



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

인공지능기술과 인간의 책임 연구 - 한스 요나스의 관점에서

김도윤

2020



석 사 학 위 논 문

인공지능기술과 인간의 책임 연구
- 한스 요나스의 관점에서 -

A Study on Artificial Intelligence Technique
and Human Responsibility
- focused on Hans Jonas's Perspective -

제주대학교 교육대학원

초등도덕교육전공

김 도 윤

2020년 8월

석 사 학 위 논 문

인공지능기술과 인간의 책임 연구
- 한스 요나스의 관점에서 -

A Study on Artificial Intelligence Technique
and Human Responsibility
- focused on Hans Jonas's Perspective -

제주대학교 교육대학원

초등도덕교육전공

김 도 윤

2020년 8월

인공지능기술과 인간의 책임 연구
- 한스 요나스의 관점에서 -

A Study on Artificial Intelligence Technique
and Human Responsibility
- focused on Hans Jonas's Perspective -

지도교수 장 승 희

이 논문을 교육학 석사학위 논문으로 제출함

제주대학교 교육대학원

초등도덕교육전공

김 도 윤

2020년 8월

김도윤의

교육학 석사학위 논문을 인준함

심사위원장 고대만 인 

심사위원 변종한 인 

심사위원 장승희 인 

제주대학교 교육대학원

2020년 8월

목 차

국문 초록	iii
I. 서론	1
1. 연구 필요성 및 목적	1
2. 연구 내용 및 방법	4
II. 한스 요나스의 책임이론	7
1. '책임'의 개념 및 이해	7
2. 한스 요나스의 책임이론에 대한 분석	13
3. 현대기술에 대한 한스 요나스의 관점	24
III. 인공지능기술 및 윤리적 쟁점	35
1. 로봇과 인공지능의 구분	35
2. 인공지능기술과 관련된 윤리적 쟁점	47
IV. 인공지능시대를 대비한 한스 요나스의 책임이론 적용	67
1. 한스 요나스 기술관에 근거한 인공지능기술 분석	67
2. 한스 요나스 책임관에 근거한 인공지능기술 행위주체별 책임	71
V. 결론 및 제언	79
참고 문헌	82
ABSTRACT	87

표 목 차

〈표 Ⅲ-1〉 로봇과의 공존에서 우리가 마주해야 하는 새로운 윤리적 질문들	40
〈표 Ⅲ-2〉 현재 사용되고 있는 인공지능 구분	48
〈표 Ⅲ-3〉 ‘시간적’ 범위와 ‘공간적’ 범위로서 인공지능	50
〈표 Ⅲ-4〉 분야별 로봇으로부터 발생 가능한 윤리적 쟁점	54
〈표 Ⅲ-5〉 제조 분야 인공지능윤리이슈(자율주행자동차를 중심으로) ...	56
〈표 Ⅲ-6〉 금융 분야 인공지능윤리이슈(로보어드바이저를 중심으로) ...	58
〈표 Ⅲ-7〉 의료 분야 인공지능윤리이슈	60
〈표 Ⅲ-8〉 국방 분야 인공지능윤리이슈	61

그 립 목 차

[그림 Ⅲ-1] 나는 로봇이야(1950)	44
[그림 Ⅲ-2] 우주소년 아톰(1952)	44
[그림 Ⅳ-1] 딥러닝 네트워크	68

국 문 초 록

인공지능기술과 인간의 책임 연구
- 한스 요나스의 관점에서 -

김 도 윤

제주대학교 교육대학원 초등도덕교육전공
지도교수 장 승 희

오늘날 인공지능기술은 더 이상 우리의 일상과 동떨어진 낯선 기술이 아니다. 예를 들어 우리에게 친숙한 스마트폰을 통해 접할 수 있는 인공지능비서는 날씨, 뉴스, 영상, 일정 등의 정보를 사용자의 취향에 따라 알려줄 뿐만 아니라, 질문에 대한 자연스러운 응답으로 사용자와 소통하기도 한다. 그리고 자동차에 탑재된 자율주행 시스템은 최적화된 차량환경을 운전자에게 제공하기도 한다. 이처럼 인공지능기술은 우리의 일상 속에서 다양한 모습으로 쓰이고 있다.

그러나 인공지능기술이 언제나 일상의 편리함만 가져오지는 않았다. 자율주행 자동차 사고가 발생함에 따라 제기된 안전 문제, 인공지능기술을 악용한 개인정보 무단 사용 및 정보 조작 피해 사례, 인공지능기술을 적용한 프로그램이 편향된 데이터로 인해 잘못된 결과를 도출한 사례 등 인공지능기술로 인해 과거에는 없었던 새로운 문제가 발생하고 있다. 여기서 우리가 주목해야 할 부분은 이러한 문제들이 인공지능 자체의 결함보다 인간의 과오나 그릇된 사용, 부주의로 인해

발생한 경우가 많다는 점이다. 따라서 오늘날 사회에서는 인공지능기술을 사용하는 인간에게 그에 걸맞은 책임이 요구되고 있다.

이에 필자는 인공지능기술과 관련된 윤리적 쟁점 중에서도 특히 '책임'과 관련된 윤리적 쟁점을 중심으로 그 내용을 살펴보고, 법률적 의미의 책임과 윤리적 의미의 책임을 비교했다. 나아가 인공지능기술을 다루는 주체에게 요구되는 행위 주체별 책임을 몇 가지로 정리해보았다. 이를 위해 '한스 요나스(Hans Jonas)'를 중심으로 책임이론을 분석하고 그의 관점에서 인공지능기술로 인한 쟁점들을 살펴보았다.

우리가 한스 요나스의 이론에서 주목해야 할 점은 한스 요나스가 과학기술의 발달로 인해 발생하는 문제점의 특징을 분석하고 이에 대응하는 기술관과 책임관을 제시했다는 것이다. 그는 우선 인간의 삶에 비교적 영향을 미치지 않았던 근대 이전의 과학기술의 특징과 삶의 방식에 근본적인 변화를 초래한 근대 이후의 과학기술의 특징을 비교하였다. 그리고 근대 이후의 과학기술을 크게 두 가지 특징으로 정리하였다. 첫째, 인간 자체가 기술의 대상이 되었다. 유전학, 생명공학, 뇌과학의 발달로 인간의 생리작용, 심리, 행위 등이 과학기술의 연구 대상이 되면서 인간에 대한 인식의 전환이 발생하였다. 둘째, 인간이 기술의 대상이 되면서 현대기술도 윤리학의 대상이 되었다. 인간의 윤리적 문제뿐만 아니라 기술의 윤리적 쓰임이나 영향력에도 주목하기 시작한 것이다.

이러한 인식의 전환을 통해 한스 요나스는 그에 걸맞은 책임을 요구하였는데, 이때 말하는 책임은 법률에서 쓰이는 책임과 성격이 다르다. 법률적 의미로서 책임은 손해배상이나 처벌의 관점으로 어떤 잘못에 대한 원인이나 행위주체, 동기 등을 면밀히 따져 그에 따른 인과적 책임을 이야기한다. 그러나 한스 요나스는 법률적 의미의 책임을 강조하기보다는 불확실한 미래에 요구되는 책임을 주장하였다. 다시 말해, 한스 요나스의 이론에서 '책임' 개념은 곧 존재 자체에 대한 인간의 책임, 인간으로서 마땅히 해야 하는 일에 대한 책임, 미래세대를 위한 책임이라고 할 수 있다.

한편 인공지능기술이 발달하면서 의료, 군사, 제조, 금융 등의 산업분야 및 이와 관련된 행위주체에 따라 인공지능기술을 둘러싼 윤리적 쟁점은 복잡해지고 세분화되고 있다. 특히 인공지능기술의 정보처리 과정이 정교해지면서 사물을 인식

하고 문제를 해결하는 일이 인간과 유사해지고 있다. 이러한 인공지능기술의 특징을 한스 요나스의 관점에 비추어 해석하면 다음과 같다. 첫째, 인공지능기술의 주요 분야 중에서 인공 신경망의 정보처리과정이 복잡해지고 있다. 인공 신경망은 그것을 구성하는 층위(Layer)가 많아질수록 보다 정확한 결과를 도출한다는 특징을 갖고 있다. 이로 인해 인공지능기술의 정보처리과정을 분석하는 것이 어렵게 되었다. 또한 알파고(AlphaGo)처럼 하나의 입력자료가 인공 신경망을 거쳐 어떤 결과로 출력될지 우리로서 예측하기가 힘들어졌다. 둘째, 인공지능기술로 인해 산업구조의 변화, 직업구조의 변화, 법제도의 변화, 생활환경의 변화가 나타나고 있다. 기술이 기술을 낳듯이, 이러한 변화에 대응하기 위해서라도 보다 효율적인 인공지능기술 개발은 불가피하다. 따라서 인공지능기술 개발은 필연성을 띠게 되었다. 셋째, 인공지능기술을 연구하는 학자들은 인간의 편의와 이익만을 위해 인공지능기술이 활용되기보다는 수질 정화, 환경문제 개선, 생태계 관리에도 인공지능기술이 활용될 수 있어야 한다고 주장한다. 이는 인공지능기술이 인간의 삶을 넘어 범지구적 차원으로 활용되어야 한다는 것을 보여준다. 넷째, 인공지능기술이 발전하면서 인공지능로봇이 인간의 사고과정과 인간의 모습을 점차 닮아가고 있다. 이로 인해 인간의 의식, 존재와 관련된 형이상학적 물음이 학계로부터 제기되고 있다.

이러한 특징의 인공지능기술을 다루는 주체는 현재로서는 인간이다. 따라서 한스 요나스가 인간의 책임을 강조한 내용을 참고하여 필자는 인공지능기술에 대한 인간의 윤리적 책임을 몇 가지로 제시해보았다. 첫째, 인공지능기술 개발자는 기술의 영향력이 사회의 제반 분야로 확대되었음을 인지하고 신중한 태도로 인공지능기술을 개발해야 한다. 이때 개발자는 환경보존, 생태계보호, 사고예방, 안전확보를 목표로 설정해야 한다. 둘째, 입법자를 포함한 정치인은 사회적 약자, 취약계층을 포함한 사회 구성원 모두가 인공지능기술을 쉽게 접하고 사용할 수 있도록 인프라를 구축해야 한다. 그리고 인공지능기술의 발달로 변화하는 사회 모습을 반영하여 법률 및 제도를 재정비해야 한다. 셋째, 사용자는 인공지능기술을 오용 또는 악용하지 않도록 절제의 미덕을 갖추어야 한다. 또한 인공지능기술을 개발하고 사용하는 것이 적절하게 이루어지고 있는지 비판적으로 살펴보고, 정부와 제조사에 사용자 의견을 제시해야 한다. 넷째, 교육자, 연수기관 등의 교육주

체는 소프트웨어 교육뿐만 아니라 윤리적 차원의 인공지능교육을 강화해야 한다. 이와 더불어 교육주체는 인공지능기술로 인한 전 지구적 변화 속에서 변하지 않는 도덕적 가치를 탐색하여, 교육받는 주체의 도덕적 정서를 일깨우고 길러줘야 한다.

주요어 : 한스 요나스, 책임, 로봇, 인공지능, 로봇윤리, 인공지능윤리

I. 서 론

1. 연구 필요성 및 목적

2016년, 스위스 다보스에서 제46회 세계경제포럼(이하 다보스포럼)이 개최되었다. 다보스포럼에서 클라우스 슈밥(Klaus Schwab)은 제4차 산업혁명을 주창하고 제4차 산업혁명과 관련된 주제와 안전을 세계 주요 기업가들과 논의하였다. 다보스포럼에 참석한 인사들은 “빅 데이터, 인공지능, 로봇공학, 사물인터넷, 무인 운송 수단, 3D 프린팅, 나노 기술 등의 7대 분야에 걸친 기술 혁신”¹⁾을 통해 경제구조부터 시작하여 사회, 문화, 교육 등의 전반적인 변화가 발생할 것이라고 보았다.

그로부터 4년 후인 2020년, 전 세계적으로 코로나바이러스감염증-19(이하 코로나19)라는 신종 감염병이 확산하면서 방역과 공중보건에 비상이 걸렸다. 코로나19는 이전의 유행병과 달리 ‘블랙스완’²⁾ 효과라 일컬어질 정도로 사회적 혼란과 충격을 가져왔다. 코로나19를 종식시키기 위한 백신이나 치료제가 승인되지 못한 상황에서, 보건당국은 우선 전파경로 차단과 방역에 온 힘을 쏟았다. 이를 위해 국내외에서는 코로나19의 확산을 방지하기 위해 인공지능기술을 활용하였다. 코로나19 알리미 앱이 개인, 단체, 기업 단위로 다수 출시되었으며, 미국이나 유럽 등 세계 주요 나라에서 인공지능기술을 활용한 진단, 발열감지, 확진자 현황, 동선 파악 등을 알려주는 다양한 솔루션이 나오고 있다. 그뿐만 아니라 드론, 진단 키트, 코로나바이러스 데이터 분석을 활용한 치료제 개발에도 인공지능 기술이 활용되고 있다.³⁾

제4차 산업혁명이 화두가 된 후 불과 5년도 되지 않아 지구적 보건 위기를 극

1) 제4차 산업혁명. <https://ko.wikipedia.org>. (검색일: 2020. 4. 4.)

