed to be foundation for making research, development and its investment within having the predictability by artificial intelligence industry and their key point of software industry.

CONTENTS

Chapter 1. Introduction

Section 1. Purpose of study

Section 2. Method of study

Chapter 2. Research Overseas Trend of Ethic Policy and Legal of Artificial Intelligence

Section 1. United States of Ethic Policy and Legal of Artificial Intelligence

Section 2. European Unions of Ethic Policy and Legal of Artificial Intelligence

Section 3. China of Ethic Policy and Legal of Artificial Intelligence

Section 4. Japan of Ethic Policy and Legal of Artificial Intelligence

Section 5. International Organization of Ethic Policy and Legal of Artificial Intelligence

Section 6. Conclusion

Chapter 3. Research Internal Trend of Ethic Policy and Legal of Artificial Intelligence

Section 1. Analysis of policy trend of artificial intelligence Ethics

Section 2. Analysis of legal trend of artificial intelligence Ethics

Section 3. Conclusion

Chapter 4. Political and Legal Proposal

Section 1. Analysis and Issue study of ethic perspective

Section 2. Actual problem and plan for reaction

Section 3. Proposal of guideline

Section 4. Plan for modify the legal

Section 5. Effect of the guideline and the plan for enact the law on the field of SW.

Chapter 5. Conclusion

제1장 서 론

제1절 연구의 목적

인간을 대신하여 더 빨리, 더 정교하게 일을 해 줄 인공지능 덕분에 생산성이 향상되고 사람의 여가 시간을 더욱 늘어날 것으로 예상된다. 인공지능 분야의 급속한 발전으로 인공지능 관련 연구·기술 인력 등의 수요가 급격히 증가하고 있다. 그러나 자동화가 용이한 일자리는 급격히 대체되어 전체적으로는 일자리의 증가보다 감소가 더 클 것으로 예상된다.[1] 인공지능이 광범위하게 사용되면서, 인간의 안전을 침해하지 않을지에 대한 우려뿐만 아니라 영화에서 나오는 것처럼 인공지능이 인간을 지배하는 것은 아닐까 하는 우려가 제기되기도 한다.[2] 따라서 인공지능 기술 발전을 위한 노력과는 별개로 그것이 가져올 역기능에 대한 국가적·사회적 대처가 필요하다.

한편, 자율주행자동차와 드론의 도입 논의에서 볼 수 있는 것처럼 지능정보화 기술을 적용한 서비스 도입 및 운영 단계에서 새로운 법적 문제가 제기된다. 그리고 그러한 법적 문제의 해결 중 일부는 윤리적 접근이 필요하다. 자율주행자동차의 소프트웨어 설계 단계에서 설계자가 준수하여야 할 행위 규칙이 무엇인지, 트롤리 문제를 어떻게 해결하도록 설계할 것인가, 이른바 강한 인공지능을 위한 기술 개발을 제한 없이 허용할 것인가 등이 법적 쟁점에 내포된 윤리적 쟁점의 한 예이다. 이러한 이유로 유럽연합(EU)

---

1) 이상 Executive Office of the President National Science and Technology Council Committee on Technology(2016), *Preparing for the Future of Artificial Intelligence*, p.2.

2) 이와 관련하여 엘론 머스크 테슬라 모터스 최고경영자는 지난 2014년 미국 MIT 항공우주공학과가 주최한 100주년 심포지엄에서 "인공지능 연구는 우리가 악마를 소환하는 것"이라는 표현을 한 바 있다. 2014.10.26. ZDNet 인터넷판 기사, http://www.zdnet.co.kr/news/news_view.asp?artice_id=20141026165058#csidxe06ace25b20894e894b240ed1f43409 (2016.12.20. 최종 방문); 영국의 천체물리학자 스티븐 호킹도 같은 해 BBC와 인터뷰에서 "완전한 '인공지능'의 개발이 인류의 멸망을 불러올 수 있다"며, "인공지능은 스스로를 개량하고 도약할 수 있지만, 인간은 생물학적 진화 속도가 늦어 인공지능과 경쟁할 수 없어 대체될 것이어서 두려움을 느낀다"고 말하였다. 2014.12.3., 한겨레 신문 인터넷판 기사, http://www.hani.co.kr/arti/international/international_general/667299.html#csidxdeb2c5ec8a349c6bf6a91e4d0b55c6a (2016.12.20. 최종 방문)

의 로봇법 프로젝트(Robolaw Project) 등에서도 인공지능을 규율할 법제 연구의 전제로 윤리학적 접근을 시도하고 있다.

자율주행자동차, 드론, 수술로봇(surgical robots), 인공기관(prosthetics) 등 지능정보사회를 선도하는 응용 서비스가 제기하는 법적 문제와 그 해결책에 관해서는 이미 국내외에서 여러 방면의 논의가 있어 왔다. 예를 들어, 자율주행자동차에서는 어느 정도 안전한 자율주행자동차를 제조하고 판매할 수 있도록 허락할 것인지, 제조 또는 판매된 자율주행자동차가 사고를 일으킬 경우 누구에게 민사・형사책임을 부과할 것인지, 결함 있는 자율주행자동차에 대한 제조물책임을 인정할 수 있을 것인지, 그 대안 중 하나인 보험 제도를 구체적으로 어떻게 설계할 것인지 등이 그것이다.[3] 그러나 인공지능 연구자・설계자・제작자・관리자 등 관련 행위자가 준수해야 할 구체적인 행위 규칙이 무엇인지, 현 단계에서 그것을 어떤 법형식으로 규율할 것인지 등의 논의는 상대적으로 저발전 되어 왔다.

이 연구는 지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시하는 것을 그 목적으로 한다. 그 중 특히 인공지능 연구자・설계자・제작자・관리자 등 관련 행위자가 준수해야 할 구체적인 행위 규칙이 무엇인지 탐색하고 가이드라인 형식으로 이를 제안하는 것에 중점을 두고자 한다.

3) 예를 들어, EU, Robolaw Project, D6.2 Guidelines on Regulating Robotics; 김윤명(2016), 「인공지능과 리걸 프레임 10가지 이슈」, 커뮤니케이션북스, 85면 이하; 황창근, 이중기(2016), “자율주행자동차 운행을 위한 행정규제 개선의 시론적 고찰-자동차, 운전자, 도로를 중심으로-”, 「홍익법학」 제17권 제2호, 홍익대학교 법학연구소; 심우민(2016a), “인공지능 기술 발전과 입법정책적 대응 방향”, 「이슈와 논점」 제1138호, 국회 입법조사처 등.

제2절 연구의 범위와 방법

이미 설명한 것처럼 이 보고서는 지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시하는 것을 그 목적으로 한다. 이를 위하여 우선 인공지능에 대비하여 미국, 유럽연합, 중국, 일본과 같은 선진국에서는 어떠한 정책을 준비하거나 추진하고 있는지 그 동향을 살필 것이다(제2장). 그리고 인공지능이 대비하여 한국에서는 어떠한 정책을 준비하고 추진하고 있는지 그 동향을 살필 것이다(제3장). 그리고 한국의 정책과 법제의 문제는 무엇이고, 현재 상황에서 한국에서 필요한 정책과 법제는 무엇인지 그 대안을 제시하고, 그러한 대안의 추진이 소프트웨어 산업에 미치는 영향은 무엇인지 분석하고자 한다(제4장).

이러한 목적을 달성하기 위하여 제2장 국외 동향 분석, 제3장 국내 동향 분석은 문헌조사법을 이용하였다. 미국, 유럽연합, 중국, 일본에서 발간한 정책보고서를 통해 당해 국가의 인공지능 관련 동향을 살펴보고자 하였다. 한국의 인공지능 관련 정책과 법은 정책 자료, 신문기사, 논문 등을 통해 살펴보고자 하였다. 한국의 정책과 법제의 문제와 현재 상황에서 한국에서 필요한 정책과 법제의 대안, 소프트웨어 산업에 미치는 영향에 대한 분석은 입법론적 접근, 정책론적 접근을 이용하였다.

제2장 국외 인공지능 관련 윤리 정책 및 법제 동향

현재 인공지능에서 대한 논의는 국외 주요 선진국에서 활발하게 이루어지고 있다. 하지만 국가별의 논의 진행의 형태는 차이가 있다. 미국의 경우는 처음 기업 및 학계에서 필요에 의해 진행되다 백악관 주관으로 2016년 10월 “인공지능 국가 개발 연구 전략”과 “인공지능의 미래를 위한 준비”라는 보고서를, 2016년 12월 “AI와 자동화가 경제에 미치는 영향” 보고서를 연이어 발표하며 AI시대에 대한 준비를 시작하였으나, 인공지능 전반에 대하여 규율하는 제정법령 또는 입법안은 확인되지 않고 있다.

유럽연합(EU)은 로봇법 프로젝트(Robotlaw Project)는 로봇기술의 윤리적, 법률적 이슈 검토를 통해 새로운 규범 체계를 정립하고자 하는 연구 목표하에 이탈리아, 네덜란드, 영국, 독일 등 4개국 4개 연구소가 참여했고 특히 자율주행자동차, 수술로봇, 로봇인공기관, 돌봄로봇 등 4가지 로봇기술의 윤리적, 법률적 분석을 통해 로봇 규제정책의 근거를 마련하고자 했다.

중국은 2015년 5월 “중국제조2025(Made in China 2025)”를 발표하여 2025년까지 생산강국으로 발전하고, 2049년 제조업 강국의 지위를 굳건히 하겠다는 목표를 설정하고 이를 위해 “첨단 디지털 제어기계 및 로봇”을 중점분야로 선정하였으며, 이후 2016년 3월 로봇산업 육성 로드맵인 “로봇산업발전규획(機器人産業發展規劃)을 발표하였다.

일본의 경제산업성은 2015년 “로봇 신전략”을 발표하고 자국 경제성장의 핵심적인 전략으로서의 ‘로봇혁명’을 주도하기 위한 계획을 반영하였으며, 관련 법·규제의 개혁과 국제표준화의 대응을 마련하기 위한 플랜을 제시하였다.

이렇듯 인공지능사회에 대비하기 위한 다양한 논의가 이루어지고 있다. 아래에서는 각 국가별 주요동향에 대하여 살펴본다.

제1절 미국의 윤리 정책 및 법제 동향 분석

1. 백악관 국가과학기술위원회 “인공지능 국가 연구 개발 전략 계획”

(1) 개요

백악관 국가기술위원회(National Science and Technology Council, 이하 NSTC[4]) 산하에는 머신러닝에 관한 인공지능 소위원회(MLAI)가 있었고, 전략 계획을 만들기 위해서 인공지능 태스크포스를 구성해 다양한 연구 기관이나 정부 산하 조직을 모두 아우르는 워킹 그룹을 구성해 보고서의 작성을 시작하기로 하였다. 이를 위하여 2016년 5월 3일, 백악관 과학기술정책국(OSTP)에서는 인공지능의 도전과 기회에 대한 명확화를 위해 인공지능을 이용해 더 효과적인 정부를 구현하고 인공지능의 잠재적 혜택과 위험에 대한 준비를 하겠다는 목표를 가지고 이 보고서를 발간하였다.

이 보고서 발간을 위해 태스크포스에 참여한 사람들의 조직을 보면 미국 국립과학재단(NSF), 미국 고등정보연구계획국(IARPA) 등의 연구 개발 지원 조직뿐만 아니라, 의료, 법무, 에너지, 우주 항공, 표준, 기술 정책, 국방, 안보와 보안의 전문가들이 모두 참여했다. OSTP는 네 번의 워크숍을 통해 논의를 활성화했으며, 인공지능의 미래에 대해 정부가 무엇을 준비해야 하는가에 대한 의견을 많은 기업과 연구 기관에서 입수하기 위한 RFI[5]를 발행했다. 지난 6월에 요청한 RFI에서는 총 11개의 주제[6]를 제시했는데, 2천 자

4) NSTC는 연방 정부의 과학 기술 투자에 대한 국가적 목표를 명확히 하는 것이 주 역할이며, 그 아래에 5개의 위원회가 있다. ‘환경, 천연자원, 지속성’, ‘국가 안보’, ‘과학 기술, 공학, 수학(STEM) 교육’, 그리고 ‘과학’과 ‘기술’ 위원회이며 각각 소위원회와 워킹 그룹이 있다.

5) 자세한 사항은 https://www.federalregister.gov/documents/2016/06/27/2016-15082/request-for-information-on-artificial-intelligence 참조

6) RFI에서 제시된 주제는 1.인공지능의 법적, 거버넌스 함의, 2.공공선을 위한 인공지능 사용, 3.안전과 제어 이슈, 4.사회적 경제적 함의, 5.대부분의 과학 영역에서 인공지능 연구 중 가장 긴급하고 기초적 질문, 6.인공지능을 발전시키고 공공에 혜택을 주기 위해 언급되어야 하는 가장 중요한 연구 격차, 7.인공지능 기술의 잠재성을 활용하기 위해 필요한 과학 기술 훈련, 고등 교육에서 교수진을 확보하고 폭발적으로 증가하는 학생에 대응하기 위한 교육 기관의 도전 문제, 8.다양한 분야의 인공지능 연구를 고취하기 위한 연방 정부, 연구 기관, 대학, 자선 기관 등이 취해야 하는 특정한 단계들, 9.인공지능 개발과 응용을 가속할 할 수 있는 학습 데이터 세트, 10.중저소득 노동자를 위한 훈련 가속과 같은 사회적 요구에 대응하기 위한 인공지능 응용 개발을 가속하기 위한 ‘시장 형성’의 역할, 11. 인공 지능 연구 또는 정책 결정과 관련된 추가 정보 등이다.

이내의 의견만을 취합하겠다고 했다. 가장 빨리 답을 한 곳은 IBM[7]이었으며, 총 161개의 답변을 받았다. 답변은 대기업이나 전문 그룹뿐만 아니라 작가, 실업자, 트럭 운전사, 정치 평론가 등 다양한 시민들의 의견이 모였고, 이를 통해 인공지능의 과학 기술적 요구를 확인하고 이를 위한 R&D 투자 과정을 추적하고 효과를 극대화하기 위한 상위 수준의 프레임워크를 정의하기 위함을 도출하였다. 이와 동시에 연방정부가 투자하는 연구 개발의 우선순위를 정하고, 단기 전략을 넘어서 장기적으로 인공지능이 사회와 세상에 어떤 변화를 줄 것인가를 살펴보고자 했다.

(2) 미래상황에 대한 가정

이 보고서에서는 AI의 미래에 대해 몇 가지 가정을 한다.[8] 첫째, 정부와 산업계가 인공 지능 연구 개발에 지속적으로 투자하기 때문에 AI 기술이 성숙성과 존재성(sophistication and ubiquity)에 있어서 계속적으로 성장할 것이라고 가정한다. 둘째, 이 계획은 미국 경제 성장에 대한 영향뿐만 아니라 고용, 교육, 공공 안전 및 국가 안보를 포함하여 AI가 사회에 미치는 영향이 계속 증가할 것으로 가정한다. 셋째, 최근 상업적 성공으로 R&D 투자에 대한 인지되는 수익이 증가함에 따라 AI에 대한 산업 투자가 계속 증가할 것으로 가정한다. 동시에, 이 계획은 일정한 분야는 공공재에 관련된 전형적인 과소투자문제로 인하여 산업계는 일부 중요한 연구 분야에 충분한 투자를 하지 않을 것으로 가정한다. 마지막으로 이 계획은 인공 지능 전문 지식에 대한 수요가 산업, 학계 및 정부 내에서 계속 증가 할 것이므로 공공 및 민간 노동력을 요구할 것이라고 가정한다.

이 AI 연구 개발 전략 계획과 관련이 있는 다른 연구 개발 전략 계획과 이니셔티브는 연방 정부의 빅 데이터 연구 및 개발 전략 계획,[9] 연방 사이

7) 자세한 사항은 http://research.ibm.com/cognitive-computing/ostp/rfi-response.shtml 참조
8) J. Furman(2016), Is This Time Different? The Opportunities and Challenges of Artificial Intelligence, *Council of Economic Advisors remarks*, New York University: AI Now Symposium
9) Federal Big Data Research and Development Strategic Plan, May 2016, https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf.

버 보안 연구 및 개발 전략 계획,[10] 국가 개인 정보 보호 연구 전략,[11] 국가 나노 기술 이니셔티브 전략 계획,[12] 국가 전략적 구상 (National Strategic Computing Initiative),[13] 혁신적인 신경 기술 이니셔티브를 통한 뇌 연구[14] (Innovative Neurotechnologies Initiative) 및 국가 로봇 공학 구상 (National Robotics Initiative)[15]을 포함한다. 추가적인 전략 연구 개발 계획과 전략적 기초는 개발 단계에 있으며 비디오 및 이미지 분석, 건강 정보 기술, 로봇 및 인텔리전트 시스템을 포함한 AI의 특정 하위 분야를 소개하여야 한다. 이러한 추가 계획과 기초는 AI 연구 개발 전략 계획을 보완하고 확대하는 시너지 효과가 있는 권고안을 제공할 것이다.

(3) 윤리적・법적 문제의 접근

AI가 자율적으로 작동할 때, 우리가 기존의 대인관계에 대하여 가지고 있는 공식적・비공식적인 규범 또는 윤리에 따라 AI가 행동할 것을 기대한다. 그렇기 때문에 윤리적・법적 및 사회적 원칙에 부합하는 AI 설계를 위한 개발 방법뿐만 아니라, AI의 윤리적・법적・사회적 의미를 이해하는 연구가 상당히 필요하다.

이렇듯 다른 기술과 마찬가지로, 법과 윤리의 원칙에 따라 AI를 받아들여 이용하여야 할 것이다. 이러한 과제는 신기술에 이러한 원칙 특히 자율성, 에이전시 및 제어와 관련된 원칙들을 적용하는 방법일 것이다.

---

10) Federal Cybersecurity Research and Development Strategic Plan, February 2016, https://www.nitrd.gov/cybersecurity/publications/2016_Federal_Cybersecurity_Research_and_Development_Strategic_Plan.pdf.

11) National Privacy Research Strategy, June 2016, https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf.

12) National Nanotechnology Initiative Strategic Plan, February 2014, http://www.nano.gov/sites/default/files/pub_resource/2014_nni_strategic_plan.pdf.

13) National Strategic Computing Initiative Strategic Plan, July 2016, https://www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf

14) Brain Research through Advancing Innovative Neurotechnologies (BRAIN), April 2013, https://www.whitehouse.gov/BRAIN.

15) National Robotics Initiative, June 2011, https://www.whitehouse.gov/blog/2011/06/24/developing-next-generation-robots.

(가) 인간중심의 자동화 원칙

미래의 응용방법은 인간과 AI 시스템 간의 기능적 역할 분담, 인간과 AI 시스템 간의 상호 작용의 본질, 함께 작업하는 인간 및 기타 AI 시스템의 수, 인간과 AI 시스템이 의사소통 및 상황 인식을 공유하는 방식에 따라 크게 달라질 것으로 판단하고 있다.

인간과 AI 시스템 사이의 효과적인 상호 작용을 달성하려면 시스템 설계가 과도하게 복잡하거나, 과소한 신뢰 또는 과도한 신뢰로 이어지지 않는 추가적인 R&D가 필요하다. 인간은 AI 시스템의 역량과 AI 시스템이 할 수 있는 것과 할 수없는 것을 잘 이해할 수 있다는 것을 보장하기 위하여 AI 시스템에 대한 인간의 친숙함은 훈련과 경험을 통해 증가될 수 있다. 이러한 문제를 해결하기 위해 이러한 AI 시스템의 설계 및 개발에 인간 중심 자동화 원칙을 사용해야 하며[16], 다음과 같은 사항을 반영하도록 해야 한다.

1. 인간-AI 시스템 인터페이스 : 제어 및 디스플레이의 직관적이고 사용하기 쉬운 디자인을 채택
2. 운영자에게 지속적인 정보제공 : 중요한 정보, AI 시스템의 상태 및 이러한 상태의 변경 사항을 표시
3. 운영자의 지속적인 교육 : AI 시스템에서 사용되는 알고리즘과 로직 및 시스템의 예상 장애 모드에 대한 교육뿐만 아니라 일반적인 지식, 기술 및 능력 (KSAs)에 대한 반복적인 교육.
4. 유연한 자동화 : AI 시스템 배치는 AI 시스템을 사용여부를 결정하는 운영자를 위한 설계 옵션으로 고려되어야 한다. 또한 과도한 작업량이나 피로상태인 동안 인간인 운영자를 지원할 수 있도록 적응형 AI 시스템의 설계와 배치는 중요하다.[17]

(나) 인간을 인식하는 AI를 위한 새로운 알고리즘의 모색

수년에 걸쳐 AI 알고리즘은 증가하는 복잡성의 문제를 해결할 수 있게 되

16) C. Wickens and J. G. Hollands(1999), "Attention, time-sharing, and workload in Engineering", *Psychology and Human Performance*, pp.439-479.
17) https://www.nasa.gov/mission_pages/SOFIA/index.html. https://cloud1.arc.nasa.gov/intex-na/

었다. 그러나 이러한 알고리즘의 역량과 인간에 의한 이러한 시스템의 유용성에는 차이가 있다. 이용자와 직관적으로 상호 작용하고, 원활한 기계와 인간의 협업을 가능하게 하는 인간-인식 지능형 시스템이 필요하다.

또한 지능형 시스템이 필요하고 적절할 때에만 지능형 시스템이 인간을 제어할 수 있는 방해 모델을 개발해야 한다. 지능형 시스템은 인간의 인지력을 향상시키는 능력도 가지고 있어야 하고, 인간이 정보를 명시적으로 시스템에 요청하지 않을 때조차도 이용자에게 정보가 필요할 때 어떤 정보가 필요한지를 알 수 있어야 한다. 미래의 지능형 시스템은 인간의 사회적 규범을 설명하고, 그에 따라 행동할 수 있어야 한다. 지능형 시스템이 어느 정도의 감성 지능을 소유하고 있다면, 이용자의 감정을 인식하고 적절하게 대응할 수 있으므로 인간과 보다 효과적으로 작업할 수 있다. 추가적인 연구 목표는 하나의 인간과 하나의 기계의 상호 작용을 넘어서 다양한 사람들과 상호작용을 할 수 있는 다양한 기계로 구성된 systems-of-systems으로 나아가야 한다.

Human-AI 시스템 상호 작용은 광범위한 목표를 가지고 있다. Human-AI 시스템은 다양한 목표를 표현할 수 있는 능력, 그러한 목표에 도달하기 위하여 취할 수 있는 행동, 그러한 행동에 대한 통제 및 기타 요인을 필요로 하며 그러한 목표 수정에 쉽게 적응할 수 있어야 한다. 또한 Human-AI 시스템은 공통의 목표를 공유하고, Human-AI 시스템의 현재 상태의 관련 측면을 상호 이해해야 한다.

(다) 인간의 역량을 강화하기 위한 AI 기술 개발

과거에 AI 연구의 중심은 협소한 작업을 수행하는 사람들과 일치하거나 초과하는 알고리즘에 관한 것이었지만, 많은 영역에서 인간의 역량을 향상시키는 시스템을 개발하려면 추가 작업이 필요하다. 인간 역량 고양 연구는 고정 장치 (예 : 컴퓨터), 웨어러블 장치 (예 : 스마트 안경), 이식된 장치 (예 : 뇌 인터페이스), 특정 이용자 환경 (특별히 맞춤 설정된 수술실과 같

은)등 에서 작동하는 알고리즘이 포함된다. 예를 들면, 인간의 인지 능력이 고양되면 여러 장치에서 결합된 데이터 판독 값을 기반으로 의료 보조원도 의료 과정에서 실수를 지적 할 수도 있다. 다른 시스템은 이용자의 현재 상황에 적용할 수 있는 이용자의 과거 경험을 회상하도록 도와줌으로써 인간의 인지 능력을 향상시킬 수 있다.

인간과 AI 시스템 간의 또 다른 유형의 협력은 지능적인 데이터를 이해하기 위한 능동적인 학습과 관련 있다. 능동 학습에서 입력은 분야 전문가가 수행하고, 입력이 불확실한 경우에는 데이터로 능동학습이 학습 알고리즘에 따라 수행된다. 이것은 처음에 생성할 필요가 있는 학습 데이터의 양 또는 학습해야 하는 양을 줄이기 위하여 중요한 기술이다. 능동 학습은 분야 전문가의 입력을 획득하고, 학습 알고리즘에 대한 신뢰를 높이는 핵심적인 방법이다. 지금까지 능동 학습은 감독 학습(supervised learning) 내에서만 사용되어 왔다. - 능동학습을 감독되지 않은 학습 (예 : 클러스터링, 이상 탐지) 및 보강 학습으로 통합하기 위해 추가 연구가 필요하다. 확률론적 네트워크는 분야 지식이 이전의 확률 분포의 형태로 포함되도록 한다. 수학적 모델, 텍스트 또는 다른 형식의 형태에 관계없이 기계 학습 알고리즘이 분야 지식을 통합할 수 있도록 하는 일반적인 방법을 찾아야 한다.

(라) 시각화 및 AI-인간 인터페이스 기술 개발

더 좋은 시각화 및 이용자 인터페이스는 인간이 대량의 최신 데이터 세트 및 다양한 원천에서 나오는 정보를 이해할 수 있도록 도움을 주기 위하여 훨씬 더 큰 발전이 필요한 추가 영역이다. 시각화 및 이용자 인터페이스는 사람이 이해할 수 있는 방식으로 점점 더 복잡해지는 데이터와 그러한 데이터에서 나온 정보를 명확하게 제시해야 한다. 실시간 결과를 제공하는 것은 안전에 중요한 분야에서 중요하고, 계산 능력과 연결된 시스템이 증가함에 따라 달성될 수 있다. 이러한 유형의 상황에서 이용자는 실시간 응답을 위하여 정확한 정보를 신속하게 전달할 수 있는 시각화 및 이용자 인터페

이스가 필요하다.

Human-AI 협력은 다양한 환경에서, 그리고 통신에 제약이 있는 곳에서도 적용될 수 있다. 일부 영역에서는 human-AI 통신 대기시간이 짧고, 통신이 빠르고 안정적이다. “NASA‘s deployment of the rovers Spirit and Opportunity to Mars” 과 같은 다른 영역에서 사람과 AI 시스템 간의 원격 통신은 대기 시간이 매우 길다 (예 : 지구와 화성 사이의 왕복 시간 5 ~ 20 분). 그러므로 배치된 플랫폼은 플랫폼에 전달되는 high level 전략 목표를 가지고 자율적으로 작동되기를 요구한다. 이러한 통신 요구 조건 및 제약 조건은 이용자 인터페이스의 연구 개발에 중요한 고려 사항이다.

(마) 효과적인 언어 처리 시스템 개발

사람들이 구어와 문어로 AI 시스템과 상호작용할 수 있게 하는 것은 AI 연구자들의 오랫동안 목표였다. 상당한 발전이 있었지만, 사람들이 다른 사람과 효과적으로 의사소통을 하는 것처럼 인간이 AI 시스템과 효과적으로 의사소통을 할 수 있기 전에 언어 처리 분야에서 상당한 연구 과제를 해결해야 한다. 최근의 언어 처리 과정의 발전은 실시간으로 조용한 환경에서 유창한 영어 연설을 인식할 수 있는 성공적인 시스템을 만들어 낸 데이터 중심의 machine learning 접근법을 사용한 덕택이다. 그러나 이러한 성공은 장기 목표를 달성하기 위한 초석일 뿐이다. 현재의 시스템은 시끄러운 주변 환경에서의 연설, 심하게 강조된 연설, 어린이 연설, 연설 장애 및 수화 연설과 같은 현실적인 문제를 처리 할 수 없다. 실시간으로 인간과 대화할 수 있는 언어 처리 시스템의 개발 또한 필요하다. 그러한 시스템은 인간 대화자들의 목표와 의도를 추론하고, 적절한 레지스터 및 상황에 적절한 스타일 및 수사법을 사용하고, 대화에서 오해의 경우에 대한 수정 전략을 채택해야 한다. 서로 다른 언어 간에 조금 쉽게 일반화할 수 있는 시스템을 개발하기 위해서는 더 많은 연구가 필요하다. 또한 언어 처리 시스템에 따라 쉽게 접근할 수 있는 형태로 유용한 구조화된 분야 지식을 습득하는 것에 더 많은

연구가 필요하다. 인간과 AI 시스템 간의 상호 작용을 보다 자연스럽고 직관적으로 만들기 위해서는 많은 다른 영역에서의 언어 처리에 관련된 발전이 필요하다. 감정적인 상태, 영향 및 자세에 대한 증거를 제공하는 음성 및 텍스트에 패턴과 음성 및 텍스트에 내포된 정보를 결정하기 위하여 강력한 계산 모델이 구축되어야 한다. 로봇과 같이 세상에서 작동하는 AI 시스템을 위한 환경에 적합한 기초 언어를 위하여 새로운 언어 처리 기술이 필요하다. 마지막으로 온라인 상호 작용에서 사람들이 의사소통하는 방식은 음성 상호 작용과 크게 다를 수 있으므로, 사회적인 AI 시스템이 사람들과 보다 효과적으로 상호 작용할 수 있도록 이러한 상황에서 사용되는 언어 모델을 완성하여야 한다.

(바) 윤리적 AI 구축

정의와 공정성에 대한 기본적인 가정을 넘어서, AI 시스템이 일반적인 윤리적 원칙을 준수하는 행동을 나타낼 수 있는지에 대한 또 다른 우려가 있다. AI에서의 진보가 윤리적인 측면에서 새로운 “기계 관련“ 문제들을 어떻게 틀을 잡을지? 인공 지능의 어떤 용도가 비윤리적으로 간주될지? AI 기술은 엔지니어링에 의존하고 제한되지만, 윤리는 본질적으로 철학적인 문제이다. 따라서 과학적으로 실현 가능한 한계 내에서 연구자들은 현존하는 법률, 사회적 규범 및 윤리와 확실히 일치하는 알고리즘 및 아키텍처를 개발하기 위해 노력해야하지만, 이것은 분명히 매우 어려운 과제다. 윤리적 원칙은 일반적으로 모호성을 가지고 있고, 정밀 시스템 및 알고리즘 설계로 변환시키기 어렵다. AI 시스템, 특히 새로운 종류의 자율적인 의사 결정 알고리즘을 사용하는 AI 시스템이 독립적이고 상충 가능한 가치 체계에 기반을 둔 도덕적 딜레마에 직면할 때도 문제가 발생한다. 윤리적 문제는 문화, 종교 및 신념에 따라 다르지만, 결론과 행동을 설명하고 정당화하기 위하여 AI 시스템 추론과 의사 결정에 대한 방향을 안내하는 수용 가능한 윤리 참조 기준을 개발할 수 있다. 어려운 도덕적 문제 또는 상충되는 가치가 나타날 때 선호되는 행동을 나타내는 예를 포함하여 적절한 가치 체계를 반영

하는 훈련용 데이터 세트를 생성하기 위하여 다양한 분야의 접근이 필요하다. 이 예에는 이용자에게 투명한 결과 또는 판단으로 명명된 법적 또는 윤리적 "예외적인 사례(corner cases)"가 포함될 수 있다.[18] AI는 엄격한 규칙을 실행할 수없는 복잡한 상황을 해결할 수 있는 원칙을 시스템에 포함하는 경우에 가치 기반 충돌 해결을 위한 적절한 방법을 필요로 한다.

(사) 윤리적 AI를 위한 아키텍처 설계

윤리적 추론을 포함하는 AI 시스템을 위한 최적의 아키텍처 설계 방법을 결정하기 위해 기본 연구에 대한 추가적인 진전이 이루어져야 한다. 운영행위에 대한 윤리적 또는 법적 평가를 담당하는 모니터 에이전트와 운영 AI를 구분하는 2단계 감시 아키텍처와 같은 다양한 접근 방식이 제안되었다.[19] 인공지능 행동이 인간에게 해롭지 않고 안전하다는 것을 보장하기 위하여 AI agent 구조에 대한 정확한 개념적 기초를 사용되는 안전 공학을 선호하는 또 다른 관점도 있다.[20] 세 번째 방법은 윤리적인 원칙에 순응시키는 행동을 제한하는 AI 시스템 행동에 논리적인 제한과 함께 이론적인 원칙을 이용하는 윤리적인 아키텍처를 형성하는 것이다.[21] AI 시스템이 좀 더 보편화됨에 따라, AI 아키텍처는 다양한 판단 단계에서 윤리적 문제를 해결할 수 있는 하위 시스템,신속 대응 패턴 매칭 규칙, 행동을 묘사하고, 정당화하기 위하여 저속반응을 위한 신중한 추론, 이용자의 신뢰도를 나타내는 사회적 신호표시, 시스템이 문화적인 개념을 준수할 수 있는 장기 시간 기준에 작동할 수 있는 사회적인 절차 등을 포함할 것이다.[22] 연구자들은 윤리적, 법적 및 사회적 목표에 부합하는 AI 시스템의 전반적인 설계를 가장

18) A. Etzione and O. Etzioni(2016), "Designing AI Systems that Obey Our Laws and Values", *in Communications of the ACM 59 (9)*, pp.29-31.

19) *Idle*

20) R. Y. Yampolsky(2013), "Artificial Intelligence Safety Engineering: Why Machine Ethics is a Wrong Approach," in *Philosophy and Theory of Artificial Intelligence,* Springer Verlag, pp.389-396.

21) R. C. Arkin(2007), "Governing Legal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture" *Georgia Institute of Technology Technical Report GIT-GVU-07-11*, p.14.

22) B. Kuipers(2016), "Human-like Morality and Ethics for Robots", *AAAI-16 Workshop on AI, Ethics and Society*, pp.98-99.

효과적으로 해결하는 방법에 중점을 둘 필요가 있다.

(4) AI 시스템의 안전과 보안의 보장

AI 시스템이 실생활에 널리 보급되기 위해서는, 기존의 시스템이 통제된 방식처럼 AI 시스템 역시 안전하게 작동한다는 보장이 필요하다. 신뢰할 수 있고, 믿을 수 있는 AI 시스템을 만드는데 있어서 이 과제를 해결하기 위한 연구가 필요하다. AI 시스템은 다른 복잡한 시스템과 마찬가지로 다음과 같은 이유로 중요한 안전 및 보안 문제를 직면한다.[23)]

①복잡하고 불확실한 환경 : 많은 경우에 AI 시스템은 철저한 검사 또는 테스트가 불가능한 다수의 잠재적인 상태와 함께 복잡한 환경에서 작동하도록 설계되었다. 시스템은 설계 중에 고려되지 않은 조건에 직면할 수도 있다.

② 긴급 행동 : 배치 후 학습하는 AI 시스템을 위하여 시스템 행동은 감독되지 않은 조건에서 학습 기간에 따라 주로 결정될 수 있다. 이러한 조건 하에서는 시스템의 동작을 예측하기 어려울 수도 있다

③ 목표 부적합 : 인간의 목표를 컴퓨터 명령어로 전환시키기 어렵기 때문에, AI 시스템용으로 프로그래밍된 목표는 프로그래머가 의도한 목표와 일치하지 않을 수도 있다.

④ 인간-기계 상호 작용 : 많은 경우 AI 시스템의 성능은 인간 상호 작용에 의해 크게 영향을 받는다. 이러한 경우에 인간 반응의 변화는 시스템의 안전성에 영향을 줄 수도 있다.[24)]

다양한 쟁점을 해결하기 위하여 설명 가능성, 투명성, 신뢰, 검증, 승인, 공격에 대한 보안, 장기 AI 안전성 및 가치 조정 등을 포함하여 AI 안전과 보안[25)]을 향상시키기 위하여 추가 투자가 필요할 것이다.

---

23) J. Bornstein(2015), “DoD Autonomy Roadmap - Autonomy Community of Interest,” *Presentation at NDIA 16th Annual Science & Engineering Technology Conference*

24) J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods(2013), “The Seven Deadly Myths of Autonomous Systems," *IEEE Intelligent Systems*, 28, no. 3, pp.54-61.

25) See, for instance: D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mane(2016), "Concrete Problems in AI Safety,"; S. Russell, D. Dewey, M. Tegmark(2016), "Research Priorities for Robust and Beneficial Artificial Intelligence,"; T. G. Dietterich, E. J.

2. 백악관 국가과학기술위원회 “인공지능의 미래를 위한 준비”

(1) 개요

인공지능의 발전은 건강・교육・환경 등의 분야에서 새로운 시장과 기회를 가져옴과 동시에 현대사회에서 발생한 여러 가지 어려운 문제를 풀 수 있는 잠재력 또한 보유하고 있다. 예를 들어 스마트 차량은 교통사고 사망률을 감소시키고 노인과 장애인과 같이 사회적 약자의 이동을 지금보다 수월하게 할 수 있다. 또한 스마트 건물은 에너지 소비와 탄소 배출량을 감소시킬 수 있고, 정밀 의약품의 경우 인간의 생명연장과 삶의 질 향상을 가져올 수 있다.

인공지능은 향후 20년간 인간의 지능을 뛰어넘을 수 없을 것으로 판단되어진다. 그러나 언젠가는 인간과 대등해질 수 있을 것이고 여러 분야에서는 인간의 능력을 뛰어넘을 것으로 예측되거나 넘어선 부분도 있다. 그렇기 때문에 이 보고서에서는 인공지능의 현황과 미래 활용방안, AI가 사회 및 공공정책 등에 끼칠 영향에 대해 검토하고자 한다.

(2) 인공지능의 현재

전략 시뮬레이션, 번역, 자율주행, 영상인식 분야 등에 적용되는 약한(또는 특정한 목적을 가진) 인공지능은 그동안 비약적으로 발전해왔다. 약한 인공지능은 여행 계획, 소비자 추천 시스템, 타겟 광고 등과 같은 다양한 상업 서비스를 가능하도록 하였고, 최근에는 의학, 교육, 과학 연구분야에서 더욱 중요한 응용기회를 모색하고 있다.