2) 블랙스완(Black swan)이란, 절대 일어날 것 같지 않은 일이 일어나는 것을 뜻하는 말로, 미국의 나심 니콜라스 탈레브(Nassim Nicholas Taleb)에 의해 쓰이기 시작했다. 그는 블랙스완의 속성을 세 가지로 정의하였다. 첫째, 일반적 기대 영역 바깥에 존재하는 관측값(극단값)이다. 둘째, 극심한 충격을 동반한다. 셋째, 존재가 사실로 드러나면 그에 대한 설명과 예견이 가능하다. (블랙스완. <https://m.post.naver.com>. 검색일: 2020. 5. 3.)

3) 코로나와 AI 스타트업. <http://www.aitimes.com>. (검색일: 2020. 4. 4.)

복하는 데 인공지능기술이 널리 활용되는 현상은 인공지능기술의 발전 속도와 파급력을 보여주는 하나의 사례라 볼 수 있다. 그러나 인공지능기술의 발전 속도와 파급력으로 인해 그것을 악용하는 사례가 발생하기도 하였다. 하나의 예가 ‘딥페이크’(Deepfake) 기술이다. ‘딥페이크’란 인공지능기술 가운데 하나인 딥러닝(Deep learning)과 가짜(Fake)의 합성어로 인공지능을 기반으로 한 이미지 합성 기술이다.⁴⁾ 여기서 딥러닝이란 다층구조 형태의 신경망을 기반으로 하는 머신러닝의 한 분야로, 다량의 데이터로부터 높은 수준의 추상화 모델을 구축하고자 하는 기법이다.⁵⁾ 딥러닝을 기반으로 한 딥페이크는 포르노 영상에 유명인이나 일반인의 얼굴을 합성하여 디지털 성범죄 논란을 가져오기도 하고,⁶⁾ 뉴스 영상을 조작하여 마치 유명 정치인이 나와 이야기하는 듯한 가짜뉴스를 생산하기도 한다.⁷⁾ 딥페이크가 적용된 영상을 보면 일반인은 구분이 힘들 정도로 정교하여 사회적으로 문제가 되고 있다. 이를 위해 딥페이크를 판별하는 인공지능 기반 딥페이크 탐지기술이 개발되기도 하였다. 하지만 이러한 딥페이크 탐지기술이 해당 콘텐츠의 삭제 여부를 판단해줄 수 없으며, 피해자에게 이미 늦은 기술 도입이라고 미국의 ‘MIT 테크놀로지 리뷰’는 부정적인 평가를 내리기도 하였다.⁸⁾ 이처럼 명(明)과 암(暗)이 섞여 있는 인공지능기술을 활용하는 데 있어 오늘날 사회에서는 인간에게 그에 걸맞은 책임이 요구되고 있다.

이러한 요구에 기업, 단체, 국가, 협의체 단위로 ‘인공지능윤리 헌장’을 발표하여 인공지능기술을 활용하는 설계자, 제작자, 사용자의 올바른 윤리의식 함양을 강조하고 있다. 안나 조빈(Anna Jobin)·마르첼로 인카(Marcello Ienca)·에피바예나(Effy Vayena)에 따르면 2019년까지 전 세계적으로 발표된 인공지능윤리 가이드라인의 개수는 84개라고 한다.⁹⁾ 이 윤리 가이드라인 중에서 책임(Res

4) 딥페이크. <https://ko.wikipedia.org>. (검색일: 2020. 4. 4.)

5) 딥러닝. <https://terms.naver.com>. (검색일: 2020. 2. 15.)

6) 딥페이크. <https://terms.naver.com>. (검색일: 2020. 4. 4.)

7) 가짜뉴스보다 더 무서운 가짜 동영상 ‘딥페이크’. <https://news.sbs.co.kr/news>. (검색일: 2020. 4. 4.)

8) 딥페이크 탐지기술, 가짜뉴스 해결사 될까?. <http://www.hani.co.kr>. (검색일: 2020. 4. 4.)

9) Anna Jobin·Marcello Ienca·Effy Vayena. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399.

possibility)을 다루는 가이드라인은 60개(Anna Jobin·Marcello Lenca·Effy Vayena, 2019, p. 395)로 그만큼 인공지능기술을 취급하는 데 있어 인간의 책임이 적극적으로 요구되고 있다고 볼 수 있다. 그러나 인공지능기술과 관련된 책임 논쟁을 살펴보면 인공지능기술로 인해 발생한 사고에 대한 책임은 누가 질 것인지에 대한 법적 책임 문제와 인공지능기술의 운용에 있어 누가 책임을 맡을지에 관한 기술 소유권의 문제 등이 주로 논의되고 있는 실정이다. 이러한 인공지능기술에 대한 명확한 법적 책임 확립은 현재 발생하고 있거나, 앞으로 발생할 수 있는 문제점을 해결하고 방지하는 데 필요한 작업이다. 하지만 필자는 법적 책임을 논하기 이전에 책임의 본질은 무엇인지, 책임이 필요한 이유가 무엇인지에 대한 고찰이 필요하다고 본다. 그 이유는 다음과 같다.

첫째, 책임의 주체에 대한 철학적 탐색이 필요하다. 예를 들어 법률상 책임의 주체는 오직 인간이었다. 하지만 2017년 영국 일간지 가디언에 의하면 유럽연합(EU) 의회에서 인공지능 로봇의 법적 지위를 '전자 인간'(Electronic Personhood)으로 지정하는 결의안이 통과되었다. 이는 일상 속에서 인공지능 로봇 사용이 확산됨에 따라 법적 기틀을 마련해야 한다는 취지에서 통과된 것이다.¹⁰⁾ 이처럼 법률상 책임의 주체에 변화가 발생하고 있는 실정이다. 만약 인간과 유사한 로봇이 출현하게 되었을 때 우리가 인간의 고유 영역이라고 여겨지는 이성, 감성 등의 특성은 어떻게 설명할 수 있을까? 비록 아직은 먼 미래의 이야기일지라도 이러한 물음에 대해서 철학계에서는 다양한 관점으로 접근하고 있다. 이러한 관점을 바탕으로 윤리적 책임의 주체를 인간으로만 볼지 인간과 로봇으로 볼지에 대한 논란이 제기되고 있다. 책임에 대한 철학적 고찰은 법률상의 책임을 논하기 위한 배경이 되면서 책임의 주체를 이해하는 데 도움을 주기 때문에, 책임이란 무엇이고 법률적 의미의 책임과 윤리적 의미의 책임은 어떻게 다른지 살펴볼 필요가 있다.

둘째, 사람들의 책임이 요구되는 사례가 다양한 사회현상을 통해 나타나고 있다. 유명인사의 무책임한 행실에 대해서 여론은 비난하고 사과를 요구하고 있으며, 특별한 책임이 요구되는 직무에 대해서는 역할의 충실한 이행을 바라고 있

10) EU, AI 로봇에 '전자인간' 지위 부여..."무조건 인간에 복종하도록 설계해야". <https://biz.chosun.com>. (검색일: 2020. 5. 3.)

다. 그리고 인공지능기술 발달로 인한 직업 구조의 변화와 산업 구조의 변화로 인해 기업에서는 협업능력과 책임감을 포함한 인성을 선발 기준으로 책정하기도 한다. 특히 비대면(非對面)이 새로운 생활양식으로 화두가 되고 있는 현재, 온라인상에서 개인의 책임 있는 자세는 더욱더 강조되고 있다. 예를 들어 개인 저작물의 사용이 급증함에 따라 나타나는 저작권 문제, 데이터 기반의 인공지능기술 플랫폼이 확대됨에 따라 나타나는 익명의 데이터 수집 및 관리, 보안 문제 등이 있다.

이에 필자는 인공지능기술과 관련된 윤리적 쟁점, 특히 '책임'과 관련된 윤리적 쟁점을 중심으로 그 내용을 살펴보고, 법률적 의미의 책임과 윤리적 의미의 책임을 비교하고자 한다. 나아가 인공지능기술을 다루는 주체에게 요구되는 행위주체별 책임을 몇 가지로 정리해볼 것이다. 이를 위해 '한스 요나스(Hans Jonas)'를 중심으로 책임이론을 분석하고 그의 관점에서 인공지능기술로 인한 쟁점들을 살펴보고자 한다.

2. 연구 내용 및 방법

한스 요나스의 책임이론을 살펴보기 전에, 국내에서 한스 요나스를 어떻게 평가하고 있는지 몇 가지 논문들을 중심으로 간략하게 검토해보았다.

먼저, 양해림(2000)¹¹⁾은 과학기술의 발달과 급격한 사회 변화로 인해 자연이 훼손되고 생태계가 파괴되는 지구적 현상에 문제의식을 가졌다. 이러한 문제의식을 토대로 양해림은 인간과 자연의 윤리적 관계에 주목하여 한스 요나스의 이론을 분석하였다. 그 결론으로 양해림은 한스 요나스에 의해 제시된 정언명법 “너의 행위의 영향이 진정으로 지상에서 인간적인 삶이 되도록 행위하라.”를 오늘날 살아가는 인류에게 필요한 격률(格率)로 인정하였다. 그러나 오늘날 환경문제를 바라보는 데 한스 요나스의 ‘존재와 당위, 원인과 목적, 가치’의 관점을 적용할 수 있을지는 의문을 가졌다.

다음으로 이유택(2005)¹²⁾은 한스 요나스가 제시한 미래윤리의 5대 원칙이

11) 양해림. (2000). 생태계의 위기와 책임윤리의 도전-한스 요나스의 책임개념을 중심으로-. 철학, 65, 241-267.

현실적으로 타당한지 살펴보았다. 여기서 한스 요나스가 제시한 미래윤리의 5대 원칙은 미래에 대한 충분한 정보와 지식(미래학), 미래 인류에 대한 두려움(염려), 두려움에 따르는 겸손함(조심스러움), 낭비하지 않는 소비습관(검소함), 걱정 한계를 설정하고 지키는 태도(절제)이다. 이에 대해 이유택은 한스 요나스의 미래윤리 원칙이 결국 '인류 전체'를 위한 마음을 기르는 것으로, 추상적이고 낯설게 느껴질 수 있다고 평가하였다. 그럼에도 불구하고 이 원칙들을 미래교육의 지표로서 작용하여 교육현장에서 실현되어야 할 가치로 인정하였다.

그리고 김은철·송성수(2012)¹³⁾는 과학기술의 윤리적 측면에 대한 논의로서 과학기술자의 책임에 관해 연구하였다. 이를 위해 한스 요나스의 책임이론을 주목하였는데, 김은철·송성수는 한스 요나스가 과학기술의 발전에 대한 윤리적 반성을 촉구하는 데 이바지하였으며 책임에 대한 몇 가지 규범을 제시하였다고 보았다. 이러한 관점에서 한스 요나스의 책임이론을 분석한 뒤, 과학기술자의 책임에 대해 시사하는 바를 도출하였다. 그 내용은 다음과 같다. 첫째, 자연이 수용 가능한 범위 내에서만 과학기술의 발전을 도모해야 한다. 그리고 미래세대가 존속할 수 있는 범위 내에서 현재 세대의 수요를 충족시키는 방향으로 과학기술의 발전을 추구해야 한다. 둘째, 과학기술자의 활동이 건강, 안전, 환경에 심각한 손상을 초래할 위험이 있는 경우에는 이에 대한 예방적 조치를 취하는 것이 필요하다. 셋째, 부모가 자식을 잘 키울 책임이 있는 것처럼, 과학기술자도 과학기술이 자연과 인류의 지속가능성을 담보할 수 있는 방향으로 발전할 수 있도록 세심한 주의를 기울여야 한다.

마지막으로 정결(2018)¹⁴⁾은 과학기술의 발달에 따른 기존 윤리학의 한계와 이에 따른 새로운 윤리학 또는 기존 윤리학의 변형을 요청하였다. 이때 새로운 윤리학으로서 거시 윤리학을 들었으며, 이론적 토대로 한스 요나스와 칼 오토 아펠(Karl-Otto Apel)의 이론을 활용하였다. 그는 한스 요나스가 과학기술시

12) 이유택. (2005). 요나스의 미래윤리와 책임. **동서철학연구**, 36(36), 113-137.

13) 김은철·송성수. (2012). 과학기술시대의 책임윤리를 찾아서: 한스 요나스의 '책임의 원칙'을 중심으로. **공학교육연구**, 15(1), 72-78.

14) 정결. (2018). 과학기술시대의 새로운 윤리학의 필요성에 대한 연구 -거시 윤리학의 관점을 중심으로-. **인문학연구**, 112(3), 87-113.

대에 윤리학의 변화가 필요하다는 것을 강조한 것으로 보고 거시 윤리의 범위에 포함시켰다. 이를 통해 현대과학의 위기라는 시대적 극복방안으로서 한스 요나스와 칼 오토 아펠의 연구가 활발하게 이루어져야 함을 주장하였다.

이러한 연구들은 한스 요나스의 '책임윤리'를 중심으로 과학기술시대 속에서 요구되는 윤리적 지향점을 살펴보고 그 타당성을 검토하였다고 볼 수 있다. 이러한 검토를 위해 한스 요나스가 말하는 '책임'이 무엇인지 분석한 뒤, 급변하는 사회 속에서 한스 요나스의 주장이 어떤 가치를 내포하고 있는지 탐구하였다. 그러면서 위의 선행연구는 책임윤리의 세부적인 연구와 적용을 요청하기도 하였다.