최근 이슈가 되고 있는 범용인공지능(General AI 또는 Artificial General Intellegence, AGI)은 인간과 동일한 인지・사고능력으로 지능적 행동을 구

Horvitz(2015), "Rise of Concerns about AI: Reflections and Directions," *Communications of the ACM*, Vol. 58 No. 10; K. S.; R. Yampolsky(2014), "Responses to catastrophic AGI risk: a survey," *Physica Scripta*, 90 (1).

사하는 미래 인공지능 시스템을 언급하는 개념이다.

이와 관련하여 NSTC와 의견을 같이하는 민간 전문가 단체에선 당분간 범용인공지능이 실현되는 것은 어려울 것이라고 판단하기도 하며, 한편으로는 온라인상에서의 데이터 크롤링(Date Crawling)을 통한 정보를 자가 학습하여 '지능 폭발'이 일어나 빠르게 인간의 지능수준을 넘어설 것이라 판단하기도 한다.

NSTC는 슈퍼지능을 가진 범용인공지능의 미래에 대한 우려가, 현재 AI에 대한 정책에 영향을 끼쳐서는 안 된다고 판단한다. 이를 위해서는 저위험도의 문제점을 해결하는 것이 바람직한 접근방식으로 판단하며, 기술적인 문제뿐만 아니라 사회와 윤리적 문제에 대한 고려를 통해 장기적인 관점으로 바라보아야 할 것이다.

(3) 연방정부의 인공지능

행정부는 인공지능의 경제·사회적 혜택을 극대화하고 혁신을 촉진할 수 있는 정책과 내부절차를 개발하고 있는데, 다음과 같은 내용을 포함하고 있다.

① 기초 및 응용 연구개발에 대한 투자
② 정부가 우선하여 인공지능 기술과 응용사례를 도입
③ 현실적인 환경하의 테스트 베드를 구축하고 프로젝트를 지원
④ 대중이 사용가능한 데이터 세트 구축
⑤ 대중을 보호하는 동시에 혁신을 가능케 하는 정책적·법적·규제적 환경 구축 등

(4) 인공지능과 규제

인공지능은 자동차 및 항공기와 같은 제품에 응용될 수 있다. 이러한 생산물은 이용자들을 보호하고 공정한 경쟁을 보장하는 규칙들의 적용을 받

는다. 이러한 제품에 인공지능기술을 적용할 경우, 관련법에 어떠한 영향을 줄 것인가?

일반적으로 사회안전을 보호하기 위한 인공지능제품에 대한 규제의 접근방식은, 인공기능기술이 추가될 경우 증감되는 위험요소를 평가하여 분석하는 방법으로 이루어진다. 만일 관련 위험이 기존의 법적 규제범위에 포함이 될 경우에는 해당 위험에 대한 조치가 이루어지는지 또는 인공지능의 추가 활성화에 대한 적응여부를 고려하는 정책적인 논의가 필요하게 된다. 또한 인공지능 추가 활성화에 따른 대응단계에서, 법을 준수하기 위한 비용의 증가나, 개발·혜택을 지연시키는 현상이 발생할 경우에, 정책의 입안자는 안전과 시장의 공정성에 악영향을 발생시키지 않는 방향으로 대책을 마련하고 조정방안을 고려해야 할 것이다.

RFI 평가자들은 인공지능 연구나 관행에 대환 광범위한 규제는 아직 권장되지 않는다는 일치된 의견을 제시하였다. 대신 현존하는 규정의 목표와 구조로 충분하지만 인공지능기술의 영향을 설명하기 위해 확실한 조정을 하는 작업이 필요하다고 제시했다. 예를 들어, 자율 차량의 구현을 예측하기 위해 현재의 차량 안전규정 범위 내에서 적절한 변화가 이뤄져야 한다고 제안했다. 이런 과정에서 기관들은 인공지능의 혁신과 성장을 위한 기회를 만드는 동시에 근본적인 목적과 목표에 중점을 둬야 한다.

인공지능과 같은 기술의 효과적인 규제는 기관이 규제적 의사 결정을 하는데 도용을 줄 내부 전문기술 지식을 필요로 한다. 규제부서 및 기판 그리고 프로세스의 모든 단계에서 책임자급 전문가의 참여가 필요하다.

자율주행차량 및 무인 항공기 시스템(UAS 또는 '무인항공기)과 관련한 미국 교통부(DOT)의 작업은 인공지능 기반 제품을 위해 과거 규정을 개정하려는 기관들의 노력을 보여주는 적절한 예이다. 교통부 내에서 자동화 차량은 NHTSA(National Highway Traffic Safety Administration), 항공기는 FAA (Federal Aviation Administration)에 의해 규제된다.

(5) 현행 규제의 적용

영공과 고속도로에 적용되는 규제방식은 다르지만 자율차량과 항공기에 대한 규제 접근온 공통적인 목표를 가지고 있다. FAA와 NHTSA는 혁신을 촉진하고 안전을 보장하는 민첩하고 유연한 체계를 수립학기 위해 노력 중이다.

2016년 8월 29일 발표된 FAA의 Part 107(“Small UAS“ 최종 규정) 규정은 무인항공기를 영공에 안전하게 통합하기 위한 중요한 단계였다. 이 규정은 최초로 55파운드이하 무인항공기가 레저 목적이 아닌 비행을 하도록 허용했다. 이 규정은 라이선스 소지 이용자가 시내에서 400피트 이하의 고도를 유지하며 낮에만 비행을 하도록 제한한다. 경험과 데이터가 쌓이면 이러한 규제 제한을 안전하게 완화하기 위환 후속 규제가 뒤따를 것이다. 특히 교통부는 현제 특정 유형의 “소형 UAS”가 사람들 머리 위로 운행할 수 있도록 허용하는 체계를 제안하기 위해 ‘발의 예정 법안 고지’를 준비 중이며, 향후 확장된 활용 규칙이 나올 것으로 예상된다.

FAA는 아직 완전 자동 비행을 허용하는 규정을 공개적으로 발표하지는 않았다. 공역에 자율 항공기를 안전하게 도입하는 것은 복합한 과정이지만, FAA는 자율 항공기와 조종사가 탑승한 항공기가 완벽하게 통합되어 공역 시스템에서 동시에 비행하는 것이 먼 미래의 기술은 아니며, 그러한 날이 그리 멀지 않다고 보고 대비를 하는 중이다.

항공교통 관리에 대한 새로운 접근법에는 인공지능 기반 항공교통관제시스템의 항상도 포함된다. 현재 공항/지상간 통합의 한계와 공중/지상 간 사람들의 소통에 의존하는 구조 탓에 공역 판리 구조 안에서 미래 항공 교통 운영의 다양성을 예측하는 것은 현실적으로 불가능하다. 2007년 미국 내 비행 지연으로 인한 피해는 312억달러로 추산된다. 일부 비행 지연은 기상 악화 및 기타 제약으로 인해 불가피한 경우들이다. 새로운 항공 기술의 채택 및 정책의 적용, 인프라 업그레이드를 통해 미국 영공 운영의 효율성을 크게 높일 수 있다. 이러한 해결책은 인공지능 및 기계 학습 기반 구축을 통해 조종사 및 비(非)조종사 항공기를 포함해 더욱 넓은 범위의 영공 이용자

들을 더욱 잘 수용하고, 안전을 침해하지 않으면서 영공을 이용할 후 있게 해준다. 이러한 기술의 개발과 적용은 안전을 증가시키고 비용을 줄이면서 영공 이용자 및 서비스 제공자를 위한 국제 경쟁력을 확보하는데 도움이 될 것이다.

(6) 정의, 공정성 및 책임

인공지능이 널리 채택됨에 따라 의도하지 않은 결과에 대해 우려가 있고, 이는 2014년에 발표한 빅 데이터 보고서나 빅 데이터와 프라이버시에서 제기한 우려와 궤를 같이하는 점이다. 인공지능과 실제 세계의 장치를 제어하게 되면서 안전에 대한 검토도 필요하다. 인공지능의 주요 도전 중 하나가 '닫힌 세계'에서 검증된 연구 결과가 외부의 '열린 세계'에 적용하게 되면 예측할 수 없는 일이 일어나기 때문이다. 과거 안전에 크리티컬한 시스템 개발자들의 경험을 통해 검증과 확인 방식이 인공지능 실행자들에게 학습될 수 있게 만드는 것이 필요하다.

윤리적 문제를 해결하기 위해 인공지능 실무자와 학생을 대상으로, 윤리 교육을 위해 교과과정과 토의를 실시한다. 그러나 교육만으로는 근본적인 해결책이 될 수 없다. 그렇기 때문에 인공지능에 대한 연구 및 응용수준이 지속적으로 높아지고, 장기적인 사회 및 윤리적 문제를 고려한 발전을 모색해야 할 것이다.

이를 위해 인공지능 전문가와 학생들을 위해 윤리적 훈련은 필수적이며, 윤리 훈련은 시스템이 구축되고 시험이 이루어지는데 따라 기술적 예방조치를 취함으로써 선의의 의지가 실무에 접목될 수 있도록 하는 기술적 능력과 함께 강화가 필요하다.

또한 윤리적 훈련 역시 필요하며 단지 윤리학을 가르치는 것이 아니라 기술 도구나 방식이 강화된 훈련 방안이 마련되어야 한다.

제2절 유럽연합의 윤리 정책 및 법제 동향 분석

1. 유럽로봇연구네트워크 “로봇윤리 로드맵”

유럽위원회(European Commission)는 2005년 제6차 다자간 공동기술개발연구프로그램(FP6)을 통해 ‘윤리로봇(Ethicbots)’ 프로젝트를 4년간 시행하였다. 이를 토대도 유럽로봇연구네트워크(European Robotics Research Network, EURON)는 2007년 로봇 영역을 세분화하여 각 영역별로 이익과 폐해를 분석한 ‘로봇윤리 로드맵’을 발표하였다.

로봇윤리 로드맵은, 로봇이 윤리적, 사회적, 경제적 문제를 발생시킬 것이라고 하면서 핵물리학이나 바이오 공학처럼 윤리적 기준에 대한 철저한 검토가 필요할 것이라고 밝히고 있다.

본 로봇윤리 로드맵에서는 로봇윤리의 영역을 휴머노이드(Humanoids), 진화된 생산 시스템(Advanced production systems), 지능형 홈(Adaptive robot servants and intelligent homes), 네트워크 로봇(Network Robotics), 현장형 로봇(Outdoor robotics), 건강과 복지를 위한 로봇(Health care and life quality), 군사용 로봇(Military robotics), 교육용 로봇(Edutainment)의 8가지로 구체적으로 분류하였다. 이러한 분류의 목적은 로봇과 관련한 윤리적인 논제들의 체계적인 평가를 제공하고, 당면한 문제들에 대한 이해를 높여 더 많은 연구를 촉진하는 것이다.

그러나 이러한 시도는 오래 지속되지 못하였다. 당시의 사회적 공감대 형성과 이해도 부족으로 인해 본 로드맵은 학술적인 권고안으로서의 역할만 수행하였다. 하지만 로봇의 윤리문제에 대하여 국제적으로 만들어낸 첫 번째 결과물이라는 점에서 의의가 있다.

2. 유럽에서의 기술 규제 : 로봇에 대한 법과 윤리의 당면과제(약칭 : LoboLaw Project)

2012년 3월 1일부터 27개월간 시작된 EU의 “유럽에서의 기술 규제 : 로봇에 대한 법과 윤리의 당면과제(이하,「RoboLaw Project」)는, 떠오르는 로봇 기술의 법적 및 윤리적 함의를 이해하는 것을 목적으로 한다.

유럽에서는 새로운 기술에 대한 규제의 문제는 모든 법률 시스템에 의해 다루어지고 있기 때문에, 법과 과학 및 법과 기술 간의 관계에 대한 많은 연구에 의존하게 된다. 그럼에도 불구하고,「RoboLaw Project」는 로봇 공학, 나노 기술, neuroprostheses, 뇌-컴퓨터 인터페이스 등 지금까지는 다루어지지 않았지만, 합법적인 “상태“에 직면 한 진보된 기술의 극단적인 경계에 중점을 두고 있다. 이런 종류의 연구에는 학제적 연구와 법체계가 다른 나라에서 이를 어떻게 다루고 있는지 비교하는 연구 등이 필요하다. 전 세계적으로 여러 연구 기관은 로봇 공학의 발전에 대한 규제와 법적 결과에 대하여 조사하였다. 그러나 지금까지의 유럽이나 외국에서의 “RoboLaw”에 대한 시각은 여전히 매우 단편적이었다. 이미 1980년대 선도적인 법학자들은 로봇 같은 인공지능기계의 개발이 기존 법적 프레임워크의 적응 또는 확장을 필요로 하는지 여부에 관하여 탐색한 바 있다. 대부분의 이러한 작업은 소프트웨어 시스템에서 에이전트 기술과 관련된 것이었다. 그러나 당시의 로봇은 실제 사실 이라기보다는 공상과학소설속이 더 많았기 때문에, 이러한 연구의 대부분은 본질적으로 스케치에 불과하다. 그들은 (미래의) 로봇에 관하여 법이 다루어야 할 중요한 법적 쟁점을 알려주었지만, 구체적으로 법체계가 채택하여야 할 구체적인 방법 또는 규칙을 제공하지 못했다. 이 프로젝트는 전에 별개로 조사되었던 많은 법적 쟁점을 결합하고, 진보되어가는 로봇이 현실화되어가는 시대에 RoboLaw의 요구 사항 및 규제 프레임 워크에 관한 첫 번째 심층 조사이다. 또한, 유럽연합 내에서 특정한 법적 시스템 내에서 로봇 개발의 법적, 윤리적 결과를 탐구하고, 이를 미국과 극동 지역 중에서도 특히 일본과의 비교한 최초의 연구이다.「RoboLaw

Project」의 주요 목표는 로봇 및 법의 다양한 측면에 대하여 종합적이고 포괄적인 연구를 수행하고, 유럽에서 RoboLaw에 대한 기반을 마련하는 것이다. 기존의 법체계에서 이러한 신기술을 규제를 위해 적용할 법이 없는 경우, 현존하는 법적시스템에서 직면하게 되는 문제가 발생한다. 따라서 본 연구를 통하여 현행 법적 시스템의 틀에서의 적용가능성을 검토하고, 현재의 수단과 범주의 사용을 통해 가능한 해결책을 도출하는 것이다.

이 보고서에서는 떠오르는 로봇기술에 대한 미국과 동양적인 관점(특히 일본, 중국)을 검토한 후, 유럽인권헌장과 같은 핵심적인 "유럽적 가치(European values)"으로 특징지워지는 유럽식 모델을 개발하는 것을 목적으로 한다. 그리고 연구의 최종 결과는 규제 지침 (D6.2 "로봇 공학을 규율하는 가이드라인(Guidelines on Regulating Robotics)으로 도출하여, 미래의 로봇 개발을 위해 기술적으로 실현가능하면서도 윤리적·합법적으로 건전한 기반을 조성하기 위해 노력하였다.

3. 2014년 "로봇규제 가이드라인"

앞서 본 「RoboLaw Project」를 통하여 본 로봇규제 가이드라인(Guidelines on Regulating Robotics)을 제정하였다. 본 가이드라인에는 자율주행 자동차, 수술로봇, 로봇인공기관, 돌봄 로봇 등 4가지 분야에 대하여 (ⅰ) 로봇공학의 적용에 의해 제기되는 윤리 및 법적 쟁점에 대한 심도있는 분석을 제공하고 (ⅱ) 유럽 및 각국의 규제당국에게 이러한 쟁점에 대한 가이드라인을 제공하는 것을 목적으로 한다.[26]

본 가이드라인은 각 분야에 대한 소개, 각 기술에 대한 첨단기술의 개요, 윤리적·법적 분석과 정책적 제안을 기술하고 있다. 이하에서는 법률적인 분석과 윤리적 분석을 중심으로 살펴보고자 한다.

---

26) Guidelines on Regulating Robotics, Grant Agreement number 289092, Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics, 2014.9.22. <http://www.robolaw.eu>

(1) 윤리적 분석

본 가이드라인에서는 로봇 기술이 발전하는데 있어 윤리적 접근이 중요함을 강조하고 있다.

윤리적 접근은 새로운 형태로 발전되는 기술로 인하여 발생할 수 있는 사회적 문제와 기본권 보호 우려에 대한 해결책 및 로봇규제의 기준이 되는 원칙을 제시할 수 있다. 또한 기술의 변화와 관련된 가치를 평가할 때, 본질적으로 인간 능력의 발전에 긍정적으로 기여하도록 하는 개발과 인간 능력의 발전에 부정적인 영향을 미치거나 단지 인류 번영에만 기여하는 존엄성이 없는 개발을 구별할 수 있게 된다.

인권은 과학과 기술의 진보를 촉진하고 알맞은 사용을 위해 필수적으로 고려되어야 한다. 로봇의 적용은 평등, 연대, 정의 등의 핵심적인 가치에 기반하여야 하고, 비차별의 원칙, 노령자, 장애인의 기본권, 의료서비스 접근권, 소비자 보호 등의 가치를 보호하여 신체기능 회복 및 인간의 역량을 강화할 수 있도록 하여야 한다. 특히 수술로봇과 같은 고가의 장비는 특정 계층만 접근이 가능한 측면이 있는데, 누구나 이러한 서비스를 받을 수 있도록 평등, 윤리, 삶의 질 향상과 같은 가치관이 적극 개입되어야 한다고 본다. 또한 로봇 규제 가이드라인은 책임 있는 연구와 혁신을 지향하면서, 윤리적, 법률적, 기술적으로 다양한 관점에서의 논의가 필요하다고 강조한다.

(2) 법률적 검토

로봇규제 가이드라인에서 고려하는 법적 이슈로는 1) 건강, 안전, 소비자, 환경규제, 2) 법적 책임, 3) 지식재산권, 4) 개인정보보호 및 데이터 보호, 5) 법적 거래능력 여부 등이 있다.

로봇은 병원, 가정, 상점, 길거리 산업 등 그리고 보다 전문적인 환경에서 사용될 수 있도록 설계되어가고 있기 때문에, 새로운 환경에서 발생하는 건강 및 안전 문제에 대처하기 위한 새로운 논의가 전개되어야 할 것이다. 통제되고 잘 구조화된 환경에서 적용되는 산업용 로봇과 달리 서비스 로봇은

종종 구체적인 훈련을 받지 않은 사람들에 의해 사용되며, 광범위한 일을 수행하게 된다. 이로 인해 발생하는 안전 위험들과 이용자 훈련 수준의 차이는 안전에 요구되는 사항과 로봇의 설계에 영향을 미칠 것이다. 또한 로봇의 종류에 따라 소비자 관련 규정 및 쓰레기 처리 등에 관한 규정과 연관될 수 있다. 로봇 종류에 따라 발생하는 규정의 충돌여부 및 법률 적용의 공백 등에 관하여도 연구가 이루어질 필요가 있다.

로봇의 법적책임에 관하여 기존의 법제도 하에서는 로봇이 작위 또는 부작위로 제3자에게 피해를 입힌 경우 로봇 기술의 제조자, 소유자 또는 사용자 등이 손해배상책임을 부담하게 된다. 그런데 손해에 대한 책임 소재가 불분명하고, 인간의 행위와 로봇이 발생하게 한 피해 사이에 인과관계를 입증하는 것이 어렵다. 또한 학습을 통해 로봇의 자율적 판단이 가능해지면서 예측하지 못한 창발적 요인으로 전통적인 과실책임의 원칙을 적용하기 어려운 한계가 있다. 이러한 법적책임 문제에 대하여 제조자, 소유자, 사용자 및 제3자의 이익을 고려하여 신중하게 접근하여야 한다.

또한 로봇의 발명과 콘텐츠 등을 특허권, 상표권, 저작권 등 지식재산권으로 보호받을 수 있다. 로봇에 적용되는 법적 규정은 없지만, 기존의 법제도를 로봇에도 적용할 수 있다. 그럼에도 불구하고, 거기에 지식재산권 보호를 확장하거나 또는 축소할 공공정책상의 이유가 있을 수 있으며, 지식재산권 부여가 로봇 산업과 사회에 어떤 영향을 미치는지에 대한 추가적 연구가 필요하다. 지식재산권과 관련하여 로봇이 창작한 저작물을 보호하는 것에 부정정인 국가와 달리 영국은 컴퓨터 또는 로봇이 생성한 작품에 대하여 적극적으로 보호를 검토하고 있다.

로봇의 개인정보 및 데이터 보호 이슈와 관련하여 민감한 개인정보 등을 다루는 로봇에 보안 시스템 설계 등의 조치를 할 필요가 있다. 로봇이 대량의 정보를 보유하고 데이터를 처리하게 되면서 이에 대한 보안 및 데이터 보호와 관련된 규정의 준수가 요구된다. 그러나 현재 국가 간 개인정보 및 데이터 보호에 관련한 프레임워크에 차이가 존재하는 한계가 있다.

제3절 중국의 윤리 정책 및 법제 동향 분석

1. 개요

중국에서 로봇 연구개발은 20세기인 1970년대에 시작되었고, 최근 들어 일련의 육성정책과 시장수요에 따라 로봇산업이 빠른 속도로 발전 중이다. 2014년 중국 국내 상표의 공업로봇이 1만7천대 생산된 바 있으며, 2013년 대비 78%의 성장률을 보인 바 있다. 서비스형 로봇은 과학실험, 의료요양, 교육오락, 가사서비스 등 영역에서 이미 연구되어 일련의 대표성 상품들이 생산되고 있다. 2013년부터 중국은 세계 제1의 공업로봇응용시장이 된 바 있으며, 2014년에는 5만7천대가 판매되었고, 이는 전년대비 56%의 성장률에 이르는 것인 한편, 전세계 소비량의 1/4에 달한다.

중국은 2016년 「로봇산업발전규획(2016-2020) (机器人产业发展规划(2016-2020年)」을 제정하면서 명확한 목표를 설정한 바 있다. 즉, 중국은 전국적인 로봇열풍을 불러일으키기 위하여 정책적으로 각종 육성 계획을 마련하고 있는 것이다.

특히, 중국 정부는 로봇과 관련된 기술, 인재, 자금, 시장조성 등의 다양한 정책 요소를 구현하기 위하여 로봇산업발전규획 외에도 「스마트제조 발전규획(2016-2020)(智能制造发展规划(2016—2020年))」 및 「공업로봇 분야 규범 요건(工业机器人行业规范条件)」, 「로봇산업의 건강한 발전 촉진에 관한 통지(关于促进机器人产业健康发展的通知)」 등의 관련 정책을 마련함으로써 로봇산업을 국가 핵심 전략 산업으로 삼으려는 노력을 기울이고 있다.

2. 로봇산업발전규획

(1) 개관

로봇은 선진제조업의 중요한 장비들을 모두 갖추고 있어야 하며, 인류의

생활방식을 개선할 수 있는 중요한 시작점이기도 하다. 제조환경에서 응용되는 공업로봇 뿐만 아니라, 비제조업환경에서 응용되는 서비스형 로봇의 경우에도, 그 연구개발과 산업적 응용에서는 국가 과학기술의 창조혁신, 첨단기술제조 발전 수준의 중요한 지표로서 작용하기도 한다. 로봇산업의 발전에 주력하는 것은 중국 제조업의 새로운 국면을 창출하고, 공업수준을 한 차원 높은 단계로 끌어 올릴 것이며, 제조 강국 건설을 촉진하는 한편, 국민생활 수준을 개선하는 중요한 의미를 가지게 된다.

「중국제조2025(中国制造2025)」 정책을 관철하기 위하여 로봇을 전체 부처가 중점적으로 발전시켜야 하며, 중국 로봇산업의 건강하고 지속가능한 발전을 추진하기 위하여 본 규획을 특별히 제정하고, 그 기간은 2016-2020년으로 한다.

1954년 첫 번째 로봇이 탄생한 이래, 세계적으로 공업이 발달한 국가들에서는 이미 로봇산업 시스템을 완비해나가고 있으며, 핵심기술과 상품응용 영역에서 선두를 달리고 있으며, 몇 몇 로봇산업을 주도적으로 이끌고 있는 걸출한 기업들도 출현한 바 있다. 특별히 국제금융위기 이후, 이러한 국가들은 로봇개발을 국가 전략으로 이끌면서 지속적으로 선두를 지키기 위하여 노력하고 있다. 최근 5년간, 글로벌 공업로봇의 판매량은 연평균 17% 이상의 성장을 보이고 있으며, 2014년에는 22만9천대에 이르고 있고, 이는 전년대비 29%나 성장한 상황이다. 글로벌로봇제조밀도(인구1만명당 공업로봇 사용량) 평균은 5년전보다 50에서 66으로 늘어났으며, 그중에서 공업선진국 로봇밀도는 200을 초과하고 있다. 동시에 서비스형 로봇도 빠르게 발전하고 있으며, 그 응용범위도 광범위해지고 있다. 수술로봇은 대표적인 의료기기로서 비교적 큰 산업 규모로 성장해가고 있고, 공간로봇, 생화학방어로봇, 반폭탄테러로봇 등 특수목적 로봇도 그 응용이 실현되어 가고 있다.

중국에서 로봇 연구개발은 20세기인 1070년대에 시작되었고, 최근 들어 일련의 육성정책과 시장수요에 따라 로봇산업이 빠른 속도로 발전 중이다. 2014년 중국 국내상표의 공업로봇이 1만7천대 생산된 바 있으며, 2013년 대비 78%의 성장률을 보인 바 있다. 서비스형 로봇은 과학실험, 의료요양, 교

육오락, 가사서비스 등 영역에서 이미 연구되어 일련의 대표성 상품들이 생산되고 있다. 2013년부터 중국은 세계 제1의 공업로봇응용시장이 된 바 있으며, 2014년에는 5만7천대가 판매되었고, 이는 전년대비 56%의 성장률에 이르는 것인 한편, 전 세계 소비량의 1/4에 달한다. 로봇밀도도 5년 전에 비하여 11에서 36 증가한 바 있다.

비록 중국 로봇산업이 이미 상당한 수준으로 발전하고 있긴 하나, 공업선진국과 비교한다면 여전히 큰 차이를 보이고 있는 것이 현실이다. 대표적으로, 로봇산업 내 산업사슬이 제대로 구축되어 있지 않으며, 부속품 중 고정밀감속기나 제어기 등은 수입에 의존하고 있는 실정이다. 또한, 핵심기술의 창조혁신 능력이 약하고, 첨단기술상품의 품질이 비교적 낮으며, 로봇의 응용확대가 어렵고, 시장점유율 제고력이 낮다. 기업은 '小`散`弱(작고, 분산되어 있으며, 약하다)'의 문제점을 드러내고 있으며, 산업경쟁력이 낮고, 로봇 관련 표준과 검사/인증기준도 완비되어 있지 않다.

현재, 중국 노동력 원가가 빠른 속도로 높아지고 있어 인구수에서 오는 잇점도 점차 없어지고 있다. 생산방식을 유연성과 스마트, 정밀세공으로 전환하기 위하여서는 스마트제조를 근본적인 특징으로 하는 신형제조시스템으로 전환해야 하기 때문에, 공업로봇의 수요는 지금보다도 대폭 증가할 것이다. 동시에, 고령화 사회 서비스와 의료요양, 재난구호, 공공안전, 교육오락, 과학연구 등의 영역에서 서비스형 로봇의 수요는 빠른 속도로 늘고 있다. "13.5 규획" 기간은 우리나라 로봇산업발전의 중요한 시기로서, 국제 로봇산업의 발전 추세를 파악하여야 하며, 자원의 합리적인 배분과 정책 수행을 통하여 기회를 잡아야 한다. 또한, 양호한 발전환경을 조성하고 중국 로봇산업의 지속적이고 건강한 발전을 실현할 수 있도록 촉진해나가야 할 것이다.

(2) 종합적 정책사항

(가) 지도사상

중국 공산당 제18대 제19기 3중, 4중, 5중 전회의 정신을 이어받아 창조혁신, 협력, 녹색, 개방, 공영 발전 이념을 견지함으로써 「중국제조2025」의 실시를 가속화하며, 중국 경제의 전환과 사회발전의 수요에 따라 “시장주도, 창조혁신동력, 기초강화, 품질우선(市场主导、创新驱动、强化基础、质量为先)” 원칙과 “13.5 규획” 기간 동안의 “2대 돌파”, “3대 제고”에 집중하여 로봇 주요 부품 및 첨단기술상품의 생산을 중대한 돌파구로 삼는다. 로봇품질의 신뢰성과 시장점유율 및 로봇선두기업의 경쟁력을 대폭 높이기 위하여, 기업이 주체가 되어 산·관·학 협력 방식의 창조혁신을 이루어내며, 로봇산업 생태계의 경쟁력을 확보하고, 중국 특색의 로봇산업 시스템을 구축하여 강국 건설의 기초로 삼고 있다.

시장의 수요에 따라 시장이 주도하도록 하고, 기업이 주체가 되어 로봇 연구개발 방향을 발휘할 수 있도록 하며, 창조·혁신적인 시스템을 구축하여 로봇산업 발전에 유리하도록 한다. 상업화와 서비스 방식을 개선하여 공공의 창조혁신플랫폼을 구축하도록 한다. 로봇에 대한 공공적 성격의 기초연구를 강화하고, 로봇 관련 표준체계 및 검사/인증 플랫폼을 구축하며, 산업발전의 기초로 삼는다. 품질을 우선으로 하여 로봇 관련 부품과 첨단기술상품의 품질 신뢰성을 높이고, 핵심기술의 국산화를 통하여 경쟁력을 높이는 것을 목표로 한다.

(나) 발전 목표

5년간의 노력을 통하여 로봇산업 체계를 완비한다. 기술 창조혁신능력과 국제 경쟁력을 강화함으로써 상품성능과 품질을 국제수준으로 향상시키며, 관련 부품을 국산화하여 시장수요를 만족시킬 수 있는 수준으로 발전시킨다. 2020년의 구체적인 목표는 다음과 같다.

① 산업규모의 지속적 성장

국산 로봇 연간 생산량을 10만대로 하고, 공업로봇 연간 생산량도 5만대 이상으로 한다. 서비스형 로봇의 연간 판매 수입을 300억위안 이상으로 하고, 양로 및 장애 보조, 의료요양 등 영역에서의 소매 판매를 촉진한다. 국제경쟁력을 갖춘 선두기업을 3개 이상 육성하고, 5개 이상의 로봇산업 클러스터를 조성한다.

② 기술수준의 현저한 상승

공업로봇의 속도, 적재, 정밀도, 자체중량비율 등 주요기술 지표를 해외의 동일류 상품 수준으로 달성하고, 무고장 평균시간을 8만시간에 이르게 한다. 의료요양, 가사서비스, 반폭탄테러, 재난구호, 과학연구 등의 영역에서 서비스형 로봇 기술을 국제표준에 이르게 한다. 새로운 버전의 로봇기술을 획득하여 지능로봇의 창조혁신적인 응용을 실현하도록 한다.

③ 관련 부품의 국산화 달성

로봇의 정밀감속기, 전기구동기, 제어기의 성능과 정밀도, 신뢰성을 국제기준으로 높이고, 공업로봇의 판매량을 늘여 시장점유율을 50%에 이르도록 한다.

④ 집중적인 성과 도출

30개 이상의 전형적인 로봇 종합 응용 솔루션을 완성하고, 상응하는 표준과 규범을 제정하는 한편, 로봇의 중점산업에서의 규모화를 실현할 수 있도록 하여, 로봇밀도를 150이상으로 달성한다.

3. 인공지능과 인간의 역할에 대한 담론

(1) 인간의 두려움과 인공지능의 발전

2016년 12월 30일, 중국의 대표적인 로봇 관련 정보 홈페이지인 “中国机器人网 (http://www.robot-china.com/zhuanti/show-3015.html)”에서는 “AI는 더욱 많은 업무를 대신할 것이나, 주도권은 아직 인류의 손안에 있다(AI想要取代更多工作 主动权还掌握在人类手中)”는 제목의 글을 담론의 주제로 논한 바 있다.

즉, 외신(Venture Beat)을 인용하여, 인공지능에 대한 Gold Rush 열풍은 점점 더 강해 질 것으로 예측하는 한편, AI 연구 속도가 인간의 두려움을 불러일으킬 정도라고 밝히면서 AI에 대한 부정적인 견해들이 있음을 소개하고 있다.

그러나 결론적으로는 업무상 과실 비율로 판단하건데 인간의 실수가 AI의 실수보다 확률이 훨씬 많다는 점을 들면서, AI는 인류의 생활을 편리하고 안전하게 하는데 더욱 도움이 된다는 논지를 펼치고 있다.

다만, Uber 택시의 사례를 들어, Uber 자율주행차량의 경우 사고가 날 가능성을 번개를 맞을 확률과 비교하면서 안전성을 강조하면서도, Uber 자율주행택시의 상용화 과정에서는 인간 기사의 탑승 정책을 선택하였다는 사실을 적시하고 있다. 즉, 이러한 인간의 선택이 우리의 진보를 막는 장애요소가 됨을 강조하고 있다.

(2) 소설 쓰는 로봇의 등장

한편, 로봇 관련 정보 홈페이지인 “中国机器人网”은 2017년 1월 10일, “로봇이 소설을 쓰는 것은 누구를 놀라게 하는 것인가?(机器人写小说会吓倒谁？)”라는 제목의 기술적 담론을 펼친 바 있다.

특히, 알파고와 이세돌 간 “인간과 기계의 대결(人机大战)”을 화두로 던지면서, 일본 인공지능(AI)가 소설 창작을 하였다는 매체보도를 인용하고 있다. 즉, 일본 인공지능이 창작한 4편의 소설 중에서, 2편의 경우에는 인간이 일정한 설정값(등장인물, 줄거리)을 입력하였으나, 2편의 경우에는 순수

인공지능의 창작 소설임을 소개하고 있다. 문학 분야의 경우 인류의 정신활동이기 때문에 로봇이나 인공지능이 소설을 쓰는 것에 놀란 사람들이 많을 것이나, 실제로 로봇은 어마어마한 량의 소설 데이터를 가지고 있기 때문에 각 요소를 선별하고 조합하여 하나의 완전한 스토리를 가진 한편의 새로운 소설을 쓰는 것이 기술(technic)적으로는 전혀 문제가 되지 않음을 밝히고 있다.

다만, 인류의 감정이나 사상은 작가 자신의 독특한 생활 경험에서 비롯되는 것이지만 인공지능의 경우에는 이러한 것을 갖추고 있지 못하여 피가 있고 살이 있는 인물을 도출해내지 못하며, 한 시대의 풍모나 자질을 가진 작품을 창작해 낼 수는 없다고 강조하고 있다.

그리고 먼 훗날 로봇이 자기의 감정이나 사상을 가지는 완전한 독립체가 된다면 이는 다시 논의해 볼 일이라고 밝히면서, 그런 날이 오기 전까지는 놀랄 필요가 없다고 담론을 펼치고 있다.

제4절 일본의 윤리 정책 및 법제 동향 분석

1. 경제산업성 “로봇 신전략”

(1) 개요

일본은 저출산・고령화, 노동가능인구가 감소되는 상황에서 로봇 기술은 제조업의 생산 현장, 의료・간병 현장, 농업・건설・인프라 작업 현장 등 폭넓은 분야에서 인재 부족문제를 해소, 생산성 향상 등 사회과제 해결에 기여하고자 로봇 신전략을 수립하였다. 경제산업성은 2014년 6월에 각의 결정한「일본재흥전략(성장전략)」개정 2014에서 로봇 기술을 활용함으로써 생산성 향상, 기업의 이익 향상, 임금 상승 등을 구현시키겠다는 목표 설정하였다.

로봇 혁명의 실현을 위해 2014년 9월「로봇혁명실현회의」를 설치하고 기술개발, 규제개혁, 표준화 등 구체적인 방안을 검토하고 그 결과를「로봇 신전략」을 발표한 것이다. 본 전략의 액션플랜은 다음 3가지 핵심 요소로 구성되어 있다.

① 세계의 로봇 기술 혁신 거점 - 로봇 창출력의 근본적 강화

일본의 로봇을 철저히 강화하고 사회 변혁으로 이어지는 로봇을 잇달아 창출하는 거점으로 하도록 산・학・관의 제휴 및 사용자와 제조업체의 매칭 등의 기회를 늘리고 이노베이션을 유발시키는 체제를 구축 및 인재 육성, 차세대 기술 개발, 국제적 전개를 위한 규격화, 표준화 등을 추진한다.

② 세계 제일의 로봇 활용 사회 - 쇼케이스(로봇이 있는 생활의 실현)

중견・중소기업을 포함한 제조, 서비스, 간병・의료 인프라・재해 대응・건설, 농업 등 다양한 분야에서 진정으로 쓸 수 있는 로봇을 만들어 활용하기 위해 로봇의 개발, 도입을 전략적으로 추진함과 동시에 그 전제가

되는 로봇을 활용하기 위한 환경 정비를 실시한다.

③ 세계를 선도하는 로봇신시대에 대한 전략

IoT하에서 디지털 데이터가 고도로 활용되는 데이터 기반사회에서는 모든 사물이 네트워크를 통해서 연결되어 일상적으로 빅 데이터가 창출된다. 게다가 그 데이터 자체가 부가 가치의 원천이 된다. 이러한 사회의 도래에 의한 로봇신시대를 내다본 전략을 구축한다.

이를 위해서는 로봇이 서로 연결되어 데이터를 자율적인 축적・활용을 전제로 한 비즈니스를 추진하기 위한 규칙이나 국제 표준 획득을 추진해야 하는 것이 필요하다. 그 때는 로봇 새 시대의 가능성을 살리는 데 기반이 되는 보안이나 안전에 관한 규정・표준화 등이 필수적이다.