요약해보면, 한스 요나스에 대한 선행연구들은 과학기술시대에서 기술의 범위가 생명과학, 천문학, 물리학, 화학, 컴퓨터과학 등, 다양하고 전문성이 있어야 하는 분야가 되면서 각 분야에 관한 구체적인 연구와 일상생활 속의 응용을 요구하고 있다. 그러나 과학기술 분야마다 추구하는 것과 연구하는 것이 다르고 그 특성이 다르므로, 한스 요나스의 관점에서 각 분야의 바람직한 발전 방향을 제시할 필요가 있다. 이에 본 연구에서는 과학기술 분야 중에서 인공지능 기술을 중심으로 그 내용을 살펴보고자 한다. 이를 위해 한스 요나스가 집필한 저서의 번역본 『책임의 원칙: 기술 시대의 생태학적 윤리』, 『기술 의학 윤리: 책임 원칙의 실천』을 분석하고, 인공지능과 로봇윤리를 다룬 문헌들을 살펴볼 것이다.

Ⅱ. 한스 요나스의 책임이론

1. ‘책임’의 개념 및 이해

가. ‘책임’의 개념과 의미

‘책임’에 대해 논하기에 앞서, 책임의 사전적 의미와 어원을 통해 낱말의 언어적 특징과 쓰임을 살펴보자. 먼저 책임의 사전적 의미는 다음과 같다.

- ① 도맡아서 하여야 할 임무. 떠맡아서 하지 않으면 안 되는 의무.
- ② 일을 담당하여 그 결과에 대한 손실이나 꾸지람·제재(制裁)를 자기가 떠맡는 일.
- ③ (법) 불법(不法)한 행위를 한 자에게 법률 상의 불이익(不利益) 내지는 제재(制裁)가 가해지는 일. 대(對)개인적인 것과 대사회적인 것, 즉 민사 책임과 형사 책임으로 나뉨. (이희승, 1994, p. 3717)¹⁵⁾

이로 보았을 때 책임의 사전적 의미는 의무 또는 마땅히 해야 하는 일, 그리고 맡은 일을 수행하지 못했을 때 수반되는 불이익 또는 제재를 의미한다.

그리고 책임에서 책(責)과 임(任)의 자원(字源)을 살펴보면 다음과 같다. 먼저 책(責)은 “원래 貝+束(音). 束(자)는 가시의 象形으로, ‘꾸짖다, 비난하다’의 뜻. 금품(金品)을 강요하다의 뜻을 나타냄”(편집부, 1998, p. 1967)¹⁶⁾이다. 다음으로 임(任)은 “人+壬(音). ‘壬(임)’은 장시간에 걸쳐 지속적(持續的)으로 어떤 무게 있는 물건을 지니다의 뜻. 사람이 짊어지다, 지탱하다의 뜻을 나타냄.”(편집부, 1998, p. 160)이다. 이러한 의미의 자원과 오늘날 일상생활 속에서 쓰이는 ‘책임’의 맥락적 의미를 고려하였을 때 책임의 책(責)¹⁷⁾은 주로 “책임을 지우다, 책임, 직책, 의무, 처벌”의 의미를 나타내고, 책임의 임(任)¹⁸⁾

15) 이희승 (편). (1994). **국어대사전**(3판). 민중서림.

16) 편집부 (편). (1998). **한한대사전**(증보판). 민중서림.

17) 책(責)은 “① 꾸짖다, ② 나무라다, ③ 책망하다(責望--), ④ 헐뜯다, ⑤ 취하다(取--), ⑥ 받아내다, ⑦ 요구하다(要求--), 바라다, ⑧ 재촉하다, ⑨ 권하다(勸--), ⑩ 책임을 지우다, ⑪ 책임, 직책(職責), ⑫ 의무, ⑬ 처벌”의 뜻으로 쓰이고 있다. (<https://hanja.dict.naver.com>. 검색일: 2020. 3. 21.)

은 “맡기다, 주다, (책임)맡다, 지다, 보증하다, 맡은 일, 책무”의 의미를 나타내는 것임을 알 수 있다. 이상의 내용을 정리해보면 사전적 의미로서 책임은 주로 의무와 부담, 또는 어떤 결과에 대해 받게 되는 것과 범을 어겼을 때 받게 되는 법률적 불이익이나 제재로서 쓰이고 있다. 그러나 책임의 사전적 의미만으로 분석해보았을 때 법적 책임과 도덕적 책임에서 책임의 의미를 명확하게 구분하기에는 어려움이 있으며, 두 가지 의미가 혼용되어 일상생활 속에서 구분 없이 두루 쓰이고 있는 것처럼 보이기도 한다.

이러한 문제의식과 관련하여 이영록(2019)¹⁹⁾은 책임 개념의 구조와 성격, 관계 등을 개념사적 측면에서 논의하였다. 그에 따르면 책임이라는 영어 명사 ‘Responsibility’가 최초로 등장한 것은 1766년이다. 그 기원은 영국에서 법률상 왕이 아닌 장관이 책임을 진다는 내용에서 발생한 헌법논쟁으로부터 시작되었다. 이러한 용어는 프랑스로 넘어가 ‘responsabilité’ 형태로 사용되어 프랑스혁명을 통해 점차 퍼지게 되었다. 또한 라틴어 ‘responderē’도 결혼 같은 혼례의식에서 사용되는 부부간의 서약 행위에서 사용된 ‘spōndēō’에서 유래한 법적 의미를 지니고 있다. 책임을 나타내는 독일어 ‘Verantwortung’도 역시 법정에서 자기 행위의 설명 또는 옹호를 의미하는 단어로 쓰였다. 이러한 법적 용어로서의 책임은 법정에서 신의 판결이라는 맥락에서 종교적 용어로 쓰이기 시작했다. 이는 서양에서 종교적 신의 영향으로서 쓰인 ‘양심’을 매개로 책임이라는 개념이 도덕과 윤리의 영역으로 점차 확대되는 계기가 되었다. 본격적으로 시민들에게 일상적으로 쓰이게 된 것은 민주주의의 발전과 함께 책임정부의 개념이 강조되면서 정치 영역에서 책임의 의미가 중요하게 다루어진 것과 관련이 있다. 이영록은 책임이 국왕과 공직자들의 권리 또는 의무와 관련된 행위를 나타낼 때 사용되었으며, 주로 법정에서 책임이라는 용어가 자주 쓰였다고 분

18) 임(任)은 “① 맡기다, 주다, ② 능하다(能--), 잘하다, ③ (공을)세우다, ④ 배다, 임신하다(妊娠·妊娠--), ⑤ 맞다, 당하다(當--), ⑥ (책임)맡다, 지다, ⑦ 견디다, 감내하다(堪耐--), ⑧ 보증하다(保證--), ⑨ 비뚤어지다, 굽다, ⑩ 마음대로 하다, ⑪ 미쁘다(믿음성이 있다), ⑫ 당해내다(當--), ⑬ 맡은 일, 책무(責務), ⑭ 짐, 부담(負擔), ⑮ 보따리, ⑯ 재능(才能), 재주, ⑰ 협기(俠氣), 사나이의 기개(氣概), ⑱ 임지(任地: 임무를 받아 근무하는 곳), ⑲ 마음대로, 멋대로”의 뜻으로 쓰이고 있다. (<https://hanja.dict.naver.com>. 검색일: 2020. 3. 21.)

19) 이영록. (2019). 책임의 의미와 성격에 관한 역사적 탐색. **일감법학**, (42), 195-221.

석하였다.(이영록, 2019, pp. 198-201)

한편 책임이 학문적, 철학적 논의의 대상이 된 것은 알렉산더 베인(Alexander Bain)에 의해 이루어졌다. 알렉산더 베인은 형벌 집행에 상관없이 일상적으로 책임이 쓰이는 사례를 통해 책임은 무조건 형벌을 받는 것이라는 사회적 통념에 대해 의문을 가졌다. 동시에 책임을 형벌과 밀접하게 연결시키는 그 당시의 해석을 부정하였다. 여기서 이영록(2019)은 자유의사에 의한 원인행위로서 “귀책(Imputation)”이라는 용어를 사용하여 책임에 대한 철학적인 분석을 시도하였다. 이를 위해 고대 그리스 철학으로부터 그 근거를 제시하였다. 이영록에 의하면, 아리스토텔레스(Aristoteles)는 행위의 원인을 행위자 외부와 행위자 자신으로 구분하여 어떤 결과의 원인을 행위자 자신으로 볼 수 있도록 이론적 틀을 마련하였다고 한다. 예를 들어 사람의 움직이는 행위는 물리적 힘이나 타인에 의해서 움직일 수 있고 자신이 움직이고 싶어서 움직일 수도 있다. 전자의 경우 행위의 원인은 행위자 외부에 있고 후자의 경우 행위의 원인은 행위자 자신에게 있다. 이 중에서 이영록은 “행위의 원인이 행위자 자신에게 있는 경우란 그 행위가 행위자의 자유로운 의지에서 비롯된 것으로 볼 수 있을 때”(이영록, 2019, p. 202)에 주목하였다. 행위자의 자유로운 의지로 인한 행위가 행위의 평가와 관련된 ‘*imputatio*’라는 용어로 개념화된 것이다. 여기서 행위의 평가는 곧 보상이나 제재, 또는 칭찬이나 비난이 따르는 것으로, 어떤 행위에 대한 평가적 반응을 행위자의 자유로운 의지와 직접적으로 관련된다.

이러한 접근은 자연과학의 발전과 함께 인간 행위에 대한 결정론(決定論)²⁰⁾과 자유의지론(自由意志論)²¹⁾ 사이의 대립으로 심화되었다. 논의 초기에는 ‘귀책’ 개념이 인간의 자유와 뗄 수 없는 관계로 논의가 전개되었는데, 자연과학의 발달로 결정론적 사고가 확산하면서 자유의지가 의심받게 되자 ‘귀책’ 개념의

20) 인간의 행위를 포함하여 이 세상에서 일어나는 모든 일은 우연이나 선택의 자유에 의하여 일어나는 것이 아니라, 일정한 인과(因果) 관계의 법칙에 따라 결정된다는 이론을 말한다.(결정론. <https://terms.naver.com>. 검색일: 2020. 2. 16.)

21) 인간의 의지를 비롯하여 모든 자연 현상과 역사 현상은, 합법칙성과 인과성에 의하여 결정되지 아니하고 자신 스스로 결정한다는 이론. (자유의지론. <https://ko.dict.naver.com>. 검색일: 2020. 2. 16.)

이론적 배경이 흔들리게 되었다. 이에 따라 ‘귀책’ 개념은 결정론과 자유의지론 사이에서 이론적 교착상태에 빠져버리게 되는데, 이영록(2019)은 이 상황을 대체할 용어로 ‘책임’이라는 용어가 나타나게 되었다고 보았다. 이때의 책임은 인간 행위와 관련된 법적, 도덕적 사회문제를 실질적으로 해결할 수 있는 개념적 도구로서 작용하였다. 그러므로 책임이라는 용어는 법적 필요에 의한 응답(Response)을 근간으로 발전해 왔으며 인과관계, 도덕적 양심의 형태로 발전하여 현재의 책임이 있게 되었다고 볼 수 있다.(이영록, 2019, p. 208)

책임 개념의 구조와 성격, 관계 등을 개념사적으로 분석한 이영록(2019)의 연구는 서양의 역사 속에서 책임이라는 용어의 탄생과 배경을 자세하게 서술하였다고 볼 수 있다. 그 기원은 법률, 법적 논쟁 등으로부터 발생하였다고 말할 수 있으며 개인의 양심, 또는 죄책감 등의 감정적 반응과 연관되어 도덕적 책임으로 확대되었다고 정리할 수 있다. 결국 이러한 책임은 행위자의 내·외부에 직접적인 동기로 인해 발생하는 특징으로 보아 책임은 법, 도덕, 감정 등의 다양한 원인으로부터 발생한 개념이라 할 수 있다.

나. 한스 요나스의 관점에서 본 법률 용어로서 책임

책임의 어원에서 살펴보았듯이 처음에 책임은 법적 용어로서 사람들에게 쓰였다. 한스 요나스에 의하면,²²⁾ 법적 용어로서 “책임의 전제 조건은 인과적 권력이다. 행위자는 자신의 행위에 대해 책임을 져야 한다.”(Hans Jonas, 1984/1994, p. 168) 이때 판단기준은 행위의 동기보다는 결과에 있다. 나쁜 의도의 동기가 아니더라도, 또는 예상하지 못한 결과일지라도, 어떤 결과로 인해 피해가 발생하게 되었다면 행위자는 그 피해에 대해 보상을 해야 한다.(Hans Jonas, 1984/1994, p. 168) 이 피해보상이 알맞게 이루어지기 위해서는 피해를 발생시킨 주체를 면밀히 따져야 하고 피해의 정도나 보상의 범위를 명확하게 밝힐 수 있어야 한다. 이러한 보상의 의미인 책임은 대체로 법률에서 사용되는데 우리나라 법률을 통해 살펴보면 다음과 같다. 민법 제 750조를 살펴보면 “고의 또는 과실로 인한 위법행위로 타인에게 손해를 끼친

22) Hans Jonas. (1994). *책임의 원칙: 기술 시대의 생태학적 윤리(Das Prinzip Verantwortung. Versuch einer Ethik Für die technologische Zivilisation)*(이진우 역). 서광사. (원저 1984 출판)

자는 그 손해를 배상할 책임이 있다.”²³⁾ 여기서 법률용어로서 고의(故意)란 “자기의 행위에 의하여 일정한 결과가 생길 것을 인식하면서 그 행위를 하는 경우의 심리 상태”²⁴⁾를 말한다. 그리고 과실(過失)이란 “부주의로 인하여, 어떤 결과의 발생을 미리 내다보지 못한 일”²⁵⁾을 말한다. 민법 제750조와 ‘고의’, ‘과실’의 의미를 살펴보았을 때 손해배상은 행위자의 의도가 어떤지에 상관없이, 즉 행위의 동기보다는 행위의 결과에 따라 결정된다. 결과의 원인 또는 동기를 따지는 것은 손해배상이 결정된 뒤 고려되는 사항으로 그것의 위법 정도에 따라 손해배상의 비용과 범위는 달라지게 된다. 그렇다고 해서 손해배상의 책임이 무효가 되는 것은 아니다. 단지 손해배상의 양, 정도만 달라질 뿐이다.