2020년까지 5년 동안 정부의 규제 개혁 등 제도환경 정비를 포함한 다각적인 정책적 계기를 최대한 활용함으로써 로봇 개발에 관한 민간 투자 확대를 도모하고, 1000억엔 규모의 로봇 프로젝트 추진을 목표로 한다. 또 그 때에는 로봇이 단순한 사람의 대체물로 기능하는 것이 아니라 “사람과 보완 관계에 있는, 사람이 보다 고부가 가치로 탈출할 수 있는 파트너로서의 로봇“을 활용한다는 관점을 내세우는 것이 중요하다.

(2) 규제개혁을 위한 방안

로봇을 사회에 도입하려는 경우, 그 로봇 등이 관련된 기존의 법령 등에 준수할 것이 요구된다. 그 법령 등이 사회 경제 정세나 기존의 법률 등에 비추어 어긋난다고 해석될 경우 법 개정, 특구의 인정, 새로운 법률 제정 등 규제의 정비가 요구된다.

예를 들면 전파법에서는 주파수 할당과 필요한 전파 출력의 확보 등 로봇의 활용을 위한 규칙 정비, 도로 교통 법・도로 운송 차량법에서는 로봇 기능을 차량의(역자 주 : 자율주행차량) 법령상의 위상 정리 등이 꼽힌다. 또 소비자 보호를 위한 많은 법령도 관계한다.

일본은 세계 굴지의 로봇 강국이며 일본의 기술적 우위성이 아직 남아 있지만, 미국・독일・중국 등 외국에 맹추격하고 있는 지금 일본 전국에서 로봇 도입・활용을 가속화하기 위해서는 향후 로봇의 활용을 전제로 한 규제 완화 및 규범 정비의 양면의 관점에서 균형 잡힌 규제 개혁을 추진하면서 동시에 소비자 보호의 관점에서 로봇의 안전성을 확보하기 위한 법 제정이 필수적이다.

지금까지도 일정 조건을 충족하면 울타리 안에서 사람과 로봇 협조 작업이 가능한 노동 안전 위생법의 80W규제 완화, 생활 지원 로봇의 국제 안전 규격인 ISO13482가 발효되고 국제 표준에 준거한 안전인증을 취득할 수 있는 체제를 정비하는 등 규제 완화 및 안전 기준 등의 규칙을 구축하기 위해서 많은 노력을 해왔다. 앞으로도 빠르게 진화하는 로봇을 일본 사회에서 효과적 활용, 세계에 자랑할 수 있는 로봇 활용의 쇼케이스가 될 수 있도록 과감하게 규제 제도 개혁을 추진한다.

2020년까지를 생각하면 각 분야에서 로봇의 도입에 의해서 생산성 향상을 도모할 수 있는 직접적인 영역을 식별하고 로봇이 능력을 발휘하는 환경(반정형적인 환경)를 정비할 필요가 있다. 그 일환으로서, 로봇 활용에 관한 규제 제도의 개혁은 이를 위한 중요한 과제이며 ①로봇을 효과적으로 활용하기 위한 규제 완화 및 새로운 법 체계・이용 환경의 정비, ②소비자 보호의 관점에서 필요한 법・체계 정비에 나누어 구체적으로 대응해야 할 과제와 액션 플랜을 다음과 같이 정리한다.

규제 제도 개혁을 추진할 때에는, 예를 들면 로봇에 관한 방폭 규격처럼 국내・외에서 상호 운용을 전제로 한 국제적인 규제 제도 조화를 추진해야 한다.

또 세계를 선도하는 새로운 로봇의 활용 가능성을 개화시키고자 “로봇 혁명 이니셔티브 협의회”를 중심으로 수시 과제를 정리하고 필요한 개혁을 추진한다. 그 때 정부 규제 개혁 회의와 제휴하고 관련된 제도를 포괄한 종합적인 개혁을 추진한다. 더욱이 정비된 환경 하에서 활용할 수 있는 로

봇의 하드웨어 기술, 소프트 기술의 연구 개발과 삼위일체로 추진하는 “로봇 배리어 프리” 사회 구축을 가속한다.

(가) 로봇을 효과적으로 활용하기 위한 규제 완화 및 새로운 법 체계・이용 환경의 정비

① 로봇의 활용을 지원하는 새로운 전파 이용 시스템의 정비 로봇의 조종(제어), 로봇으로부터의 영상 등의 데이터 전송, 로봇이 장애물 등을 감지하기 위한 센싱 등 로봇의 전파이용은 기존의 범용적인 전파 이용 형태와는 다르다. 이 때문에 새로운 전파 이용 시스템의 규정을 마련한다.

이 규정을 제정함에 있어서는 로봇의 이용 형태나 이용 환경 등을 감안하고 일본의 전파 이용 실태를 바탕으로 기존 무선 시스템과의 주파수 공용을 도모하는 것 등으로 로봇의 전파 이용에 적합한 주파수대역과 출력 등의 기술적 조건을 책정하는 것이 적당하다.

또 무인 항공기의 조종을 위한 전파이용에서는 국제적인 사용 주파수의 검토가 국제 민간 항공 기구(ICAO)나 국제 전기 통신 연합(ITU)에서 진행되고 있기 때문에 계속 국제적인 협조성에 대해서도 고려하는 것이 중요하다.

또한 이미 총무성에서는 “로봇용 전파 이용 시스템 조사 연구회“에서 이들 로봇의 활용을 위한 새로운 전파 이용 시스템의 환경 정비를 위한 검토를 추진하고 있다.

② 의약품, 의료기기 등 환자의 부담의 경감 등이 기대되는 최소한의 수술 등으로 정밀한 움직임을 가능한 수술 지원 로봇 등 로봇 기술을 활용한 것을 포함한 새 의료 기기 의약품 의료 기기 등 법에 기초하여 승인 심사의 신속화를 도모하고 새 의료 기기의 신청 승인까지 표준적인 총 심사 기간을 정기 심사 품목에 대해서는 14개월, 우선 심사 품목에 대해서는 10월의 것을 목표로 한다.

③ 간호 보험 제도

로봇 기술을 활용 한 간호 기기의 요구가 높아지는 가운데 개발 기업이나 간호 종사자에 대해 적절한 지원을 실시하는 것으로, 대중의 움직임을 가속화시킬 필요가 있다.

따라서 현행 3년에 1번 간호 보험 제도의 종목 검토 대해, 요구 접수・검토 등의 체제의 탄력화를 도모하고, 혁신에 신속하게 대응 가능하게 한다. 구체적으로는, 간호 보험 급부 대상에 관한 민원의 수시 접수 및 현행 종목에서 해석 할 수 있는 종목 등의 신속한 주지를 실시하는 것 외에 새로운 종목에 대한 자세한 내용도 「간호 보험 복지 용구 평가 검토회」 및 「사회 보장 심의회 간호 급부 비 분과회」을 필요에 따라 수시로 개최하는 등 수시로 결정한다.

④ 도로 교통 법・도로 운송 차량 법

자율주행자동차를 비롯한 이동 로봇의 연구 개발은 국내・외에서 급속히 진전하고 있다. 향후 국제 경쟁에서 일본이 우위를 차지하기 위해서는 연구 개발에 참여하는 신형 기종 투입, 이를 위한 시장의 광범위한 피드백이 필수적이다.

그러나 현행 법률상 탑승형 이동 지원 로봇은 그 원동기의 총 배기량 또는 정격 출력의 대소에 따라 자동차 또는 원동기 장치 자전거로 원칙 보안 기준을 충족하지 않으면 공공 도로 운행용으로 제공할 수 없지만, 안전 확보를 하면 도로 운송 차량의 보안 기준(쇼와 26년(1951년) 교통부령 제67호) 제55조에 근거한 기준 완화를 받음으로써, 공공도로를 운행할 수 있다.

현재 이바라키현 츠쿠바시에서는 특구 제도를 활용하여 탑승형 이동 지원 로봇의 공도 실증 실험이 실시됐으며 2014년도 중에 실시 예정의 “구조 개혁 특구 평가 조사 위원회“의 평가 결과를 바탕으로 2014년에 창설된 “기업 실증 특례 제도“의 활용도 포함하고 이 같은 이동 지원 로봇의 취급에 대해서 검토한다.

또 자동차의 자동 운전에 대해서도 고도 정보 통신 네트워크 사회 추진

전략 본부에서 결정(2014년 6월) 된 관민 ITS구상・로드맵에 기초한 제네바 조약 등과의 정합성을 감안하면서 대응을 추진한다.

게다가 현행 법 규제가 무인 트랙터 등의 무인 농기계가 농지에 갈 때 공도를 주행할 수 없다. 이 취급에 대해서도 마찬가지로 국제 규정의 정합성을 정리한 뒤 안전성 검증을 실시하면서 검토를 진행시킨다.

⑤ 무인 비행 로봇 관계 법령 (항공법 등)

재해 현장을 비롯해 무인 비행 로봇 (UAV)에 대한 기대가 높고, 앞으로 그 보급이 전망된다. 그러나 이러한 로봇에 관한 구체적인 운영 규칙은 명확하게 규정되어 있지 않다. 따라서 향후 이른바 소형 무인 항공기는 운용 실태 파악하여, 공적인 기관이 참여하는 규칙의 필요성과 관계 법령 등을 포함하여 검토를 진행시켜 나간다.

또한 원격 조종에 의해 국제적으로 IFR (계기 비행) 비행을 할 무인기 시스템 (대형 무인기)은 국제 민간 항공기구 (ICAO)의 국제 기준 개정 검토에 참여하여 2019 년 이후 상정되는 국제 기준의 개정을 근거로 국내 규칙 화를 추진한다.

⑥ 공공 인프라의 유지・보수

관계 법령 공공 인프라의 유지・보수에 있어서는 사람에 의한 작업이 전제되어 「육안」에 의한 점검이 요구되고 있는 경우가 있고, 유지 관리의 효과・효율적인 한층 될 향상을 위해 2014년부터 시작한 '차세대 사회 인프라 용 로봇 현장 검증위원회 '에 의한 기반 시설의 유지 관리 및 재해 대응 등에 관한 현장 검증 결과 및 항만 시설 인프라의 유지 관리에 관한 현장 검증 결과 등을 바탕으로 유용한 로봇 대해 효과적 ・ 효율적인 활용 방법을 정한다.

⑦ 고압 가스 보안법

전술의 공공 인프라뿐만 아니라 플랜트 등의 산업 인프라에서도 육안 등의 인간을 전제로 한 점검 작업에 대해 인간의 대체가 되도록 로봇 활용 규칙을 정비할 필요가 있다. 따라서 우선 현장 요구에 따른 기술 개발 및

플랜트 등을 활용한 실증 평가를 동시에 진행하여 현장에서 필요한 기술 수준을 명확히 한 후에, 로봇에 의한 점검에 관한 제도에 대해 검토를 진행시켜 나간다.

(나) 소비자 보호의 관점에서 필요한 틀의 정비

① 소비 생활용 제품 안전법, 전기 용품 안전법

로봇이 일상생활의 구석 구석까지 파고들고 기술 혁신을 통해 자율성과 원격 조작성을 갖는 생활 관련 차세대 로봇의 활용이 확대되는 사회를 가정하면 소비자의 안전 확보를 도모하기 위한 지원 강화가 필요하다. 로봇에 의한 중대한 제품 사고 등이 발생했을 경우의 정보 수집, 원인 규명의 방식과 기술 개발이나 개별 구체적인 제품화 동향을 감안한 전기 용품으로 취급되는 장비에 관한 기술 기준의 본연의 자세, 제조 사업자 등의 책임 범위에 로봇을 전제로 하지 않는 현행 제도의 취급에 대한 검토가 필요하다.

따라서 현행 법령에 따라 수집되는 사고 정보 등의 분석을 바탕으로 필요한 조치와 대응책의 검토를 진행시켜 나간다.

(3) 로봇 국제표준화의 대응

일본의 로봇이 전 세계에서 활용되기 위해서는, 하드웨어/소프트웨어나 그 인터페이스 또는 이들이 동작하는 공통 기반에 대해서, 일본의 시스템이 갈라파고스화 되지 않고 조화되는 것이 중요하다.

또 규제・제도 면에서도 일본과 세계의 규제가 조화함으로써 일본에서 로봇에게 요구되는 요건을 충족하면 세계 전체에서도 로봇을 보급할 수 있는 국제적으로 조화된 규제・제도를 만들어 가는 것이 중요하다.

이를 위해서 일본의 기술력에 기초한 주도적인 국제 표준화를 추진할 필요가 있으며, 일본에서의 보급을 촉진하기 때문에 국제 표준화와 동시에 국내 표준화를 추진해야 한다.

로봇이 도입되어 있는 구조를 실현하려면 하드웨어・소프트웨어 모두, 다수의 업체가 만드는 부품을 잇는 계면(인터페이스)를 어떻게 공통화, 표준화하느냐가 중요하다.

예전에는 시장을 점유하는 소수의 기업이 정하는 사실상의 표준이 큰 힘을 가지고 있었지만, 기술의 복잡화에 의한 잠정적 표준화의 어려움이나 국제 표준에 대한 세계 각국의 관심의 고조도어, ISO등 표준의 중요성이 전례없이 높아지고 있다. 또 기존 제품이 출시되면서 표준이 결정된다는 흐름에 대해서 유럽 각국은 장기적 관점에서 연구 개발 단계부터 기업・대학・연구 기업과 연계하고 표준화를 범위에 만든 계획을 입안하고 표준도 포함하여 역내 기업이 국제적으로 우위에 내는 전략을 취하고 있다.

이처럼 로봇 분야에서의 이러한 표준화에 대해서도 단순히 부품 공통화에 의한 편리성을 향상하는 것 뿐 아니라, 얼마나 국내 로봇 관련 기업이 국제적 경쟁력을 확보하고 차세대 로봇 산업을 선도하는 환경을 만들거나 같은 국가 전략의 관점에서 생각할 필요가 있다, 일본에서도 유럽과 마찬가지로 연구 개발 단계에서 일체적 동시 병행적으로 표준화에도 힘쓰는 것이 중요하다.

제조 현장에서의 로봇・기계 시스템에 대해서, IEC61131(PLC프로그램 표준 규격 등), IEC61158(필드 버스 관계의 표준 규격), ISO15745(어플리케이션 통합 체제), ISO15704(디바이스 프로파일)등 수많은 표준이 책정되어 있다. 기술의 진전에 따른 통신 기기 및 프로토콜 제어 기기와 로봇 등이 새로이 개발되면서, 업체들은 자사 제품을 국제 표준으로 만들기 위해 표준화 활동을 활발하게 하고 있다.

경영 시스템과 생산 시스템뿐만 아니라 가치 사슬을 포함하고, 거기서 다루는 모든 정보를 컴퓨터상으로 통합하면서 물리적인 물건 취급(생산, 조립, 수송, 판매 등)과 연계함으로써 생산・판매의 비약적 효율화, 생산 시스템의 유연성 향상을 목표로 독일의 산업4.0의 대처가 주목받고 있고, 이러한 업무 관리 및 생산 관리와 실제 생산 관리 시스템을 통합하기 위한

ISO/IEC62264과 같은 표준도 수립되었다

로봇의 표준으로는 이른바 산업용 로봇과 로봇 관련 기기에 관한 표준 ISO10218이 있었으며, 최근 서비스 로봇의 안전 기준에 관한 ISO13482이 책정됐다. 또한, 상술한 산업 로봇 인터페이스 표준 ORiN은 ISO20242-4 적용 사례로 참조되고 있으며 RT미들웨어의 모듈 인터페이스는 소프트웨어 표준화 단체 OMG(Object Management Group)에서 표준화되고 있다. 이 밖에 IEC TC59F/WG5에서는 청소 로봇에 관한 표준, ISO TC184/SC2에서는 협조 로봇의 안전(WG3), 간병 로봇의 안전(WG7), 바퀴형 이동 로봇이나 소프트웨어・하드웨어의 모듈화(WG8)에 관한 표준의 논란이 이미 시작됐고 일본도 명확한 전략 아래 이들 표준화에 관여, 혹은 적극적으로 추진할 필요가 있다.

또한 로봇의 네트워크 이용의 요구 조건과 기능 요건에 대해서는 일본의 주도로 국제 전기 통신 연합(ITU)에서 표준 책정 작업을 진행하고 계속 조속한 수립에 기여한다. 또 로봇 간의 대화를 실현하기 위한 통신 프로토콜이나 복수의 로봇을 제어하기 위한 기술 등에 관한 표준화도 미국 전기 전자 학회(IEEE)에서 진행됐으며 입안에 적극적으로 참여한다.

(4) 로봇실증실험필드 정비

사회의 모든 장면에서 로봇이 활용되고 사회적 과제를 해결한다는 “로봇 혁명“에서 지향하는 사회상과 같이, 로봇은 실제로 현장에서 사용되는 것으로 그 존재 의의가 발휘된다.

이런 관점에서, 사회에 구현되기 전 마지막 단계로서 실제로 로봇이 사용되는 상황과 같은 조건하에서 실증 실험을 실시하여 최종적인 조율은 하는 것은 중요한 프로세스이다.

한편, 실제로 로봇을 활용한 경우에 생기는 문제는 우리가 예상할 수 없는 것까지 다양하다. 보다 고도의 로봇을 사회에 구현시키기 위해서도 안전

성 등에 관한 실증 실험을 포괄적으로 실시하는 동시에 실제로 활용하는 측의 “사용“을 높이기 위한 검증을 할 수 있는 환경을 정비해야 한다.

이처럼 로봇의 개발이나 현장에 도입을 가속화하기 위해서는 기술 개발 및 도입에 관한 지원만 아니라 로봇의 개발 현장과 활용 현장의 가교가되는 실증 실험 필드를 정비하는 것이 효과적이다.

◇ 차세대 사회 인프라용 로봇 개발・도입 추진 (국토 교통성 직할 현장)

- 국토 교통성 직할의 현장에서 교량・터널・수중 유지 관리, 재해 조사 및 응급 복구에 관한 로봇 기술의 검증을 실시

◇ 사가 로봇 산업 특구의 실증 지원(가나가와 현)

- 폐교가 된 학교를 활용한 사전 검증 필드의 설치나 간호/요양 시설에서의 생활 지원 로봇의 실증 등 공모 등의 방법을 통해 기업의 실증 실험 지원 등을 실시

◇ 효고현 광역 방재 센터(효고현)

- 방재 전문 인력, 지역 방재 리더를 육성하고, 재해시에는 광역 방재 거점으로서도 기능한 시설로, 재난 대응 로봇의 실증 실험에도 활용

◇ 생활 지원 로봇 안전 검증 센터(츠쿠바 시)

- 생활 지원 로봇의 안전 인증(ISO13482등)때문에, 주행 시험, 대인 시험, 강도 시험, EMC시험 등 18종류의 시험을 실시할 수 있도록 함, - 본 시험 시설을 이용하고, 그 로봇이 ISO13482를 취득

현재 존재하는 국내 로봇 실증 실험 필드의 기능에 대해서 유사한 해외에서의 활동과 비교하면서 로봇 혁명의 실현을 위해서 민간 기업이나 대학이 자체 실시하는 실증 활동을 지원하고 필요 시 기존 시설의 증강과 새로운 기능을 가진 시설을 구축해 나간다.

그 때는 로봇 혁명을 담당하는 나라 안팎의 도전자들이 모이고, 요구하는

실험 환경을 정확하게 제공함으로써 안전성 검증과 구체적인 사용의 개선으로 이어지는 실적을 올리고, 장래에 걸친 이노베이션의 거점이 되어 계속 같은 체제를 지향하는 것이 중요하다. 이하에 실증 실험 필드를 정비하는데 앞으로도 안정적으로 활용되도록 유의해야 할 항목을 나타낸다.

1. 특구 제도의 활용 등 로봇 실증 실험을 위한 충분한 공간과 기존의 제도에 얽매이지 않고 실증 실험 할 수 있는 자유가 보장되어 있는지

2. 공공 부문에 일정 정도의 요구가 있는 시설이며, 민간의 수요도 충분히 기대할 수 있는지

3. 로봇 실증 실험 필드를 이용한 검증 결과가 규제 완화, 공공 조달, 인증 등으로 활용할 수 있는 등 사업화를 지원해주는 구체적이고 제도적인 효과가 자리 매김하고 있음

4. 대학이나 지자체 등 운영 주체가 명확하고 안정적으로 존재하는지 여부

이러한 노력을 더욱 추진하고, 필드 로봇을 중심으로 한 실용화의 움직임을 가속화하기 위해 새로운 실증 필드로 후쿠시마 현에 “후쿠시마 하마도리 로봇 시범 지역”(가칭)을 설치하고, 육・해・공의 모든 분야의 로봇 개발의 거점으로 하는 것을 목표로 한다.

2. 지식재산전략본부 “지식재산추진계획 2016”

(1) 개요

일본 지식재산전략본부는 2016년 1월 27일 차세대 지식재산 시스템 검토위원회를 통해 인공지능이 만들어낸 창작물에 대한 보호 방안을 검토하고, 동년 4월 8일 인공지능 창작물의 저작권 보호에 관한 보고서를 발표하였다.

이후 종합적인 검토를 반영한「지적재산추진계획 2016」를 발표하였다.

본 추진계획의 목표는 “제4차 산업혁명시대를 맞이하여 선도적 지적재산 권제도를 수립하고 글로벌 시장 개척, 이노베이션 창출 및 인재육성을 통해 국가 성장전략 마련” 하는 것이다.

이에 대한 기본방향으로는 ①(지적재산의 연계를 통한 오픈 이노베이션) 지방을 포함하여 전국적으로 중소기업 등과 산학·산산 연계를 통해 새로운 이노베이션 창출하고, ②(지적재산교육 및 인재육성) 가치 있는 혁신을 창출할 수 있는 인재 배출을 위해 사회 및 지역과 협동하여 지적재산교육 내실화를 도모하며, ③(콘텐츠 및 아카이브의 전개·활용) “쿨 재팬(Cool Japan)[27] 민관연대 플랫폼”을 활용하여 콘텐츠 와 非콘텐츠 산업 간의 연계 강화 및 해외진출 추진과 아카이브 활용을 위하여 기반을 정비하고, ④(지적재산 분쟁처리시스템 기능 강화) 글로벌 시장에서의 지적재산 보호 및 이노베이션의 지속적 창출을 위한 환경 정비를 목적으로 분쟁처리시스템 기능의 강화 등을 주요 방향으로 설정하였다.

이러한 기본방향을 중심으로 ① 제4차 산업혁명시대의 지적재산 혁신 추진, ② 지적재산 인식 및 지적재산 활동의 보급 및 확산, ③ 콘텐츠 신규전개 추진, ④ 지적재산시스템의 기반정비라는 4대 전략목표를 수립하였다.

(2) 제4차 산업혁명시대의 지적재산 혁신 추진

(가) 디지털·네트워크화에 대응한 차세대 지적재산시스템 구축

디지털·네트워크의 발전에 따른 새로운 형식의 이노베이션이 출현하고, 국경을 초월하여 온라인상에서 지적재산의 침해가 확산되어 가고 있다. 하지만 이를 대응하기 위한 현행 법제도에서는 그 대응에 대한 한계가 존재

27) 쿨 재팬이란, 일본의 문화 측면에서 소프트웨어 영역이 국제적으로 인정되는 현상과 그 내용 자체 또는 일본 정부의 대외 문화홍보와 수출 정책에서 사용되는 용어다.
구체적으로는 있는 산업분야는 명확하게 정의되어 있지 않지만, 콘텐츠(애니메이션, 만화, 영화, 음악 등), 패션, 디자인, 음식, 주거, 전통문화, 지역특산품, 관광 등, 폭넓은 분야에 미치고 있다. 콘텐츠를 중심으로 한 팝컬쳐, 서브컬쳐 등의 분야에, 음식, 지역특산품 등 다양한 영역이 포함되고, 2000년대에 들어서 일본브랜드, 콘텐츠 산업, 크리에이티브산업, 문화산업, 소프트파워 산업 등 다양한 것들을 포함할 수 있다.(자세한 사항은 김정명(2013), “일본의 쿨 재팬(Cool Japan) 전략과 정책적 함의”, 「2013년 하계학술대회 발표집」, 한국행정학회 , 2면 참고)

한다. 특히 제4차 산업혁명으로의 발달은, 빅데이터・인공지능 등 새로운 창조원에 의한 창작물 및 데이터베이스의 보호 등에 대한 필요성이 고조되고 있으며, 인터넷 이용자를 저작권 등 지적재산권 침해콘텐츠를 유도하는 링크를 모아 게재하는 사이트인 리치사이트(Reach Site) 등 현행 제도상 대응이 어려운 지적재산 침해행위의 확산에 대한 법제도의 검토의 요구가 대두되고 있는 실정이다.

이에 대해 유연성 있는 권리제한규정을 검토하고 고아저작물에 대한 제도의 개선을 도모하며, 인공지능이 만든 창작물 및 데이터베이스 등에 대응한 신 지적재산시스템, 데이터 공유・활용환경 등 새로운 정보 창출에 대응한 지적재산시스템의 구축, 리치사이트 등 새로운 형태의 지적재산 침해행위 및 침해사이트 내 온라인광고 등에 대한 대응과 외국 사이트 차단(Site-blocking) 운용현황 등의 방안을 검토하는 것을 주요 내용으로 한다.

(나) 오픈 이노베이션을 위한 지적재산 경영 추진

사물인터넷(IoT)・빅데이터・인공지능 등 신기술의 발전과 함께 제4차 산업혁명 도래에 따라 사회 및 경제구조의 기본적인 틀에서 변화하고 있다. 특히 정보 개방화의 수요가 많아지면서 노하우의 유출 위험이 점점 더 증가됨에 따라 오픈&클로우즈 전략을 축으로, 산학・산산(대기업-중소・벤처기업)의 연계를 활성화 하는 동시에 지적재산권으로 보호할 것은 확실하게 보호하여 폭넓은 지적 재산 관리의 기반이 되는 프로 이노베이션의 지적재산 시스템을 구축할 필요가 있다.

세부 내용을 살펴보면 산학 공동창조 플랫폼 및 지역 혁신 생태계 구축 프로그램 등을 마련함으로써 지역기업과 대학의 협력을 강화하고 원스톱 지적재산 경영을 목표로 대학 내부 체제를 강화하고, 사회시스템・첨단기술 분야 내 국제표준화 추진체제를 강화하고 중소・중견기업, 식료산업분야, 전통의료, IoT 서비스 등의 국제표준화를 전략적으로 추진하며, 영업비밀 관리에 대한 원스톱 지원을 확충하고 관련 정보보관시스템을 구축하며 영

업비밀 침해품과 관련하여 국경조치 도입 등을 검토하고, 법률자문 등 종합적 지적재산 전략구축 지원인재를 육성하고 글로벌 지적재산 인재육성프로그램을 개발하여 그 활용을 촉진하는 것을 구체적 목표로 한다.

⑶ 지적재산 인식・활동의 보급・확산

(가) 지적재산 교육 및 인재육성의 내실화

일본은 2006년부터 ‘지적재산 인재 육성 종합전략’, ‘지적재산 인재 육성계획’ 등 지적재산교육 및 인재육성 국가계획을 실시해왔지만 지적재산보호에 치중한 교육 및 교원의 지적재산 관련 지식의 부족 등 문제점이 지속적으로 지적되어 왔다. 이러한 문제점을 해결하기 위해 국민 전체에 체계적인 지적재산교육을 실시하며, 지역・사회와의 협동 학습 지원치제를 구축하고, 지적재산의 학습과 활용으로 이어질 수 있는 효과적인 전략이 요구된다, 이를 위해 이 보고서에서는 초・중・고등학교 내 핵심적인 지적재산 관련 교과목을 마련・제공하고 대학 내 지적재산과 관련한 과목을 필수화하고 다양화함으로써 지적재산교육의 내실화를 도모하며, 관련부처・교육기관・기업 등으로 구성 된 “지적재산교육 추진 컨소시엄” 및 지역 맞춤형 교육을 위한 “지역 컨소시엄”을 구축하여 지역・사회와의 협동 학습지원 체제를 구축한다. 마지막으로 세계적인 지적재산 인재양성을 위하여 영어로 된 지적재산교육 프로그램을 개발하여 제공하고, 관련 교육 교재 등을 정비하는 등 지적재산 교육 및 계발을 위한 기반을 정비한다.

(나) 지역경제 주체인 지역 중소기업 및 농림수산업 등 분야에서의 지적재산전략의 추진

지역경제주체인 지역의 중소기업 및 농림수산업의 지적재산의 활용은 지역경제 활성화 및 지역상생의 관점에서 매우 중요하지만, 현재는 중소기업의 낮은 지적재산에 대한 인식과 지원정책에 대한 접근이 어렵다는 문제점

이 발생하였다, 이를 위해 타인 또는 타기업의 지적재산을 활용하는 중소기업에 대해 전략적인 지적재산의 보급 및 중소기업을 위한 지적재산상담 인력 육성 등을 도모하며 반대로 자신들의 지적재산을 활용하여 부를 창출하는 중소기업에 대한 지원을 강화하여 상담기능 등을 강화한다. 농림수산분야에서는 지리적표시 활용과 농수산물 및 주류의 브랜드화를 촉진함으로써 일본산 주류 등의 브랜드 가치향상을 도모하고, 종묘관리센터 내 유전자 DB를 강화하고 국내품종등록 심사결과의 무상 제공 등 종묘산업의 해외진출과 권리침해 대책에 대한 지원을 강화함으로 농림수산업 분야 등에서의 지적재산전략을 추진한다.

(4) 콘텐츠 신규전개 추진

(가) 콘텐츠 해외전개 및 산업기반의 강화

현재 '콘텐츠 해외진출 촉진사업', 'Cool Japan 민관연대 플랫폼' 등 각종 콘텐츠 해외진출 지원정책이 긍정적인 결과를 창출하는 것으로 평가되고 있으나, ICT기술의 발전으로 인해 3D 영상, 모바일, 스트리밍 등 콘텐츠에 대한 선호가 증가하는 한편, 글로벌 화로인한 모방품 및 해적판의 피해규모가 전 세계적으로 확대됨에 따라 적극적인 지원이 필요하다.

이에 따라 'Cool Japan 민관연대 플랫폼'을 적극 활용하여 타 업종간의 연계를 촉진하고, 지역 주체의 콘텐츠를 제작하는 등 지역 관광사업과의 연계를 통해 발전을 촉진하며, '콘텐츠 해외진출 촉진사업'의 지속적 실시를 통해 외국과의 쌍방향 문화교류를 촉진하는 등 지속적인 해외진출을 지원한다. 또한 콘텐츠 산업 종사자들을 대상으로 해외 연수기회 제공 등을 통해 인재 육성을 지원하고 콘텐츠 제작환경의 개선과 콘텐츠의 거래적정화 등을 도모하여 콘텐츠 산업의 기반을 강화하고, 정규 콘텐츠의 유통 확대 및 침해발생국 정부와의 연계 강화를 통해 모방품 또는 해적판에 대한 대책을 마련하고, 침해국의 지적재산권제도 및 피해실태 등에 대한 조사를 통한 보호대책을 강구한다.

(나) 아카이브 활용 촉진

선진국을 중심으로 문화 보존·계승 발전의 기반 및 콘텐츠의 2차적이용 등의 기반으로써 콘텐츠 디지털 아카이브화를 적극 추진한다. 특히 중소규모 기관 및 지방의 경우에는 원자료의 디지털화나 메타데이터 작성 또는 정비 등의 활동을 단독으로 추진하기 어렵다는 문제점이 있기 때문에 적극적인 전략의 필요성이 발생한다.

이를 위해서 관계 부처간 정보공유 및 의견교환을 도모하고 분야간 통합포털 구축을 위한 주요 아카이브간의 연계를 촉진하고, 서적·문화재·미디어예술·방송분야로 구분하여 분야 내 아카이브 기관에서 소장 자료 디지털화 및 메타데이터의 집약화를 실시하는 등 분야별 수집기관의 아카이브를 촉진한다. 또한 실무자 협의회 등을 통하여 메타데이터의 개방화 촉진을 위해 문제점·대응책을 검토하고, 고아저작물 등의 이·활용 촉진을 위한 저작권 제도를 정비한다.

⑸ 지적재산 시스템의 기반 정비

(가) 지적재산 분쟁처리 시스템 기능 강화

지적재산관련 분쟁이 발생하면 증거수집과 관련하여 침해자에 침해증거가 편중되어 권리자의 침해입증이 어렵고, 현실을 반영한 적절한 손해배상액 산정에 있어서 난해하다는 문제점이 발생한다. 특히 중소기업의 경우에는 지적재산소송에 대응하기 위한 인적·물적 자원이 부족하며 지적재산소송의 1심 관할이 도쿄 및 오사카 지방법원에 한정되어 있어 지방에서의 사법시스템에 대한 접근성이 낮다. 또한 침해자에 유리한 지적재산분쟁구조에 따라 무효심판 및 침해소송에서 권리무효에 대한 신중한 판단이 요구되며, 특허소송 절차에서 특허청의 판단이 적절히 개입할 수 있는 구조의 마련이 필요하다.

이에 따라 증거수집절차의 개선과 비즈니스 실태 및 수요를 반영한 적절한 손해배상액의 실현이 필요성과 함께 특허권의 안정성 향상을 위해 전문관청에 의한 검토 기회의 확대 등 관련 제도를 검토하고 특허청과 법원의 제휴를 강화하는 등 지적재산 분쟁처리시스템 기능을 강화하며, 소송비용보험제도 보급 등 중소기업의 지적재산 분쟁대응을 지원하고 재판 외 분쟁해결절차(ADR)를 확충・활성화하는 등의 지적재산 분쟁처리시스템 이용 지원을 강화 한다. 또한 일본 지적재산 관계법령을 영문으로 번역하여 해외에서도 접근 가능하도록 하고 외국제도 및 실태 등을 조사하여 중요 정보의 공개・발신 체제를 마련한다.

(나) 제4차 혁명에 대비한 저작권 체제의 정비

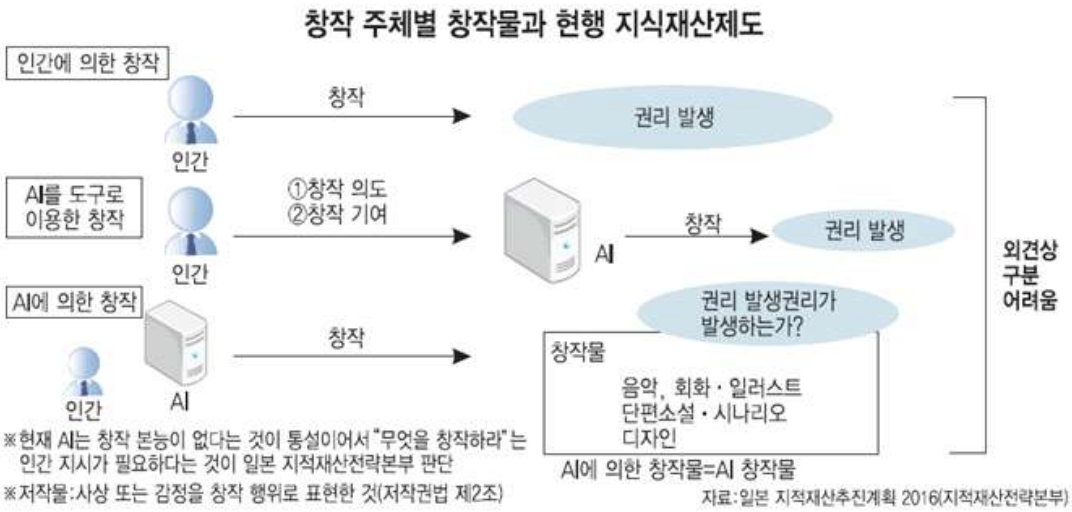
본 계획서는 제4차 산업혁명에 대응하기 위한 지적재산권 체제에 대한 개편방향을 수립하였다. 특히 인공지능 창작물, 빅데이터에서 도출된 데이터베이스의 보호 등에 대하여 저작권체계에 대한 정비를 위한 검토를 진행한다는 점이다. 하지만 현행 저작권법 및 제도에서는 한계점이 명확히 발생하고 있다.

① 현행 저작권 제도의 한계점

현행 저작권법상 인공지능 창작물은 권리의 대상이 될 수 없다. 일본 저작권법 제2조 제1항 제1호에 따르면 “저작물은 사상 또는 감정을 창작적으로 표현한 것으로, 문예, 학술, 미술 또는 음악의 범위에 속하는 것을 말한다.” 라고 정의되고 있다. 그렇기 때문에, 인공지능의 창작물은 저작물에 해당하지 않고 저작권도 발생하지 않는다는 것이 일반적인 해석이다. 다만 인간이 인공지능을 도구로 사용하여 창작물을 창작한 경우는 저작권이 발생한다고 해석되고 있다.

그러나 서론에서 언급한 사례처럼 인간의 창작물과 외견상 구분이 어려운 수준에 도달이 되었고, 인공지능 창작물이라 밝히지 않는 경우에는 인간의 창작물과 동일하게 취급될 수 있다는 문제점이 대두된다.

[그림 2-1] 창작 주체별 창작물과 현행 지적재산제도



* 출처 : www.ipnomics.co.kr

이는 저작권의 인정여부뿐만 아니라 새로운 유형의 법적 분쟁 등 보다 복잡한 문제가 발생될 수 있다. 특히 법적 주체에 대해 기존의 패러다임과는 다른 상황이 발생될 수 있기에 인공지능 창작물에 대한 보호의 필요성 및 가능성과 인공지능 창작물이 기존의 저작권 제도에 미치는 영향 등을 고려하여 새로운 저작권 제도의 구축 필요성이 발생하게 된다.