한스 요나스는 이러한 법적 책임에 대해 다음과 같은 문제점을 제기하였다. “그런데 예전부터 법적인 속죄의 이념과 처벌의 이념은 혼합되어 왔는데, 도덕적 의미를 지니는 후자는 원인적 행위를 도덕적으로 죄 있는 것으로 평가한다.”(Hans Jonas, 1984/1994, p. 168) 우리는 어떤 잘못을 하거나 법을 어겼을 때 ‘죄가 있다, 죄를 지었다.’라는 표현을 사용하곤 한다. 이렇게 사용되고 있는 ‘죄(罪)’는 크게 세 가지의 의미로 쓰인다. 첫째, “양심이나 도리에 벗어난 행위”, 둘째, “잘못이나 허물로 인하여 벌을 받을 만한 일”, 셋째, “법률에 위반되어 처벌을 면하지 못하는 불법 행위”이다.²⁶⁾ 이때 ‘죄’는 위에서 말한 ‘손해배상’의 판단기준과는 다른 방향으로 나아간다. 죄를 이야기할 때 행위의 결과뿐만 아니라 행위자의 의도, 동기, 숙고 따위를 모두 고려해야 하는 것이다. 만약 법정에서 어떤 행위를 유죄로 판결 내린다면 그에 마땅한 처벌로 책임을 물을 것이다. 이는 도덕적 책임이 법률 형태로 작용한 것이다. 한편 한스 요나스는 이러한 도덕적 책임에서도 특정한 행위가 현실에서 행해지거나 최소한 시작되어야 한다고 하였다.(Hans Jonas, 1984/1994, p. 169) 어떠한 의도에서든지 우선 행위가 외적으로 드러나야 한다는 것이다. 어떤 사람이 단지 끔찍한 범행을 생각만 했다고 해서 법적으로 그 사람에게 책임을 물을 수는 없다.

23) 민법 제750조. <http://www.law.go.kr>. (검색일: 2020. 2. 27.)

24) 고의. <https://ko.dict.naver.com>. (검색일: 2020. 2. 27.)

25) 과실. <https://ko.dict.naver.com>. (검색일: 2020. 2. 27.)

26) 죄. <https://ko.dict.naver.com>. (검색일: 2020. 2. 27.)

외적인 행위가 실행되었을 때 그로부터 법적인 책임 논의는 시작되는 것이다.

이렇게 법률적 의미의 책임 안에서 손해배상 차원의 책임과 처벌 차원의 책임을 각각 살펴보았다. 정리해보면 손해배상 차원의 책임은 “개인 사이의 재산, 신분 따위의 관한 법률관계를 규정”²⁷⁾한 사법의 형태로, 처벌 차원의 책임은 “범죄와 형벌에 관한 법률 체계”²⁸⁾인 형법의 형태로 구분하여 볼 수 있다. 그리고 실제 법률에서는 이러한 구분이 명확하지 않고 혼용되어 쓰이고 있다. 따라서 법적 의미의 책임은 행위자의 행위에 따라 손해배상이나 처벌 등의 실제적 형태로 책임이 부과된다. 그러나 이러한 과정에서 수반되는 감정이 있다. 죄책감, 속죄의 마음, 후회, 반성 따위가 그것인데, 한스 요나스는 이러한 감정의 수반을 반동적이라고 보았다.(Hans Jonas, 1984/1994, p. 169) 행위에 따라서 수반되는 감정의 정도가 달라지는 것이다. 이렇게 행위에 따라 달라지는 감정에 대해 한스 요나스는 형식적이라고 보았다.

간단히 말해서 이렇게 이해된 ‘책임’은 스스로 목적을 설정하는 것이 아니라, 사람들 사이의 모든 인과적 행위에 대해서 해명을 요구할 수 있다는 전적으로 형식적인 책임 부담이다. 그것은 도덕 전제 조건이지만 도덕 그 자체는 아니다. 책임과 동화되는 감정은 - 예감뿐만 아니라 추후의 감정도 - (자신의 행위에 책임을 지는 자세로서) 비록 도덕적이기는 하지만, 그것이 가지고 있는 순전한 형식성으로 말미암아 궁극적으로 인간선에 대한 긍정적 목적의 제시, 확인, 동기 유발 등과 관계가 있는 윤리 이론의 감정적 원칙을 제공할 수 없다. (Hans Jonas, 1984/1994, pp. 169-170)

결국 한스 요나스는 법적 용어로부터 유래된 책임과는 다른 책임을 말하고자 한다. 그렇지만 책임의 인과관계, 즉 ‘어떤 것’에 대한 책임이라는 형식은 유지하면서 윤리적 책임이론으로 발전시키고자 하였다. 이러한 흐름 속에서 한스 요나스는 일상적으로 활용되는 책임의 쓰임을 분석하여 불확실한 미래에 요구되는 책임을 제시하였다. 어떠한 과정으로 미래윤리로서의 책임윤리를 제시하였는지 그 과정을 살펴보자.

27) 사법. <https://ko.dict.naver.com>. (검색일: 2020. 2. 27.)

28) 형법. <https://ko.dict.naver.com>. (검색일: 2020. 2. 27.)

2. 한스 요나스의 책임이론에 대한 분석

가. 한스 요나스의 책임에 대한 이론적 배경

이러한 문제의식에서 한스 요나스는 무엇보다 존재에 대한 책임을 주장하였다. 이러한 주장을 뒷받침하기 위해 한스 요나스는 존재 자체를 의무의 대상으로 생각해보는 “존재가 과연 당위를 근거지을 수 있는가”(Hans Jonas, 1984/1994, p. 142)라는 의문을 갖고 목적의 문제를 가치의 문제로 전환하고자 시도하였다. 이때 한스 요나스가 말하는 존재란 무엇인가? 한스 요나스는 ‘존재’를 이야기하기 위해 ‘인간은 존재해야만 하는가?’라는 물음을 설정하고 논의를 전개한다. 다음은 한스 요나스가 제시한 설명이다.

인간은 존재해야만 하는가? 이 물음에 올바르게 대답하기 위해서 우리는 우선 다음의 물음에 대답해야 한다. 그 어떤 것에 관해 그것이 존재해야 한다고 말하는 것은 무엇을 의미하는가? 물론 이 물음은 도대체 어떤 무엇이 - 무(無) 대신에- 과연 존재**해야만 하는가** 하는 물음으로 되돌아간다.

앞의 마지막 두 물음의 차이는 적지 않다. 이런저런 사물이 존재해야만 한다는 존재 당위에 관련된 첫째 질문은 주어진 존재의 범위 안에서 제기되는 대안 가능성을 비교함으로써 상대적으로 대답할 수 있다. 무엇인가가 존재해야 한다면, 그것은 이것이 저것보다 더 좋기 때문이며, 따라서 이것이 존재해야만 한다. 대안이 다른 어떤 존재가 아니라 무(無) 자체인 둘째 물음은 단지 절대적으로 대답할 수 있다. 예를 들면 무(無)와는 정도에 있어서의 어떤 비교도 불가능하기 때문에 존재 그 자체는 “선하다”고 할 수 있다. 따라서 실존은 이에 대한 (반대적 대립이 아닌) 어떤 모순적 대립에 대해서도 우선권을 “가져야 한다.” (Hans Jonas, 1984/1994, p. 95)

한스 요나스는 존재를 설명하기 위해 존재의 반대항 무(無)와 비교하여 무(無)보다 존재가 우선함을 주장하였다. 그렇지만 한스 요나스가 더욱 중요하게 여긴 것은 어떤 것이 존재해야만 하는 이유, 즉 존재의 당위에 대한 것이다. 이를 위해 한스 요나스는 행복의 추구를 예로 들어 설명하였다.(Hans Jonas, 1984/1994, pp. 142-143) 여기에 ‘행복을 추구해야 한다.’와 ‘행복을 추구해도 된다./추구할 수 있다.’라는 말이 있다. ‘행복을 추구해야 한다.’라는 마땅히

해야 하는 '의무'로 볼 수 있고, '행복을 추구해도 된다./추구할 수 있다.'라는 행복을 행할 수 있거나 요구할 수 있는 '권리'로 볼 수 있다. 대체로 사람들은 '행복을 추구해야 한다.'라는 말에 대해 동의하거나 동의하지 않는다. 왜냐하면 '사람은 행복을 추구해야 한다.'라고 할 때 '반드시' 또는 '필연적으로'의 의미가 담겨 있다는 것을 생각하면 사람에 따라 행복을 추구하거나 행복이 아닌 다른 것을 추구할 수 있기 때문이다. 이처럼 개인에 대한 의무는 개인 또는 상황에 따라 긍정 또는 부정의 판단이 내려진다.

한편 '행복을 추구해도 된다./추구할 수 있다.'라는 말에 대해서 사람들은 대체로 동의를 한다. 이때 말의 의미를 좀 더 깊게 살펴보면 행복을 추구해도 되고 안 해도 되는 경우를 모두 포함하고 있다. 다시 말해, 이 말 속에는 필연적인 의미가 있지 않다. 이처럼 행복의 추구가 개인의 권리로서 인정되면 타인의 행복 추구 또한 마땅히 인정할 수 있어야 한다. 왜냐하면 개인의 권리는 '나'에 대해서만 가지고 있는 것이 아니라 다른 사람들도 마찬가지로 권리를 가지고 있기 때문이다. 이때 '나'와 '타인'은 동등한 도덕적 지위를 갖고 있다는 전제가 설정되어 있다. 내가 가지고 있다면 다른 사람도 마땅히 가질 수 있는 것이다. 이러한 전제에서 자신의 권리는 인정하는데 다른 사람의 권리는 인정하지 않는 것은 불공평한 것이다. 이는 곧 '타인의 권리를 인정해야 한다.'라는 의무로 발전하게 된다. 이때 다른 사람의 권리를 인정하는 것은 개인의 의무와는 달리 필연적인 것이 된다. 이와 비슷한 방식으로 논증하면 '타인'에 대한 의무, '자연'에 대한 의무로 확장하여 '존재'에 대한 의무를 논할 수 있는 지점을 마련해준다.

이러한 논증에서 한 가지 의문이 발생한다. 타인은 권리를 가지고 있다고 볼 수 있지만 자연은 무엇을 가지고 있는가? 한스 요나스는 이에 대해 자연을 비롯한 모든 존재는 권리 이상의 의미인 목적이 있다고 주장한다. 한스 요나스에게 목적(目的)이란 "그것을 위하여 어떤 사태가 실존하고, 또 그것을 산출하거나 보존하기 위하여 어떤 과정이나 행위가 실행되는 것이다."(Hans Jonas, 1984/1994, p. 102) 여기서 '그것'은 목적이라고 볼 수 있다. 한스 요나스에게 목적 자체는 좋고 나쁨의 가치판단의 대상이 아니다.(Hans Jonas, 1984/1994, p. 102) 단지 하나의 특성이다. 예를 들어 스피커는 소리를 출

력하기 위해 존재하고 연필은 글이나 그림을 표현하기 위해 존재하는 것이다. 또한 물체의 이동은 위치를 변화시키기 위해 나타나는 것이다. 우리가 가치판단 하는 것은 '목적' 그 자체가 아닌 목적을 가진 대상의 '특징'에 대한 가치판단이다. 스피커의 경우, 우리는 스피커의 모양, 색깔, 사양, 소리, 가격 등 다양한 요소들을 고려하여 스피커의 가치를 판단하곤 한다. 사람의 경우, 우리는 사람의 행동이나 생각, 됨됨이를 보고 그 사람을 도덕적으로 판단한다. 이때 가치의 판단기준은 개별 주체의 개인적인 목표와 의지, 또는 그들의 주관에 따라 달라지는 것으로, 절대적이지 않다. "세계에 대해 '아니오'라고 말할 수 있는 것은 인간적 자유의 특권이다."(Hans Jonas, 1984/1994, p. 144)라고 말한 한스 요나스의 생각은 가치판단의 상대성을 보여준다. 결국 인간의 인식은 어떤 대상에 대한 가치를 인식하는 것으로, 이러한 '가치'의 관점에서 세계를 바라본다면 존재에 대해서 긍정 또는 부정을 임의대로 판단하게 되어 일관성이 없어진다. 나름의 기준에 따라 대상의 가치를 판단하기 때문에 평가가 제각각이고 자의적인 특징을 띠게 되는 것이다. 따라서 한스 요나스는 '가치'의 개념과 동일하지 않은 '선'의 개념을 요청하였다. 이를 위해 "어떤 의무의 구속력을 존재 내에서의 선의 근거로 정당화할 수 있다면 그것은 바로 선과 존재의 관계이다."(Hans Jonas, 1984/1994, p. 145)라고 설명하였다. 여기서 '선'은 존재 자체를 실제적이고 의무적인 긍정으로 정당화하는 요소로 변하지 않는 고유의 '목적' 그 자체를 의미한다.