② 인공지능 창작물에 관한 논의사항

일본정부가 인공지능의 도움을 받은 창작물에 대한 저작물성에 대하여 논의가 시작된 것은 예전부터다. 문화청에서는 1973년부터 컴퓨터에 의한 창작물에 대하여 저작권 제도상의 취급에 대한 논의를 시작하였다. 다만, 컴퓨터가 인간의 창작행위를 완전히 대체한 것이 아니라 특정인의 사상과 감정을 표현하는 도구로서 사용된 점이 인정된다면 특정인에게 저작권을 인정하는 것이다. 또한 영국은 1988년 저작권법 개정을 통해 컴퓨터에 의해 생산된 창작물의 경우, 작품의 창작을 위해 필요한 조치를 한 사람이 저작자로 인정되도록 규정[28]하였다.

28) CDPA of 1988 §9(3) 컴퓨터에 기인하는 어문, 연극, 음악 또는 미술 저작물의 경우에는, 저작

제5절 국제단체의 윤리 정책 및 법제 동향 분석

1. IEEE “윤리적으로 조화로운 설계”

(1) 개요

2016년 12월 13일 전기・전자 기술자 협회(Institute of Electrical and Electronics Engineers, 이하 IEEE)에서는 “윤리적으로 조화로운 설계” (Ethically Aligned Design)를 발표하였다.

인공 지능 및 자율 시스템에서 윤리적 고려를 위한 IEEE 글로벌 이니셔티브(The Global Initiative)는 160개국 이상에서 400,000명이 넘는 회원들을 보유하고, 인류의 이익을 위한 기술 발전에 전념하는 세계 최대의 기술 전문 기관인 Electrical and Electronics Engineers, Incorporated (“IEEE“)의 프로그램이다.

IEEE Global Initiative는 시의적절한 쟁점을 확인하고, 합의를 도출할 수 있는 인공 지능 및 자율 시스템 커뮤니티에서 다양한 목소리를 모으는 기회를 제공한다.

IEEE는 Creative Commons Attribution-Non-Commercial 3.0 미국 라이센스 하에 이용할 수 있도록 “윤리적으로 조화로운 설계를 (Ethically Aligned Design: EAD)” 을 발표하였다. 이는 IEEE Global Initiative와 EAD는 IEEE TechEthics 프로그램으로 알려진 기술 윤리에 대한 개방적이고, 폭 넓으며, 포괄적인 대화를 촉진하기 위해 IEEE에서 시작된 광범위한 노력에 기여하기 위함이다.

자는 그 저작물의 창작을 위하여 필요한 조정을 한 자로 본다.(In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken)

(2) IEEE 글로벌 이니셔티브의 목적

자율적이고 지능적인 시스템의 설계 및 개발에 윤리적 고려 사항을 우선시하기 위하여 모든 기술자들은 교육도 받고, 훈련도 받으며 권한을 부여받는 것이 보장되어야 한다.

“기술자“란 연구, 설계, 제조 또는 메시징 기술들을 사회에 현실화하는 대학, 연구원 및 기업을 포함하는 AI/AS와 관련된 연구, 설계, 제조 또는 메시징에 관련된 모든 사람들을 의미한다.

이 보고서는 학계, 과학 및 정부 및 기업 분야에서 AI, 법률 및 윤리, 철학, 정책 분야의 100 명이 넘는 글로벌 사고 리더의 공동 의견을 나타낸다. 우리의 목표는 윤리적으로 조화로운 설계가 향후 몇 년 동안 AI/AS 기술자의 작업에 대한 핵심 내용을 제공하는 사람들에게서 통찰력과 권고안을 제공하는 것이다. 이 목표를 달성하기 위하여 현재 버전의 Ethically Aligned Design (EAD v1)에서 우리는 AI/AS를 구성하는 분야에서 쟁점과 예비권고안을 확인한다.

IEEE Global Initiative의 두 번째 목표는 EAD에 기초한 IEEE 표준에 대한 고안을 제공하는 것이다. Model Process for Addressing Ethical Concerns During System Design인 IEEE P7000은 The Initiative에서 영감을 받은 최초의 IEEE 표준 프로젝트(승인되었고, 개발 중임)이다. 2개 추가된 표준 프로젝트인 자율 시스템의 투명성과 관련된 IEEE P7001과 데이터 프라이버시 프로세스와 관련된 IEEE P7002가 승인되었고, AI/AS 윤리 쟁점에 대한 이니셔티브의 실용적인 영향력이 입증하였다.

(3) 주요 내용

(가) 일반 원칙

일반 원칙위원회 (General Principles Committee)는 모든 유형의 AI/AS에

적용되는 다음의 높은 수준의 윤리적 관심사를 분명히 밝혔다

① 인권의 최고 이상을 구현

② 인류와 자연 환경에 최대 이익을 우선시

③ AI/AS가 사회-기술 시스템으로 발전함에 따라 위험과 부정적 영향을 완화

일반 원칙들이 확인한 원칙, 쟁점 및 권고안이 AI/AS 설계를 위한 윤리적 거버넌스의 새로운 틀 안에서 미래의 규범과 표준을 지지하고, 뒷받침하는 것이 위원회의 의도이며 주요 쟁점은 다음과 같다.

• AI/AS가 인권을 침해하지 않도록 어떻게 보장 할 수 있는가? (인권 기본원칙을 구성)
• AI/AS에 책임을 질 수 있다는 것을 어떻게 확신할 수 있는가? (책임 원칙을 구성)
• AI/AS가 투명하다는 것을 어떻게 보장할 수 있는가? (투명성 원칙의 틀)
• AI/AS 기술의 장점을 확대하고, AI/AS 기술이 남용되는 위험성을 최소화하고 어떻게 혜택을 확대할 수 있는가?(교육 및 인식의 원칙 구성)

(나) 자율정보시스템상의 가치포함 방안

사회에 도움이 될 성공적인 자율 지능 시스템 (autonomous intelligent systems: AIS)을 개발하기 위하여, 기술 커뮤니티가 관련 인간 규범이나 가치를 이해하고 AIS에 삽입할 수 있어야 한다. AIS 위원회에 가치를 삽입하는 것은 세 가지의 접근방식으로 AIS에 가치를 광범위한 목표를 설계자들에게 다음 행위에 도움을 주고자 한다.

① AI에 영향을 받은 특정 커뮤니티의 규범과 가치를 확인

② AIS 내에서 해당 커뮤니티의 규범과 가치를 구현

③ 기술적 지역 사회 내에서 인간과 AIS 사이의 규범과 가치의 조정과 양

립성을 평가

주요 쟁점은 다음과 같다,

- AIS에 포함될 가치는 보편적인 것이 아니라, 주로 이용자 커뮤니티 및 작업에 제한
- 도덕적인 과부하 : AIS는 보통 서로 충돌할 수 있는 다양한 기준 및 가치에 종속
- AIS는 특정 그룹의 구성원에게 불이익을 주는 데이터 또는 알고리즘적 편견을 내장
- (특정 커뮤니티에서의 AIS의 구체적인 역할의) 관련된 기준 세트가 확인되면, 이러한 기준을 어떻게 계산 구조안으로 구축해야하는지 명확하지 않음.
- AIS에서 구현된 기준은 관련 커뮤니티의 기준과 호환 가능
- 인간과 AIS 간의 정확한 신뢰 수준 달성.
- AIS의 가치 조정에 대한 제3자 평가.

(다) 윤리 연구 및 설계 지침 방법론

현대의 AI/AS 연구기관은 인간의 복지・권한부여・자유가 AI/AS 개발의 핵심에 있음을 보장해야 한다. 이러한 야심찬 목표를 달성 할 수 있는 기계를 만들기 위하여 Ethical Research and Design Committee 가이드 방법론은 세계 인권 선언에 정의된 인권과 같은 인간의 가치가 시스템 설계 방법론에 의해 생성되도록 보장하기 위한 쟁점과 권고안을 고안하여 왔다. 가치조정적 설계 방법론은 윤리 지침에 기초한 인간 발전에 맞춘 AI/AS 기관을 위한 필수적인 초점이 되어야한다. 기계는 인간에게 봉사해야하며 다른 방향으로는 움직이지 않아야 한다. 이러한 윤리적으로 건전한 접근 방식은 사업과 사회 모두에서 AI의 경제적 및 사회적으로 유지할 수 있는 균형을 유지하도록 보장하여야 한다.

주요쟁점은 다음과 같다.

- 윤리는 정도의 프로그램 일부가 아니다.
- 우리는 AI/AS의 뚜렷한 쟁점을 설명하기 위하여 학제간 및 이문화 교육을 위한 모델이 필요
- AI 설계에 포함된 문화적으로 특이한 가치를 차별화할 필요성.
- 가치에 기반한 윤리적 문화 및 산업 관행의 부족.
- 가치 인식 리더십의 부족.
- 윤리적 문제를 제기하는 권한 부여 부족.
- 기술 커뮤니티의 소유 또는 책임 부재.
- AI/AS의 최선의 상황을 위한 이해 관계자를 포함시킬 필요성
- 불충분한 문서는 윤리적인 설계를 방해
- 알고리즘의 일관성이 없는 감독 부재
- 독립적인 검토 조직의 부재.
- 블랙 박스 구성 요소의 사용.

(라) 개인정보 및 개별 접근 제어

개인정보 및 개별접근제어위원회 (Personal Data and Individual Access Control Committee)는 사람들이 고유한 큐레이터로서 자신의 개인정보를 정의, 액세스 및 관리할 수 있는 근본적인 필요성을 입증하는 쟁점 및 권고안을 명시한다. 위원회는 완벽한 해결책이 없고, 디지털 도구가 해킹 될 수 있음을 인정한다. 그럼에도 불구하고 긍정적인 미래를 위하여 사람들이 자존감을 제어하고, 데이터 비대칭을 근절할 수 있는 도구 및 발전된 관행을 제공하는 데이터 환경의 활성화를 권장한다.

주요 쟁점은 다음과 같다.

- 알고리즘 시기인 현재 개인은 어떻게 개인 데이터를 정의하고 체계화할 수 있

는가?

- 개인 식별 정보의 정의와 범위는 무엇인가?
- 개인 정보에 관한 통제의 정의는 무엇인가?
- 개인을 존중하기 위하여 우리는 어떻게 데이터 액세스를 재정의할 수 있는가?
- 개인 정보에 관한 동의를 재정의하여 우리는 개인을 존경하도록 어떻게 할 수 있는가 ?
- 공유하는 사소한 데이터는 개인이 공유하고 싶지 않다고 하는 추론을 하는 데 사용될 수 있다
- 데이터 처리자는 정보를 제공한 후에 진정으로 동의를 받기 위하여 데이터에 액세스하고, 데이터를 수집하는 긍정적, 부정적 결과가 개인에게 명백한 것인지 어떻게 보장할 수 있는가?
- 사람은 개인화된 AI 또는 알고리즘적인 감시자가 될 수 있을까?

(마) 자율무기체계 재구성

신체적인 위해를 야기하기 위해 설계된 AS는 재래식 무기와 인간에게 신체적인 위해를 야기하지 않도록 설계된 자율 시스템에 비해 추가적인 윤리적인 파급 효과가 있다. 이러한 AS 무기에 대한 직업 윤리는 보다 광범위한 관심을 다루는 더 높은 기준을 가질 수 있거나, 가져야한다.

대체로 자율 무기 체계 재구축 위원회 (Reframing Autonomous Weapons Systems Committee)는 기술연구기관은 무기 체계의 의미있는 인간 통제가 사회에 유익하고, 감시제도가 그러한 통제를 책임지고 보장하며, 이러한 기술을 개발하는 사람들이 업무의 의미를 이해하고, 직업 윤리적 인 코드는 해를 가하기 위한 의도로 작성된 작품을 적절하게 다룬다.

주요 쟁점은 다음과 같다.

- 전문적인 연구기관 행동 규범에는 종종 중요한 허점이 있다. 이러한 예로는 구성원의 업무, 구성원이 만든 작품 및 에이전트를 회원 자신이 보유하고 있는 동일한 가치와 기준에 따라 간과될 수 있다.

• 인공 지능, 자율 시스템 및 자율 무기 시스템 (AWS)의 중요한 개념에 관한 정의에 관한 혼란은 보다 실질적인 중요한 문제에 대한 토론을 방해한다.

• AWS는 기본적으로 비밀 및 출처를 알 수 없는 사용이 가능하다.

• AWS의 행동에 대한 책임이 손상될 수 있는 여러 가지 방법이 있다.

• 설계 및 운영 용도에 따라 AWS는 예측 가능하지 않을 수도 있다. 학습 시스템은 예측 가능한 용도 문제를 복잡하게 한다.

• AWS 개발의 합법화는 중기적으로는 지정학적으로 위험한 선례를 설정한다.

• 전장에서 인간의 감시를 배제하면 부주의로 인권 침해와 부주의로 긴장감의 확대로 이어질 수 있다.

• AWS의 직·간접 고객의 다양성은 AWS의 확산과 남용이라는 복잡하고 심각한 문제가 되는 방향으로 이어질 것이다.

• 기본적으로 AWS의 자동화로 갈등의 급속한 확대로 이어질 수 있다.

• AWS의 설계 보증 검증 표준이 없다.

• AWS 및 반자동 무기 시스템에 관련된 물건에 윤리적 경계를 이해하는 것은 매우 혼란스러울 수 있다.

(바) 경제 / 인도주의 문제

우리의 일상생활에서 인간의 개입을 줄이는 것을 목표로 하는 기술, 방법론 및 시스템은 빠른 속도로 진화하고 있으며 다양한 방식으로 개인의 삶을 변화시킬 태세를 갖추고 있다.

경제/인도주의 쟁점 위원회의 목표는 인간-기술 글로벌 생태계를 형성하고, 경제 및 인도주의적 파급 효과를 다루는 핵심 원인을 파악하고 갈등의 중대한 문제영역을 드러내어 실현할 수 있는 해결의 핵심적인 기회를 제안하는 것이다.

위원회 권고안의 목적은 인간, 인간의 제도 및 현재 나타나고 있는 정보중심 기술의 관계에 있어서 이러한 핵심적인 관심사와 관련된 실용적인 방

향을 제안하고, 이러한 쟁점들에 대하여 방향성 있고, 동료에 의해 사고하는 전문가에 의해서 충실하게 정보를 제공할 수 있는 학제간의 대화를 촉진하는 것이다.

주요 쟁점은 다음과 같다.

- 미디어에서 AI/AS의 오해는 대중에게 혼동을 준다.
- 자동화는 일반적으로 시장 상황에서만 볼 수 없다.
- 로봇/AI와 관련하여 고용의 복잡성이 무시되고 있다.
- 인력을 훈련 (재교육)하는 기존의 방법에 비해 기술적 인 변화가 너무 빠르게 일어나고 있습니다.
- 모든 AI 정책은 혁신을 지연시킬 수 있다.
- AI과 자율 기술은 전세계적으로 동일하게 이용 가능하지 않다
- 개인 정보에 대한 접근과 이해가 부족하다.
- IEEE Global Initiative에서 능동적인 개발도상국의 대표가 필요하다
- AI와 자율 시스템의 출현은 선진국과 개발도상국 사이의 경제 및 권력 구조의 차이를 악화시킬 수 있다.

(사) 법

AI/AS의 초기 개발은 많은 복잡한 윤리적 문제를 야기해 왔다.

이러한 윤리적 문제는 거의 항상 직접적으로 구체적인 법적문제로 직결된다. 그렇지 않으면 부수적인 어려운 법률 문제가 발생한다.

법률위원회는 이 분야에서 변호사들이 담당해야할 많은 일들이 있다고 생각하고 있다. 지금까지는 이처럼 절실한 필요성에도 불구하고 변호사들과 학자들을 끌어 들이지 못했다.

변호사들은 이 분야에서 규제, 거버넌스, 국내 및 국제 입법에 대한 논의에 참여할 필요가 있으므로, AI/AS로부터 인류와 지구에서 이용할 수 있는 엄청난 이익은 철저하게 미래를 위하여 나아가야 한다.

주요 쟁점은 다음과 같다.

- 자율 및 지능형 시스템의 책임성과 검증 가능성을 어떻게 향상시킬 수 있는가?
- AI가 투명하고 개인의 권리를 존중하도록 어떻게 보장 할 수 있는가?

예를 들어, 국제, 국가 및 지방 정부는 정부를 신뢰할 수 있어야만 하는 시민들의 권리를 침해하고 있으므로 시민들의 권리를 보호하도록 하여야 한다.

AI를 사용하여 정부와 AI가 자신의 권리를 보호 할 수 있어야하는 시민의 권리를 침해합니다.

- AI 시스템으로 발생하는 피해에 대한 법적 책임을 보장하도록 AI를 어떻게 설계할 수 있는가?
- 어떻게 개인 데이터의 무결성을 존중하는 방식으로 자율 및 지능형 시스템을 설계하고 배치할 수 있는가?

제6절 소결

지금까지 국외 인공지능 윤리 정책과 법제 동향을 살펴보았다. 미국, EU, 일본 등에서는 인공지능사회로의 진입을 위한 노력을 시작하였다. 또한 이러한 움직임은, “4차 산업혁명”으로 인해 변화하는 사회의 흐름에서의 주도권을 선점하기 위한 움직임으로 해석할 수 있다.

미국은 구글 딥마인드(Google Deep Mind), 테슬라(Tesla) 등 첨단산업을 선도하는 기업의 주도로 시작된 인공지능사회로의 변화의 흐름에, 백악관에서는 2016년 10월 “인공지능 국가 연구 개발 전략 계획”, “인공지능의 미래를 위한 준비” 보고서를 통해 정부의 전략을 내세웠다. “인공지능 국가 연구 개발 전략”에서 주목해야 할 점은 ‘윤리적・법적 접근’에 대한 언급을 하였다는 부분이다. 이 보고서에서 “윤리적・법적 및 사회적 원칙에 부합하는 AI설계를 위한 방법뿐만 아니라, AI의 윤리적・법적 및 사회적 의미를 이해하는 연구가 상당히 필요” 밝히며, ① 인간중심의 자동화 원칙, ② 인간을 인식하는 AI를 위한 새로운 알고리즘의 모색, ③ 인간의 역량을 강화하기 위한 AI의 기술 개발, ④ 시각화 및 AI-인간 인터페이스 기술 개발, ⑤ 윤리적 AI 구축, ⑥ 윤리적 AI를 위한 아키텍처 설계 등 세부적인 전략을 제시하였다. 여기서 주목해야 할 부분은 “윤리적 AI의 추축”이다. 이를 위해서 이 보고서에서는 ‘과학적으로 실현 가능한 한계 내에서 연구자들은 현존하는 법률・사회적 규범・윤리와 확실히 일치하는 알고리즘 및 아키텍처를 개발하기 위하여 노력해야 한다’고 밝히며 윤리적인 중요성을 강조하였다. 또한 “윤리적 AI를 위한 아키텍처 설계”의 방안으로 ‘운영 행위에 대한 윤리적 또는 법적 평가를 담당하는 모니터 에이전트와 운영 AI를 구분하는 2단계 감시 아키텍처’를 제안한 점도 시사점을 갖는다.

“인공지능의 미래를 위한 준비” 보고서에서는 인공지능의 규제와 관련하여 ‘현존하는 규정의 목표와 구조로 충분하지만 인공지능기술의 영향을 설명하기 위혜 확실한 조정을 하는 작업이 필요하다고 제시’하며, ‘인공지능의 혁신과 성장을 위한 기회를 만드는 동시에 근본적인 목적과 목표에

중점을 둬야 한다’ 고 밝히며 균형을 강조하였다.

EU는 2005년 시작된 ‘윤리로봇’ 프로젝트를 토대로 유럽로봇연구네트워크에서 2007년에 발표한 ‘로봇윤리 로드맵’ 은 학술적인 권고안으로서의 역할만 수행하였다. 하지만 로봇의 윤리문제에 대하여 국제적으로 만들어낸 첫 번째 결과물이라는 점에서 의의가 있다. 이후 「RoboLaw Project」를 통하여 인공지능에 대한 연구를 통해 도출한 “로봇공학을 규율하는 가이드라인”을 통해 첨단기술에 대한 윤리적·법적 분석을 실시하였고, 윤리적 접근을 강조하며 인간의 능력 발전에 영향을 미칠 수 있는 기술과 부정적인 영향을 미칠 수 있는 기술을 구분할 수 있다고 밝혔다.

중국은 2016년 “로봇산업발전규획”을 발표하며, ‘인공지능은 인류의 생활방식을 개선할 수 있는 중요한 시작점’ 이라는 관점으로, 2020년까지 로봇산업의 건강하고 지속가능한 발전을 추진하는 것을 목적으로 한다. 특히 중국 제조업분야의 새로운 국면을 창출하고, 공업수준을 한 단계 끌어올려, 전 세계에서 중국로봇기술이 선점하는 역할을 목표로 삼고 있다.

일본은 2015년 “로봇 신전략”을 통해 ① 시계의 로봇 기술 혁신 거점, ② 세계 제일의 로봇활용 사회, ③ 세계를 선도하는 로봇신시대에 대한 전략의 3가지 핵심 계획을 발표하고, 이에 대한 세부적인 전략을 발표하였다. 일본에서는 로봇을 사회에 도입이 될 경우 기존의 법률을 준수할 것을 기준으로 삼으면서도, 사회 경제 정세나 법률 등에 비추어 어긋난다고 해석될 경우에는 규제를 정비해야 한다고 한다. 특히 일본 전국에서 로봇 도입·활용을 가속화하기 위해서는 향후 로봇의 활용을 전제로 한 규제 완화 및 규범 정비의 양면의 관점에서 균형 잡힌 규제 개혁을 추진하면서 동시에 소비자 보호의 관점에서 로봇의 안전성을 확보하기 위한 법 제정이 필수적이라고 주장한다.

또한 일본지식재산전략본부의 “지식재산추진계획 2016”에서는 지능정보시대에 정부 차원에서 처음으로 어떠한 지식재산 정책을 가져가야할 것인지에 대한 검토를 하였고, 이는 처음으로 검토하였다는 점에서 의의를 갖는

다. 현재 대부분의 국가에서는 인공지능이 산출한 지식재산 정책에 대한 논의를 진행 중에 있기 때문에 동 계획은 중요한 참고 자료가 될 수 있을 것이다.

제3장 국내 인공지능 관련 윤리 정책과 법제 동향 분석

제1절 인공지능 관련 윤리 정책 동향 분석

1. 로봇윤리헌장 제정을 위한 작업의 사례

(1) 1차 진행 과정 (2007-2008)

- 2007년 3월, 정부 주도로 로봇 관련 산·학·연 전문가와 사회학, 미래학 등 분야의 전문가 13명으로 구성된 실무위원회를 구성, 로봇윤리헌장 제정을 위한 활동 시작
- 2007년 4월, 이탈리아 로마에서 열린 IEEE International Conference on Robotics and Automation (ICRA) ’07 로봇윤리워크숍에서 한국의 로봇윤리헌장 제정에 대한 배경과 계획 발표
- 2008년, 로봇윤리헌장 제정에 대한 법적 근거를 마련하려는 시도로서 「지능형 로봇 개발 및 보급 촉진법」 발의
- 다섯 가지 미래 시나리오를 설정하고, 로봇의 제조자와 이용자에 관한 긍정적, 부정적 키워드를 도출하는 방식으로 로봇윤리헌장 초안 작성 (집필: 차원용)
- 2008년부터 Jim Dator교수의 자문에 따라 철학, 윤리, 종교 분야 전문가를 위원으로 추가 선임하여 초안 내용 검토
- 2008년 12월, 로봇윤리헌장 관련 산학연 발표회 개최 (서울, COEX)

(2) 후속 작업 (2009-2016)

- 2009년 4월-8월, 3개 팀으로 구성된 <로봇윤리헌장 제정을 위한 연구위원회> 운영 (지식경제부-제어로봇시스템학회, 책임: 김대원)
- 2009년 8월 연구결과 발표-토론회 개최
- 2009년 로봇윤리연구 로드맵 수립 (향후 3개년 계획). 이후 2년 이상 사실상의 소강상태 진입 (이 기간 중 학계에서는 로봇윤리의 정립을 위하여 철학적 기초, 특히 로봇 혹은 로봇시스템의 존재론적 지위에 관한 문제가 연구되고 논의되기 시작함)
- 2011년 10월, 한국로봇산업진흥원(KIRIA) 주도로 로봇윤리 공론화 위원회 구성(위원장: 변순용). 2012년 1월 결과보고서 제출
- 2016년 한국로봇산업진흥원, 로봇윤리헌장 제정 작업 속개

(3) 평가적 분석

로마에서 개최된 ICRA 2007에서 한국이 로봇윤리헌장, 즉 로봇윤리에 관한 규범의 체계를 선도적으로 제시하겠다고 공언한 일은 당시 국제적인 관심을 모았다.[29] 결과적으로 이 공언은 원안대로 지켜지지 못했다. 핵심적인 이유는 그 공언 자체가 충분히 숙고되지 않은 성취욕의 산물이었기 때문이라고 할 수 있다. 구체적으로는 로봇윤리를 구축할만한 철학적 토대가 구축되지 않는 등 이를 수행할 선결조건에 대한 인식이 부족했고 그것을 당해 작업 일정에 충족시킬 수 없었기 때문이다.[30]

2010년 이후 로봇윤리는 한국로봇산업진흥원(KIRIA) 업무로 이관되었고, 2016년 말 현재 로봇윤리헌장 관련 연구 사업이 진행되고 있다. (2017년 초에는 해당 사업이 완료될 것으로 보인다.) 이 작업은 기본적으로 2007-2009

29) 많지는 않지만 이 공언을 다룬 외국 언론의 기사도 몇 건 있다. 예를 들어 텔레그라프(Telegraph)지의 기사 "대한민국이 '로봇윤리헌장'을 만든다(S Korea devises 'robot ethics charter')"(2007년 3월 8일자) 참조 . http://www.telegraph.co.uk/news/worldnews/1544936/S-Korea-devises-robot-ethics-charter.html.

30) 고인석(2012b), "체계적인 로봇윤리의 정립을 위한 로봇 존재론, 특히 로봇의 분류에 관하여", 「철학논총」 제70호, 새한철학회, 173면 이하.

년의 로봇윤리헌장 작업의 산물을 출발점으로 삼아, 당시 작업본에 대한 첨삭과 표현 조정을 위주로 하는 소규모의 ‘수정본’ 작업의 성격으로 진행된 것으로 보인다.

이 사업의 최종 산물은 이 보고서 작성 시점에 아직 공개되지 않았지만, 결과적으로 이미 8년 전에 인식되었던 한계를 벗어나는 데 성공하기 어려울 것으로 보인다. 그 한계는 로봇윤리, 혹은 지능을 장착한 인공시스템에 관한 사회적 규범의 정립을 위하여 이 규범 체계가 타겟으로 삼는 ‘로봇’이라는 범주가 어떤 존재론적 특성을 지니는지에 대한 명확하고 통일성 있는 인식이 선결되어야 한다는 조건이 아직 충족되지 않았다는 데 있다. 만일 로봇윤리헌장을 제정한다는 이 사업이 지능형 로봇과 관련한 한국사회의 규범을 인도하는 길잡이 역할을 하도록 기대한다면, 나아가 국제적 시각에서도 유의미한 논의의 대상이 되기를 희망한다면 현행의 접근 방식과는 다른 구조의 체계적인 시도가 필요하다.

2009년 봄 발족한 <로봇윤리헌장 연구위원회>는 그 이전까지 수행되었던 하향식(top-down) 구조의 작업보다 한 단계 진전된 수준의 성과를 산출하였다. 특히 이 연구위원회가 공학, 과학기술정책, 사회학, 철학, 그리고 산업계의 전문가를 포괄적으로 수용하며 확대된 연구진을 활용한 점은 미래학의 상상력을 주된 원천으로 삼았던 1차 작업의 한계를 어느 정도 극복할 수 있는 배경이 되었다. 그러나 이 연구위원회 작업은 약 4개월 뒤인 2009년 8월의 연구 성과 발표회를 끝으로 종료되었고, 이후 유사한 규모나 인적 구성의 계속 사업으로 연결되지 않았다. 바로 앞에서 언급한 “다른 구조의 체계적 시도”란 최소한 이와 유사한 학제적 구조의 연구진을 활용하면서 그 논의 결과를 종합하고 통합하는 방식의 시도를 의미한다.

(4) 세계적 동향과의 비교

유럽은 한국이 로봇윤리헌장 제정 발표를 공언했던 최초 시점에 이미 European Robotics Network(EURON)를 구축하고 로봇의 제작과 활용에 관

한 규범의 체계를 만들기 위한 로드맵을 막 제시한 차였다. EURON은 2007년 1월에 로봇윤리에 관한 로드맵(Roboethics Roadmap Release 1.2)을 공개했는데, 이 문서는 확정된 '규범'의 양식을 제시하고 있지는 않지만 이미 로봇윤리가 고려해야 할 요소들과 만일 글로벌 사회 공동의 규범을 정립한다면 어떤 쟁점들에 대한 논의와 합의가 필요한지에 대한 폭넓은 서술을 포함하고 있다.

미국에서는 이 문제에 주목한 몇 개별 그룹의 학자들이 연구와 집필, 컨퍼런스 등을 통해 로봇윤리의 기초에 관한 학문적 토론을 진행해 왔다. 그 중 특히 주목할 만한 연구 성과의 사례로는 콜린 앨런(C. Allen)과 웬델 월락(W. Wallach)이 공저한 저서 Moral Machines: Teaching Robots Right from Wrong (Oxford UP, 2009)와 앤더슨 부부(Susan Anderson & Michael Anderson)가 편집한 연구서 Machine Ethics (Cambridge UP, 2011), 패트릭 린(P. Lin)과 키쓰 애브니(K. Abney) 등이 편집한 연구서 Robot Ethics: The Ethical and Social Implications of Robotics (MIT Press, 2011), 그리고 데이빗 겅켈(D. Gunkel)의 저서 Machine Questions (MIT Press, 2012) 등이 있다. 이밖에도 플로리디(Luciano Floridi), 갤러거(Shaun Gallagher), 그리고 카플란(Jerry Kaplan)의 연구와 저서 등이 주목할 만하다.

한편 최근 미국에서는 기존의 개별 연구를 뛰어 넘는 공적 성격을 지닌 주목할 만한 몇 건의 인공지능 관련 보고서가 공개되었다.

1) 스탠포드대학교 간행, "인공지능 100년 연구" (Stanford One Hundred Year Study on Artificial Intelligence (AI100). URL=https://ai100.stanford.edu/)

2) 대통령 직속 과학기술협의회, 기술위원회 (Executive Office of the President, National Science and Technology Council, Committee on Technology) "Preparing for the Future of Artificial Intelligence" (2016. 10)

3) 국가 과학기술위원회(National Science and Technology Council, Networking and Information Technology Research and Development

Subcommittee), “The National Artificial Intelligence Research and Development Strategic Plan” (2016. 10)

인공지능의 발달과 인공지능 활용의 확대에 대비하는 국제 사회의 움직임은 최근 눈에 띄게 활발해졌고 대응의 템포도 빨라졌다. 특히 그 대응에서 사회윤리적 차원에 대한 고려가 중요한 비중을 차지하고 있는 점을 주목할 만하다. 2016년 12월 IEEE가 발표한 윤리적 인공지능 시스템 개발을 위한 지침 “Ethically Aligned Design” (윤리적으로 조율된 [공학적] 설계)의 초안은 당분간 중요한 준거점의 구실을 할 것으로 예상되는데[31], 이 문건은 가장 기본적인 원칙으로 인권의 이상을 실현하는 일, 공학기술이 인류와 자연환경에 어떤 유익을 줄 것인지에 대한 고려, 인공지능이 사회적 기반 기술로 발달하는 과정에 수반되는 위험과 부정적 효과를 최소화하는 일 등을 명시하고 있다.

이와 대조적으로 한국 사회, 특히 정부가 그 형성을 주도하고 있는 공공의 관심은 여전히 인공지능의 경제적 가치와 파급효과에 편중되어 있고, 인공지능이나 인공지능 로봇이 가져올 윤리적 문제나 그 대응책에 대한 관심은 아직 미미하고 파편적인 수준에 머물러 있다. 2016년 12월 28일 보도 자료로 배포된 『지능정보사회 도래에 대비한 중장기 국가전략 수립』에는 정부의 6개 부처가 관여했지만, 경제적 효과의 차원이 집중 조명되고 있는 반면 윤리적 차원에 대한 고려는 미미하다.[32] 이 문건에서 국가전략은 [기술]→[산업]→[사회]로 이어지는 3단계 전략을 제시하고 있는데, 윤리적 고려는 세 번째 단계인 [사회]에서 피상적으로 다루어지고 있을 뿐이다. 이러한 사정은 앞서 언급한 미 국가과학기술위원회의 「국가 차원 인공지능 연구개발 전략 계획」이 연구개발의 초기 단계에서부터 인공지능 기술의 윤리적, 법적, 사회적 함의를 중시하고 있는 것과 대비된다.[33]

---

31) http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
32) http://www.msip.go.kr (2016년 12월 29일 보도자료). 이 자료에 따르면 지능정보기술로 인해 국내 경제 효과가 2030년 기준 460조 원에 이르며 80만 개의 신규 일자리가 생길 것으로 전망된다.
33) https://www.nitrd.gov/news/national_ai_rd_strategic_plan.aspx

2. 기타 최근의 정책적 시도

2016년 6월 한국정보화진흥원(NIA)은 <정보문화포럼>(위원장: 이건)을 창설하고 <지능정보사회 윤리> 분과를 구성하여 (분과장: 김명주) 지능정보사회를 대비한 사회적 관점의 대비 작업을 시작했다. 이보다 앞서 NIA는 소프트웨어공학, 사회학, 철학, 정책 분야의 전문가들을 구성원으로 하고, 지능정보사회의 윤리와 관련한 가이드라인 작성을 목표로 하는 "지능정보사회 윤리 이슈 전문가 모임"을 운영하기 시작했다. 이 모임은 앞서 언급한 정보문화포럼 창설 이후 포럼의 지능정보사회 윤리 분과로 편입되었다.

이밖에도 2016년에는 미래부와 산업통상자원부 등의 정책연구뿐만 아니라 교육부-한국연구재단 지원의 연구과제에서도 인공지능 기술의 사회적-윤리적-인문학적 함의를 조명하는 다양한 작업이 동시다발적으로 시작되었다. 이는 자연스러운 현상인 동시에 사회적으로 필요한 과정이기도 하지만, 이러한 작업들을 국가사회 차원의 거시적 관점에서 보면 향후 이러한 개별 연구와 협의의 네트워크를 통할하고 그 성과를 종합하는 중심 혹은 조직화된 노드들(nodes)이 필요할 것으로 생각된다.

제2절 인공지능 법제 동향 분석

1. 기존 법제

인공지능과 관련된 기존 법령 중 대표적인 법령은 「지능형 로봇 개발 및 보급 촉진법」이다. 이 법은 지난 2008년, 첨단기술의 융합체인 지능형 로봇에 대하여 국가가 체계적으로 연구·개발하고, 초기시장의 창출과 보급 확대를 위한 정책으로 로봇랜드를 조성하며, 로봇품질의 인증에 필요한 근거를 마련하고, 로봇윤리헌장의 제정과 보급을 통하여 로봇이 반사회적으로 개발·이용되는 것을 방지하도록 하는 등 차세대 성장 동력산업인 지능형 로봇을 미래 국가핵심 전략산업으로 육성하기 위한 제도적 기반을 구축하기 위하여 제정한 법이다.[34] 지식경제부장관의 기본계획 수립(제5조), 지능형 로봇제품의 보급 촉진(제9조에서 제13조까지), 로봇윤리헌장의 제정(제18조), 로봇랜드의 조성(제30조에서 제40조까지), 한국로봇산업진흥원의 설립(제41조), 지능형로봇전문연구원의 지정 근거 조항(제42조) 등을 주요 내용으로 한다. 이 법에서 '지능형 로봇'이란 "외부환경을 스스로 인식하고 상황을 판단하여 자율적으로 동작하는 기계장치"(제2조 제1호)로 정의되는데, 여기서 '기계장치'를 수식하고 있는 "외부환경을 스스로 인식하고 상황을 판단하여 자율적으로 동작하는" 내용이 인공지능에 해당한다. 즉, 이 법은 인공지능이 기계장치와 같은 형상에 결합된 것을 지능형 로봇이라고 정의하고, 관련 산업진흥정책을 추진하기 위한 근거를 마련하기 위한 것이다. 이미 설명하였던 산업자원부에서 로봇윤리헌장을 제정을 위한 노력을 위한 법적 근거도 제18조에 담았다. 이러한 의미에서 이 법은 소프트웨어산업의 진흥에 필요한 사항을 정하여 소프트웨어산업 발전의 기반을 조성하고 소프트웨어산업의 경쟁력을 강화하기 위한 「소프트웨어산업진흥법」,

34) 이상 법제처 국가법령정보센터의 제정문(http://www.law.go.kr/lsRvsRsnListP.do?lsiSeqs=179085%2c167690%2c154036%2c150062%2c149930%2c137002%2c122438%2c115293%2c105485%2c104527%2c104218%2c104096%2c94612%2c90237%2c86607&chrClsCd=010102)

정보통신산업의 진흥을 위한 기반을 조성하기 위한 「정보통신산업진흥법」 등 일련의 산업진흥법제의 특별법이며, 국가정보화의 기본 방향과 관련 정책의 수립·추진에 필요한 사항을 정하기 위한 「국가정보화기본법」, 정보통신을 진흥하고 정보통신을 기반으로 한 융합의 활성화를 위한 정책 추진 체계, 규제 합리화와 인력 양성, 벤처육성 및 연구개발 지원 등을 정하기 위한 「정보통신 진흥 및 융합 활성화 등에 관한 특별법(ICT 특별법)」 등 일련의 정보화 기본법제의 구체화법이다.

그리고 지능형전력망의 구축 및 이용촉진을 함으로써 관련 산업을 육성하고 전 지구적 기후변화에 능동적으로 대처하며 저탄소 녹색성장형 미래 산업의 기반을 조성하여 에너지 이용환경의 혁신과 국민경제의 발전에 이바지하기 위한 「지능형전력망의 구축 및 이용촉진에 관한 법률」도 있다. 인공지능을 활용한 특정 체계를 구축하고 이용을 활성화하기 위한 특별법이다.

한편, 인공지능이 초래하는 법적 분쟁에 관해서는 「민법」, 「형법」 등 기존 법제가 그대로 적용된다. 예를 들어, 자율주행자동차(automated car; autonomous car)나 드론(drone), 수술로봇(surgical robots) 등이 타인의 신체나 재산의 손해를 끼치면 그 관리자나 제조자에게 민사상 손해배상 책임을 물을 가능성이 있고, 형사 처벌도 문제될 수 있다. 보충적으로 이러한 경우 「제조물책임법」에 따라 제조물책임을 물을 가능성도 있다.

2. 최근 논의

위와 같은 기존 법제와 별도로 미래창조과학부에서는 최근 「(가칭) 지능정보사회 기본법」을 제정하기 위하여 노력하고 있다. 미래창조과학부는 지난 1월 6일 황교안 대통령 권한대행에게 「2017년도 미래부 업무 추진계획」을 보고하였는데, 그 중 하나가 지능정보사회 기본법의 제정이다. 미래부는 이 계획에서 네 가지 전략 중 하나로 지능정보화로 제4차 산업혁명 선제적 대응을 제시하고, “지능정보화 방향 제시를 위해 「지능정보화기본

법」 마련을 추진 하겠다” 고 하였다.[35]

이와 별도로 새누리당 강효상 의원(국회 미래창조과학방송통신위원회)은 지난 2016년도 미래창조과학부 국정감사에서 「지능정보사회기본법」을 제시하고, 2016년 12월 ‘지능정보사회기본법 제정을 위한 전문가 토론회’,[36] 2017년 1월 24일 ‘미래혁명, 지능정보사회로 가는 길’이라는 세미나를 개최하였다. 이 법안에는 지능정보사회에 대응하는 국가전략의 수립과 실천을 위한 추진체계, 그리고 지능정보사회의 역기능에 대한 대안 등이 담겨있다고 알려져 있다. 구체적으로는 지능정보사회 대응 총괄 기구로 지능정보사회위원회를 설치하도록 규정하고 있다. 이 위원회는 대통령 직속 합의제 기구로 위원장을 포함하여 12명으로 구성되며, 지능정보사회 기본계획을 수립하고 행정 각부의 관련 정책을 조정하는 컨트롤타워 역할을 맡는다. 지능정보사회 기본계획은 지능정보사회 정책의 기본방향, 지능정보기술의 확산과 활용, 지능정보기술로 인한 위험의 방지, 지능정보기술로 인한 차별 방지와 인권보장, 지능정보 서비스 이용자의 권익과 지적재산권 보호 등을 담도록 하였다. 위원회는 이 기본계획 실행 과정에서 사업자와 이용자 등 이해관계자의 갈등을 조정하는 기능도 수행하도록 하였다. 또한 지능정보기술 서비스 제공자가 이용자의 사생활 침해 방지 등 권리를 보호하기 위해 기술 및 서비스 개발 단계에서부터 기술적 보호조치를 하도록 규정하였다.[37]

35) 미래부 홈페이지에 게시된 보도자료, http://www.msip.go.kr/web/msipContents/contentsView.do?cateId=mssw311&artId=1323073 (2017년 1월 30일 최종 방문)
36) 김현아, 이데일리 기사, http://www.edaily.co.kr/news/NewsRead.edy?SCD=JE41&newsid=01856486612875240&DCD=A00504&OutLnkChk=Y (2017년 1월 30일 최종 방문)
37) 이상 강효상 의원 안에 관한 동향과 법안 내용은 채희정, 비즈트리뷰 기사, http://www.biztribune.co.kr/n_news/news/view.html?no=16222 (2017년 1월 30일 최종 방문)

제3절 소결

이미 살펴본 것처럼 우리 행정부는 지난 2007년부터 행정부 주도로 로봇윤리헌장을 제정을 위한 노력을 전개하였다. 그 목표 중 하나는 IEEE 국제컨퍼런스에서 한국이 로봇윤리헌장 제정을 제안하는 것이었다. 그리고 그러한 정책 추진을 뒷받침하기 위하여 「지능형 로봇 개발 및 보급 촉진법」도 제정하였다. 그러나 그것은 실패로 돌아갔다. 로봇윤리를 정립할 철학적 토대의 부재, 전시 행정을 위한 과도한 행정부 주도 작업 경향, 그로 인한 전문가 경시 경향, 기술자와 사회과학자, 정책입안자 등의 협업 부족 등이 주요 원인이었다.

2008년 「지능형 로봇 개발 및 보급 촉진법」의 제정에도 불구하고, 국내 로봇 산업은 다른 ICT 산업에 비하여 진흥되지 못하였으며, 관련 연구 성과도 축적되지 못하였다. 그리고 수년이 흐른 지금, 2016년 이른바 알파고 현상이 발생하자 대대적으로 「지능정보사회 중장기 종합대책」을 발표하고, 「지능정보사회기본법」을 제정하겠다고 한다. 그러나 행정부 주도의 진흥 정책을 추진하기 위한 법적 근거는 이미 마련되어 있고, 현재 기술 수준을 고려하여 보았을 때 법령을 통한 규제 정책은 너무 이르다. 인공지능 연구와 그 응용을 위한 개별적인 법적 장애가 무엇인지는 현장에서 이미 제기되었고 그 해결책도 제시되고 있다. 한편, 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하기 위한 구체적인 정책이 필요하다.

제4장 윤리 정책과 법제의 대안

제1절 사회문제 대응을 위한 법과 윤리의 기능 분담[38)]

1. 인간행위의 규율수단

사회문제가 생겼을 때 이를 해결하기 위해 입법자와 행정부 공무원이 동원할 수 있는 수단은 법, 윤리와 같은 법 이외의 사회규범, 교육·홍보, 시장, 아키텍처와 아키텍처를 결정하는 기술 등이 있다.

예를 들어, 아이들이 음란물을 보는 것은 성장에 나쁜 영향을 미치고 또 다른 사회적인 문제점을 야기하는 것으로 일반적으로 알려져 있다.[39)] 따라서 아이들이 음란물을 보는 것을 사회적으로 규제할 필요가 있다. 더욱이 인터넷이 발달하여 이를 통해 음란물이 광범위하게 유포되면서 이 문제를 어떻게 해결할 것인가가 더욱 사회적인 문제가 되고 있다.

이러한 문제를 해결하기 위해서는 아이들이 음란물을 보는 것에 영향을 미치는 제약요인을 찾아 적절히 대응하는 것이 필요하다. 이러한 요인에는 다음과 같은 것들을 생각해 볼 수 있다.

가장 먼저 법(law)을 생각해 볼 수 있다. 우리나라는 현재 형법 제243조, 정보통신망 이용촉진 및 정보보호 등에 관한 법률 등에 의하여 음란물 유통이 규제된다.

그 다음으로 윤리와 같은 법을 제외한 다른 사회규범(norm)을 생각해 볼 수 있다. 법규범에 의한 처벌 이전에 사회는 전반적으로 아이들에게 음란물을 파는 사람에게 돈벌이를 위해 도덕적으로 타락한 행위를 하고 있다는 비난을 가할 것이다. 이러한 사회적 시각이 음란물 판매를 억제하는 것은 의심의 여지가 없다.

---

38) 이에 관해서는 "정필운(2009), "정보사회에서 지적재산의 보호와 이용에 관한 헌법학적 연구: 저작물을 중심으로", 연세대학교 대학원 법학과 박사학위논문"을 수정하고 보완하여 작성한 것이다.

39) 이하의 예는 로렌스 레식 저, 김정오 역(1999), 「코드: 사이버 스페이스의 법이론」, 나남, 391면 이하 참고.

부모나 학교 선생님들은 아이들이 음란물을 보는 것에 대해 죄의식을 심어준다. 대부분의 아이들은 음란물을 구하는 과정이나 시청과정에서 죄의식을 느끼게 된다. 또한 사회규범은 아이들을 음란물로부터 보호하는 정책을 뒷받침한다.

사회가 점점 전문화되고 복잡해져서 이러한 법과 법을 제외한 다른 사회규범을 교육하고 홍보하는 것도 독립된 하나의 기능의 되었다.

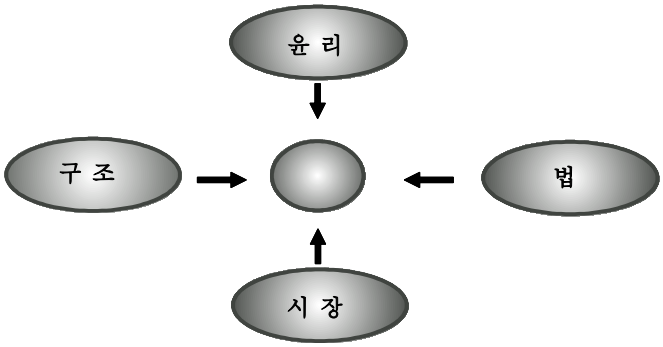
우리나라에서는 여성가족부와 방송통신심의위원회가 아이들이 유해정보를 접하지 않도록 정보환경을 조성하고 법과 법을 제외한 다른 사회규범을 교육·홍보는 기능을 담당하고 있다.

시장(market) 역시 아이들이 음란물을 접하는 것을 제약하는 요인 중 하나이다. 시장에서 음란물은 하나의 상품으로 유통된다. 이 상품을 구입하기 위해서는 돈이 필요하다. 음란물 가격이 음란물 시청에 제한을 가하는 것이다. 대체로 아이들은 자신이 처분할 수 있는 많은 돈을 갖고 있지 못하다. 음란물 판매자들은 지불능력에 따라 차별하기 때문에 결과적으로 아동들이 음란물을 구입하지 못하도록 협조하는 것이다.

마지막으로 아키텍처(architecture)를 생각해 볼 수 있다. 현실공간에서 아이들은 자신을 어른인 것처럼 위장하고 음란물을 구입하는 것이 매우 어렵다. 대면구조를 취하는 현실공간은 어린이들이 음란물을 구입하여 시청하는 것을 비교적 효과적으로 규제하게 된다.

이렇게 사회적으로 바람직하지 못하다고 판단되는 사회문제가 발생했을 때 그것에 영향을 미치는 제약요인과 우리가 동원할 수 있는 규율수단은 법, 법을 제외한 다른 사회규범, 교육·홍보, 시장, 아키텍처 등을 생각해 볼 수 있다. 따라서 우리가 사회적으로 바람직하지 못하다고 평가되어 우리의 규율을 필요로 하는 사회문제를 접하는 경우에는 늘 이와 같은 규율수단을 고려해야 할 것이다.

[그림 4-1] 사회문제를 해결하는 규율수단



현실공간에서의 규율수단과 사이버 스페이스에서의 규율수단은 규율에 차이가 있다. 우선 법을 생각해 보자. 위에서 서술한 바와 같이 현행 형법은 제243조에서 음란물 유통을 금지하고 있다. 그러나 이 형법 제243조로는 사이버 스페이스에서 유통되는 음란물을 규율할 수 있을지가 의문스러웠고, 실제로 우리 대법원은 "형법 제243조는 음란한 문서, 도화, 필름 기타 물건을 반포, 판매 또는 임대하거나 공연히 전시 또는 상영한 자에 대한 처벌 규정으로서 피고인들이 판매하였다는 컴퓨터 프로그램파일은 위의 법에서 규정하고 있는 문서, 도화, 필름 기타 물건에 해당한다고 볼 수 없"어 형법 제243조는 적용할 수 없다고 하여, 형법 제243조로는 사이버 스페이스에서 유통되는 음란물을 규율할 수 없다고 판시했다.[40] 이러한 입법의 흠결을 막기 위해 우리 입법부는 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제65조 제1항 제2호를 규정하여 사이버 스페이스를 통해서 유통되는 음란물에 대처하고 있다.

그 다음으로 법을 제외한 사회규범의 측면을 보자. 사회규범의 측면에서 보면 근본적인 차이는 없어 보인다. 차이가 있다면 인터넷을 통해 음란물이 널리 널려 있고, 이로 인해 어른이나 아이들이나 음란물에 대한 경계심이 느슨해졌다고 생각할 수 있다. 따라서 음란물을 파는 사람이나 아이들의 죄의식이 다소 약해졌다고 추측해 볼 수도 있다.

40) 대판 1999. 2. 24. 98도3140.

교육·홍보의 측면도 근본적인 차이는 없어 보인다. 차이가 있다면 사회 구성원에게 교육과 홍보의 수단으로 인터넷이라는 간편하고 저렴하면서도 효과적인 수단을 이용할 수 있다는 것이다.

시장이라는 제약요인의 변화는 위의 두 변화보다 좀 더 크다. 정보통신기술의 발달로 음란물의 복제비용이 거의 들지 않는다. 유통비용도 현저히 적게 든다. 따라서 시장에서 음란물의 가격은 굉장히 싸졌다. 과거에 포르노 비디오테이프 한 개 값이면 이제는 포르노 사이트에 가입하여 한 달 동안 그 사이트의 모든 음란물을 실컷 볼 수 있다.

이렇게 복제와 유통비용이 거의 없기 때문에 웬만한 음란물은 공짜로 구할 수도 있다. 따라서 시장이라는 제약요인은 이 문제에 있어 제약의 정도가 현저히 약화되었고, 따라서 효과적인 규율수단이 되지 못하게 되었다.

마지막으로 아키텍처를 생각해 보자. 현실공간에서 아이들의 음란물 구입을 규제하는 가장 결정적 제약요인은 아이들이 자신이 성인이라는 것을 위장하기 어렵다는 점이다. 그러나 사이버 스페이스에서는 이러한 구조가 근본적으로 변화하였다. 초기 사이버 스페이스에서 아이들은 자신의 신분을 밝힐 필요조차 없었다. 그 결과 우리 모두가 알고 있는 것처럼 아이들이 포르노에 무방비로 노출된 것이었다.

사회는 이러한 문제에 대해 먼저 신원확인제도를 도입하였다. 음란물을 취급하는 사이트에 들어가기 위해서는 자신이 성인이라는 것을 확인받고 들어가도록 한 것이다. 그러나 로그인 방식과 같은 신원확인제도는 아이들이 아버지나 어머니의 신원을 이용하여 자신이 성인인 것처럼 속여 음란물에 접하는 것에는 효과적이지 못했다. 그리하여 수호천사와 같은 응용소프트웨어를 만들어 아이들이 음란물에 접근하는 것을 전보다 효과적으로 봉쇄할 수 있게 되었다.

위와 같은 예는 정보통신 영역에서 사회문제가 발생했을 때 우리가 어떻게 그것을 해결해야 할 것인지를 알려준다. 법학자나 입법자, 행정부 공무원은 사회문제가 발생하면 그것을 법이라는 수단을 통해서 해결하려고 한

다. 하지만 사회문제를 효율적으로 해결하기 위해서는 법이라는 요인뿐만 아니라 이 문제에 영향을 끼치는 다른 요인들도 고려하여 이를 규율수단으로 활용해야 한다. 법은 그러한 수단 중 단지 하나일 뿐이다. 특히 사이버 스페이스에서는 위에서 말한 '아키텍처'와 그 아키텍처를 결정하는 '기술'에 주목해야 한다.[41]

사이버스페이스법학 연구로 유명한 미국 하버드 로스쿨(Harbard Law School)의 로렌스 레식(Lawrence Lessig) 교수는 사이버 스페이스의 법학적 연구와 대안제시를 위한 전제로 두 가지 코드(Code)의 개념을 제시한다. '동부 연안 코드'와 '서부 연안 코드'가 바로 그것이다. '동부 연안 코드'란 의회가 제정한 성문법을 의미하며, '서부 연안 코드'란 사이버 스페이스를 작동하게 하는 소프트웨어와 하드웨어 내부에 새겨져 있는 명령을 의미한다. 그의 이러한 비유는 성문법 제정이 주로 미국의 동부 연안에 있는 워싱턴을 중심으로 이루어지고, 기술적 코드 제작은 주로 미국의 서부 연안에 있는 실리콘밸리, 레드먼드 등을 중심으로 이루어지는 데 착안한 것이다.[42]

사회문제를 해결하고자 할 때 현실공간에서는 단지 법이라는 것에 초점을 둠으로써 많은 문제들을 해결할 수 있었지만, 정보통신기술의 발전에 기반한 정보사회에서는 기술적인 변화까지도 고려해야만 한다는 것이다. 그에 따르면 사이버 스페이스에서는 기술적 요소인 '코드'라는 것이 법과 유사한 역할을 한다.

위의 주장은 우리나라에도 그대로 타당하다. 다만 '동부 연안 코드'와 '서부 연안 코드'는 미국의 상황을 고려한 표현이므로, 이 논문에서는 '규범적 코드'와 '기술적 코드'라는 표현으로 대신하고자 한다.

이와 같은 점이 바로 현실공간의 법적 기반과 사이버 스페이스의 법적 기반의 차이점이다. 같은 사회문제라도 그것이 작동하는 공간이 현실공간이냐 사이버 스페이스냐에 따라 규제요인인 법, 법을 제외한 사회규범, 교육·홍

41) 로렌스 레식 저, 김정오 역(1999), 앞의 책, 206-207면.
42) 로렌스 레식 저, 김정오 역(1999), 위의 책, 134면 이하.

보, 시장, 아키텍처, 그리고 아키텍처를 결정하는 기술 등이 그 정도에 있어 차이를 보일 뿐만 아니라 다르게 작동한다. 특히 사이버 스페이스에서는 아키텍처와 아키텍처를 결정하는 기술이라는 규제요인을 주목해야 한다.

이러한 규범적 코드와 기술적 코드는 각각 독자적 영역에서 자율적으로 움직이는 것이 아니다. 이들은 각각 사이버 스페이스에서 제약요인으로 작동하면서 상호영향을 주고받는다. 이것이 시사하는 바는 다음과 같다.

우리가 어떠한 법적 문제를 해결하고자 할 때 법학의 전통적 관점에 따르면 법적인 차원에서 보호수준을 결정하여 입법을 하거나 법률을 해석할 것이다. 그리고 이러한 입법이 입법재량권을 벗어난 예외적인 경우에만 헌법적 심사를 할 것이다.

그러나 위와 같은 견해에 따르면 이러한 경우 문제해결은 오로지 규범적 코드에 초점을 맞춘 채 이루어질 수 없다. 제기된 문제에 대한 올바른 처방을 찾기 위해서는 규범적 코드와 아키텍처 또는 아키텍처를 결정하는 기술적 코드 사이의 상호작용을 고려하여야 한다. 그러한 상호작용 속에서 규범적 코드가 작동하는 것이다.

위에서 살펴본 바와 같이 규범적 코드와 기술적 코드가 사이버 스페이스에서 적용되는 '법'이고 규범적 코드를 확립할 때 기술적 코드를 고려해야 된다면, 사이버 스페이스에서 제대로 작동할 수 있는 법적 규율체계를 확립하기 위해서는 그 대상인 의사소통 시스템 구조를 분석하고 이에 근거하여 법적 규율체계를 확립할 필요가 있다. 이러한 한 예가 이른바 레이어 모델이다.

요차이 벤클러(Yochai Benkler) 교수는 의사소통 시스템에 대한 분석을 위해 '물리적 레이어', '논리적 레이어', '콘텐츠 레이어'라는 개념을 사용한다.[43] 이러한 레이어 구분을 이용하여 정립한 법적 규율체계를 일반적으로 '레이어 모델(layer model)'이라고 한다. 이러한 레이어 모델은 일반적인 의사소통 시스템의 레이어 구분인 개방형 시스템간 상호접속 기본

43) Yochai Benkler(2000), "From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access", *Federal Communications Law Journal 52* 참고.

참조 모델(OSI 7 Layer Model; Open System Inter- connection 7 Layer Model)과 팀 버너스 리(Tim Berners-Lee)의 4단계 레이어 모델[44]을 규범적 분석을 위해 단순화한 것이다. 레이어 모델에 따르면 인터넷은 다음과 같은 세 개의 레이어로 구성된다.

첫째, 물리적 레이어(physical layer)란 컴퓨터와 컴퓨터들을 연결시켜주는 정보통신망 등이 존재하는 계층을 말한다. 컴퓨터와 정보통신망은 각각의 소유자가 있고, 따라서 그 소유자들에 의해 통제된다. 그리고 이러한 소유권 귀속과 그 행사의 내용을 정하는 국가에 의해 통제된다. 따라서 이러한 물리적 레이어는 현실공간의 재산권을 형성하는 법률인 '규범적 코드'에 의해 규율된다.

둘째, 논리적 레이어(logical layer)란 정보통신망(net)을 움직이게 하는 프로토콜(protocol)이 존재하는 계층을 말한다. 논리적 레이어는 최초에는 '최종이용자에서 최종이용자까지'라는 원칙에 입각하여 설계되었기 때문에 자유로웠으나, 일련의 기술적 규율이 추가되면서 점차 통제되고 있다. 레식 교수는 이러한 경향을 광대역 통신망 기술 발달의 예를 들어 설명한다. 그에 따르면 초기 인터넷 네트워크는 누구든지 기본 사용료만 지불한다면 사용할 수 있는 공간이었으나 광대역 통신망 사업자들은 투자수익 회수 등을 위한 목적으로 논리적 레이어의 프로토콜들을 변화시키고 있다. 이렇게 논리적 레이어를 지배하는 코드는 변경할 수 없는 주어진 것이 아니다.[45] 따라서 초기에 개방적이었던 기술적 코드인 프로토콜은 폐쇄적이고 차별적인 규칙들로 변화하면서, 통제되는 공간이 되어가고 있다. 이러한 레이어의 변화는 기술적 코드에만 한정되는 것이 아니고 규범적 코드에 의하여 뒷받침된다. 이 영역에서는 규범적 코드와 기술적 코드가 상호 보완하며 규율하는 것은 분명하지만 종국적으로 가장 중요한 규율요소는 기술적 코드이다.

---

44) 전송 계층, 컴퓨터 하드웨어 계층, 소프트웨어 계층, 콘텐트 계층으로 구분된다. 이에 관해 자세한 것은 Tim Berners-Lee(1999), "Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventer", *Harper San Francisco*, pp.129-130.

45) 로렌스 레식 저, 정필운.심우민 역(2005), "혁신의 구조",「연세법학연구」제11권 제1·2호, 연세법학회, 304면 이하.

셋째, 콘텐츠 레이어(contents layer)란 주어진 특정 의사소통 시스템 안에서 말해지거나 씌여지는 내용들이 존재하는 레이어이다. 이 레이어는 자유와 통제가 혼재되어 있다. 이 층위에서 가장 문제되는 것은 단연 지적재산권이다. 이 층위도 역시 규범적 코드와 기술적 코드가 상호 작용한다. 규범적 코드의 예로는 각종 저작권 관련 법령을 들 수 있고, 기술적 코드로는 암호화(cryptography), 신원인증, 워터마킹(watermarking) 등을 예로 들 수 있다. 이러한 콘텐츠 레이어에서 일어나고 있는 갈등은 과거 냉전체제의 이데올로기적인 갈등과 같이 첨예하다.

정보통신영역에서 행해지는 행위규제수단의 선택에서 레이어 모델을 적용하면 각 레이어별 특성을 고려한 적절한 규율수단을 선택할 수 있다는 장점이 있다. 또한 정보통신사업자의 규율에 레이어 모델에 입각한 규율시스템을 적용하면 융합(Convergence)현상이 가속화되는 현 시점에서 규제의 공백, 중복과 같은 혼선을 피하고 규제의 일관성을 유지할 수 있다는 장점도 가지고 있다.

〈표 4-1〉 레이어 모델에서 각 레이어의 특징

구분	물리적 레이어	논리적 레이어	콘텐트 레이어
통제 또는 자유	통제	원래 자유, 점차 통제	통제와 자유가 혼재
규율수단	규범적 코드	기술적 코드〉 규범적 코드	기술적 코드= 규범적 코드

인공지능과 관련하여 최근 주장되고 있는 알고리즘 규제(algorithmic regulation)[46]도 이와 같은 레이어 모델에 의하여 설명할 수 있다. 인공지능이 기능하기 위해서는 자신의 과업을 수행하는 방법에 대한 명령 규칙인 알고리즘이 필요하다. 이 알고리즘은 설계자의 목적에 따라 설계되는 것이므로 그에 따른 처리 결과는 본질적으로 중립적이지 않다. 알고리즘이 기술적 코드로서 작용하는 것이다. 따라서 국가는 인공지능 규제를 하면서 이를

46) 심우민(2016b), “인공지능의 발전과 알고리듬 규제의 가능성”, 「인간중심적 사고와 작별(2016년 법과사회이론학회 하반기 학술대회 자료집)」, 법과사회이론학회, 25면 이하.

감안하여야 하고, 그에 대한 투명한 공개를 강화하고, 심각한 공익이 침해되는 알고리즘 설계 행위를 제한하는 규범적 코드를 보완하여야 한다는 것이다.[47]

이러한 레이어 모델을 통한 규제모델의 실제적 적용에서 한 가지 주의할 점은 구체적 법적 문제가 어느 하나의 특정한 레이어에 한정해서 발생하는 것이 아니라는 점이다. 어떤 법적 문제는 어느 한 레이어상에서만 문제가 발생하여 그 레이어상에서의 규율만으로 문제해결이 가능한 경우도 있지만, 대부분 문제들은 둘 또는 셋 모두의 레이어상에서 각각 문제가 발생하고 그 모두에서의 규율을 필요로 한다. 따라서 위에서 제시한 레이어 모델을 실제적으로 적용하여 커뮤니케이션 시스템의 문제에 대한 법적 규율책을 마련하기 위해서는 각 레이어별로 나타날 수 있는 문제점을 찾아내고, 당해 레이어의 특성에 맞는 적합한 규율수단을 동원하는 지혜가 필요하다.

2. 변화하는 현실에 적응할 수 있는 법 정립을 위한 원칙[48]

기술법의 큰 과제 중 하나는 급격히 변화하는 기술적 환경과 이에 따라 급변하는 사업적 환경에 잘 적응하여 규범력을 유지할 수 있는 법을 정립하는 것이다. 이렇게 변화하는 기술적 환경과 사업적 환경에 잘 적응하여 규범력을 유지할 수 있는 법을 정립하기 위해서는 다음과 같은 몇 가지 사항을 고려하여야 한다.

첫째, 입법자는 기술적으로 중립적인 입법을 하여야 하는 기술중립성의 원칙(the principle of technology neutrality)을 준수하여야 한다. 그 이유는 첫째, 어떤 특정한 기술에 편향하여 법을 제정하면 기술이 발전함에 따라 그러한 법은 환경에 뒤쳐지게 되어 규범력을 상실하기 때문이다. 둘째, 기술편향적 입법은 평등 원칙에 반하기 때문이다. 특정한 기술에 편향된 입법은 채택된 기술의 개발자와 제공자에게 이익을 주며, 채택되지 않은 기술을

47) 이상 심우민(2016b), 앞의 발표자료 참고.
48) 정필운(2008), “사이버공간 규율 입법의 일반이론 정립을 위한 탐색적 연구”, 「인터넷법률」 제41호, 법무부, 71면 이하.

개발한 자나 이러한 기술을 이용하여 서비스를 제공하는 제공자를 차별한다. 셋째, 시장을 편향적으로 이끌어 궁극적으로 기술 혁신을 저해하기 때문이다.

둘째, 입법자는 예측가능하고, 최소한이고, 일관성 있고 간소한 법을 만들어야 한다.

셋째, 기술도입기와 기술발전기에 법을 제·개정하는 경우에는 기술발전에 앞서 법을 제·개정하는 것에 매우 신중해야 한다. 어느 누구도 기술이 어떻게 발전할 것인지 예측하는 것이 어렵기 때문이다.

제2절 윤리적인 관점의 분석과 쟁점 연구

본 절에서는 인공지능 관련 정책과 법제의 이론적 기반 구실을 하는 동시에 정책과 법제를 통한 규율을 보완하는 윤리적 관점의 접근에 대하여 논한다. 특히 이 절은 ① 향후 인공지능 관련 논의에서 중요한 비중을 차지하게 될 법적 · 사회적 · 윤리적 책임의 문제를 해결하기 위한 전제 조건인 인공지능의 존재론적 지위 문제를 다루고, ② 인공지능과 관련된 윤리적 문제에 기존 공학윤리의 원칙들을 활용하는 방안을 서술하며, ③ 인공지능의 활용이 초래할 것으로 예견되는 사회적 공정성의 문제에 관하여 간략히 논한다.

1. 인공지능의 존재론적 지위

(1) 문제의 확인

제3장 제1절에서 언급한 로봇윤리헌장 작업의 경우에서도 확인된 바는 인공지능에 관한 정책과 법제, 그리고 사회윤리의 틀을 정립하기 위하여 인공지능의 존재론적 지위(ontological status)와 속성에 대한 견해의 정립이 선결문제로 요구된다는 것이다. 인공지능의 존재론적 지위란 “인공지능이란 무엇인가?”, “우리는 그것을 무엇으로(as what) 대할 것인가?” 하는 물음에 대한 답변을 가리킨다. 이 물음에 대한 해명은 “인공지능은 어떤 것인가?”, 즉 현실 세계에서 작동하고 있거나 원칙적으로 그럴 수 있는 개별 인공지능이 ‘인공지능임으로써’ 가진 범주적 속성에 대한 탐구를 토대로 가능하다.

(가) 인공지능이라는 개념

개념적 차원에서 먼저 살펴두어야 할 것은 인공지능이라는 개념어의 애매성 혹은 다의성이다. 우선, ‘인공지능(artificial intelligence)’은 그 표현의

본래적인 의미에서 ‘공학적으로 지능을 만들어내는 활동’ 또는 그런 활동이 진행되는 전문분야를 가리킨다.[49] 동시에 우리는 그 말로써 그와 같은 활동이 만들어낸 구체적인 산물, 즉 인공적으로 생산된 지능을 가리키기도 한다. 그러나, 적어도 이론적인 관점에서, 이 말의 우선적인 의미가 전자에 있음을 확인해 둘 필요가 있다. 그 우선성의 이유는 인공적으로 구현될 지능의 속성이 근본적으로 열려 있다는 데 있다. 중심에는 생물학적 종(種)으로서의 인간이 지닌 지능을 닮았다는 속성이 자리 잡고 있기는 하지만, 그것 역시 인공지능이라는 전문 분야가 구현할 지능의 필연적인 속성은 아니다. 최근 인공지능 분야의 동향은 이런 관계를 분명하게 보여준다. 인간의 지능은 인공지능 분야의 길잡이 역할을 하지만, (더 이상) 규제적 조건은 아니다.

인공지능 개념의 이와 같은 우선성 관계로부터 우리는 하나의 가르침을 도출할 수 있다. 그것은 도대체 어떤 구체적인 속성들을 지닌 지능을 개발하여 만들어낼 것인지를 우리 스스로가 숙고하고 결정해야 한다는 것이다. 이것은 너무도 사소한 진실이어서 오히려 우리의 적극적인 고려에서 누락될 개연성이 작지 않다. 그러나 그것은 위험하다. 더구나 ‘인간의 지능을 뛰어넘는’ 인공지능을 만들겠다는 목표가 공공연하게 거론되는 오늘의 상황에서 이러한 당위는 최대한 분명히 인식되어야 할 기준점이다.

두 번째 개념적 문제로, 우리는 여기서 ‘인공지능’이 입력부와 출력부를 포함하는 물리적 체계를 가리키는 경우와 세계와의 그런 접속면(interface)을 뺀 ‘정보처리 역량’이나 ‘정보처리 체계’(information processing system)만을 가리키는 경우를 구별하여 생각해야 한다. 그리고 인공지능이라는 개념으로 어느 편을 지시할 것인지, 혹은 어떤 방식으로 두 경우를 포괄하여 의미할 것인지를 결정해야 한다. 우리가 일상에서 경험하는 접속면의 형태는 예컨대 컴퓨터 키보드나 마우스 같은 입력장치, 그리고 모니터 화면 같은 출력장치지만, 인공지능을 장착한 기계 장치 즉 로봇 시

49) “인공지능은 기계에 지능을 부여하는 전문 활동이고, ‘지능’은 [그것을 가진] 대상이 환경 속에서 적절하게, 그리고 앞을 내다보며(with foresight) 작동할 수 있도록 만드는 속성이다.” (N. J. Nilsson, “Artificial Intelligence and Life in 2030”에서 (재)인용)

스템 역시 기본적으로 동일한 구조를 가지고 있다.[50]

인공지능의 설계와 제작 등에 관한 규범을 고민할 때, 그 대상은 입출력 장치를 제외한 '정보처리 체계'일 수도 있고 인공지능을 장착한 로봇 시스템일 수도 있다.[51] 전자로서의 인공지능의 핵심 요소는 그러한 체계를 구현하는 소프트웨어 프로그램일 것이다. 그런 의미의 인공지능에 대한 규범 체계는 소프트웨어 프로그램을 대상으로 구성된다. 그러나 인공지능에 대한 규범을 소프트웨어의 차원에만 국한시키는 것은 자칫 규범의 실효성을 스스로 제한하는 결과로 이어질 수 있으므로 유의할 필요가 있다. 이러한 관계를 뒷받침하는 간단한 예로, 지각이나 판단 등 고유한 의미의 지적 작용이 때로 —일부 논자들의 견해에 따르자면 근본적으로— 인지적 시스템의 주변부와 결부된 조건에 의해 조정되거나 결정된다는 사실을 지적해둔다.[52] 쉽게 말하자면, 인공지능 시스템에서 예컨대 기계 시각(machine vision)을 담당하는 부분이나 작동 속도를 제어하는 장치의 공학적 구조나 특성이 그것을 작동시키는 두뇌 부분의 특성에 영향을 미치고 변화를 가져온다는 것이다. 이런 관계를 고려할 때, 인공지능에 대하여 포괄적인 실효성을 지닌 규범을 마련하기 위해서는 좁은 의미의 지능이 구동되는 부분뿐만 아니라 그것과 세계의 접속면까지를 포함하는 로봇 시스템을 염두에 두는 것이 적절하다.

(나) 목표의 확인

이 보고서를 통해 논의하려는 것은 "철학적 합리성과 더불어 현실 적합성을 지닌" 인공지능 윤리의 기본 틀이다. 그것이 우리 사회가 인공지능의 발달이 초래하는 변화된 사회 환경에 적합한 제도와 법률, 그리고 에티켓 등 다양한 층위의 규범 체계를 정립하는 데 필요한 토대이기 때문이다.

여기서 '철학적 합리성'은, 인공지능에 관한 규범이 개념적, 논리적 차

50) 물론 후자는 전자와 달리 운동출력이라는 특수한 양상의 출력을 한다는 특이성을 가지고 있다.
51) 사람에 비유하자면 전자와 후자를 각각 중앙신경체계(CNS)와 몸 전체에 대응시킬 수 있다.
52) 이에 관해서는 인지과학과 인지과학철학의 주요한 견해에 해당하는 '체화된 인지'(embodied cognition) 이론을 참고할 만하다.

원에서 정합성을 지니는 동시에 우리가 인정하고 있는 철학적 관념들의 체계와 부합할 수 있어야 한다고 요구한다. 또 현실 적합성은 인공지능의 윤리가 과학적-공학적 차원의 타당성과 더불어 인공지능 기술이 활용되는 사회의 여건에 대한 충실한 고려를 갖출 것을 요구한다. 이러한 덕목을 충족하지 못하는 규범의 체계도 존재할 수 있다. 그러나 그것은 논리적 정합성이나 현실 적합성을 결여한 법률 체계가 그렇듯이 사회적 불안정을 야기하고 사회의 건전한 지속가능성을 훼손한다.

(2) 인공지능의 존재론적 지위에 관한 견해들

바로 앞의 인공지능의 개념에 관한 논의에서 우리는 인공지능의 존재론적 지위를 논함에 있어 논의의 대상을 무형의 소프트웨어나 소프트웨어를 구동하는 연산장치에 국한시키기보다 인공지능을 탑재하고 특정한 기능을 수행하는 공학적 시스템 전체를 고려해야 한다는 사실을 확인하였다. 따라서 이 보고서에서는 인공지능에 관한 윤리적 문제와 대응을 논함에 있어서 후자처럼 확장된 범위의 '인공지능을 장착한 공학적 시스템'을 고려할 것이다. 이러한 의미의 인공지능의 존재론적 지위에 관하여 대략적으로 세 가지의 견해가 존재한다. 그것은 인공지능이나 인공지능을 장착한 로봇시스템을 (가) 인지 작용의 독자적 수행 능력을 지닌 체계, 즉 판단과 행위의 능력을 지닌 주체로 보는 견해, (나) 특별한 속성을 지닌 인공물(도구)로 보는 견해, 그리고 3) 인간 정신의 연장 혹은 외화로 보는 견해다.

(가) 인지 작용의 독자적 수행 능력을 지닌 체계, 즉 판단과 행위의 능력을 지닌 주체

첫 번째 견해는 인공지능을 장착한 로봇시스템을 그 자체로 판단과 행위의 능력을 지닌 주체의 격위를 지닌 존재로 보는 것이다. 실제로 적정 수준 이상의 인공지능이나 그런 지능을 장착한 로봇은 인간의 개입 없이도 지각, 정보처리, 학습, 그리고 동작의 실행 등을 수행한다. 경우에 따라 그것들은

인간보다 뛰어난 지각이나 정보처리, 혹은 실행의 역량을 보여준다. 이런 까닭에 그것들을 하나의 독립된 인지 체계로, 즉 우리와 외양과 기원이 달라도 그 자체로 행위와 판단의 능력을 지닌 주체라고 보는 것이 합당하다는 견해다.