한스 요나스는 이러한 '선'의 개념을 도입함으로써 '선'의 요청을 감지하고 실행할 수 있는 의지를 말하고자 한다.(Hans Jonas, 1984/1994, p. 147) 다시 말해, "도덕적 '지시명령'이 예를 들어 인격신과 같은 명령하는 의지로부터만 나오는 것이 아니라 그 자체로 선한 즉자적 선이 자신의 현실화를 요구하는 내면적 요청으로부터 나온다고 우리는 말하고자 한다."(Hans Jonas, 1984/1994, p. 148) 보충하여 설명하면 개인의 욕망, 필요, 선택에 상관없이 존재 자체인 선은 현실로 드러날 것, 존재할 것을 요구한다. 이때 선은 개인의 처지, 입장, 상황에 따라 상대적인 착함과 나쁨 따위를 말하는 것이 아니다. 존재 고유의 목적을 가진다는 것 자체가 한스 요나스가 말하는 '선'이다. 한스 요나스는 존재하는 모든 것은 목적을 가지기 마련이며 목적이 없는 무(無)

목적과 비교했을 때 목적을 가지는 것은 직관적으로 보았을 때 언제나 우월하다고 보고 있다.(Hans Jonas, 1984/1994, p. 149) 한스 요나스는 이를 존재 내에서 목적 자체가 스스로를 증명하는 존재론적 공리로 전제하였다.(Hans Jonas, 1984/1994, p. 150) 생명이라는 존재를 생각해보면 한스 요나스 주장의 요지를 이해하는 데 도움을 준다. “생명의 존재 양태는 행동을 통한 보존이다.”(Hans Jonas, 1984/1994, p. 152) 다시 말해, 생명의 존재 목적은 행동을 통한 자기보존이다. 모든 생명은 삶과 죽음 앞에서 불의(不意)의 사고나 천재지변, 전쟁 따위의 사회적 요소, 그에 따르는 우울감, 괴로움 등 외부요인의 작용을 제외하고는 삶을 선택한다. 곧 모든 생명은 죽음보다는 삶을 선택한다. 모든 존재에는 목적, 이유가 있는 것이다. 다만, 이러한 관점에서 자살(自殺)은 예외적이면서 옳지 못한 행동이다. 자살은 외부요인이 작용한 부정적인 경우라고 볼 수 있다. 이 경우를 제외하고, 모든 존재는 비(非)존재보다 우선한다. 이를 한스 요나스는 존재론적 공리로 전제하는 것이다.

존재의 목적 자체는 ‘선’이지만, 그럼에도 불구하고 사람들은 가치 있는 목적과 가치 없는 목적을 구별하려는 경향이 있다고 한스 요나스는 보았다.(Hans Jonas, 1984/1994, p. 156) 욕구, 이익, 성향 등의 경향성 있는 우리의 판단은 옳든 그르든 간에 ‘무엇이 어떠하다.’라는 판단을 할 수 있는 가능성을 함축하고 있다. 그러므로 우리의 판단과 존재의 목적 사이에 일치하는 경우도 있게 되고 일치하지 않는 경우도 있게 된다. 만약 가치와 목적 사이의 관계가 우리의 경향성에 상관없이 언제나 타당하다면, 한스 요나스는 그것을 선하다고 본다.(Hans Jonas, 1984/1994, p. 156)

선한 인간은 자신을 잘 만든, 즉 행위를 잘한 사람이 아니라 선을 위하여 행한 사람이다. 그런데 선은 세계 내의 사태이며, 바로 세계의 사태이다. 도덕성은 결코 자기 자신을 목표로 가질 수 없다. 그러므로 제일 중요한 문제는 형식이 아니라 행위의 내용이다. 이런 의미에서 도덕은, 종종 자아의 상태를-다시 말해서 세계 사태에 속하는 의무에 적합한 상태-대상으로 가지기도 하지만, “자아가 없다.” 물론 무자아성 자체가 도덕적인 것일 수는 없다. 의무 자체가 대상은 아니다. 도덕법이 도덕적 행위의 동기를 유발하는 것은 아니다. 오히려 나의 의지에 대립하고 나의 경청을 요구하는 세계 내의 즉자적 선의 호소가 도덕법에 **적합한** 것이다. 이 호소를 경

칭하라는 것이 바로 도덕법이 명령하는 바이다. ... 이 도덕법이 나에게 다가와 나를 촉발시켜서 나의 의지를 움직일 수 있으려면 나는 그와 같은 것에 대해 촉발될 수 있어야만 한다. 우리의 감정적 부분이 일종의 역할을 해야만 하는 것이다. 통찰력이 호소를 매개하는 바와 마찬가지로 호소가 우리의 감정 속에서 일종의 대답을 발견한다는 것은 우리가 가지고 있는 도덕적 본성의 본질에 속한다. 그것은 책임성의 감정이다. (Hans Jonas, 1984/1994, pp. 157-158)

다시 말해, 선하다는 것은 개인의 욕구, 의지, 행위에 의해 결정되는 것이 아니라 존재해야 한다는 의무 그 자체에 대한 우리의 경청인 것이다. 이것은 도덕법, 즉 도덕적 본성의 본질로 나타난다. 이러한 경청으로 반응하는 것은 감정적으로 나타나는데, 이러한 감정을 한스 요나스는 “책임성의 감정”으로 표현한다. (Hans Jonas, 1984/1994, p. 158) 정리해보면 한스 요나스의 책임이란 존재론적(存在論的) ‘선’의 호소에 경청하는 것으로 존재해야 한다는 의무를 실행하고자 하는 의지를 말한다.

한스 요나스의 책임이론은 이성적인 측면과 감정적인 측면을 동시에 고려한다. 이성적인 측면은 객관적인 측면으로 분석적인 성격이고 책임의 타당성을 논증하는 역할을 한다. 감정적인 측면은 의지를 움직일 수 있게 하는 요소로 경험적이고 주관적이다. 한스 요나스의 주장을 살펴보면 대체로 전통윤리는 감정적인 측면보다 이성적인 측면을 중시했다는 점을 지적하였는데, 자신은 이성과 감정, 두 가지 요소를 모두 고려하면서 감정적인 측면의 필요성을 주장하였다. (Hans Jonas, 1984/1994, p. 158) 이를 위해 한스 요나스는 전통윤리에서 쓰인 감정적 요소와 대비하여 책임의 주목할 만한 특징을 제시하였다. 플라톤의 ‘에로스’, 유태인의 ‘신에 대한 경외’, 기독교적 ‘사랑’, 칸트의 ‘존경’ 등 전통윤리에서 감정은 최고선을 지향하는 종류의 것들이다. 이때 목적은 최고선에 도달하는 것, 최고의 대상에 가까워지는 것을 말한다. (Hans Jonas, 1984/1994, p. 161) 책임의 감정 또한 이런 방식으로 보았을 때 최고선을 향하여 개인의 의지가 수동적으로 움직이는 것처럼 보인다. 하지만 책임의 감정은 전통윤리에서 말하는 도덕적 이상, 선을 지향하는 것이 아닌 어떤 존재의 요청으로부터 자연스럽게 나오는 것이다. 그 요청에 응답하여 따르는 것이 한스 요나스가 말하는 ‘선’이고 ‘책임’이다. (Hans Jonas, 1984/1994, p. 166)

나. 한스 요나스의 책임 이해

법률 용어로서 책임은 행위자의 행위와 그로 인한 결과에 대해서 주로 논의되었다. 그러나 한스 요나스는 법적 용어로부터 유래된 책임과는 다른 책임을 말하고자 하였다. 그가 강조하는 책임은 앞으로 행위 되어야 할, 우리가 해야 하는 일들에 대한 책임이다. 이를 한스 요나스는 “나의 행위에 대한 요청권을 제기하는 사태에 대해 책임이 있다.”라고 표현하였다.(Hans Jonas, 1984/1994, p. 170) 내 권력 또는 행위의 영향을 받은 어떤 사태에 대해서는 그 모습이 눈에 당장 보이지 않더라도 책임이 있다고 보는 것이다. 이러한 형태의 책임은 자신의 감정과도 결합된다. 한스 요나스는 어떤 일이 행위자의 권력에 영향을 받게 되어 그 일이 행위자에게 의존하게 된다면 그 행위자는 어떤 사태에 대해 감정적으로 참여하게 된다고 보았다.(Hans Jonas, 1984/1994, p. 171) 예를 들어 환경보호에 대한 책임의 경우 인간의 무분별한 개발과 소비로 환경이 많이 파괴되었다. 파괴된 환경에 대해서 자연 스스로 치유할 수 있는 범위는 이미 넘어서고, 따라서 인간이 개입함으로써 환경보호를 위해 앞장서고 있다. 이때 우리는 형식적인 의무로서 환경보호에 힘쓰는 것이 아닌 자연을 생각하는 마음, 자연친화적 사고, 과거 인간이 저지른 잘못에 대한 반성 등 감정적으로 참여하는 경향을 생각해볼 수 있다. 이러한 책임의 형태는 법률 용어로서 책임과는 다르지만 도덕적 의지의 형태와 유사한 성격을 띠게 된다. 한스 요나스는 이에 대해 경험적으로 접근하면서 “비호혜적 관계에 대한 책임, 계약적 책임, 부모의 책임, 정치인의 책임”을 중심으로 설명하였다.

한스 요나스에 의하면 “비호혜적 관계에 대한 책임”은 어떤 조직이나 단체 등의 공동체에서 주로 발생한다. 한스 요나스는 조직 내에서 책임의 대상은 궁극적으로 조직의 목표 달성이나 성공이라고 하였다. 조직 내에 있는 동료들의 안녕을 고려하는 것과는 별개의 책임이다. 다시 말해, 집단 내 상호관계 속에서 문제가 발생했을 때 비호혜적 관계에 대한 책임과 다른 형태의 책임이 발생하게 되는 것이다.(Hans Jonas, 1984/1994, p. 173) 예를 들어 어떤 프로젝트를 수행할 때 프로젝트에 참여한 사람들은 프로젝트 자체에 대해 우선적으로 책임이 있다. 하지만 프로젝트를 수행하는 과정 중에 갈등이나 문제가 발생하게 되었을 경우 프로젝트 목표에 따르는 책임과는 다른 타인에 대한 상호 책임

이 발생하게 된다. 타인에 대한 상호 책임은 용기, 결단력 등의 미덕으로 나타난다. 대체로 이러한 상호 책임의 경우 문제에 따라 특수하며 비호혜적 관계에 대한 책임과 비교했을 때 일시적이다. 그러나 프로젝트가 부정적인 상황으로 악화될 경우 개인 간 대립이나 갈등 상황으로 인해 프로젝트의 진행이 원활하지 못하게 된다. 이와 달리 비호혜적 관계에 대한 책임이 작용하게 되면 목표 달성을 위해 상호 협동과정에서 지속적인 교류를 추구하는 기틀을 마련해준다. 만약 어떤 사람이 공동체의 목표에 어긋나는 행위를 하는 경우, 또는 공동체에 위협이 되는 행위를 하는 경우, 한스 요나스는 이러한 행위를 보고 ‘무책임한 행위’라고 하였다.(Hans Jonas, 1984/1994, p. 173)

한스 요나스에게 “계약적 책임”이란 직무의 책임과 같이 지시의 부과와 수용을 통해 ‘인위적’으로 제도화된 책임이다.(Hans Jonas, 1984/1994, p. 174) 직무, 담당, 역할 등의 다양한 형태로 어떤 계약이 맺어졌다면 그 계약은 우선 자신이 스스로 선택한 것이다. 이러한 관점에서 계약적 책임은 계약을 반드시 이행해야 한다는 의무로부터 자유로워 보인다. 마음만 먹으면 얼마든지 계약을 파기할 수 있는 것처럼 보이기도 한다. 하지만 사회적 구조 속에서 어떠한 계약을 맺는다는 것은 단순히 개인적 차원의 의무를 넘어서 집단, 공동체로 확대되어서 자신의 행위나 수행능력을 사회적으로 이바지한다는 의미를 내포하고 있다. 또한 계약의 대상은 개인적 차원이 아닌 구성원의 합의로부터 발생한 것으로 보고 있다. 그래서 계약적 책임은 “어떤 사전적 동의에 의존하지 않으며, 취소할 수 없으며, 파기할 수 없는”(Hans Jonas, 1984/1994, p. 174) 자연적 책임과 성격이 유사하다고 한스 요나스는 말하고 있다.

“책임을 획득하기 위하여 권력을 추구하고, 최고의 책임을 목적으로 최고의 권력을 추구하는 정치인”(Hans Jonas, 1984/1994, p. 175)의 책임은 계약적 책임에서 한 단계 더 나아간다. 한스 요나스에 따르면, 정치인의 경우 정치 권력을 획득하기 위해 민주적인 절차를 거쳐 정당한 방식으로 획득해야 한다.(Hans Jonas, 1984/1994, p. 176) 이때 정치권력을 획득하기 위해서는 정치인, 또는 정책, 공약에 대한 시민들의 지지가 필요하다. 그리고 이 과정을 통해 최종적으로 정치권력을 획득한 경우 정치인은 정치권력을 자신이 아닌 시민들을 위해 활용해야 한다. 여기서 정치인이란 위선적이고 이기적인 폭정을 피

하는 정치인이 아니다. 정당하게 정치활동에 임하고 지역사회나 국가를 위해 힘써 일하는 “진정한 정치인(Homo Politicus)”(Hans Jonas, 1984/1994, p. 176)이다. 진정한 정치인의 “책임의 대상은 국가(Respublica), 즉 공동적 사태이다.”(Hans Jonas, 1984/1994, p. 176) 정치라는 직무에 진정으로 뜻이 있는 정치인은 국가 구성원인 시민의 행복과 만족을 위해 힘써 일할 것이다. 이때의 정치인은 정치권력을 악용하여 명성이나 쾌락, 부를 추구하는 것이 아니라, 자신의 역할을 임명해준 시민들에게 바람직한 정책 실현으로 보답해야 진정한 정치인이라고 할 수 있는 것이다. 정치인의 책임은 다른 계약적 책임과 비교했을 때 그 대상(국가)이 비교적 무겁고 광범위하므로 한스 요나스는 자연적 책임인 부모의 책임과 견줄만한 책임으로 보고 있다.(Hans Jonas, 1984/1994, p. 176) 정치인의 책임과 부모의 책임을 비교하면서 미래 윤리적 책임의 근본적인 속성을 제시했는데, “총체성”, “연속성”, “미래”가 그것이다. 이 세 가지 속성은 인공지능기술을 다루는 데 있어 요구되는 인간의 책임을 논하는데 필요한 직접적인 근거를 마련해준다.