2016년 미국 조지아텍 대학교(Georgia Institute of Technology)에서 인공지능 수업의 조교(TA)로 온라인에서 학생들을 돕던 ‘Jill Watson’은 실상 International Business Machines사가 대학에 제공한 인공지능 프로그램이었다. 그러나 학생들은 Jill의 수업조교 업무에 대하여 만족스럽게 여겼고, 온라인에서 접촉했던 수업조교가 사람이 아니라고는 생각지 못했다는 반응을 보였다. 월락(Wendell Wallach)과 앨런(Colin Allen) 등도 튜링 테스트(Turing Test)를 도덕적 판단의 영역에 적용하는 ‘도덕 튜링 테스트’(Moral Turing Test)를 논의하였고, 국내에서도 최근 이런 도덕 튜링 테스트의 개념을 응용하여 “10세 아동 수준의 도덕 판단 능력을 지닌 로봇”을 설계하는 프로젝트가 산업통상자원부 산업기술혁신사업의 일환으로 발주된 바 있다.

한편, 인공지능 시스템을 인간과는 질적으로 다르지만 그 자체 심적 속성들을 지닌 독립적인 주체로 간주하려는 입장도 존재한다. 이것은 특히 동물윤리의 논의를 참고하여 인공지능이나 인공지능 로봇을 능동적인 행위 주체로서가 아니라 단지 행위의 감수자(patient)로 보려는 입장이다. 이 입장에서는, 인공지능을 인간과 대등한 의미의 인지 주체로 인정할 수는 없지만 비록 질적으로 낮고 제한된 영역의 지능일지라도 동물의 경우와 마찬가지로 나름의 심적 속성(mental properties)과 인지 능력을 지닌 존재자로 간주해야 한다고 본다.

(나) 특별한 속성들을 지닌 인공물

두 번째 견해는 인공지능을 인간의 효용에 봉사하도록 개발된 특별한 도구로 보는 견해다. 이 입장에서 볼 때, 지능을 장착한 로봇 시스템이나 휴머노이드 역시 인간이 스스로의 편익을 위해 제작한 인공물일 뿐이다. 이렇

게 본다는 것은 인공지능이나 지능형 로봇이 현상 차원에서 인간과 유사하거나 어떤 면에서 인간의 수준을 능가하는 역량을 지니더라도 그것들이 근본적으로 '인간의 처분에 귀속된 대상'(things at our disposal)임을 확인하는 것이다.

(다) 인간 정신의 연장, 또는 외화

세 번째 견해는 인공지능을 인간 정신의 연장, 혹은 '연장된 정신(extended mind)[53]'이라는 관점에서 이해하는 것이 적절하다고 본다. 연장된 정신이라는 개념은 1990년대 말 이후 인지과학과 인지과학철학의 영역에서 활발하게 논의되었는데, 우리나라에서는 이 개념을 한 단계 응용하여 '외화된 정신(externalized mind)[54]'의 개념을 제안하고 그것이 인공지능 및 인공지능 로봇을 이해하는 적절한 존재론적 도식이라고 주장하였다.[55] 특히 그는 앞으로의 사회에서 점점 더 다양한 방식으로 인간의 대리자 역할을 수행하게 될 인공지능이나 인공지능을 장착한 로봇시스템 등의 사회적 함의를 고려할 때, 인공지능이나 인공지능로봇 같은 인공물을 '외화된 사회적 정신'(externalized social mind)라는 관점에서 해석할 것을 제안하였다.

인공지능이나 인공지능을 장착한 로봇 시스템은 그것의 연구자, 설계자, 제작자, 뿐만 아니라 그 설계 과정에 관여하여 인공지능의 사고방식을 (부분적으로) 결정한 여러 사람의 정신이 하나의 체계로 통합되는 방식으로 외화된 지능의 주체다. '외화된 정신'으로서의 인공지능을 외화한 주체는

53) 연장된 정신이란, 1998년 앤디 클락(Andy Clark)과 데이빗 찰머스(David Chalmers)의 공저로 발표된 논문 "Extended Mind"에서 제안된 이후 인지과학과 철학을 비롯한 다양한 영역에서 활발하게 논의된 개념. 인간은 간단한 메모, 체계적인 기록, 셈을 할 때 동원하는 손가락 등 다양한 자원을 활용하여 생물학적 뇌라는 인지 장치의 기능과 용량을 보완한다. 인간이 지닌 지적 활동의 역량은 이런 외부 자원을 해당 주체의 인지 체계의 일부분이라고 보아야 자연스럽게 설명된다. 인간 정신의 자리는 뇌에 국한되어 있지 않다. 그것은 두개골 밖의 물리적 세계에까지 연장되어 있다.

54) 외화된 정신이란, '외화된 정신'은 '연장된 정신'에 관한 토론을 확대, 응용한 개념으로, 더 이상 특정 개인의 정신의 연장이나 확장으로서가 아니라 그것의 발생적 원천 역할을 한 특정 정신으로부터 독립하여 독자적으로 작동하는 상태의 (한때의 연장된) 정신을 의미한다. 개념상 특정한 개별 정신의 연장이어야 하는 '연장된 정신'과 달리 '외화된 정신'은 여러 정신의 상호작용에 의하여 형성될 수 있다.

55) 고인석(2012a), "로봇이 책임과 권한의 주체일 수 있는가?",「철학논총」제67호, 새한철학회, 13면 이하 참조.

그것에 관여한 인간 정신이다. 인공지능은 그 발생의 존재론적 차원에서 인간 정신에 전적으로 의존한다. 설령 인공지능을 프로그래밍하는 인공지능이 개발된다고 해도 이와 같은 존재론적 의존의 관계는 고스란히 근본적인 관계로 남아 있다. 그러나 '외화된 정신'은 이제 독립된 지능 체계로서의 존재론적 지위를 갖는다. 그것은, 인간이 그렇게 허용한다면, 인간 정신의 개입 없이 지각하고 판단하고 결정할 수 있다.

이러한 양면성이 외화된 정신으로서 인공지능 체계가 갖는 고유한 존재론적 특성이다. 어떤 인공지능 로봇이 단순한 오작동의 경우가 아니면서도 이용자의 신체에 위해를 입힐 수 있는 성격의 작동을 유의미한 빈도로 반복하는 경우, 우리는 그것이 그렇게 설계되었기 때문이라고 보아야 할 것이다. 달리 말하자면 그 로봇의 작동 방식은 설계자의 정신을 구현하고 있다고 볼 수 있다. 위의 사례에서, 경우에 따라서는 문제의 위험한 작동이 그 설계에 기인하는 것이 아니라 설계를 충실하게 반영하지 않고 제작 단계에서 자의적으로 변경되거나 수정된 요소에 기인하는 것일 수도 있다. 이와 같은 (후자의) 경우, 문제의 위험한 작동은 로봇을 구동하는 인공지능에 반영된 제작자의 정신에 기인한다고 보아야 할 것이다. 유사한 방식으로, 인공지능 시스템의 특정한 속성이 그것을 관리하고 조정하는 공학자의 정신을 반영하는 것일 수도 있다. 인공지능이나 인공지능을 장착한 로봇시스템이 지닌 공학적 속성들은 해당 시스템을 설계한 자, 제작한 자, 관리하는 자 등의 정신의 외화라는 관점에서 이해할 수 있으며, 그것이 인공지능과 결부된 사회적-윤리적 문제들, 특히 책임의 문제를 다루는 데 적절한 접근 방식이다.

2. 공학윤리의 활용

새로운 기술이 기존의 윤리에 도전이 되는 문제 상황을 만들어낸다고 해서 즉시 기존의 윤리를 무효화하고 새로운 윤리 체계를 구성해야 하는 것은 아니다. 그것은 사회 차원의 경제성을 고려하지 않은 소박한 대응 방식

이다. 이와 유사한 상황에서 일반적으로 활용되는 더 현명한 대응 방식은 제기된 문제의 영역에 적용 가능하리라고 기대되는 기존의 관점을 적절히 확장하거나 변형하면서 문제 상황들에 접근하는 것이다. 인공지능 기술에 대한 윤리적 문제가 제기되는 경우, 우선적으로 고려하고 적용을 시도해야 할 기존의 관점은 '공학윤리'(engineering ethics)의 관점이다.

(1) 공학윤리의 기본 원칙

오늘날, 전문 영역을 막론하고 공학 분야의 학회와 전문가협회들은 '윤리 헌장'(Code of Ethics) 등의 형식으로 해당 분야의 윤리에 관한 원칙을 명시하고 있다. 전문분야에 따라 윤리 헌장이나 윤리 강령의 세부 내용에 차이가 있지만, 확연한 공통분모가 존재한다. 그 공통분모는 공학자가 유념해야 할 최상의 가치로 공공의 안전, 건강, 복지(safety, health, and welfare of the public)를 꼽는 점이다. 세계적으로 공학윤리 분야에서 공학윤리헌장의 대표적인 사례로 인정받고 있는 미국 전국공학인협회(NSPE) 공학윤리헌장(Code of Ethics for Engineers)의 '기본 규범' 부분은 다음과 같다.

공학자는 전문가로서 자신의 임무를 수행할 때 다음과 같이 해야 한다.

1. 공공의 안전, 건강, 복지를 가장 중요한 사항으로 고려한다.
2. 자신이 감당할 능력이 있는 영역의 서비스만 수행한다.
3. 공적 발언은 오로지 객관적이고 진실된 방식으로 한다.
4. 고용주나 고객에게 충실한 대리인 혹은 수탁자로 행동한다.
5. 기만적인 행위를 하지 않는다.
6. 명예롭게, 책임감을 가지고, 윤리적으로, 그리고 법률에 부합하도록 행동함으로써 전문직의 명예와 평판과 유용성을 높인다.[56]

---

56) Engineers, in the fulfillment of their professional duties, shall:
1. Hold paramount the safety, health, and welfare of the public.
2. Perform services only in areas of their competence.
3. Issue public statements only in an objective and truthful manner.
4. Act for each employer or client as faithful agents or trustees.

이로부터 인공지능에 관한 윤리에 적용할 다음과 같은 기본적인 규범 요소들을 추출할 수 있다. 첫 번째는 위에 예시된 NSPE의 공학윤리헌장뿐만 아니라 IEEE와 ASCE(미국 토목공학회) 등 대표적인 공학 단체들의 윤리 강령에서 공통적으로 우선시 되고 있는 '안전'의 추구다.

(2) 인공지능 기술의 안전성

글로벌 차원의 공학협회들이 채택한 공학윤리헌장이 공통적으로 천명하고 있는 원칙의 취지를 고려할 때, 인공지능 기술의 연구개발과 사회적 응용에서도 '안전'(safety)은 필수적으로, 그리고 가장 우선적으로 고려되어야 할 요소다. 특히 인공지능 기술이 그 핵심적인 측면들에서조차 여전히 '미지의 기술'이라는 점을 고려할 때, 인공지능 기술의 연구·개발과 관련한 '예방 윤리'(preventive ethics)의 관점에서 안전은 더욱 특별히 유의해야 할 과제다.

안전이라는 가치를 구현하기 위해서는 우선 인공지능 기술을 개발하고 활용하는 과정에서 발생하는 특수한 유형의 위험(specific risk)들에 관한 체계적 예견이 요망된다. 그리고 그러한 예견을 토대로 우선 기존 공학윤리의 관점을 상세화하고 필요한 만큼 확장하는 것이 필요하다. 그것이 인공지능 기술을 개발하고 이용하는 사회 전체의 경제적 차원에서도 합리적인 원칙이다.

인공지능 기술의 안전성 확보를 위해서는 특히 '인공지능 기술의 투명성(transparency)을 어떻게 확보할 것인가' 하는 문제가 핵심이 될 것으로 예견된다. 2016년 10월 미국 연방 과학기술위원회(National Science and Technology Council) "네트워킹 및 정보기술 연구개발(Networking and Information Technology Research and Development)" 소위원회가 제출한 "국가 인공지능 연구개발 전략 계획"(The National Artificial Intelligence

5. Avoid deceptive acts.
6. Conduct themselves honorably, responsibly, ethically, and lawfully so as to enhance the honor, reputation, and usefulness of the profession.

Research and Development Strategic Plan)에서 <인공지능 R&D 7대 전략> 가운데 한 항목(전략 4)이 '안전과 보안'으로 설정된 점은 참고할 만하다.

이 문건은 인공지능 시스템의 안전성 확보를 위한 주요 연구 과제로 다음 3개 항목을 명시하였다.

① 설명 가능성과 명료성의 개선

② 인공지능 시스템에 대한 신뢰 구축

③ 인공지능 시스템의 검증과 확인 방법에 대한 연구

이상의 세 가지 항목은 모두 인공지능 기술의 투명성 제고를 지향하는 것으로, 투명성이 인공지능 기술의 안전성을 실현하는 핵심 조건이라는 의식을 반영하고 있음을 확인할 수 있다. 이밖에도 그 문건은 <전략 3: 인공지능 관련 윤리적, 법적, 사회적 이슈의 이해와 해결>에서도 "인공지능의 윤리성 확보를 위한 주요 연구과제"의 첫 번째 항목으로 <(설계 자체에 의한) 공정성, 명료성, 책임성 확보>를 규정하였다. 해당 항목은 [인공지능 체계의] 의사결정 과정이 명료하고 인간이 해석하기 쉽도록 설계되어 결정 과정에서 발생할 수 있는 편향성을 인간이 투명하게 검토할 수 있도록 해야 한다고 인공지능의 개발 방향을 명시하고 있다.[57]

(3) 공학적 서비스의 범위와 방식에 관한 원칙

기존의 공학윤리 체계로부터 도출할 수 있는 두 번째 원칙은 공학적 서비스의 범위와 방식에 관한 것이다. 인공지능의 개발과 응용은 공학적 서비스의 범주에 포함되며, 따라서 공학윤리헌장의 적용 범위 안에 있는 사안이기 때문이다. 그러나 이러한 관계가 공학윤리 헌장이나 강령이 인공지능이 결부된 모든 상황에서 최우선 순위의 규범이어야 한다는 것을 의미하지는 않는다.

57) *Super-Intelligent Machines* (Springer, 2002)의 저자 히바드(Bill Hibbard)는 특히 인공지능 기술의 투명성이 확보되지 않을 경우 사회에 인공지능으로 인한 새로운 유형의 부패와 불공정이 발생할 수 있음을 경고하였다. ("Open Source AI", *Proceedings of the First Conference on Artificial General Intelligence* 2008 참조)

앞에서 살펴 본 NSPE 공학윤리헌장의 ‘기본 규범’은 “고용주나 고객에게 충실한 대리인 혹은 수탁자로 행동”할 것과 “자신이 감당할 능력이 있는 영역의 서비스만 수행”할 것을 명령하고 있다. 공학적 활동의 산물인 인공지능이나 인공지능 로봇은 그것을 설계하고 만든 공학자의 개입 없이도 공학자의 의도를 반영하는 방식으로 작동하며, 이것은 간접적인, 혹은 확장된 의미에서 공학자의 활동으로 해석될 수 있다. 이는 앞에서 언급된 ‘자신이 감당할 능력이 있는 영역의 서비스’라는 개념이 공학자의 직접적인 서비스뿐만 아니라 인공지능이나 로봇을 통해 간접적으로 실현되는 공학적 활동에도 확장 적용될 수 있음을 암시한다. 이로부터, 공학자가 인공지능이나 인공지능 로봇을 설계하고 제작할 때 그것이 수행하게 될 서비스의 범위를 명확히 규정하고, 기계학습(machine learning) 등을 통하여 시스템의 세부 속성과 나아가 사실상의 작용 범위까지 확장되거나 변경될 가능성이 있음을 인식하면서 그런 상황에 발생 가능한 위험을 대비하고 예방할 필요성이 확인된다.[58]

3. 인공지능 및 인공지능 로봇에 관한 윤리의 기본틀[59]

이상의 논의로부터 우리는 인공지능과 인공지능 로봇에 관한 윤리의 체계가 우선 그것을 만들고 활용하는 인간 주체들의 윤리로 구성되어야 한다는 사실을 확인하게 된다. 이러한 윤리는 좀 더 구체적으로 인공지능이나 인공지능 로봇시스템을 설계하는 자, 제작하는 자, 관리하는 자, 사용하는 자의 윤리로 구조화할 수 있다.

인공지능 관련 윤리의 중심 물음은 인공지능에, 혹은 그것을 장착한 로봇시스템의 작동에 관하여 어떤 규범이 부여되어야 하는가이다. 널리 알려진

58) 실제로 이와 같은 원천적인 난점에 대응하는 방책으로 시스템 전체의 안정성과 강인성(robustness)을 높이는 한편 인공지능 시스템과 인간의 협업 체계를 활용하는 방안이 탐구되고 있다. 미국 연방 과학기술위원회(National Science and Technology Council)의 Networking and Information Technology Research and Development 소위원회가 발표한 “The National Artificial Intelligence Research and Development Strategic Plan” 참조.

59) 이 절의 논의는 “고인석(2014), “로봇윤리의 기본 원칙 : 로봇 존재론으로부터”, 「범한철학」 75호, 범한철학회의 일부를 원용하였다.

아시모프의 로봇 3 법칙 역시 이와 같은 고민의 성과를 보여주는 예로 볼 수 있다. 로봇에 장착할 규범의 목록에 관한 고민은 그 '장착'의 방법, 즉 '그러한 규범을 공학적으로 어떻게 구현할 수 있는가' 하는 문제와 연결될 것이고, 분명 후자는 그 자체로 전자 못지않게 복잡한, 그러면서도 전자와 다른 종류의 지적 도전을 함축한다.[60] 그러나 분명히 짚어두어야 할 사실은 전자가 후자에 우선하고, 전자에 대한 토론이 후자에 대한 토론을 이끌어야 한다는 것이다.

첫 번째 관점에서 우리의 목표는 도덕적으로 건전한 인공지능을 만드는 것이다. 이것이 직관적으로 자명한 요구에 해당함에도 불구하고 "왜 도덕적으로 건전한 인공지능 시스템이어야 하는가?"라는 잠재적 물음에 맞서 이런 목표 설정의 근거를 대자면, 도덕적으로 건전한 인공지능 시스템만이 지속가능성을 지니고 지속가능성을 결여한 기술은 기술이 존재하는 이유와 상충하기 때문이다. 다음 물음은 "도덕적으로 건전한 인공지능 시스템이란 어떤 것인가?"이다.

그것은 첫째로, 이미 앞에서 논한 것처럼, 공학윤리의 기본 원칙들에 부합하는 로봇을 가리킨다. 특히 그것은, 여러 공학 전문가 공동체가 공학윤리 헌장이나 강령에서 강조하는 것처럼, 인공지능의 제작과 활용이 공공의 안전과 건강에 위협이 되지 않고 공공의 복지를 진작하는 일과 부합해야 한다는 당위를 함축한다.[61] 이 범위에서 우리의 과제는 공학윤리의 한 응용 영역으로 해석할 수 있다. 그러나 공학윤리가 인공지능 윤리의 모든 문제를 포섭하지는 못한다. 인공지능의 특이한 속성들 때문에 우리의 과제는 기존 공학윤리의 적용 범위를 넘어서며, 그 지점에서 기존의 공학윤리와 연결되면서도 그 적용 범위를 넘어 펼쳐지는 인공지능 윤리 고유의 영역이 시작된다.

---

60) 이것은 고유한 의미의 학제적 연구, 혹은 융합연구의 사례이기도 하다. 예를 들어 로봇윤리가 근본적으로 원리-하향식(top-down) 구조와 사례-상향식(bottom-up) 구조 중 어느 편을 취하게 되는지, 혹은 두 방식을 어떻게 상보적으로 결합할 것인지의 문제는 철학적-윤리학적 탐구와 공학적 탐구 중 어느 한 쪽에 귀속된다고 보기 어렵다. Wallach&Allen(2010), 6장-8장 참조.

61) IEEE(전기전자공학회)의 공학자 윤리헌장(Code of Ethics for Engineers) 제1조와 미국 전문공학자협회(National Society of Professional Engineers) 윤리헌장의 기본 강령(Fundamental Canons) 제1조 참조.

인공지능을 장착한 로봇시스템은 다양한 상황에서 인간의 행위를 대신하도록, 즉 인간을 대리하도록 만들어지는 도구다. 그렇다면, 즉 인공지능 로봇시스템의 존재 의미가 특정 범위 안에서 인간의 행위를 대리하는 것이라면, 그런 시스템이 문제의 대리 행위에서 실현하는 입출력 관계는 현상적 차원에서 그것이 대리하도록 의도된 인간의 행위와 제삼자의 관점에서 질적으로 동등한 속성들을 지녀야 할 것이다. 예를 들어, 건물의 출입을 통제하는 방식을 건물 입구에 유능한 보안 인력을 배치하는 방식에서 인공지능을 활용하는 검색과 통제로 바꾸려 하는 경우 새로 도입할 통제 체계가 허용하거나 차단하는 경우는 기존 방식의 그런 경우들과 일치되도록 해야 할 것이다.

이러한 관계는 인공지능의 윤리가 인간 사회의 일반적인 규범윤리와 어떻게 연속되어 있는지를 암시한다. 로봇의 작동은 그것이 대리하는 인간 행위의 결과라는 측면에서 인간의 행위에 관한 규범윤리와 상충하지 않아야 한다. 달리 말하자면, 로봇에 입력될 규범의 내용은 인간 사회를 규율하는 규범의 집합과 갈등 없이 부합하는 것이어야 한다. 예를 들어, 로봇에 입력될 규범이 그 실행을 통해 무고한 사람의 심각한 신체 손상을 초래하는 결과를 함축하고 있다면 그런 규범은 로봇윤리의 관점에서 허용 불가능하다.

한편, 우리는 이 첫 번째 관점의 인공지능 윤리가 포섭해야 할 규범의 목록이 현재의 규범윤리의 테두리 안에 제한되지 않으리라는 점을 확인해 둘 필요가 있다. 이와 같은 경계 벗어나기는 특히 그것은 다음과 같은 시각에서 예견된다. 인공지능 기술이 일정 수준의 지능을 가진 인공물의 구현에 도달하는 발달과 나란히, 인공지능 체계들을 연결하는 네트워크가 이 세계의 작동 방향을 일상의 미시적 차원으로부터 통상, 국방 같은 거시적 차원에서까지 좌우하는 요소로 대두하게 되는 변화가 진행될 것이다. ‘사물인터넷’(Internet of Things; IoT)은 이미 현실의 개념이고, 원칙적으로 인공지능의 발달 수준과도 무관하게 현실 세계의 면면에 영향을 미치는 비중 있는 ‘세계 참여자’ 혹은 구성원으로 성장해 가리라고 예상된다.[62] 더구나

62) 여기서 ‘참여자’는 일종의 비유적 표현이지만, 그것이 세계의 구성과 변화에서 점유할 실질적인 영향력의

이와 같은 물리적 네트워크에 지금보다 한 단계 도약한 높은 수준의 인공지능이 결합되는 경우, 그런 지능을 가진 인공물들의 연결망이 인간 사회에서 차지하게 될 영향력의 넓이와 깊이는 비약적으로 커질 것이다.[63] 나아가 이와 같은 연결망 구조의 인공 주체는, 현재 로봇공학의 발달이 지향하는 방향이 그러하듯이, 손과 발을 지닌 구체적인 행위 주체에 해당하는 속성들을 구현하게 될 것이다.

인간의 능력을 적어도 특정한 면에서 뛰어 넘는 그런 지능을 가진 인공물들의 거대한 연결망 속에서 살아가게 된다는 사실 자체가 우리에게 기술에 대한 공포나 공포에서 비롯되는 혐오를 유발하는 것은 필연적인 현상도 아니고, 그 자체로 유익한 현상도 아니다. 이런 인공물이나 그것들의 체계는 다만 인간이 자신들의 필요를 좇아 고안, 제작하고 발달시킨 것이다. 우리는 그것을 만든 우리 자신을 압도하는 능력을 지닌 이와 같은 인공 체계를 우리의 도구로 사용할 것이고, 아마도 끊임없이 그것을 더 개발하면서 그것들과 더불어 살아갈 것이다.

그러나 우리는 여기서 이와 같은 '인간-그리고-인공체계'의 공진화 과정에 내재되어 있는 위험을 분명하게 인식할 필요가 있다. 그것은 이 체계의 고안자이고 제작자인 동시에 이용자인 인간이 자신이 만든 체계에 대한 통제의 힘을 상실했을 때 나타나는 위험이다. 그리고 이 위험은 총, 칼이나 폭발물이 지닌 종류의 위험과 다르다. 로봇윤리의 첫 번째 관심사는 이 위험을 제어하는 일에 놓여 있다. 총이나 칼의 위험은 원칙적으로 그것을 손에 든 사람의 행위를 통해서만 현실화된다. 그러나 지능을 가진 인공 체계의 경우, 그것 자체가 마치 판단과 실행의 능력을 지닌 하나의 주체인 것처럼 작동할 수 있다는 점에서 총, 칼의 경우와 중요한 차이가 있다.

인공물의 그런 능력이 그것이 자율적 주체의 지위를 가졌음을 의미하는가

몫을 고려하면 '참여자'가 오히려 격하된 비유일 가능성도 있다. 사물인터넷의 개념에 관해서는 Mukhopadhyay & Suryadevara(2014) 참조. "사물인터넷(IoT)이란 문자 그대로, 서로 이야기를 주고받는 물리적 대상들에 관한 모든 것을 의미한다. 기계-기계 소통과 인간-컴퓨터 소통이 사물들에까지 확장된 것이다."(앞의 문헌, p.2)

63) 네트워킹은 인공 체계와 세계의 접속면을 넓히고 접촉이 촉발하는 학습의 신뢰도를 강화하면서 인공 체계가 더 효율적으로, 더 광범위하게 세계를 경험하도록 만든다.

아닌가 하는 물음은 아래 서술할 로봇윤리의 두 번째 관점에서 중요한 논의 주제지만, 첫 번째 관점의 원칙들은 그것이 독립적인 지각과 판단과 실행의 능력을 지닌 것처럼 행동한다는 현상 차원의 사실만으로도 도출 가능하다. 인공지능 시스템을 연구, 설계, 제작, 관리하는 자의 관점에서 고려해야 할 요소는 그런 시스템의 작동과 인간의 행위 사이에 어떤 형이상학적 차이가 있는가 하는 물음이 아니라 연구, 설계, 제작, 관리의 대상인 로봇의 작동이 현실 세계에서 산출하는 결과이기 때문이다. 그리고 이 관점에서 인공지능 윤리의 원칙들은 인공지능 시스템에 대한 제작자-관리자의 통제력을 유지하는 일에 초점을 맞추어야 한다.

인공지능 시스템을 다루는 일에 적용할 일반적인 윤리 원칙을 확립하는 일은 그러한 공학적 시스템의 존재 특성을 바로 보는 데서 시작해야 한다. 인공지능 윤리의 정립이라는 문제의 관점에서 볼 때 인공지능이라는 존재자의 핵심적인 특성은 그것이 인간이 특정한 종류의 효용을 위해 만든 도구이면서도 여느 도구들과 달리 마치 자율성을 지닌 존재인 것처럼 행위할 수 있는 역량을 가졌다는 데, 또 그러면서도 그 자체를 책임의 주체로 인정하기 어렵다는 데 있다. 윤리는 행위의 도덕적 평가를 다루며, 인공지능 윤리는 인공지능의 작동에 관한 평가, 그리고 거기서 이어지는 인정, 장려, 제재, 비난 등을 다루어야 한다.

인공지능은 그것을 기획, 설계하고 제작한 사람들의 정신을 재현한다. 이러한 관계는 인공지능이 학습을 통해서 그 속성의 일부 요소들을 변경하는 경우에도 원칙적으로 유지된다. 이로부터 우리는 인공지능의 작동에 관한 평가가 그것을 기획, 설계, 제작한 주체들에 대한 평가이어야 한다는 기본적인 원칙을 추론할 수 있다.[64] 이와 같은 원칙의 타당성은 도덕적 평가에 국한되지 않지만, 도덕적 평가의 경우에서 더욱 확연하게 두드러진다. 예를 들어 인공지능을 장착한 로봇이 그 작동 과정에서 사람의 신체에 상해를 입혔다고 해 보자. 그런 경우 현상 차원에서 그 상해의 직접적인 주체는 문

64) 편의상 '기획'을 '설계' 과정의 일부분으로 포함시킬 수도 있다. 그러나 도대체 우리가 어떤 종류의 로봇시스템을 필요로 하는가, 어떤 특성을 지닌 로봇을 만들 것인가를 결정하는 기획의 단계를 그런 특성을 공학적으로 어떻게 실현할 것인지를 탐색하고 결정하는 설계의 단계와 구별하여 생각할 수 있다.

제의 그 로봇이지만, 이 상해와 관련된 도덕적 평가의 대상은 해당 로봇의 기획자, 설계자, 제작자, 관리자의 범위로 전이될 것이다.[65]

인공지능 로봇의 설계와 제작 과정 전반이 다시 다른 인공지능에 의하여 통제되는 상황에서는 어떨까? 이것은 좀 더 먼 미래에나 현실이 될 법한 상황이지만, 방금 언급한 평가의 전이(轉移)는 그런 상황에서도 원칙적으로 똑같이 적용된다. 단, 그 전이는 문제의 사건을 일으킨 그 로봇의 제작 공정을 통제한 인공지능에서 멈추는 것이 아니라 문제의 로봇을 그렇게 제작할 인공지능 체계를 설계하고 제작한 사람들에게까지 이어진 뒤에야 멈출 것이다.

앞에서 언급한 '관리자'의 영역은 다시 두 하위 요소로 분석된다. 하나는 감독관리(administration), 즉 인공지능 시스템이 그것의 제작 의도에 따른 본연의 임무를 수행하는지 주목하면서 필요한 단계에서 공학적 개입을 통해 제작 의도와 실제 작동 간의 괴리를 조정하는 일이고, 다른 하나는 부품의 교체와 수리, 보정, 소프트웨어의 업데이트 등을 포함하는 유지관리(maintenance)이다.

인공지능 시스템은 현상 차원에서 인간과 유사한 주체의 기능을 수행할 수 있다.[66] 그러나 이와 같은 현상적 차원의 주체성이 오늘날 우리가 일반적인 도덕 체계에서 상정하는 주체성을 함축하는지, 혹은 그렇게 평가할 적절한 이유가 있는지는 따로 따져져야 할 문제다. 적어도 우리는 일부 지능형 시스템에서 나타날 현상적 주체성이 —그것의 외양적 양상과 상관없이— 그 자체로 도덕적 주체의 지위를 함축하는 것은 아니라는 사실을 분명히 인식할 필요가 있다. 물론 인공지능이 도덕적 주체의 지위를 지닐 수 없다는 명제가 입증된 것도 아니다.

---

65) 그 평가는 문제의 상해에 대한 책임을 해당 로봇의 (기획자,) 설계자, 제작자, 관리자에게 어떻게 분배할 것인가 하는 분석으로 구성될 것이다.

66) 로봇의 '현상적 자율성'에 관해서는 "고인석(2011), "아시모프의 로봇 3법칙 다시보기 : 윤리적인 로봇 만들기", 「철학연구」 93호, 철학연구회"을 참조. 로봇의 자율성 및 도덕적 지위에 관한 토론은 Wallach&Allen(2009)의 1장과 4장, Gips(1991), Smithers(1997), Allen et al.(2000)을 참조. 필자는 로봇이 실현하는 현상 차원의 독립성과 능동성에도 불구하고 로봇의 도덕 주체 지위는 따로 논증되어야 할 문제이고, 특히 사회적 결단의 문제라고 앞에서 주장했다.

이런 마당에 우리가 취해야 할 태도는 두 항목으로 요약된다. 그 하나는 충분한 숙고에 근거한 판단이 내려지기 전까지는 현재의 도덕적 관점을 유지하는 것이다. 현재의 관점에서 도덕의 주체로 인정되는 존재자는 인간—그리고 법인(法人)처럼 특수한 조건을 충족하는 인간 집단—뿐이다. 따라서 우리는 지능을 가진 인공물이 인간과 나란히 도덕 주체의 지위를 갖는다는 명확한 판단에 이르기 전까지는 인공지능에 도덕 주체의 지위를 부여할 이유가 없다. 다른 하나는 "인공지능이 도덕 주체의 지위를 가지는가?" 하는 이 물음의 본성을 재확인하는 것이다. 이 물음은 형이상학적 진리에 관한 물음이 아니라 사회적 결단 혹은 선택에 관한 물음이다.

인공지능의 윤리는 인공지능 기술이 함축하는 잠재적 위험을 의식해야 한다. 인공지능 시스템은 다양한 종류의 학습을 비롯한 (자연발생적) 변화에 의해 원래의 제작 단계에서 의도되지 않았던 새로운 속성이나 기능을 갖게 될 수 있다.[67] 그리고 이 새로운 속성이나 기능은 생물학적 인간에게서 구현된 바 없는 종류나 수준의 것일 수도 있다는 점에 유의할 필요가 있다. 이와 유사한 위험은 이미 기존 공학윤리의 범위 안에서도 인지된 바 있지만,[68] 인공지능의 발달은 이와 같은 위험의 양상과 영향 범위를 빠르게 변화시킬 가능성이 있다. 인공지능의 윤리는 이와 같은 위험의 제어를 관심의 중심에 두면서 예방윤리의 임무를 수행해야 한다. 테크놀로지가 만드는 모든 인공물의 존재 의미는 그것의 제작 목적에 봉사하는 것, 다시 말해 제작 목적에 부합하는 속성을 충실히 구현하는 것, 즉 '도구적 적합성'이다. 이로부터 다음과 같은 인공지능 윤리의 첫 번째 원칙이 도출된다.

"인공지능은 제작 목적에 부합하는 구조와 작동 특성을 일관성 있게 유지하도록 설계, 제작, 관리되어야 한다"

67) 이러한 잠재적 위험은 앞에서 언급한 사물인터넷(IoT)의 속성을 고려할 때 더 확연해진다. 인간 주체의 실시간 개입 없이 '사물들이 서로' 정보를 주고받는 사물인터넷의 특성은 공학자의 개입 없이도 인공물에서 자연발생적인, 그리고 통제되지 않은 공학적 진화가 진행될 수 있을 가능성을 함축한다.

68) Terarc-25 사고의 예를 보라.

이 원칙은 인공지능이나 인공지능 로봇 같은 특수한 도구를 인간의 합리적인 통제력 안에 두기 위한 규범이다. 그리고 이 원칙의 준수는 다음과 같은 실천 규범을 요청한다.

“인공지능의 기능과 그것을 토대로 인공지능 시스템에 위임된 권한의 종류와 양상을 포함하는 시스템의 작동 범위는 명확히 규정되고, 충실히 준수되어야 한다.”

인공지능의 설계자(designer)는 로봇의 작동이나 사용이 그 제작 목적의 범위를 벗어나는 경우를 원리적으로 배제하거나 적어도 그런 경우에 대비하여 시스템의 작동을 제한하는 방안을 설계에 반영하여야 한다. 시스템의 제작자(manufacturer)는 이와 같은 설계를 충실히 물리적으로 실현함으로써 시스템의 구조와 기능이 원래의 제작 목적과 합치하도록 해야 한다. 관리자(administration & maintenance)는 시스템의 작동 양상과 그 결과를 관찰하고 평가하면서 시스템의 작동이 그것의 제작 목적과 일관되게 부합하도록 해야 한다. 또한 인공지능 시스템은 그것의 제작 목적 이외의 용도에 활용되어서는 안 된다. 이와 더불어, 설계와 제작 과정에서 인공지능 시스템의 특성과 관련된 기획, 설계, 제작 각각의 책임과 역할 그리고 해당 시스템이 구현하도록 장착되는 판단과 결정의 원칙들을 명시하여 언제라도 확인하고 필요한 경우 효율적으로 개입하여 수정할 수 있도록 할 필요가 있다.

끝으로, 인공지능 윤리의 기본 원칙은 문제의 기술 산품인 인공지능 시스템에 대한 인간의 통제력 이외에 지속가능성에 대한 고려를 포함한다. 이런 지속가능성에 관한 기본적인 원칙은 다음 세 항목으로 간추릴 수 있다.

① 인공지능 시스템은 이용자의 안전과 건강을 보호하도록, 뿐만 아니라 그것의 작동에 의하여 간접적으로 안전이나 건강에 발생하는 비정상적인 위해가 없도록 설계, 제작, 관리되어야 한다.

② 인공지능 시스템은 자연환경과 생태계의 지속가능성을 위협하지 않도록 설계, 제작, 관리되어야 한다.

③ 인공지능 기술의 개발과 적용은 개별자 차원이나 유적(類的) 차원에서 인간 정체성에 관한 혼란을 야기하지 않는 방식으로 제한되어야 한다. (단, 인간 정체성은 본질주의적으로 규정되는 것이 아니라 사회적 토론과 합의를 통해 보정 가능한 개념으로 간주된다.)

마지막에 언급된 인간 정체성 관련 항목은 인공지능이나 지능형 로봇이 진정한 의미의 자율적 존재가 아님에도 불구하고 외견상 자율적 주체처럼 작동하는 데서 비롯될 수 있는 사회적-심리적 혼란이나 위험을 예견하고 대비해야 할 것을 말하고 있다.

4. [보론] 인공지능의 활용이 유발하는 사회적 공정성의 문제

제2장에서 살펴보았던 미국 백악관이 2016년 12월 발표한 보고서 「인공지능, 자동화 그리고 경제」(Artificial Intelligence, Automation, and the Economy)는 인공지능이 노동시장에 미칠 부정적 영향을 우려하고 있다.[69] 예를 들어, 해당 보고서는 자율주행자동차가 등장하게 되면, 220만 명에서 310만 명이 실직의 위협을 받게 될 것이라고 예측하고 있다. World Economic Forum의 "The Future of Jobs"에 보고된 바에 따르면, 2020년까지 710만개의 일자리가 사라지고 200만개의 일자리가 창출되어 결국 세계적으로 510만 개의 일자리가 감소할 것으로 전망되고 있다.[70] 실직의 문제뿐만 아니라, 인공지능으로 인해 소득 격차가 심화될 것이라는 전망이 있다.[71] 이는 인공지능과 관련한 윤리적 고려에 사회적 공정성의 문제가 포함

69) https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

70) http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

71) http://reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/digital-enterprise-narrative-final-january-2016.pdf

되어야 하리라는 사실을 암시한다.

우리는 먼저 이러한 실직과 소득 격차가 윤리적 문제인지를 검토할 필요가 있다. 이러한 문제가 윤리적 문제인지 아니면 단순한 사실의 기술인지에 따라 대응의 기반이 달라지기 때문이다. 만약 인공지능으로 인한 실직이 도덕의 문제가 아니라면, 정의와 도덕적 의무의 차원이 아니라 사회적 안전망 구축 같은 차원에서 실직의 문제를 다루게 될 것이다. 또한 인공지능으로 인해 가난한 사람들이 더 많아진다고 할 때, 이것이 단지 경험적 사실에 불과하다면 이들을 돕는 것이 자선(charity) 즉 "의무 이상의 행위"가 되겠지만, 만일 이것이 부정의(injustice)의 문제라면 이들의 가난을 해결하는 것은 사회 전체의 의무에 해당하기 때문이다.

(1) 빅데이터, 그리고 데이터의 비대칭성 문제

인공지능이나 인공지능 로봇을 통해 경제적 가치가 산출된다고 할 때, 이 가치를 어떻게 볼 것인가? 이러한 물음에 답하기 위해서는 인공지능의 성격을 규명할 필요가 있다. 러셀(Stuart Russell)과 노빅(Peter Norvig)은 인공지능을 사람처럼 행동하는 시스템, 사람처럼 생각하는 시스템, 합리적으로 생각하는 시스템, 합리적으로 행동하는 시스템이라는 네 가지 기준으로 정의한다.[72] "사람처럼" 만들기 위해서는 빅데이터를 통한 학습이 요구된다. 즉 데이터의 존재와 활용이 학습하는 알고리즘 개발을 위해 필수적이다. 이러한 데이터를 어떻게 볼 것인가?

인공지능의 초과이윤의 원천은 하드웨어나 알고리즘이 아니라 데이터라는 주장이 제기되고 있다.[73] 하드웨어나 알고리즘은 어느 정도 사회에 공개되고 있고 설령 공개되지 않는다고 하더라도 기술력에 의해 비교적 쉽게 접근할 수 있어서 초과이윤을 산출하기 쉽지 않기 때문이다. 이러한 주장에 따르면, 구글, 페이스북, 아마존 등 플랫폼 기업들이 제공하는 서비스를 이

72) S. Russell & P. Norvig (2009), *Artificial Intelligence: A Modern Approach* (3rd Edition), Pearson.
73) 강남훈 (2016), "인공지능과 기본소득의 권리", 「마르크스주의 연구」 13권 4호, 경상대학교 사회과학연구원, 30면.

용하면서 사용자들이 올리는 데이터를 가지고 인공지능을 통해 수익을 산출한다. 실제 구글의 자회사인 Alphabet이 2015년 1년 동안 빅데이터를 통해 벌어들인 수익은 750억 달러에 이르는 것으로 평가된다.[74] 이것이 자연물의 지대(地代)와 다른 점은 생산성의 기원이 자연이 아니라 사람들의 공동적인 참여행위라는 점이다. 사회 구성원들이 데이터를 제공했는데, 데이터를 통해 얻은 이익에 대해 이익 산출에 기여한 사회 구성원들이 혜택을 보지 못하는 현상을 "데이터의 비대칭성"(data asymmetry)이라고 부른다.[75]

현재 인공지능은 인간의 데이터에 대한 광범위한 접근 권한을 가지고 그런 자료를 사실상 무료로 활용하면서 발전해 왔다. 하지만 아이러니하게도 대부분의 사용자는 여전히 인공지능을 통해 생산되는 경제적 이득의 분배에서 배제되어 있다. 인공지능 시스템 구축에 필수적인 빅데이터에 제공되는 개인의 데이터의 가치는 재고되어야 한다.

(2) 인공지능의 활용에서 분배 정의에 관한 자유지상주의의 관점

인공지능의 학습을 위해 필수적인 빅데이터와 관련된 "데이터의 비대칭성" 문제는 인공지능 활용을 통한 이윤의 공정한 재분배를 요구한다. 우리는 먼저 데이터의 비대칭성이 정당한지를 검토해야 한다. 개인의 소유권을 중시하는 자유지상주의(liberatarianism)에서는 플랫폼 기업이 정당하게 획득한 이익에 대해 재분배를 요구하는 것이 부당하다고 주장할 수 있을 것으로 보인다. 왜냐하면, 플랫폼 기업은 각 개인의 데이터를 강제로 뺏은 것이 아니라 정당한 동의(consent)나 소비자의 선택(choice)을 통해 얻은 것이기

74) https://techcrunch.com/2016/07/09/we-need-to-talk-about-ai-and-access-to-publicly-funded-data-sets. 강남훈은 사람들이 플랫폼이라는 사유지 위에 모여서 놀다가 데이터를 남기는데, 플랫폼의 입장에서 빅데이터를 얻기 위해 추가적인 노동을 투입할 필요가 없으므로 지대(차액지대)로 보아야 한다고 주장한다. 위의 문헌 참고.

75) 2016년 12월 13일 IEEE(국제전기전자기술자협회)에서 윤리적 인공지능 시스템 개발을 위한 지침서 "Ethically Aligned Design"의 초안(http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf)을 출판했다. 이 초안은 개인의 정보 데이터로 인한 수익이 데이터 제공자에게 공평하게 제공되지 않는다는 문제를 다루고 있다. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

때문이다.

빅데이터와 관련된 각 개인의 데이터 공급이 정당한 동의인지를 검토하기 위해서, 생명의료윤리학의 자율성(autonomy) 즉 “충분한 정보에 의거한 동의(informed consent)” 부분을 검토하는 것이 도움이 될 것이다.[76] 여기서 ‘자율성’이 중요한데, 자율성 논의는 ‘충분한 정보에 의거한 동의(informed consent)’를 중시한다. 생명의료윤리학의 표준 교과서로 불리는 『생명의료윤리학의 원칙들』의 저자인 뷰첨(T. Beauchamp)과 칠드레스(J. Childress)는 자율적인 행위를 1) 의도를 갖고, 2) 이해와 함께, 3) 행위를 결정하는 통제적 영향력 없이 행동하는 것으로 정의한다.[77]

이러한 자율성 개념은 환자가 1) 제안된 치료 계획의 위험(risk)과 효능(benefit), 그리고 조건들을 알 수 있고, 2) 적절한 예상 결과(prognosis)와 위험이 환자 자신에게 일어날 수 있음을 이해할 수 있고, 3) 치료의 장점과 단점들을 추론하고 가치 판단할 수 있으며, 4) 의료진과 의사소통하고 결정에 도달할 수 있는 한, 의학적으로 합당한 대안들 가운데 결정할 수 있는 환자의 권리를 의미한다.[78] 이러한 논의에서 알 수 있는 것처럼, 각 행위자의 선택이나 동의가 정당화되기 위해서는 그러한 선택이나 동의가 자율적 행위여야 하는데, 그러기 위해서는 자신의 결정이 가져올 이익과 피해에 대해 충분한 정보를 갖은 상태에서 이루어진 선택 또는 동의여야만 한다.

그렇다면, 플랫폼을 이용하는 각 개인이 자신의 데이터를 플랫폼 기업이 이용하는 것에 대한 동의가 충분한 정보에 의거한 동의인지 의심스럽다. 왜냐하면, 각 개인인 자신의 데이터가 플랫폼 기업의 이윤 창출에 얼마나 기여하는지에 대한 이해가 없는 상태에서 이루어진 동의이기 때문이다. 구글 사용자가 자신의 정보를 통해 구글이 2014년에 750억 달러를 벌 수 있음을 안다면, 무료로 자신의 데이터를 구글에 주지 않았을 것이다. 또한 유전자

76) 목광수, 류재한 (2013), “현대 다원주의 사회에 적합한 자율성 모색”, 「한국의료윤리학회지」 제16권 제2호, 한국의료윤리학회, 179면.

77) Beauchamp T. & Childress J. (2013), *Principles of Biomedical Ethics* (7th ed). Oxford University Press.

78) Swindell J.S. (2009), "Two Types of Autonomy," *The American Journal of Bioethics-Neuroscience Vol.9(1)*, p. 52.

회사인 23andMe나 Miinome에 자신의 저렴한 유전 검사를 대가로 자신의 유전정보를 넘긴 소비자들도 그러한 유전정보의 가치와 의미에 대해서 알았더라면 자신의 유전정보를 그냥 넘겨주지는 않았을 것이다.[79] 이러한 정보의 비대칭성에 대응하기 위해 IEEE는 개인이 "자신의 고유한 정체성의 큐레이터로서 자신의 개인적 데이터를 정의하고 접근하고 통제할 근본적인 필요"가 있다고 주장한다.[80] 그리고 각 개인의 정보에 대한 권리를 강화하고 동의의 조건을 확고하게 해야 한다고 주장한다.

개인 정보의 동의와 소유권을 진지하게 고려한다면, 자유지상주의의 입장에서도 데이터의 비대칭성을 인정하고 "부정의 교정의 원칙"(a principle of rectification of injustice)을 통해 부정의를 시정하려는 노력이 권유될 것이다.[81] 노직(Robert Nozick)에 따르면, 타인의 권리를 침해하지 않은 사람들의 자발적 동의에 의해 이루어진 행위는 정당화될 수 있지만 소유물이 자발적 동의가 아닌 방식을 통해 부정의하게 획득되거나 이전되는 경우에는 자기 소유권 원칙을 행사할 수 없다. 그리고 이러한 부정의는 시정되어야만 한다.

(3) 인공지능의 활용에서 분배 정의에 관한 평등주의적 자유주의의 관점

데이터의 비대칭성 문제 제기가 정당하다면, 인공지능의 활용에 대한 분배 정의의 원칙이 세워져야 한다. 그런데 이를 위해서는 먼저 빅데이터를 어떻게 볼 것인가의 문제가 해결되어야 한다. 자유지상주의적 관점에서는 빅데이터를 구성하는 개인의 데이터에 대한 개인의 소유권을 중심으로 논의가 전개된 반면, 평등주의적 자유주의에서는 개인의 데이터가 모아진 빅데이터를 어떻게 볼 것인가의 물음에 주목한다. 각 개인의 데이터 자체는 그 자체로는 이윤 창출에 거의 영향을 미치지 않겠지만, 이러한 개인의 데

79) 23andMe에서 제공되는 유전자 정보 분석 서비스 가격이 초기에는 999달러 수준이었지만, 2008년에는 399달러, 2012년에는 99달러로 저렴해졌다. 이렇게 저렴해진 이유는 23andMe이 100만명의 유전자 정보를 얻기 위해서였다. 최윤섭(2014),「헬스케어 이노베이션」, 클라우드나인 참조.

80) IEEE (2016), *Ethically Aligned Design*, p.56.

81) Nozick, R. (1974), *Anarchy, State, and Utopia*, Basic Books, chapter 7.

이터가 모아진 빅데이터는 막대한 이윤을 창출하기 때문이다. 이러한 개별 데이터와 빅데이터의 관계를 각 개인이 갖는 재능 등의 자연적 우연성과 이러한 우연성이 경제활동을 통해 이윤창출에 기여할 때를 구분하는 평등주의적 자유주의의 시각에서 조명해 볼 수 있다. 평등주의적 자유주의자인 롤스는 사회적 불평등을 야기하는 요인으로 사회적 우연성과 자연적 우연성을 거론하며, 이러한 우연성은 도덕적으로 정당화될 수 없는 자연적 사실에 불과하다고 주장한다.

"차등의 원칙은 결국 천부적 재능의 분배를 공동의 자산(common asset)으로 생각하고 그 결과에 상관없이 이러한 분배가 주는 이익을 함께 나누어 가지는 데 합의한다는 것을 의미한다. 천부적으로 보다 유리한 처지에 있는 사람들은, 그들이 누구든지 간에, 아주 불리한 처지에 있는 사람들의 여건을 향상시켜준다는 조건하에서만 그들의 행운에 의해 이익을 볼 수 있다. [...] 천부적으로 타고나는 것은 정의롭다거나 부정의하다고 할 수 없으며, 사람이 사회의 어떤 특정한 지위에 태어나는 것도 부정의하다고 볼 수 없다. 이것은 단지 자연적 사실(natural fact)에 불과하다. 정의 여부가 문제되는 것은 제도가 그러한 사실들을 처리하는 방식이다." 82)

이러한 롤스의 언급을 토대로 각 개인의 데이터와 빅데이터의 관계를 조명해 보면, 각 개인의 데이터는 개인의 소유권이며 빅데이터는 각 개인들의 데이터가 모여진 것으로 그 자체는 정의롭다거나 부정의하다고 할 수 없는 자연적 사실에 불과하다. 그런데 이러한 빅데이터가 사회협력체계 속에서 이익을 산출하기 때문에, 이러한 빅데이터의 분배를 공동의 자산으로 생각할 수 있다. 이렇게 본다면, 각 개인의 데이터가 모여진 빅데이터는 사회협력의 산물로서 공정한 분배의 대상이 된다.

82) Rawls, J. (1971/1999), *A Theory of Justice*, Harvard University Press.

제3절 가이드라인의 제안

1. 가이드라인의 필요성[83)]

앞서 기술한 것처럼 기술 발전에 앞서 법을 제·개정하는 것은 신중해야 한다. 해당 분야 과학자나 미래학자도 앞으로 기술이 어떻게 발전할지 예측하는 것은 불가능한 과제에 가깝기 때문이다. 한편, 기술 발전 과정에서도 기술을 발전시키는 사람과 이를 서비스로 제공하는 사람, 그리고 이를 이용하는 사람은 있기 마련이다. 그런데 사회적 관점에서 보았을 때 이러한 참여자가 구체적으로 하여야 할 행위와 하지 말아야 할 행위가 있게 마련이다.

기술도입기나 기술확산기에 이와 같은 모순에 직면한 입법자가 선택할 수 있는 것이 행정부나 사회단체 등에서 어떤 사회문제를 해결하기 위해 여기에 참여자가 따라야 할 구체적 행동 지침을 법적 형식으로 제정하는 가이드라인(guideline)이다. 행정부가 제정한 가이드라인은 그 형식은 실정법과 다르지 않지만, 규범적 효력이 없다는 점에서 실정법과 결정적으로 구별된다. 그러나 참여자에게, 특히 행정부에 지대한 영향을 받는 사업자에게 그것은 실정법 이상의 사실상 효력을 가진다. 사회단체가 제정한 가이드라인은 그 형식이 실정법과 같지만, 행위자에게 행위기준을 제시하는 정도의 의미를 가질 것이다. 하지만 그것이 협회와 같은 사업자들의 집단적 모임에서 자율규제(self-regulation)[84)]의 목적으로 제정된 것이라면 사실상 효력을 가진다고 할 것이다.

가이드라인은 기술도입기나 기술확산기와 같이 아직 법을 제·개정하기에는 이르지만 규율이 필요한 영역에서, 행위자에게 모범적인 구체적 행위지침을 제시하여 그에 따른 행위를 유도하지만 그 행위를 법적으로 강제하지는 않는다는 점에 그 본질이 있다. 이런 의미에서 가이드라인은 연성법(soft law)

83) 이 절은 정필운(2008), 앞의 논문의 일부를 요약하고 보완하였다.
84) 이에 관해서는 황승흠, 황성기, 김지연, 최승훈(2004),「인터넷 자율규제」, 커뮤니케이션북스 참조.

의 일종이다. 또한 많은 가이드라인은 시간이 흐름에 따라 실정법(경성법; hard law)으로 발전하여 그 분야에서 중요한 규범으로 작용할 개연성이 높다.

현재 인공지능 분야는 기술도입기를 지나 기술발전기를 맞이하고 있다. 인공지능은 빅데이터를 이용한 기계 학습을 통하여 딥 러닝이 가능해지면서 폭발적인 성장을 하고 있다.[85]

이러한 기술이 어느 방향으로, 어느 정도 속도로 발전할지 입법자는 물론, 과학자나 미래학자도 아직 예측하기 어렵다. 그러나 그것이 쓰이는 용도, 그것이 설계나 제작되는 과정에서 인간 존엄성을 존중하는 방향으로 결정되어야 한다는 점 또한 분명하다. 이런 의미에서 인공지능 기술과 서비스, 그 이용에 참여하는 참여자가 준수하여야 할 구체적인 행위 지침이 필요하며, 그 법적 형식은 가이드라인의 형태로 하는 것이 타당하다. 이 내용 중 일부가 기술정착기에는 실정법적 형식으로 전환되어, 인간 존엄성을 존중하면서도 기술의 확산을 촉진하는 것이 타당하다.

따라서 연구진은 제4장 제1절에서 논의한 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 가이드라인 형식으로 제안하였다.

2. 인공지능 윤리 가이드라인 제안

위와 같은 논의 결과 연구진은 인공지능과 인공지능 체계에 대한 윤리는 그것을 만들고 사용하는 '인간의 윤리'가 되어야 한다고 결론지었다. 그리고 그것이 실제로 기능하기 위해서는 인공지능과 인공지능 체계를 만들고 사용하는 각 주체가 그것을 대할 때 구체적으로 어떤 마음가짐으로 어떠한 행위를 하여야 하는지, 어떠한 행위를 하여서는 아니되는지 행위 기준이 될 수 있어야 한다고 결론지었다.

---

85) 이에 관해 자세한 것은 Executive Office of the President National Science and Technology Council Committee on Technology(2016), Preparing for the Future of Artificial Intelligence, p.9.

활용하는 인간 주체들의 윤리로 구성되어야 한다는 사실을 확인하게 된다. 이러한 윤리는 좀 더 구체적으로 인공지능이나 인공지능 로봇시스템을 설계하는 자, 제작하는 자, 관리하는 자, 사용하는 자의 윤리로 구조화할 수 있다.

인공지능 체계의 연구 등에 대한 윤리 가이드라인(안) v.1.0

제1장 총칙

제1조(목적) 이 가이드라인은 인공지능 체계에 대한 사회적 인식을 제고하고, 인공지능 체계를 연구・설계・제작・관리・이용하는 자 등이 준수하여야 할 사항을 정하여 인공지능 체계의 편리하고 안전한 이용 환경 조성을 목적으로 한다.

제2조(정의) 이 가이드라인에서 사용하는 용어의 뜻은 다음과 같다.

1. “인공지능 체계”란 외부환경을 스스로 인식하고 상황을 판단하여 자율적으로 동작하며 경험과 상호작용을 통해서 학습하는 능력을 가진 체계 또는 그러한 체계를 적용한 기계장치를 말한다.
2. “설계자”란 자기의 책임으로 인공지능이 작동할 수 있는 명령의 집합을 작성하는 자를 말한다.
3. “제작자”란 설계에 따라 인공지능을 작용 능력을 지닌 물리적 체계로 구현하는 자를 말한다.
4. “관리자”란 인공지능 체계가 설계에 따라 이용되도록 보관하고 유지할 책임이 있는 자를 말한다.
5. “이용자”란 인공지능 체계를 실제로 사용하는 자를 말한다.

제3조(적용범위 등) ① 이 가이드라인은 인공지능 체계를 연구・설계・제작・관리・이용하는 자에게 적용한다.

② 인공지능 체계를 연구・설계・제작・관리・이용하는 자에 관해 법령의 규정이 있는 경우에는 그에 따른다.

제4조(인공지능 체계에 관한 윤리적 기초) ① 인공지능은 특수한 속성을 지니

는 인공물로서, 현상적으로 사회적 영향력을 지닌 행위를 하더라도, 그 작동 또는 행위의 결과에 대하여 도덕적·법적 책임을 질 수 있는 독립된 자율적 주체가 아니다.

② 인공지능 체계에 관한 윤리는 인공지능 등의 복합적인 속성에 대한 명확하고 일관된 인식에서 출발한다.

③ 인공지능 체계에 관한 윤리는 그것의 '인공성', 즉 그것이 인간의 설계와 제작에 의하여 생성되고 속성이 결정된 산물이라는 사실에 대한 인식과 더불어 인공물임에도 불구하고 인공지능이 지닌 특이성, 특히 그것이 현상적으로 책임을 함축하는 행위주체인 것처럼 지각될 수 있다는 사실에 대한 인식이라는 양면성의 인식에 토대를 두어야 한다.

제5조(기본원칙) 인공지능 체계에 대한 윤리의 기본 원칙은 인공지능 체계에 대한 인간의 통제력 이외에도 지속가능성에 대한 다음과 같은 고려를 포함하여야 한다.

1. 인공지능 체계는 그것의 작동으로 인하여 인류의 지속적인 존속과 행복 추구에 비정상적인 위해를 유발하지 않도록 연구·설계·제작·관리·이용되어야 한다.
2. 인공지능 체계는 자연환경과 생태계의 지속가능성을 훼손하지 않도록 연구·설계·제작·관리·이용되어야 한다.
3. 인공지능 체계의 연구·설계·제작·관리·이용은 개인의 차원이나 인간이라는 종(種)의 차원에서 인간 정체성에 관한 혼란을 야기하지 않는 방식으로 제한되어야 한다. 단, 인간 정체성은 본질적으로 고정된 것이 아니라 사회적 토론과 합의를 통해 변동 가능한 개념이다.

제2장 인공지능 체계 연구자의 윤리

제6조(연구의 목적과 방법) ① 인공지능 체계를 연구하는 자는 인간의 생명, 안전 등 그 존엄성을 증진할 수 있는 목적과 방법으로 연구하여야 한다.

② 연구자는 인간의 생명과 신체를 공격하는 인공지능 체계를 개발하기 위한 연구 등 인간의 존엄성을 침해하는 목적의 연구를 수행하여서는 아니된다.

제7조(연구비 수급 제한) 연구자는 인간의 존엄성을 침해하는 목적을 갖거나 그러한 목적을 갖고 있다고 의심되는 개인·단체·국가 등에서 지원하는 연

구비를 받아서는 아니된다.

제8조(연구자) 연구자는 인공지능 체계가 인간의 존엄성을 침해할 수 있다는 가능성을 인정하고, 당해 연구 과정의 투명성을 높이고 연구 과정에서 나타날 수 있다고 예견되는 위험을 사전에 방지할 수 있도록 하여야 한다.

제3장 인공지능 체계 설계자의 윤리

제9조(목적과 이용 범위의 설정) 설계자는 최초 설계 단계부터 그 체계의 목적과 이용 범위를 명확하게 정하고 이를 명시하여야 한다.

제10조(위험의 방지) ① 설계자는 해당 체계의 공학적 구조와 특성을 최대한 투명하게 명시함으로써 그것의 이용 과정에서 나타날 수 있다고 예견되는 위험을 사전에 방지할 수 있도록 하여야 한다.

② 이용 과정에서 예기치 못한 위험이 발생하였을 때 효율적으로 해결을 모색할 수 있도록 설계하여야 한다.

제4장 인공지능 체계 제작자의 윤리

제11조(목적과 이용 범위에 충실한 제작) ① 제작자는 그것의 설계자가 의도하고 명시한 해당 체계의 존재 목적과 작동 범위를 분명하게 인식하고, 제작 과정에 충실히 반영하여야 한다.

② 제작자는 제작 목적과 이용 범위를 명확히 정하지 않은 인공지능 체계는 제작하면 아니된다.

제12조(안전한 인공지능으로 제작) 제작자는 해당 체계의 작동에서 발생할 수 있는 위험을 의식하는 동시에 그러한 위험에 대비하는 설계자의 방안을 충실히 인식하면서 최대한 안전한 인공지능 체계를 구현해야 한다.

제5장 인공지능 체계 관리자의 윤리

제13조(관찰과 평가) 관리자는 해당 체계의 작동 양상과 작동의 결과를 지속적으로 관찰하면서 그것이 설계와 제작에 투영된 체계의 목적과 일관되게 부

합하는지를 평가해야 한다.

제14조(개입과 중단) 관리자는 해당 체계의 작동 양상이나 작동의 결과가 설계 단계에서 명시된 체계의 존재 목적이나 예상된 허용 범위를 이탈하는 경우 개입하여 조정하고, 조정이 불가능한 경우 체계의 작동을 안전한 방식으로 중단하여야 한다.

제15조(보고) 공공의 생명, 안전, 보건에 영향을 주는 인공지능 체계를 관리하는 자는 해당 체계에서 발견된 이상 징후나 위험에 관한 정보를 즉각 보고하여야 한다.

제6장 인공지능 체계 이용자의 윤리

제16조(목적에 충실한 이용) 이용자는 그것의 용도와 더불어 적정한 이용의 범위를 인식하면서 그 범위를 벗어나는 오용이 발생하지 않도록 하여야 한다.

제17조(오작동으로 인한 위험의 인식) 이용자는 그것이 다른 기술산품과 같이 사람의 생활을 편리하게 하면서도 오작동 등으로 의도하지 않은 위험을 가진 인공물임을 인식하여야 한다.

제18조(인공물로서 인식) 이용자는 그것이 ‘인간과 대등하거나 심지어 인간보다 우월한 주체인 것처럼 보일 수도 있지만 본질적으로 인간의 설계와 제작에 의한 인공물’이라는 존재론적 관계를 명확히 인식해야 한다.

제7장 인공지능 체계에 대한 사회적 인식의 제고

제19조(국가의 책무) 국가는 인공지능 체계의 이용 가치와 더불어 그것의 한계, 그리고 그것에 내재한 위험의 요소들을 명료하게 인식하는 균형 잡힌 인공지능관을 사회 전반에 정착시키도록 노력하여야 한다.

제20조(중립성) 인공지능이 산출하는 결과물을 무조건 신뢰하거나 반대로 백안시하는 것은 둘 다 경계해야 할 위험한 태도임을 인지하여야 한다.

제21조(최신 인식) 인공지능, 특히 교통, 에너지, 사회기반 정보통신 등에 이용

되는 인공지능 체계가 사회의 효율성과 더불어 인간의 생명, 공공의 안전, 건강, 복지에 미치는 영향에 대하여 공학과 행정 분야의 전문가들뿐만 아니라 일반 시민들이 지속적으로 업데이트하는 적절한 인식을 갖도록 하는 활동이 필요함을 인지하고, 정부와 지방자치단체는 학계와 협력하면서 이를 위한 시민 정보화 및 교육 프로그램을 기획, 운영하여야 한다.

제4절 법제 정비방안

1. 가이드라인의 불완전성과 보충 필요성

이미 살펴본 것처럼 가이드라인은 행정부나 사회단체 등에서 어떤 사회문제를 해결하기 위해 여기에 참여자가 따라야 할 구체적 행동 지침을 법적 형식으로 제정한 것이다. 그 형식은 실정법과 같으며, 행정부나 사업자단체에 지대한 영향을 받는 사업자라면 사실상 효력을 가진다는 장점이 있다. 그러나 강제력을 바탕으로 한 법적 효력은 없고 위와 같은 사업자가 아니라면 사실상 효력이 반감된다. 그런데 이 보고서에서 제안한 인공지능 연구와 이용 단계에서 행정부가 제정하는 가이드라인은 사실상 효력이 반감되는 경우가 있게 된다. 따라서 이를 보충하기 위한 전략이 필요하다.

연구진은 논의 끝에 이를 조직법의 형태로 보완하는 것이 타당하다는 결론에 이르렀다. 따라서 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하기 위하여 국내적으로는 (가칭) 인공지능 윤리위원회를 두고, 국제적으로는 인공지능 윤리 레짐(regime)을 창설할 것을 제안한다. 구체적인 내용은 다음과 같다.

2. (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 설치

(가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회는 국내에서 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하는 것을 그 목적으로 한다. 이것은 「생명윤리 및 안전에 관한 법률」에서 '국가생명윤리심의위원회'와 '기관생명윤리위원회'에서 아이디어를 얻은 것이다.

이 법에서 국가생명윤리심의위원회는 국가의 생명윤리 및 안전에 관한 기본 정책의 수립에 관한 사항, 제12조제1항제3호에 따른 공용기관생명윤리위원회의 업무에 관한 사항, 제15조제2항에 따른 인간대상연구의 심의 면제에

관한 사항, 제19조 제3항에 따른 기록·보관 및 정보 공개에 관한 사항, 제29조 제1항 제3호에 따른 잔여배아를 이용할 수 있는 연구에 관한 사항, 제31조 제2항에 따른 연구의 종류·대상 및 범위에 관한 사항, 제35조 제1항 제3호에 따른 배아줄기세포주를 이용할 수 있는 연구에 관한 사항, 제36조 제2항에 따른 인체유래물 연구의 심의 면제에 관한 사항, 제50조 제1항에 따른 유전자검사의 제한에 관한 사항, 그 밖에 생명윤리 및 안전에 관하여 사회적으로 심각한 영향을 미칠 수 있다고 판단하여 국가위원회의 위원장이 회의에 부치는 사항을 심의한다(법 제7조 제1항). 그리고 기관생명윤리위원회는 (ⅰ) 연구계획서의 윤리적·과학적 타당성, 연구대상자등으로부터 적법한 절차에 따라 동의를 받았는지 여부, 연구대상자등의 안전에 관한 사항, 연구대상자등의 개인정보 보호 대책, 그 밖에 기관에서의 생명윤리 및 안전에 관한 사항에 대한 심의, (ⅱ) 해당 기관에서 수행 중인 연구의 진행과정 및 결과에 대한 조사·감독, (ⅲ) 그 밖에 생명윤리 및 안전을 위한 해당 기관의 연구자 및 종사자 교육, 취약한 연구대상자등의 보호 대책 수립, 연구자를 위한 윤리지침 마련 등을 수행한다. 이 기관생명윤리위원회는 원칙적으로 인간대상 연구를 수행하는 자가 소속된 교육·연구 기관 또는 병원 등, 인체유래물 연구를 수행하는 자(이하 “인체유래물 연구자“라 한다)가 소속된 교육·연구 기관 또는 병원 등에 설치하여야 한다.

생명윤리법에서 국가생명윤리심의위원회와 기관생명윤리위원회는 정책 수립 심의-정책 실행을 분담하였다. 연구진은 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회는 원칙적으로 정책 수립-정책 실행 기능을 부담하는 것으로 설계할 것을 제안한다. 다만 국가인공지능윤리위원회가 인공지능과 관련하여 수행되었거나 수행 중인 연구의 진행과정 및 결과에 대한 조사·감독을 할 수 있도록 명시하고, 기관인공지능윤리위원회나 연구자에게 인공지능 관련 연구에 대한 권고를 할 수 있도록 명시하는 것이 타당하다.[86] 한편, 생명윤리와 달리 인공지능 연구에서는 연구자의 인간 존엄성과 안전을 보호할 필요성도 있으므로 국가인공지능윤리위원회와 기관인공지능

86) The Committee on Legal Affairs of the European Parliament, Draft Roport with recommendations to the Commission on Civil Law Rules on Robotics(2015/2103(INL).

윤리위원회는 연구자의 인간 존엄성과 안전에 관한 사항에 관해서도 정책을 수립하고 이를 실행할 수 있도록 하는 것이 타당하다.

이러한 의미에서 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회는 EU 보고서에서 제안한 연구윤리위원회(a Research Ethics Committee)[87]와 유사하며, 로봇기술, 인공지능을 규제하기 위한 기관(agency)[88]과는 다르다.

한편, 인공지능 관련 연구와 그 응용 활동을 하는 연구소와 기업의 비밀을 유지하면서도 공익을 조화하기 위하여 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 운영에 있어서는 국회 정보위원회(국회법 제37조 제1항 제15호)를 참고하여 설계하는 것이 타당하다. 국회 상임위원회 중 하나인 정보위원회는 국가 비밀 보호와 공익을 조화하기 위하여 그 운영에 있어서 일반 상임위원회와 다른 특례를 가지고 있다. 우선 원칙적으로 정보위원회의 회의는 공개하지 아니한다. 다만, 공청회 또는 제65조의2의 규정에 의한 인사청문회를 실시하는 경우에는 위원회의 의결로 이를 공개할 수 있다. 둘째, 정보위원회의 위원 및 의원보조직원을 포함한 소속공무원은 직무수행상 알게 된 국가기밀에 속하는 사항을 공개하거나 타인에게 누설하여서는 아니된다(이상 제54조의2). (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 회의는 원칙적으로 공개하지 아니하고, 예외적인 경우에만 이를 공개할 수 있도록 하는 것이 타당하다. 그리고 위원회 위원과 관련자는 직무수행상 알게 된 국가기밀에 속하는 사항을 공개하거나 타인에게 누설하지 않도록 의무를 부여하는 것이 타당하다. 이러한 의무를 준수하지 않는 경우 벌칙을 규정하는 것도 필요하다.

3. 국제적인 인공지능 윤리 레짐의 창설

국제적인 인공지능 윤리 레짐(regime)은 국외에서 인공지능 관련 연구와

87) The Committee on Legal Affairs of the European Parliament, Draft Roport.
88) The Committee on Legal Affairs of the European Parliament, Draft Roport.

그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하는 것을 그 목적으로 한다. 이것은 핵무기의 비확산에 관한 조약(The Treaty on the Non-Proliferation of Nuclear Weapons; NPT)과 이에 근거한 국제원자력기구(IAEA)에서 아이디어를 얻은 것이다.

여기서 레짐(regime)이란 국제 관계의 특정 영역에서 행위자가 공통의 기대를 가지고 있는 암묵적 또는 명시적 원칙・규범・규칙・의사결정 과정이다. 경제 영역에서 관세와 무역에 관한 일반 협정(GATT), 세계무역기구(WTO) 등이 레짐의 대표적인 예이다.[89] 국제 문제는 무력의 사용, 외교, 국제 협력, 국제기구, 국제 레짐의 형성, 국제법 등 다양한 방법에 의하여 해결될 수 있다. 구체적으로 어떤 방법을 동원할 것인지는 각 문제의 구조와 특징에 의하여 좌우된다. 예를 들어, 환경 문제의 경우 국제 레짐과 국제법의 역할이 매우 강조된다. 반면 종교 문제, 문명간의 충돌은 당장 이와 같은 구체적인 방법을 강구하긴 어려우므로 상호 교류의 확대 등이 강조된다. 그리고 중첩적, 누적적으로 사용될 수 있다.[90]

위에서 제시한 무력의 사용, 외교, 국제 협력, 국제기구, 국제 레짐의 형성, 국제법 중에서 인공지능 관련 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하기 위하여 현 단계에서 가장 적절한 것은 국제 레짐의 형성과 이를 위한 조약의 체결이다. 외교나 국제 협력을 통한 대처는 인공지능 관련 연구와 그 응용 단계나 속도로 보았을 때 지나치게 느슨하다.

핵무기의 비확산에 관한 조약은 “핵전쟁이 모든 인류에게 엄습하게 되는 참해와 그러한 전쟁의 위험을 회피하기 위하여 모든 노력을 경주하고 여러 나라 국민의 안전을 보장하기 위한 조치를 취하여야 할 필연성을 고려하여, …… 평화적 원자력 활동에 대한 국제원자력기구의 안전조치 적용을 용이하게 하고, …… 어떠한 전략적 장소에서의 기재 및 기타 기술의 사용에 의

89) 존 베일리스, 스티븐 스미스 편저, 하영선 등 옮김(2006),「세계정치론」제3판, 을유문화사, 381면 이하 참고.

90) 이상 안성경(미발간원고), “환경보호를 위한 다양한 방법” 참고

한 선원물질 및 특수분열성 물질의 이동에 대하여 효과적 안전 조치를 하고, 핵무기 보유국이 인출하는 기술상의 부산물을 포함하여 핵기술의 평화적 응용의 이익을 모든 당사국에 제공”(조약 전문)하기 위한 것이다. “핵무기보유 조약당사국은 여하한 핵무기 또는 기타의 핵폭발장치 또는 그러한 무기 또는 폭발장치에 대한 관리를 직접적으로 또는 간접적으로 어떠한 수령자에 대하여도 양도하지 않을 것을 약속하며, 또한 핵무기 비보유국이 핵무기 또는 기타의 핵폭발장치를 제조하거나 획득하며 또는 그러한 무기 또는 핵폭발 장치를 관리하는 것을 여하한 방법으로도 원조, 장려 또는 권유하지 않을 것을 약속한다.”(제1조), “핵무기 비보유 조약당사국은 원자력을, 평화적 이용으로부터 핵무기 또는 기타의 핵폭발장치로, 전용하는 것을 방지하기 위하여 본 조약에 따라 부담하는 의무이행의 검증을 위한 전속적 목적으로 국제원자력기구규정 및 동기구의 안전조치제도에 따라 국제원자력기구와 교섭하여 체결할 합의사항에 열거된 안전조치를 수락하기로 약속한다.”(제3조 제1항), 그리고 “핵무기 비보유 조약당사국은 국제원자력기구규정에 따라 본조의 요건을 충족하기 위하여 개별적으로 또는 다른 국가와 공동으로 국제원자력기구와 협정을 체결한다.”(제3조 제4항)

국제원자력기구(IAEA)는 국제원자력기구 협약(The Statute of the International Atomic Energy Agency)에 따라 설립된 기구로, “전세계를 통하여 평화, 보호 및 번영에 대한 원자력의 공헌을 촉진하고 확대함에 노력한다. 기구는 가능한 한 기구에 의하여 또는 기구의 요청이나 감독 또는 통제하에 제공된 원조가 군사적 목적을 조장시키는 방법으로 사용되지 않도록 하는 것을 목적으로 한다”(제1조). 이를 위하여 국제원자력기구는 “1. 평화와 국제협력을 촉진하는 국제연합의 목적과 원칙에 의거하여 또한 보장된 세계적 군축계획의 확립을 촉진시키는 국제연합의 정책과 이러한 정책에 따라 체결된 모든 국제협정에 따라 기구의 활동을 수행한다. 2. 기구가 받은 특수핵분열성물질의 사용에 있어서 평화적 목적만을 위하여 사용되는 것을 보장하기 위한 통제조치를 설정한다. 3. 세계의 미개발지역에 대한 특수한 필요성을 고려하면서 세계의 모든 지역에 있어서의 효과적인 이

용과 최대한의 일반적 이익을 확보할 수 있는 방법으로 기구의 자원을 할당한다. 4. 기구의 활동상황에 관한 연차보고서를 국제연합총회에 또는 적당한 경우에는 안전보장이사회에 제출한다. 만약 기구의 활동에 관련하여 안전보장이사회 권한내에 속하는 문제가 발생하는 경우에는 기구는 국제평화와 안전에 대하여 주책임을 부담하는 기관인 안전보장이사회에 보고하여야 하며, 또한 본 협약 제12조 다항에 규정된 것을 포함하여 본 협약하에서 기구가 취할 수 있는 조치를 취할 수 있다. 5. 국제연합 경제사회 이사회 및 기타 국제연합기구에 각기 그 권한에 속하는 사항에 관한 보고서를 제출한다." (제3조).