다. 책임의 속성: 총체성, 연속성, 미래

한스 요나스는 자연적 책임인 부모의 책임과 인위적 책임인 정치인의 책임을 비교하면서 책임의 주요 속성이라고 말할 수 있는 특징을 다음과 같이 제시하였다.

첫째, 책임의 총체성이다. 책임의 총체성이란 “대상의 총체적 존재, 즉 적나라한 실존으로부터 최고의 관심에 이르는 모든 양상을 포괄”(Hans Jonas, 1984/1994, p. 185)하는 것이다. 다시 말해, 대상과 관련된 모든 것에 대해 관심을 두는 것이다. 이러한 책임의 속성을 확인할 수 있는 경우는 자식에게 온갖 정성을 기울이는 부모의 모습이다. 부모의 경우 부모가 양육하고 있는 ‘아이’가 책임의 대상이다. 한스 요나스는 일반적으로 부모는 아이의 모든 것을 책임진다고 하였다. 아이가 태어난 그 순간부터 부모는 아이의 신체적 발달을 책임진다. 식생활, 수면, 운동, 청결, 건강 등 갓 태어난 생명에 온 정성을 기울인다. 시간이 지나고 아이가 성장하면서 부모의 책임은 능력, 관계, 지식, 성격, 가치관 등 교육과 사회화 측면에까지 확대된다. 그리고 좀 더 나아가 그

아이가 성인이 되어서도 부모의 책임은 언제나 유효하다고 보았다. 그러나 부모와 자식의 관계가 사회, 관습, 문화 등 외부요인에 따라 그러지 않는 경우가 오늘날 종종 발생하고 있다. 경제적 여건이나 사고 등 부득이한 사정으로 인해 버려지거나 방치되는 아이들, 법적·행정적인 절차에 의해 입양되거나 파양되는 경우를 예로 들 수 있다. 하지만 순수하게 존재론적인 측면에서 보았을 때 아이에 대한 부모의 책임은 절대적이다. 이는 부모가 한 사회의 구성원으로 존재하기 이전에, 한 생명으로서의 자기 자식을 보호하는 인간의 본성에 근거한 것이다. 이러한 측면을 전제로 한다면, 자식에게 정성을 기울이는 부모의 모습을 책임의 원초적인 형태로 볼 수 있다. 이 책임의 원형은 한 아이가 태어남과 동시에 그 아이가 부모로부터 받게 되는 책임의 형태이자 부모로서 드러내는 아이에 관한 모든 것의 책임이다.

이러한 부모의 책임은 다른 형태로도 나타나는데 그것은 바로 한스 요나스가 말하는 정치인의 책임이다. 부모가 책임지고 있는 대상이 아이라면 정치인이 책임지고 있는 대상은 국민이다. 한스 요나스의 말을 빌리자면, “인간적 삶이 가능하고 인간 생명의 실존이 지속되고 좋은 삶이 가능하기 위해 국가는 실존하게 된다.”(Hans Jonas, 1984/1994, p. 185) 정책을 수립하거나 결정을 할 때 단순히 한 가지 입장만을 고려하는 것이 아니라 재화, 서비스, 복지, 외교 등 국민의 삶과 관련된 복합적인 측면을 고려하는 정치인의 모습이 한스 요나스의 주장을 뒷받침해준다. 여기서 진정한 정치인은 오로지 자기 자신을 위해 권력을 사용하는 것이 아니라 한 국가에 거주하는 국민의 안전, 삶, 행복, 건강, 복지 등 국민의 총체적인 것을 책임져 자신의 의무를 수행한다. 이러한 특징은 부모의 책임과 어느 정도 유사하다. “물리적 실존으로부터 최고의 이익에까지, 실존의 안전으로부터 충족에까지, 그리고 안녕으로부터 행복에까지 이르는”(Hans Jonas, 1984/1994, p. 186) 것처럼 아이에 대한 부모의 총체적인 책임과 국민에 대한 정치인의 총체적인 책임은 유사한 속성을 가지고 있다.

둘째, 책임의 연속성이다. 책임의 연속성은 간단하게 말해서 “책임의 실행이 중단되어서는 안 된다.”(Hans Jonas, 1984/1994, p. 191) 부모의 경우 아이의 생명이 중단 없이 계속되는 한, 그리고 정치인의 경우 국민이 존재하는

한 그들의 책임은 중단되어서는 안 된다. 여기서는 시시각각 달라지고 개인의 성향에 따라 달라지는 어떤 문제 상황이나 사태에 주목하기보다는 존재로서의 ‘아이’ 그 자체, 존재로서의 ‘국민’ 그 자체에 주목할 필요가 있다. 만약 어떤 문제 상황이나 사태에 주목하게 되면 그 문제가 해결됨으로써 책임도 끝난다. 그러나 내 자식이 존재하는 한, 국민이 존재하는 한 부모와 정치인의 책임은 영원히 지속되어야 한다. 이렇게 대상의 실존 자체에 주목하게 되면 실존에 대해 역사적으로 접근하는 것이 가능해진다. 한스 요나스는 이러한 접근에 주목하기 위해 다음과 같은 물음을 제기한다. “다음에는 무슨 일이 발생하는가? 그것은 어떤 결과를 가져오는가? 이전에는 무슨 일이 발생하였는가? 지금 일어나는 일은 이 실존의 전체적 형성 존재와 어떻게 결합되는가?”(Hans Jonas, 1984/1994, p. 192)

예를 들어 의사의 경우 입원한 환자의 병을 치료함으로써 그 환자에 대한 의사의 책임은 잠정적으로 끝이 난다. 여기서 병을 치료한다는 것은 의사의 진료가 종결된 상황으로 환자가 예전처럼 일상생활을 하는 경우를 말한다. 이때 수술이나 약물투여 후 부작용이 발생하거나 병이 재발한 경우에는 의사는 그 환자에 대해 새로운 사태, 즉 부작용, 재발한 증상에 대해 치료할 책임을 지게 되는 것이다. 이처럼 어떤 문제 상황이나 사태를 중심으로 책임을 논한다면 책임이 일시적이고 특수적일 것이다. 하지만 부모의 경우 아이가 다치거나 힘들어할 때 그 일을 해결해줌으로써 책임이 끝나는 것이 아니다. 한스 요나스에 의하면 부모는 아이가 경험하는 모든 과정을 역사적으로 접근하여 아이의 정체성과 일대기, 자아의 형성과정에 참여하게 된다.(Hans Jonas, 1984/1994, p. 194) 오늘의 아이 생활이 어제와 다르더라도, 내일의 생활을 예상할 수 없더라도 부모에게 있어 아이는 유일무이한 것이다. 이 점을 고려한다면 아이에 대한 부모의 무조건적인 책임은 대상에 대한 연속적인 속성을 갖는다.

셋째, 책임의 미래이다. 이는 책임의 연속성에 대한 연장으로 볼 수 있다. 이것은 부모의 책임, 정치인의 책임뿐만 아니라 어떤 형태의 책임에 대해서도 유효한 속성이다. 한스 요나스는 책임의 미래와 관련하여 역사적 사례와 전통윤리 비판, 기존의 이데올로기의 한계점을 논증하여 불확실한 미래를 대하는 인간의 책임을 서술하였다. 여기서 미래는 앞으로 일어날 일이다. 우리는 미래의

모습을 지금 바로 확인할 수 없다. 단지 우리는 미래의 모습은 어떻게 예측하고 예상하고 계획할 뿐이다. 이는 시간의 흐름과 관련 있는 것으로 내일의 날씨를 어떻게, 다음날의 해 뜨는 시각과 지는 시각, 바닷물의 만조 시각, 친구들과의 약속, 일상적인 출퇴근 등 일상 속에서 일어날 일과 관련이 있다. 그렇다고 이러한 일이 반드시 일어날 것이라고 우리 스스로 보장할 수 없다. 예기치 못한 일이 발생하거나 미처 생각지 못한 부분에서 변수가 생겨 예상을 벗어날 가능성도 충분히 있다. 또한 시간의 흐름과 상관없이 인간의 자유의지에 의해서, 알 수 없는 원인으로 인해 새로운 일이 생겨나거나 우리 삶의 변화가 일어나기도 한다. 그러한 변화의 범위는 시간의 범위를 넘어 우리들의 삶 자체를 변화시키는 범위까지 확대된다. 감염병으로 인해 우리 삶의 모습과 사회 모습에 변화가 오게 될 줄 누가 알았겠는가? 한스 요나스는 이러한 미래의 예측 불가능성을 우리가 인정해야 할 “당위(當爲)(Hans Jonas, 1984/1994, p. 231)”로 인식하고 인간이라면 마땅히 이러한 미래를 책임져야 한다고 주장하였다.

그렇다면 왜 하필 인간이 미래를 책임져야 할까? 한스 요나스는 책임의 대상을 “실제적 생명 또는 잠재적 생명, 그리고 무엇보다도 인간의 생명”(Hans Jonas, 1984/1994, p. 185)이라 하면서 생명의 존엄과 실존 그 자체를 강조하였다. 또한 생명은 “불안정하고, 훼손될 수 있으며, 취소될 수 있는 것”(Hans Jonas, 1984/1994, p. 179)이라고 하면서 생명의 한계를 인정하기도 하였다. 모든 생명체는 종(種)으로서 좋고 나쁘고 우열을 가릴 수 없는 존엄한 존재이다. 그리고 수많은 생명체 중에서 유일하게 인간만이 다른 생명체를 고유한 영역과 목적으로 인식할 수 있다. 다른 말로 모든 존재 중에서 적어도 현재까지는 인간만이 다른 존재에 대해 이성적으로 생각할 수 있고 윤리적으로 판단할 수 있다는 것이다. 그렇기 때문에 다른 생명체에 대해 책임을 질 수 있다. 여기서 한스 요나스는 하나의 생명체로 존재한다는 것 그 자체를 대상에 있어 책임의 첫 번째 조건으로 제시하였다.(Hans Jonas, 1984/1994, p. 180) 또한 인간만이 책임을 질 수 있기 때문에 인간의 실존 역시 보장되어야 한다. 이러한 조건은 곧 인간이 미래에도 실존해야 한다는 하나의 명령으로 이어지게 된다.(Hans Jonas, 1984/1994, p. 181)

3. 현대기술에 대한 한스 요나스의 관점

가. 전통윤리의 한계와 새로운 윤리의 필요성

그렇다면 한스 요나스가 이토록 책임을 강조하게 된 이유는 무엇일까? 이는 한스 요나스의 기술관을 통해 알 수 있다. 한스 요나스 이론의 주요 특징은 현대 사회에서 나타나는 문제점을 분석하고 그 문제점을 해결하기 위한 방법으로 윤리적 가치를 제시한 것이다. 이때 한스 요나스는 현대 사회에서 나타나는 문제점의 원인으로 현대기술의 무분별한 개발이라고 보았다. 이를 뒷받침하기 위해 전통윤리와 현대윤리의 모습을 비교하였다.

한스 요나스는 우선 인간과 자연의 관계에 주목하였다. 특히 인간의 권력이 자연에 미치는 영향력에 주목하였는데, 과거 인간은 자연에 크게 영향을 미치지 않았다. 땅과 바다와 하늘 등 다양한 모습의 자연환경 속에서 인간이 제멋대로 굴더라도 이러한 자연환경들을 포괄하는 자연의 본성은 변화하지 않았으며, 자연의 생산력은 추후도 감소하지 않았다.(Hans Jonas, 1984/1994, p. 16) 그 당시 인간에게 있어 자연은 거대한 존재였고, 인간은 거대한 자연 속에서 살아가고 의지하는 존재였던 것이다. 따라서 그 당시 논의된 윤리도 이러한 자연의 틀 속에서 제한적인 모습을 띠었는데, 한스 요나스는 종래의 모든 윤리학이 다음과 같은 전제 조건들을 공유한다고 하였다.

- (1) 인간의 본성과 사물의 본성을 통해 주어진 인간의 상태는 그 근본 특성에 있어서 단연코 확정되어 있다.
- (2) 인간 선은 이러한 토대를 근거로 별반 어려움 없이 분명하게 규정될 수 있다.
- (3) 인간 행위와 인간 책임의 범위는 좁게 서술될 수 있다. (Hans Jonas, 1984/1994, p. 22)

여기서 한스 요나스는 전통윤리의 범위를 아리스토텔레스의 덕윤리와 칸트, 헤겔, 마르크스 등 고대의 서양철학부터 근대의 서양철학까지 설정하였다. 인간이 자연을 이용하기 위해서는 도구, 즉 기술이 필요하다. 따라서 한스 요나스는 인간과 기술과의 관계와 관련하여 전통윤리의 특징을 정리하였는데, 이를

기술의 중립성, 인간중심적 윤리, 인간 본질의 불변성, 한정적 범위의 행위라는 특징으로 요약할 수 있다.(Hans Jonas, 1984/1994, pp. 29-32)

첫째, 한스 요나스의 관점에서 전통윤리는 기술의 모든 영역에 대해 윤리적으로 중립적이다. 이때 사람을 직접 다루는 의학만 예외로 두었다. 한스 요나스에 의하면, 과거에는 기술의 사용으로 인해 불특정 대상에게 지속적으로 윤리적 침해가 가하지 않았다. 그리고 사람들은 기술을 사용하는 것에 대해서 윤리적 가치판단을 내리지 않았다. 자연으로부터 발생하는 천재지변을 막기 위해 사용된 공물이나 인간의 노동을 보완하기 위한 간단한 도구, 자신의 생명을 보호하기 위한 도구로써 기술은 존재할 뿐, 기술을 진보의 대상이나 가치판단의 대상으로 삼지 않았던 것이다. 기술(技術)의 사전적 의미 그대로 과거에는 “과학 이론을 실제로 적용하여 사물을 인간 생활에 유용하도록 가공하는 수단”²⁹⁾일 뿐이었다.

둘째, 한스 요나스는 전통윤리를 인간중심적 윤리로 바라보았다. 그가 바라보는 전통윤리의 “윤리적 의미는 자기 자신과의 교섭을 포함하여 인간과 인간의 직접적인 교섭에 속하였다.”(Hans Jonas, 1984/1994, p. 29) 이때의 윤리적 의미는 인간 내면의 성찰과 인간과 인간 사이의 관계 속에서 존재하는 것으로, 윤리적 주체와 윤리적 객체가 모두 인간뿐이었다. 그는 그 당시 사람들의 격률이나 도덕적 명령을 예로 들면서 나, 가족, 이웃, 계급 등 인간관계에서 지켜야 할 준칙 또는 윤리적 주체로서 지향해야 할 도덕적 가치를 중심으로 논의가 되었다고 주장하였다. 이에 대해 그는 자연의 모습이나 성질이 본질적으로 변하지 않는다는 믿음이 옛사람들에게 강하게 작용했다고 보았다. 또한 자연의 상태가 천재지변 등으로 인간생활에 영향을 미쳐도 인간은 자연에게 어떠한 영향도 미치지 못한다는 의미로 과거에는 자연의 영역을 인간으로서 침범할 수 없는 영역으로 보았다고 그는 설명하였다. 따라서 그는 인간과 자연과의 관계보다는 일상생활 속에서 발생하는 인간 사이의 문제를 중심으로 전통윤리가 발전하였다고 보았다.

셋째, 기술의 영역에서 이루어지는 행위에 대해 ‘인간’이라는 실체는 본질적으로 ‘불변적’인 것이고 따라서 전통윤리에서 인간은 기술의 대상이 되지 않았

29) 기술. <https://stdict.korean.go.kr>. (검색일: 2020. 3. 7.)

다고 한스 요나스는 주장하였다. 인간이 기술개발을 위한 수단으로 존재하는 것이 아니라 인간이 기술을 사용하는 주체가 되어 다양한 도구들과 과학적 원리들을 적재적소에 활용하기만 하면 되었다. 기술이 인간사회에 미치는 영향이나 자연에 미치는 영향을 생각할 필요가 없을 만큼 기술은 단지 인간의 필요에 의해 쓰이는 수단일 따름이었다. 따라서 그는 과거에는 인간이 기술의 발전을 위해 희생되거나 기술개발 연구의 대상으로 되지 않았다고 보았다.

넷째, 한스 요나스는 과거에 인간의 모든 행위로 인한 영향이 시간상으로 비교적 짧았다고 보았다. 그가 바라보는 전통윤리에서 “결과의 장기적 진행 과정은 우연, 운명 또는 예견에 맡겨졌다.”(Hans Jonas, 1984/1994, p. 30) 그에 의하면, 과거 인간의 행위가 미치는 영향은 시간적으로나 공간적으로나 행위자의 기준에서 멀지 않았다. 그 당시 사람들은 자신과 상관없는 먼 미래나 과거보다는 그 당시 그들의 삶 속에서 일어나고 있는 상황에서 고려되는 윤리적 요소를 중시한 것이다. 한스 요나스는 이웃에 대한 사랑, 가족으로 해야 할 도리, 이익추구, 개인의 행복 추구 등을 예로 들어 설명하였다. 또한 그는 개인 또는 집단의 행위 안에서 발생하는 윤리적 격률은 언제나 지금의 ‘나’와 ‘타인’이 만나는 영역에서 그 의미를 드러낼 뿐이라고 하였다.

이에 대해 김종국(2005)³⁰⁾은 전통윤리학의 이론적 특징을 비판하고 있는 한스 요나스의 주장을 비판적으로 검토하였다. 김종국은 한스 요나스가 말하는 전통윤리를 칸트윤리학으로 보고 한스 요나스의 전통윤리 비판, 즉 칸트윤리학에 대한 한스 요나스의 비판을 도덕의 원칙론과 적용론의 관점으로 나누어 고찰하였다.(김종국, 2005, p. 90) 칸트윤리학의 경우 김종국은 현재에 대해서만 타당한 법칙이 아니고 시간을 초월한 보편적인 인류의 법칙이라 평가하였다. 또한 칸트윤리학뿐만 아니라 다른 전통윤리학이 인간중심적 윤리라는 한스 요나스의 주장에 대해 그가 제시한 책임이론 역시 인간의 책임을 강조하는 것으로 인간중심적 윤리를 벗어나지 못한다고 하였다.(김종국, 2005, p. 93) 그럼에도 불구하고 한스 요나스의 ‘책임의 원칙’의 가치는 미래의 사태, 행위와의 적용에 있다는 측면에서 ‘도덕적 선을 어떻게 실천할 것인가’에 대한 실천철학적 의의가 있다고 평가하였다.(김종국, 2005, p. 99)

30) 김종국. (2005). 미래와 힘, 한스 요나스의 ‘책임의 원칙’ 그 이후. **사회와 철학**, (10), 89-104.

김종국의 평가처럼 한스 요나스의 주장은 전통윤리의 또 다른 해석으로 인한 비판을 피할 수는 없을 것이다. 다만 한스 요나스는 환경파괴, 핵무기개발 또는 핵전쟁 등 오늘날 인간의 생존 자체에 위협이 되는 현상의 주요 원인으로 위와 같은 서양의 전통윤리 특징을 제시한 것이다. 대체로 한스 요나스는 오늘날에 발생하거나 인류에게 위협이 되는 요소들을 먼저 파악하여 위험요소를 최소화하는 방향으로 자신의 주장을 전개하였다. 이를 한스 요나스는 “공포의 발견술”(Hans Jonas, 1984/1994, p. 60)이라고 하였다. 그만큼 한스 요나스는 전통윤리의 우월성보다는 현재 발생하고 있는 문제점 해결에 초점을 맞추었으며 그 문제점을 해결하기 위해서는 전통윤리만으로는 부족하다는 것을 강조하기 위해 위와 같이 해석하였다고 볼 수 있다. 한스 요나스의 관점에서 만약 전통윤리가 현실 속에서 제대로 기능을 했다면 오늘날의 환경파괴나 핵무기개발 등 전 지구적 위협은 발생하지 않았을 것이다. 다시 말해, 오늘날 발생하고 있는 문제의 주요 원인 중에 하나를 전통윤리적 사고로 보는 것이다.

요약해보면 한스 요나스는 자연과 인간의 관계에서 과거와 비교해 오늘날 인간이 자연에 미치는 영향력이 확대되었음을 강조하였다. 이를 한스 요나스는 ‘자연의 가침성’으로 표현하였다.

전래된 세계관의 첫 번째 큰 변화로서 사람들은 인간의 기술적 간섭에 의한 자연의 **가침성**을 예로 든다. 그것은 이미 저질러진 피해와 훼손을 통해 인식되기 전에는 전혀 상상할 수 없었던 가침성이다. 그 충격으로 인하여 막 시작된 환경 연구(생태학)라는 개념과 학문을 초래했던 이 발견은 사물의 체계에 있어서 인과적 요소로서의 우리 자신에 관한 전체 생각을 변화시켰다. 그것은 인간 행위의 본성이 사실상 **변화하였다**는 사실을 그 결과를 통해 보여주었으며, 또 전적으로 새로운 질서의 대상이, 지구의 전체 생명권만큼이나 우리가 그것에 대해 권력을 가지고 있기 때문에 추가로 책임져야 한다는 사실을 보여주었다. (Hans Jonas, 1984/1994, p. 33)

이러한 이유에서 윤리이론은 인간뿐만 아니라 동물, 생태계, 자연, 우주 등 다른 대상을 포괄하여 심사숙고해야 되는 대상이 되었으며, 단지 인간만을 고려한 윤리는 오히려 오늘날 발생하고 있는 문제점을 해결하는 데 걸림돌이 되는 것이다. 그렇다면 이렇게 환경에 미치는 인간의 영향이 변화된 외적인 원인

은 무엇일까? 한스 요나스가 말했듯이 그것은 바로 기술의 급속한 발전이라고 할 수 있다.

나. 인간: 기술의 대상

한스 요나스에 따르면 기술이 발전하면서 점차 인간 자신은 기술의 대상이 되었다고 한다.(Hans Jonas, 1984/1994, p. 51) 인간의 본질을 호모 파베르(Homo Faber)³¹⁾의 관점에서 바라보는 것은 이러한 주장을 뒷받침한다. 기술의 발전은 농작물을 수확하거나 재배하는 차원을 넘어서 문명을 이끌고 인간의 생물학적 한계를 뛰어넘는 데 중요한 역할을 하였다. 이에 따라 인간에 의해 만들어진 제작물은 오로지 생산을 위한 효율적인 수단에서 점차 인간의 가치와 윤리적 사유까지 영향을 미치게 되었다. 특히 과학이론과 기술이 발전하면서 기술의 대상은 건축물, 측량, 교통수단, 산업뿐만 아니라 인간 자체로 확대되었다. 한스 요나스는 크게 생명의 연장, 행동 통제, 유전자 조작을 예로 들어 기술의 대상이 된 인간을 설명하였다.

첫째, 과학기술이 발전함에 따라 생명연장의 가능성이 확대되었다. 생명연장을 위한 기술논의가 이루어지지 않았던 시대에는 인간의 삶과 죽음은 인간 스스로 제어할 수 없는 영역이었다. 질병이나 전쟁 등 인간의 생존에 위협이 되는 것으로부터 보호받기에는 기술만으로는 한계가 있었으며 오래 사는 것은 단지 '운'에 맡겨졌을 뿐이다. 하지만 세포 생물학과 의학기술이 발전함에 따라 오늘날 우리는 죽음을 피할 수는 없어도 죽음 자체를 늦출 수 있다. 이에 따라 기대수명이 높아졌으며 인간 생존에 위협이 되는 것으로부터 보호받을 수 있게 되었다. 이는 곧 신의 영역, 또는 자연의 섭리로만 여겼던 '죽음'이 새로운 시각에서 다루어질 수 있게 되었다. '웰다잉(Well-dying) 산업'의 주목이 그 예라 할 수 있다.³²⁾ 웰다잉 산업은 기대수명은 높아졌으나 인구의 고령화로 인해 가족 구성원의 해체, 독거노인의 증가, 치매, 고혈압 등 노인성 질환의 증가로 호

31) 호모 파베르(Homo Faber)는 도구의 인간을 뜻하는 용어이다. 인간의 본질을 도구를 사용하고 제작할 줄 아는 점에서 파악하는 인간관으로 베르그송에 의해서 창출되었다. 인간은 유형, 무형의 도구를 만드는 동시에 자기 자신도 만든다고 보았다. (<https://ko.wikipedia.org>. 검색일: 2020. 2. 13.)

32) '웰다잉 산업'이 뜬다. <http://www.nexteconomy.co.kr>. (검색일: 2020. 3. 20.)

스피스, 장례, 상속, 사후관리 등 노인들의 '죽음의 질'을 향상시키기 위한 산업을 말한다. 하지만 이러한 변화는 전에는 생각조차 할 수 없었던 새로운 윤리적 문제를 제기하였다. 안락사 문제가 대표적인 예인데, 생사(生死)의 갈림길에 선 환자에 대해 어떤 윤리적 판단을 해야 하는가에 관한 물음이다. 이러한 문제는 환자, 의사, 환자의 가족, 사회, 법률 등 다양한 각도에서 접근할 필요가 있기 때문에, 단지 전통윤리의 어떤 원칙이나 원리로써만 판단할 수 없게 되었다. 이렇듯 기술이 발전함에 따라 죽음을 대하는 인간의 태도는 변화되었다.