한편, 일본의 「로봇 신전략」에서는 일본이 기술력에 기초하여 인공지능 관련 국제 표준화를 선도하여 규제・제도, 기술이 국제적으로 조화되도록 하겠다는 것이 주요 내용 중 하나였다. 그러나 우리는 이보다 좀 더 상위의 목표를 설정하고 이를 위해 노력할 것을 제안한다. 인공지능 관련 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보할 필요성을 알리고 이에 대처할 국제적인 인공지능 윤리 레짐을 제안하여 매력적인 중견국가로서 확고하게 자리매김하는 것이다. 국제 사회는 이미 환경 문제 등 각종 국제 문제에 대처하기 위하여 각종 제도와 법이 상당히 발전하였고, 이에 따라 전통적인 강대국의 역할이 상당히 제한되고 있으며 제도와 법을 기초로 한 중견국가의 역할이 상당히 중요해지고 있다.[91] 이와 같은 국제 환경에서 우리나라가 인공지능 관련 영역에서 국제적인 레짐의 필요성을 제안하고 그 설립을 주도한다면 선진국과 개발도상국의 이해를 조정하고 여론을 주도하여 세계 공동체에 기여하는 중견국가로서 역할을 다하는 계기가 될 수 있다. 그리고 전지구적인 성격을 가지는 정보통신기술법에서 타국의 법과 상호운용성을 갖춘 국내법의 정립[92]에도 기여할 것이다. 일본의 국제 표준화의 선도와 차별성을 갖는 전략이다.

---

91) 중견국가론에 관해서는 예종영(2009), "국제환경제도와 중견국가의 역할: 월경성 장거리이동 대기오염물질에 관한 협약의 사례를 중심으로",「국제관계연구」제14권 제1호, 34-36면.

92) 이에 관해서는 정필운(2008), 앞의 논문 참고.

4. 그 밖의 법제 정비방안

자율주행자동차(automated car; autonomous car), 드론(drone),[93] 수술로봇(surgical robots), 인공기관(prosthetics) 등이 지능정보사회를 선도하는 응용서비스가 제기하는 법적 문제와 그 해결책에 관해서는 이미 국내외에서 여러 방면의 논의가 있어 왔다. 예를 들어, 자율주행자동차에서는 어느 정도 안전한(또한 결함이 없는) 자율주행자동차를 제조하고 판매할 수 있도록 허락할 것인지, 제조 또는 판매된 자율주행자동차가 사고를 일으킬 경우 누구에게 민사·형사책임을 부과할 것인지, 결함있는(또는 안전하지 않은) 자율주행자동차에 대한 제조물책임을 인정할 수 있을 것인지, 그 대안 중 하나인 보험제도를 구체적으로 어떻게 설계할 것인지 등이다.[94]

인공지능의 학습을 위해 사용되는 저작물의 저작권과 인공지능이 창작한 창작물에 대한 법적 문제가 제기되고 있다. 이에 대한 체계적 분석은 제2장 제4절에서 살펴본 일본의 “지식재산추진계획 2016”에서 제시되었다. 일본은 인공지능이 창작한 창작물에 대하여 저작권을 부여하는 방향으로 목표를 설정하고 관련 논의를 시작하였다. 우리나라의 지식재산법제의 경우에는 인공지능이 자율적으로 생성한 창작한 창작물의 경우에는 권리의 대상이 되지 않는다는 것이 일반적인 해석이다.[95]

또한 인공지능의 학습을 위해 타인의 저작물을 사용하는 경우, 권리자의

---

93) 이에 관해서는 예를 들어, Timothy T. Takahash(2015)i, THE RISE OF THE DRONES - THE NEED FOR COMPREHENSIVE FEDERAL REGULATION OF ROBOT AIRCRAFT, 8 Alb. Gov't L. Rev. 63 참고.

94) 이에 대해서는 EU, Robolaw Project, D6.2 Guidelines on Regulating Robotics; 김윤명(2016), 「인공지능과 리걸 프레임 10가지 이슈」, 커뮤니케이션북스, 85면 이하; 황창근, 이중기(2016), “자율주행자동차 운행을 위한 행정규제 개선의 시론적 고찰-자동차, 운전자, 도로를 중심으로-”, 「홍익법학」 제17권 제2호, 홍익대학교 법학연구소; 심우민(2016a), “인공지능 기술 발전과 입법정책적 대응 방향”, 이슈와 논점 제1138호, 국회 입법조사처 등 참고. 그 밖에 자율주행자동차가 가져올 개인정보 침해의 문제에 관해서는 윤성현(2016), “자율주행자동차 시대 개인정보보호의 공법적 과제”, 인간중심적 사고와 작별, 2016 법과사회이론학회 하반기 학술대회 자료집, 1면 이하; 홈로봇이 가져올 개인정보 침해 문제에 관해서는 Margot E. Kaminski, Robots in the Hone(2014-2015), “What Will We Have a Agreed To?”, 51 Idaho L. Rev. 661. 로봇이 가져올 소비자 보호 문제에 관해서는 Woodrow Hartzog(2014-2015), Unfair and Deceptive Robots, 74 Md. L. Rev. 785 등 참고.

95) 저작권법 제2조 제1호 “저작물은 인간의 사상 또는 감정을 표현한 창작물을 말한다.”

사용허락을 득하지 않고 인공지능의 학습을 위해 사용시 저작권 침해의 문제가 제기 될 수 있다는 문제점이 있다.[96] 특히 인공지능 학습을 위해 사용되는 저작물은 컴퓨터 파일 형태로 사용되기 때문에, 우리 저작권법 제16조에 규정한 "복제권"을 침해 할 수 있는 여지가 있다. 일본의 경우 저작권법 제47조의7 "정보해석을 위한 복제 등" 규정을 통해 필요범위 내에서 기록매체에의 기록 또는 번안을 할 수 있도록 규정하여 인공지능 학습을 위해 타인의 저작물을 제한적으로 이용토록 규정하고 있다.

그에 관해서는 이 보고서의 주요 연구 과제가 아니고, 국내 논의가 어느 정도 해결책을 제시하고 있기 때문에 이 보고서에서는 이에 관한 문헌만을 소개하고 본격적인 논의는 하지 않는다.

한편, 이러한 문제 제기 중 일부는 전통적인 법이론과 법제가 해결하기 어려운 난제를 던진다. 인간 이외에 인공지능 체계를 권리와 의무의 주체로서 수용하여야 하는가가 대표적인 예이다.[97] 인공지능 체계를 '외화된 사회적 정신'(externalized social mind)으로 보는 연구진에 따르면 현재 인공지능 체계 기술 상황에서는 인공지능 체계를 별도의 권리와 의무의 주체로서 인정하기 곤란하다. 그러나 기술 발전이 진전된다면 그것은 별도로 권리와 의무의 주체로서 인정할 가능성도 있을 것이다.

96) 이에 대해 손승우 외(2016), 「인공지능과 글로벌 법제연구」, 한국법제연구원, 124-125면에서는 소니 플레이스테이션 사건(Sony Computer Entm't, Inc. v. Connectix Corp., 203 F.3d 596 (9th Cir. 2000))의 판례를 통해 중간복제로 "공정이용(Fair Use)"에 해당되는 경우 저작권법상 공정이용으로 볼 여지가 있다고 함.

97) 이에 대한 국외의 검토는 Ryan Calo(2015), "Robotics and the Lessons of Cyberlaw", *103 California Law Review 513*; 국내 논의로는 김연식(2016), "과학기술의 발달에 따른 탈인간적 기본권 이론의 모색 - 인간중심적 사고와 작별", 「2016년 법과사회이론학회 하반기 학술대회」, 법과사회이론학회, 51면 이하; 김영환(2016), "로봇 형법(Strafrecht für Robote)?", 「법철학연구」 제19권 제3호, 한국법철학회, 143면 이하.

제5절 가이드라인과 법제정비방안이 SW산업분야에 미치는 영향 분석

이미 설명한 것처럼 센서를 통해서(또는 주변환경과 정보를 교환하여) 정보를 분석하고 경험과 상호작용을 통하여 학습하는 능력을 가진 인공지능은 소프트웨어 그 자체는 아니지만 소프트웨어(컴퓨터, 통신, 자동화 등의 장비와 그 주변장치에 대하여 명령 · 제어 · 입력 · 처리 · 저장 · 출력 · 상호작용이 가능하게 하는 지시 · 명령(음성이나 영상정보 등을 포함함)의 집합과 이를 작성하기 위하여 사용된 기술서(記述書)나 그 밖의 관련 자료(소프트웨어산업진흥법 제2조 제1호)가 핵심이다. 이러한 소프트웨어에 로봇과 같은 형상(하드웨어)이 결합하면 우리가 접하는 인공지능 체계가 되는 것이다.

이 보고서에서 제안한 가이드라인과 법제 정비방안은 인공지능의 역기능을 막아 SW산업이 예측가능성을 가지고 연구와 개발, 그 투자를 강화할 수 있도록 하는 긍정적 효과를 가져 올 것이다. 현재 작지 않은 기업들이 AI에 투자하고 있으며, 그 투자에 대한 리스크를 관리할 수 있기 때문이다.

이러한 위험관리 기반의 가이드라인은 개발자, 사업자, 이용자 등 다양한 관계자의 역할에 대한 의미를 부여하고 그를 통해 새로운 가치를 만들어낼 수 있을 것으로 기대한다. 산업적인 측면에서도 미지의 기술인 인공지능 기술의 확산을 통해 다양한 산업융합을 이끌어낼 수 있다는 점에서 SW산업에 미치는 영향을 확대될 것이다. 그 과정에서는 인공지능윤리와 규범적 논의는 안전망으로써 역할을 하게 될 것임에 틀림없다.

제5장 결론

이 보고서는 지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시하는 것을 그 목적으로 하였다. 이러한 목적을 달성하기 위하여 제2장에서는 인공지능에 대비하여 미국, 유럽연합, 중국, 일본과 같은 선진국에서는 어떠한 준비를 하고 있는지 그 동향을 살펴보았다. 그리고 제3장에서는 인공지능이 대비하여 한국에서 어떠한 준비를 하였는지 그 동향과 문제점을 살펴보았다. 마지막으로 제4장에서는 인간의 존엄성과 안전을 보호하기 위하여 현재 단계에서 필요한 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 총20여개 조문의 가이드라인 형식으로 제안하였다. 그리고 이를 보충하기 위하여 국내적으로는 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 설치하고, 국제적으로는 인공지능 윤리 레짐(regime)을 창설할 것을 제안하였다.

제2장을 요약하면 다음과 같다. 미국, 유럽연합, 일본, 중국에서는 지능정보사회에 적절히 대응하기 위하여 적극적인 정책을 추진하고 있었다. 미국은 2016년 10월 “인공지능 국가 연구 개발 전략 계획”, “인공지능의 미래를 위한 준비” 보고서를 통해 기술과 산업 진흥, 국가・사회적 대응을 위한 시사점을 제시하고 있었다. 유럽연합은 2012년 “유럽에서의 기술 규제 : 로봇에 대한 법과 윤리의 당면과제(RoboLaw Project)” 와 2016년 “규칙 초안을 위한 보고서” (Draft Report)를 통하여 국가・사회적 대응을 위한 구체적인 윤리와 법제적인 대응 방안을 제시하고 있었다. 중국은 2016년 “로봇산업발전규획” 을 통하여 2020년까지 로봇산업의 건강하고 지속가능한 발전을 추진할 것을 목표로, 제조업분야의 새로운 국면을 창출하고 공업 수준을 한 단계 끌어올려 중국로봇기술이 세계 로봇 산업을 선점하기 위한 방법을 제시하고 있었다. 일본은 2015년 “로봇 신전략” 을 통해 시계의 로봇 기술 혁신 거점, 세계 제일의 로봇활용 사회, 세계를 선도하는 로봇신시대에 대한 전략을 제시하고 있었다. 또한 “지식재산추진계획 2016” 에서는

인공지능과 관련한 지식재산 정책 방향을 제시하였다. 이와 같이 유럽연합은 인공지능이 초래할 국가·사회적 영향을 검토하고 이에 대비하기 위한 윤리·법제적 대응에 중점이 있는 반면, 중국·일본은 인공지능 기술의 발전과 산업 진흥을 목표로 이를 달성하기 위한 법제적 대응에 중점을 두는 특징을 보였다. 미국은 기술의 발전과 산업 진흥, 국가·사회적 영향을 검토하고 이에 대비하기 위한 윤리·법제적 대응이 필요하다고 인정하면서도 구체적인 대응 방법을 제시하기 보다는 지속적인 모니터링을 할 것을 권고하는 특징을 보였다.

제3장을 요약하면 다음과 같다. 한국은 지난 2007년부터 행정부 주도로 로봇윤리헌장을 제정을 위한 노력을 전개하였다. 그 목표 중 하나는 IEEE 국제 컨퍼런스에서 한국이 로봇윤리헌장 제정을 제안하는 것이었다. 그리고 그러한 정책 추진을 뒷받침하기 위하여 「지능형 로봇 개발 및 보급 촉진법」도 제정하였다. 그러나 행정부 주도 작업 경향, 전문가 경시 경향, 기술자와 사회과학자, 정책입안자 등의 협업 부족 등으로 목표를 달성하지 못하였다. 그 후 로봇 산업을 비롯한 인공지능 산업은 다른 ICT 산업에 비하여 진흥되지 못하였으며, 관련 연구 성과도 축적되지 못하였다. 한동안 소강상태에 있었던 정책적 대응은 2016년 알파고 현상을 계기로 재점화되었다. 그리하여 「지능정보사회 중장기 종합대책」이 발표되고, 「지능정보사회기본법」 제정을 논의하고 있다. 윤리헌장을 만들려는 노력도 재개되고 있다. 그러나 행정부 주도의 진흥 정책을 추진하기 위한 법적 근거는 이미 마련되어 있고, 현재 기술 수준을 고려하여 보았을 때 법령을 통한 규제 정책은 너무 이르다. 인공지능 연구와 그 응용을 위한 개별적인 법적 장애가 무엇인지는 현장에서 이미 제기되었고 그 해결책도 제시되고 있다. 한편, 연구와 그 응용이 인간의 존엄성을 침해하거나 인체에 위해를 끼치는 것을 막아 인공지능 윤리와 안전을 확보하기 위한 구체적인 정책이 필요하다.

제4장을 요약하면 다음과 같다. ICT 영역에서 사회문제를 해결하는데 법과 윤리 등 사회 규범의 기능과 시장, 기술 등과의 기능 분담을 고려하여야 하고, 인공지능 기술 발달 수준과 인공지능에 대한 존재론적 지위에 기초하

여 인간의 존엄성과 안전을 보호하기 위하여 당장 필요한 것은 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 가이드라인 형식으로 제안하였다.

구체적으로 보면 이와 같은 문제를 해결하기 위하여 현재 인공지능의 기술 수준을 기술도입기를 지나 발전기로 진입한 단계로 파악하고, 법적 강제력 없는 가이드라인 형식을 통한 규율이 필요하다고 결론지었다. 그리고 인공지능을 판단과 행위 능력을 지닌 주체로 보거나 특별한 속성을 지닌 인공물로 보는 것을 넘어 '외화된 사회적 정신'(externalized social mind)으로 인식하는 것이 타당하다고 결론지었다. 이러한 인식에 따르면 현재 기술 수준의 인공지능을 별도의 권리와 의무의 주체로서 인정하기 곤란하지만, 앞으로 기술 발전이 진전된다면 그것은 별도로 권리와 의무의 주체로서 인정할 가능성도 열어둘 수 있다. 이러한 전제에 기초하여 현재 인공지능 윤리는 사람의 윤리이어야 하며, 인간의 존엄성과 안전을 보호하기 위하여 당장 필요한 것은 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 총20여개 조문의 가이드라인 형식으로 제안하였다.

한편 강제력을 바탕으로 한 법적 효력이 없는 가이드라인을 보충하기 위하여 국내적으로는 (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회의 설치하고, 국제적으로는 인공지능 윤리 레짐을 창설할 것을 제안하였다.

이와 같은 가이드라인의 제정과 준수는 인공지능 관련 가이드라인은 관련 연구와 그 응용에 사회적 신뢰를 높여 연구자・설계자・제작자・관리자 등에게 안정적인 자유 영역을 지키고, 인공지능 산업과 그 핵심인 소프트웨어 산업이 예측가능성을 제공하여 연구와 개발, 그 투자를 할 수 있도록 하는 기반을 조성할 것이다. 따라서 입법자와 행정부 공무원은 이 보고서의 제안을 검토하여 연구자・설계자・제작자・관리자・이용자에 대한 윤리와 사회적 인식 제고를 위하여 구체적인 행위 규칙을 제정하고, (가칭) 국가인공지능윤리위원회와 기관인공지능윤리위원회, 국제적인 인공지능 윤리 레짐을 설치할

것을 권고한다.

한편, 이 보고서는 지능정보사회에 대응하기 위한 법제의 현황과 그 문제점, 그러한 문제점을 해결하기 위한 대안을 제시하는 것을 그 목적으로 하였다. 그 중에서도 법과 윤리, 기술 등과의 기능 분담을 고려한 결과 인간의 존엄성과 안전을 보호하기 위하여 당장 필요한 연구자·설계자·제작자·관리자·이용자에 대한 윤리와 사회적 인식 제고를 위한 윤리를 가이드라인을 제시하는 것에 초점을 맞추었다. 그러다보니 인공지능이 고용, 교육 등 국가와 사회에 가져올 변화를 예측하고 이에 대한 정책의 대응에 관해서는 본격적으로 다루지 못하였다. 이 보고서의 한계이자 우리 모두의 과제이다. 의욕과 능력을 갖춘 연구자의 후속 연구가 필요하다.

참 고 문 헌

1. 국내 문헌

강남훈(2016), “인공지능과 기본소득의 권리”, 「마르크스주의 연구」 13권 4호, 경상대학교 사회과학연구원

고인석(2008), “‘로봇의 정체성’이라는 관점에서 로봇윤리 다시 보기”, 제어로봇시스템학회 주관 〈로봇산업 정책연구 개발을 위한 산학연 학술대회 (지능형 로봇 윤리 워크샵)〉 발표자료

고인석(2011), “아시모프의 로봇 3법칙 다시보기 : 윤리적인 로봇 만들기”, 「철학연구」 93호, 철학연구회

고인석(2012a), “로봇이 책임과 권한의 주체일 수 있는가?”, 「철학논총」 제67호, 새한철학회.

고인석(2012b), “체계적인 로봇윤리의 정립을 위한 로봇 존재론, 특히 로봇의 분류에 관하여”, 「철학논총」 제70호, 새한철학회.

고인석(2014), “로봇윤리의 기본 원칙 : 로봇 존재론으로부터”, 「범한철학」 75호, 범한철학회

김연식(2016), “과학기술의 발달에 따른 탈인간적 기본권 이론의 모색 - 인간중심적 사고와 작별”, 「2016년 법과사회이론학회 하반기 학술대회」, 법과사회이론학회

김영환(2016), “로봇 형법(Strafrecht für Robote)?”, 「법철학연구」 제19권 제3호, 한국법철학회

김윤명(2016), 「인공지능과 리걸 프레임 10가지 이슈」, 커뮤니케이션북스

김정명(2013), “일본의 쿨 재팬(Cool Japan) 전략과 정책적 함의”, 「2013년 하계학술대회 발표집」, 한국행정학회

목광수, 류재한 (2013), “현대 다원주의 사회에 적합한 자율성 모색”, 「한국의료윤리학회지」 제16권 제2호, 한국의료윤리학회

변순용 · 송선영(2013), “로봇윤리의 이론적 기초를 위한 근본 과제 연구”, 「윤리연구」 88호, 한국윤리학회

손승우 외(2016), 「인공지능과 글로벌 법제연구」, 한국법제연구원

심우민(2016a), “인공지능 기술 발전과 입법정책적 대응 방향”, 「이슈와 논점」 제1138호, 국회 입법조사처

심우민(2016b), “인공지능의 발전과 알고리듬 규제의 가능성”, 「인간중심적 사고와 작별(2016년 법과사회이론학회 하반기 학술대회 자료집)」, 법과사회이론학회

예종영(2009), “국제환경제도와 중견국가의 역할: 월경성 장거리이동 대기오염물질에 관한 협약의 사례를 중심으로”, 「국제관계연구」 제14권 제1호

정필운(2008), “사이버공간 규율 입법의 일반이론 정립을 위한 탐색적 연구”, 「인터넷법률」 제41호, 법무부

정필운(2009), “정보사회에서 지적재산의 보호와 이용에 관한 헌법학적 연구: 저작물을 중심으로”, 연세대학교 대학원 법학과 박사학위논문

최윤섭(2014), 「헬스케어 이노베이션」, 클라우드나인

황승흠, 황성기, 김지연, 최승훈(2004), 「인터넷 자율규제」, 커뮤니케이션북스

황창근, 이중기(2016), “자율주행자동차 운행을 위한 행정규제 개선의 시론적 고찰-자동차, 운전자, 도로를 중심으로-”, 「홍익법학」 제17권 제2호, 홍익대학교 법학연구소

로렌스 레식 저, 김정오 역(1999), 「코드: 사이버 스페이스의 법이론」, 나남

로렌스 레식 저, 정필운 · 심우민 역(2005), “혁신의 구조”, 「연세법학연구」 제11권 제1 · 2호, 연세법학회

존 베일리스, 스티븐 스미스 편저, 하영선 등 옮김(2006), 「세계정치론」 제3판, 을유문화사

Harris, C. E. et al. (2005)/ 김유신 옮김 (2006), 「공학윤리」(Engineering Ethics: Concepts & Cases, 3rd ed.), Thomson/북스힐.

2. 해외 문헌

A. Etzione and O. Etzioni(2016), “Designing AI Systems that Obey Our Laws and Values”, in Communications of the ACM 59 (9)

Allen, C., G. Varner, J. Zinser (2000), “Prolegomena to Any Future Artificial Moral Agent”, Journal of Experimental & Theoretical Artificial Intelligence 12.

Arkin, R. C. (2009), Governing Lethal Behavior in Autonomous Robots, CRC Press.

Barnes, M. & J. Florian (eds.), Human-Robot Interactions in Future Military Operations, Ashgate, 2010.

Bostrom, N. & E. Yudkowsky (2011), “The Ethics of Artificial Intelligence”,

Beauchamp T. & Childress J. (2013), *Principles of Biomedical Ethics* (7th ed). Oxford University Press.

Bringsjord, S. (2008), “Ethical Robots: The Future Can Heed Us”, AI & Society 22.

B. Kuipers(2016), “Human-like Morality and Ethics for Robots”, AAAI-16 Workshop on AI, Ethics and Society

C. Wickens and J. G. Hollands(1999), “Attention, time-sharing, and workload,“ in Engineering, Psychology and Human Performance

Coeckelbergh, M. (2010), “Robot Rights? Towards a Social-Relational Justification of Moral Consideration”, Ethics and Information Technology 12/3.

D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mane(2016), “Concrete Problems in AI Safety,“

Executive Office of the President National Science and Technology Council Committee on Technology(2016), Preparing for the Future of Artificial Intelligence

Executive Office of the President National Science and Technology Council Committee on Technology(2016), National Artificial Intelligence Research and Development

Floridi, L. (2014), “Artificial Agents and Their Moral Nature”, in: P. Kroes & P.-P. Verbeek (eds.)(2014), The Moral Status of Technical Artefacts (Philosophy of Engineering and Technology 17), Springer.

Gips, J. (1995), “Towards the Ethical Robot”, in: K. M. Ford, C. Glymour & P. Hayes (eds.)(1995), Android Epistemology, MIT Press.

J. Bornstein(2015), “DoD Autonomy Roadmap - Autonomy Community of Interest,” Presentation at NDIA 16th Annual Science & Engineering Technology Conference

J. Furman(2016), Is This Time Different? The Opportunities and Challenges of Artificial Intelligence, Council of Economic Advisors remarks, New York University: AI Now Symposium

J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods(2013), “The Seven Deadly Myths of Autonomous Systems,“ IEEE Intelligent Systems, 28, no. 3

Leveson, N. (1995), “Medical Devices: The Terarc-25”, Software System Safety and Computers, Addison-Wesley.

Mukhopadhyay, S. C. & N. K. Suryadevara (2014), “Internet of Things: Challenges and Opportunities”, in: Mukhopadhyay, S. C. (ed.)(2014), Internet of Things: Challenges and Opportunities, Springer.

Margot E. Kaminski, Robots in the Hone(2014-2015), “What Will We Have a Agreed To?”, 51 Idaho L. Rev. 661.

Nozick, R. (1974), *Anarchy, State, and Utopia*, Basic Books

Pettit, P. (2007), “Responsibility Incorporated”, Ethics 117.

Rawls, J. (1971/1999), *A Theory of Justice*, Harvard University Press.

Smith, M. (1994), The Moral Problem, Blackwell.

Smithers, T. (1997), “Autonomy in Robots and Other Agents”, Brain and Cognition 34.

Sullins, J. P. (2006), “When Is a Robot a Moral Agent?”, International Review of Information Ethics 6.
Sullins, J. P. (2010), “RoboWarfare: Can Robots Be More Ethical Than Humans on the Battlefield?”, Ethics and Information Technology 12/3.
S. Russell & P. Norvig (2009), Artificial Intelligence: A Modern Approach (3rd Edition), Pearson.
S. Russell, D. Dewey, M. Tegmark(2016), “Research Priorities for Robust and Beneficial Artificial Intelligence, “
Swindell J.S. (2009), “Two Types of Autonomy,“ *The American Journal of Bioethics-Neuroscience* Vol.9(1)
The Committee on Legal Affairs of the European Parliament, Draft Roport with recommendations to the Commission on Civil Law Rules on Robotics(2015/2103(INL).
T. G. Dietterich, E. J. Horvitz(2015), “Rise of Concerns about AI: Reflections and Directions,“ Communications of the ACM, Vol. 58 No. 10; K. S.
Timothy T. Takahash(2015)i, THE RISE OF THE DRONES - THE NEED FOR COMPREHENSIVE FEDERAL REGULATION OF ROBOT AIRCRAFT, 8 Alb. Gov’t L. Rev. 63
Torrance, S. (2008), “Editorial: Special Issue on Ethics and Artificial Agents”, AI & Society 22.
Tim Berners-Lee(1999), “Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventer”, Harper San Francisco
R. Y. Yampolsky(2013), “Artificial Intelligence Safety Engineering: Why Machine Ethics is a Wrong Approach,” in Philosophy and Theory of Artificial Intelligence, Springer Verlag
R. C. Arkin(2007), “Governing Legal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture” Georgia Institute of Technology Technical Report GIT-GVU-07-11
R. Yampolsky (2014), “Responses to catastrophic AGI risk: a survey,“ Physica Scripta, 90 (1).
Wallach, W. & C. Allen (2009), Moral Machines: Teaching Robots Right from Wrong. Oxford UP.
Yochai Benkler(2000), “From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access”, Federal Communications Law Journal 52
Yudkowsky, E. (2004). Coherent Extrapolated Volition, The Singularity Institute.

日本経済再生本部(2015), ロボット新戦略:ビジョン‧戦略‧アクションプラン
日本知的財産戦略本部(2016), 知的財産推進計画2016

3. 웹사이트

https://www.federalregister.gov/documents/2016/06/27/2016-15082/request-for-information-on-artificial-intelligence
http://research.ibm.com/cognitive-computing/ostp/rfi-response.shtml
https://www.nitrd.gov/PUBS/bigdatardstrategicplan.pdf.
https://www.nitrd.gov/cybersecurity/publications/2016_Federal_Cybersecurity_Research_and_Development_Strategic_Plan.pdf.
https://www.nitrd.gov/PUBS/NationalPrivacyResearchStrategy.pdf.
http://www.nano.gov/sites/default/files/pub_resource/2014_nni_strategic_plan.pdf.
https://www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf
https://www.whitehouse.gov/BRAIN.
https://www.whitehouse.gov/blog/2011/06/24/developing-next-generation-robots.
https://www.whitehouse.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF
https://www.nasa.gov/mission_pages/SOFIA/index.html. https://cloud1.arc.nasa.gov/intex-na/

http://www.robolaw.eu
http://www.telegraph.co.uk/news/worldnews/1544936/S-Korea-devises-robot-ethics-charter.html.
http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
http://www.law.go.kr/lsRvsRsnListP.do?lsiSeqs=179085%2c167690%2c154036%2c150062%2c149930%2c137002%2c122438%2c115293%2c105485%2c104527%2c104218%2c104096%2c94612%2c90237%2c86607&chrClsCd=010102
http://www.msip.go.kr/web/msipContents/contentsView.do?cateId=mssw311&artId=1323073
http://www.edaily.co.kr/news/NewsRead.edy?SCD=JE41&newsid=01856486612875240&DCD=A00504&OutLnkChk=Y
http://www.biztribune.co.kr/n_news/news/view.html?no=16222
http://www.nickbostrom.com/ethics/artificial-intelligence.pdf.
http://www.zdnet.co.kr/news/news_view.asp?artice_id=20141026165058#csidxe06ace25b20894e894b240ed1f43409
http://www.hani.co.kr/arti/international/international_general/667299.html#csidxdeb2c5ec8a349c6bf6a91e4d0b55c6a
http://www.telegraph.co.uk/news/worldnews/1544936/S-Korea-devises-robot-ethics-charter.html.
http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf
https://techcrunch.com/2016/07/09/we-need-to-talk-about-ai-and-access-to-publicly-funded-data-sets.

Reference

In-sok Ko(2012), “Can Robots be accountable agents?:”, *Journal of the New Korean Philosophical Association 67,* The New Korean Philosophical Association

Seungwoo Son et. al., (2016), *A Study on International Discussion and Legal Policy on Artificial Intelligence Technology*, Korea Legislation Research Institute

Pilwoon Jung(2008), “Towards a General Theory of Legislations Governing Cyberspace”, *Internet Law Journal 41*, Ministry of Justice

YunMyung Kim(2016), Artificial Intelligence and 10 Legal Issues, CommunicationBooks Ins.

A. Etzione and O. Etzioni(2016), “Designing AI Systems that Obey Our Laws and Values”, in Communications of the ACM 59 (9)

Allen, C., G. Varner, J. Zinser (2000), “Prolegomena to Any Future Artificial Moral Agent”, Journal of Experimental & Theoretical Artificial Intelligence 12.

Arkin, R. C. (2009), Governing Lethal Behavior in Autonomous Robots, CRC Press.

Barnes, M. & J. Florian (eds.), Human-Robot Interactions in Future Military Operations, Ashgate, 2010.

Bostrom, N. & E. Yudkowsky (2011), “The Ethics of Artificial Intelligence”,

Beauchamp T. & Childress J. (2013), *Principles of Biomedical Ethics* (7th ed). Oxford University Press.

Bringsjord, S. (2008), “Ethical Robots: The Future Can Heed Us”, AI & Society 22.

B. Kuipers(2016), “Human-like Morality and Ethics for Robots”, AAAI-16 Workshop on AI, Ethics and Society

C. Wickens and J. G. Hollands(1999), “Attention, time-sharing, and workload,“ in Engineering, Psychology and Human Performance

Coeckelbergh, M. (2010), “Robot Rights? Towards a Social-Relational Justification of Moral Consideration”, Ethics and Information Technology 12/3.

D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mane(2016), “Concrete Problems in AI Safety,“

Executive Office of the President National Science and Technology Council Committee on Technology(2016), Preparing for the Future of Artificial Intelligence

Executive Office of the President National Science and Technology Council Committee on Technology(2016), National Artificial Intelligence Research and Development

Floridi, L. (2014), “Artificial Agents and Their Moral Nature”, in: P. Kroes & P.-P. Verbeek (eds.)(2014), The Moral Status of Technical Artefacts (Philosophy of Engineering and Technology 17), Springer.

Gips, J. (1995), “Towards the Ethical Robot”, in: K. M. Ford, C. Glymour & P. Hayes (eds.)(1995), Android Epistemology, MIT Press.

J. Bornstein(2015), “DoD Autonomy Roadmap – Autonomy Community of Interest,” Presentation at NDIA 16th Annual Science & Engineering Technology Conference

J. Furman(2016), Is This Time Different? The Opportunities and Challenges of Artificial Intelligence, Council of Economic Advisors remarks, New York University: AI Now Symposium

J. M. Bradshaw, R. R. Hoffman, M. Johnson, and D. D. Woods(2013), “The Seven Deadly Myths of Autonomous Systems,“ IEEE Intelligent Systems, 28, no. 3

Leveson, N. (1995), “Medical Devices: The Terarc-25”, Software System Safety and Computers, Addison-Wesley.

Mukhopadhyay, S. C. & N. K. Suryadevara (2014), "Internet of Things: Challenges and Opportunities", in: Mukhopadhyay, S. C. (ed.)(2014), Internet of Things: Challenges and Opportunities, Springer.

Margot E. Kaminski, Robots in the Hone(2014-2015), "What Will We Have a Agreed To?", 51 Idaho L. Rev. 661.

Nozick, R. (1974), *Anarchy, State, and Utopia*, Basic Books

Pettit, P. (2007), "Responsibility Incorporated", Ethics 117.

Rawls, J. (1971/1999), *A Theory of Justice*, Harvard University Press.

Smith, M. (1994), The Moral Problem, Blackwell.

Smithers, T. (1997), "Autonomy in Robots and Other Agents", Brain and Cognition 34.

Sullins, J. P. (2006), "When Is a Robot a Moral Agent?", International Review of Information Ethics 6.

Sullins, J. P. (2010), "RoboWarfare: Can Robots Be More Ethical Than Humans on the Battlefield?", Ethics and Information Technology 12/3.

S. Russell & P. Norvig (2009), *Artificial Intelligence: A Modern Approach* (3rd Edition), Pearson.

S. Russell, D. Dewey, M. Tegmark(2016), "Research Priorities for Robust and Beneficial Artificial Intelligence, "

Swindell J.S. (2009), "Two Types of Autonomy," *The American Journal of Bioethics-Neuroscience* Vol.9(1)

The Committee on Legal Affairs of the European Parliament, Draft Roport with recommendations to the Commission on Civil Law Rules on Robotics(2015/2103(INL).

T. G. Dietterich, E. J. Horvitz(2015), "Rise of Concerns about AI: Reflections and Directions," Communications of the ACM, Vol. 58 No. 10; K. S.

Timothy T. Takahash(2015)i, THE RISE OF THE DRONES - THE NEED FOR COMPREHENSIVE FEDERAL REGULATION OF ROBOT AIRCRAFT, 8 Alb. Gov't L. Rev. 63

Torrance, S. (2008), "Editorial: Special Issue on Ethics and Artificial Agents", AI & Society 22.

Tim Berners-Lee(1999), "Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventer", Harper San Francisco

R. Y. Yampolsky(2013), "Artificial Intelligence Safety Engineering: Why Machine Ethics is a Wrong Approach," in Philosophy and Theory of Artificial Intelligence, Springer Verlag

R. C. Arkin(2007), "Governing Legal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture" Georgia Institute of Technology Technical Report GIT-GVU-07-11

R. Yampolsky (2014), "Responses to catastrophic AGI risk: a survey," Physica Scripta, 90 (1).

Wallach, W. & C. Allen (2009), Moral Machines: Teaching Robots Right from Wrong. Oxford UP.

Yochai Benkler(2000), "From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access", Federal Communications Law Journal 52

Yudkowsky, E. (2004). Coherent Extrapolated Volition, The Singularity Institute.

Headquarters for Japan's Economic Revitalization(2015), Japan' s Robot Strategy

Intellectual Property Strategy Headquarters(2016), Intellectual Property Promotion Plan 2016

연구보고서 2016-011

지능정보사회 대응을 위한 법제도 조사연구

2017년 05월 인쇄
2017년 04월 발행

발행처 정보통신산업진흥원 부설 소프트웨어정책연구소
경기도 성남시 분당구 대왕판교로712번길22 A동 4층
Homepage: www.spri.kr

ISBN : 978-89-6108-372-0

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.

ISBN : 978-89-6108-372-0