둘째, 과학기술이 발전함에 따라 인간의 행위 자체에 변화를 꾀할 수 있게 되었다. 이에 따라 개인의 행위는 약물 따위의 다른 대상으로부터 물리적 또는 화학적인 통제를 받을 수 있게 되었다. 이러한 특징은 치료·치유의 관점에서 보았을 때는 긍정적인 것처럼 보인다. 일상생활이 어려운 정신질환자를 대상으로 투여되는 약물이나 전기적 치료는 뇌를 자극하여 환자가 겪는 고통을 줄이거나 개선하는 데 도움이 되기도 한다. 특히 주의력결핍 과잉행동장애(Attention Deficit Hyperactivity Disorder, ADHD) 환자에 대해서 상담과 병행되는 지속적인 약물치료는 그 필요성과 효과를 입증하여 원활한 대인관계와 일상생활을 하는 데 도움을 준다. 하지만 여전히 우리는 약으로 인한 부작용을 우려하기도 한다. 구체적으로 ADHD 치료제의 주요성분인 메틸페니데이트(Methylphenidate)와 아토목세틴(Atomoxetine)은 식욕 저하, 불면증, 신경과민, 두통, 복통, 혈압·심박동수 증가, 어지럼증, 기분 변화 형태로 부작용을 일으킬 수 있다.³³⁾ 한스 요나스는 물리적 또는 화학적인 통제에 의한 더욱 심각한 문제로 다음을 제시하였다. “다루기 힘든 개인적 행동에 대한 사회의 부담을 줄이는 방향으로 이행 과정이 눈에 띄지 않게 일어날 수 있다. 다시 말해, 의료적 적용으로부터 사회적 적용으로서의 이행 과정이 일어나며, ... 미지의 장(場)을 열어 놓는다.”(Hans Jonas, 1984/1994, p. 55) 예를 들어 약물투입으로 학업 향상을 유도하는 것, 전기적 자극으로 행복감과 쾌락을 유발하는 것, 두뇌를 자극함으로써 공격성을 극복하는 것 등 사회적 부담을 줄이기

33) [약 이야기] 집중력 높이고 성적 올리려 ADHD약 먹는다?. <https://jhealthmedia.joins.com>. (검색일: 2020. 2. 13.)

위해 인간의 존엄과 관련된 기술의 사용을 허용해도 되는가에 관한 물음이 제기될 수 있는 것이다. 이러한 물음 이외에도 약물을 개발하는 데 필요한 임상 실험의 경우에도 새로운 윤리적 문제를 초래했으며 이는 사회적 쟁점으로 발전하기에 이르렀다.

셋째, 인간의 유전자 조작이다. 인간 게놈 프로젝트(Human Genome Project)라 불리는 연구는 1990년부터 시작해서 2003년 분석에 성공하였다.³⁴⁾ 이 연구를 시작으로 우리나라를 포함한 세계 각국은 인간의 유전자 정보 뿐만 아니라 동식물을 포함한 생명체의 유전자 정보를 분석하고 있다. 이는 곧 인간의 기술로 유전자를 취급할 수 있는 길이 열린 것이라고 볼 수 있다. 이는 유전적 질병에 대한 새로운 치료법 개발과 유전성 질환에 대한 예측 또는 예방 까지 가능하게 하였다. 그러나 유전 조작에 대한 우려의 목소리도 발생하고 있다. 한 가지 예로 한스 요나스는 이러한 현상에 대해 “자신의 구상에 따라 인류를 개선하고 변화시키려는 목표를 가지고 이제 인간은 스스로 진화의 작업을 인수하고자 한다는 말로 요약되고 있다.”라고 설명하였다.(Hans Jonas, 1984/1994, p. 56) 불가능하다고 믿었던 생명의 탄생에 대한 인간의 관여가 이론적으로 가능해진 것이다. 또한 유전자 조작 기술을 활용한 유전자 변형 작물의 개량은 유전자 조작 기술 활용에 대한 찬반 논쟁을 일으킨다. 찬반에 대한 견해를 뒷받침하기 위해 유전자 변형 작물이 인체에 미치는 영향부터 시작하여 환경, 생태계에 미치는 영향까지 복합적인 측면에서 논의가 필요하게 된 것이다. 나아가 인간의 유전자 조작은 인간의 존엄성과 인권, 인간 자체에 대해서 어떻게 바라봐야 할지 근원적인 물음을 제기하기에 이른다.

이러한 접근으로, 인공지능기술의 대상도 인간이라고 볼 수 있을 것이다. 인공지능기술은 인간의 지능 모방을 목표로 연구가 이루어진다. 다시 말해, 인공지능기술의 대상은 학습, 추론, 인식 따위의 인간의 사고 활동이라고 할 수 있다. 예를 들어 인공지능기술의 여러 분야 중에서 심층학습(Deep Learning) 연구³⁵⁾는 수많은 데이터를 기반으로 엄청난 수준의 추상화를 시도하는 알고리즘이다. 알고리즘의 특성상 데이터가 많으면 많을수록 그 정확도는 100%에

34) 게놈지도. <https://if-blog.tistory.com/5368>. (검색일: 2020. 2. 13.)

35) 심층학습. <https://ko.wikipedia.org>. (검색일: 2020. 2. 13.)

가까워지는데 이 데이터는 사람들의 성향과 행동, 선호도 따위를 반영한 정보이다. 따라서 사람들의 정보를 기반으로 심층학습을 거쳐 결과를 산출하게 된다. 그리고 심층학습의 주요 방식인 신경망구조는 인간의 뇌를 기반으로 한 정보처리과정을 모방하고자 한 것이다. 비록 현재의 기술로 인간의 뇌 구조를 완벽히 모방하지는 못하고 있으나 그럼에도 불구하고 현재 인간과 유사한 방식으로 정보를 처리하는 심층학습으로 인해 우리의 삶에 적지 않은 변화를 가져오고 있다. 이러한 변화로 인공지능기술은 기술 자체의 개선이나 품질향상을 넘어 인간의 사유방식, 행동원리, 인간의 고유한 특성에 대한 철학적 고찰을 요구하게 했다.

다. 현대기술: 윤리학의 대상

기술의 개발과 사용이 윤리적 관점에서 논의 대상이 된다는 접근은 기술로 인해 인간의 권력이 미치는 대상과 범위가 확대되었다는 특징과 기술과 관련된 모든 인간의 행위는 도덕적 검토에 놓여 있다는 특징에서 비롯된다. 또한 인간이 기술의 대상이 되면서 윤리도 인간의 영역에서 기술의 영역으로 확대되었다고 이야기할 수 있다. 한스 요나스는 현대기술이 과거의 모든 행위에 적용하기에 적절한 윤리학과는 전혀 다른 종류의 사유를 발생시킨다고 하였다.³⁶⁾ 이를 설명하기 위해 현대기술의 특징을 결과의 모호성, 적용의 강제성, 시공간적 광역성, 인간중심주의의 파괴, 형이상학적 물음의 제기로 설명하였다.

첫째, 한스 요나스에 의하면 현대기술로 인한 결과는 예측하지 못하거나 모호해졌다. 어떠한 능력을 사용하는 것에 대해 윤리적 사고는 이러한 행위가 윤리적으로 옳은지 그른지 판단하게 해준다. 윤리학에서 '선'과 '악'이 우리의 행동양식에 대해 도덕적으로 판단할 수 있는 하나의 기준을 마련해준다는 것이 그 예이다. 물론 윤리학에서 이러한 이분법적 접근은 윤리담론에 극히 일부분에 지나지 않지만 이러한 접근법은 우리의 도덕적 사고 판단과 실천을 위한 접근으로 유용하게 쓰이기도 한다. 하나의 예로 뛰어난 언어 구사능력을 가진 사람은 그 자체로 훌륭한 능력이지만 그 능력으로 다른 사람을 속이거나 위협에 처

36) Hans Jonas. (2014). *기술 의학 윤리(Technik, Medizin und Ethik: Praxis des Prinzips Verantwortung)*(이유탉 역). 서울출판사. (원저 1987 출판)

하게 하는 것은 나쁜 것이다. 하지만 우리가 처음에는 좋은 의도로 행하였으나 결국 나쁜 결과를 초래한 경우에는 어떻게 볼 수 있을까? 또는 우리가 의도하지 않고 도덕적인 가치판단이 전혀 들어가지 않았음에도 불구하고 결과적으로 예상한 방향과는 다르게 전개되었을 때는 윤리적으로 어떻게 평가할 수 있을까? 이러한 종류의 문제의식은 현대기술이 발전하면서 제기되는 전형적인 문제 중 하나이다. 모든 현대기술이 그렇다는 것은 아니지만 어떤 기술이 장기간에 걸쳐 우리의 삶에 영향을 미치게 되었을 때 과연 인간이 의도한 대로 발전할지 아니면 예기치 못한 위험, 또는 부작용이 나타날지는 우리가 쉽게 예측할 수 없다. 이러한 현대기술의 특징에서 한스 요나스는 기술의 발전은 근본적으로 효율성과 편리성, 그리고 공익을 추구하기 때문에 기술의 이로움에 내재된 그 이면을 따질 수 있어야 한다고 하였다.(Hans Jonas, 1987/2014, p. 42)

둘째, 한스 요나스에 의하면 현대기술의 적용은 필연적이고 강제적이다. 보통 어떤 능력 또는 권력을 갖게 되는 것은 필연적으로 그것을 사용하게 요구하지는 않는다. 능력과 권력은 기회가 주어지면 소유하고 있는 주체의 의지와 욕구에 따라 행동으로 옮겨진다고 한스 요나스는 주장하였다.(Hans Jonas, 1987/2014, p. 43) 예를 들어 내가 말주변이 뛰어나다고 해서 말하기만 선호하지는 않을 수 있으며, 오히려 말하기에 대해 신중하다 못해 과묵할 수도 있다. 하지만 현대기술은 그렇지 않다. 어떤 기술이 사람들에게 받아들여지면 그 기술은 하나의 생활양식으로 자리 잡혀 언제나 그 기술을 보완해야 하는 필연성을 띠게 된다. 왜냐하면 기술보완의 중단은 제품의 오작동, 기술적 오류 등의 사회적 문제를 일으킬 수 있기 때문이다. 현대기술의 발전은 사회적 변화와 밀접한 관련이 있다. 과학적 연구 또는 의료의 발달, 정보통신의 발달로 인해 우리 삶에 획기적인 변화를 초래할 가능성이 열릴 때 사회는 그 변화를 전적으로 거부하지 않고 비판적인 과정을 통해 받아들이곤 한다. 개인의 입장에서는 최신 기술 또는 제품의 새로운 기능을 자신의 생활에 접목시키고 싶은 욕구가 일어날 수 있다. 다만 현실적으로는 우리가 금전적 문제, 사회적 제도 및 환경의 제한으로 인해 실제로 최신 기술을 접목하는 것은 어려울 수도 있다. 그러나 최신 기술을 사용하고 싶다는 욕구가 발생했다는 것은 부정할 수 없게 되었다. 이 욕구를 충족시키기 위해 우리는 기술을 좀 더 유용하고 편리한 방

향으로 개발하게 된다. 이처럼 현대기술의 적용은 인간의 욕구로 필연적이고 강제적인 성격을 띠게 되었다.

셋째, 한스 요나스에 의하면 현대기술이 미치는 범위는 시공간적으로 매우 광범위하다. 도덕적 중요성을 획득하게 해주는 기술의 또 다른 측면은 그 결과의 크기와 범위에 있다. 기술의 사용이 전 영역으로 확대되면서 기술이 사용되지 않은 일을 찾기가 어려울 정도이다. 그리고 오늘날 이렇게 대규모의 사용을 추구하는 현대기술은 지역을 넘어 사회, 국가, 전 지구적으로 그 영향력이 확대되며 우리의 후손에게까지 영향을 미치게 되었다. 예를 들어 인터넷의 발달 초기에는 군사적 목적으로 쓰였으나 점차 대중화가 이루어지면서 오늘날에는 없어서는 안 될 중요한 통신기술 분야로 자리 잡혔다. 미래에는 이 통신기술이 새로운 분야에 활용되고 스타트업으로서 고부가가치를 형성하는 기술로 전문가들은 전망하고 있다.³⁷⁾ 이 점에서 한스 요나스는 기술의 영향력을 장기적인 관점에서 평가해야 한다고 주장하였다. 우리의 행위가 미래의 후손들에게 정당한 것인지, 밝은 미래 도약을 근거로 실제로 예기치 못한 문제가 초래되었을 때 반성할 기회조차 사라지는 위험요소가 있는 것은 아닌지 주의 깊게 살펴볼 수 있어야 한다.(Hans Jonas, 1987/2014, p. 44)

넷째, 한스 요나스는 현대기술로 인한 새로운 형태의 논의가 기존의 인간중심주의를 파괴하였다고 보았다. 한스 요나스는 기존의 윤리가 인간적 선을 장려했고 타인의 권리 또는 타인의 관심 존중, 우리에게 일어나는 불의(不義)를 개선하고 고통을 완화하는 방향으로 발전했다고 보고 있다. 그리고 의무나 권리는 인간의 문제로 먼저 인식한 뒤, 그다음에 윤리적 대상을 점차 동물, 자연으로 확대되었다고 보았다.(Hans Jonas, 1987/2014, p. 45) 하지만 오늘날에 동물과 자연에 대한 윤리적 고려는 인간 다음으로 이루어지는 것이 아니라 동시에 고려되어야 하는 대상이 되었다. 기술이 발전함에 따라 우리의 권력이 미치는 영향 또한 커지게 되었다고 앞서 말하였다. 한스 요나스의 말을 빌리면 기술의 사용은 의지의 연장(延長)으로서, 즉 어떤 대상이나 현상의 변화를 일으키게 하는 인간의 권력으로서 작용한다.(Hans Jonas, 1987/2014, p. 46) 이러한 이유로 권력이 다른 대상에게 영향을 미치는 것에 대해서 인간

37) 인터넷의 역사. <https://ko.wikipedia.org>. (검색일: 2020. 2. 13.)

스스로 심사숙고할 필요가 있다. 기술의 영향력이 다른 생명체의 삶을 위협할 정도로 커지면서 인간의 책임 또한 기술적 오남용 앞에 그대로 노출되어 있는 지구의 미래로까지 확대된 것이다. 한스 요나스는 이에 대해 인간중심적인 사고를 버리고 범지구적인 사고로 확장하여 우리 주변의 문제를 다양한 안목에서 접근할 것을 요구한다.(Hans Jonas, 1987/2014, p. 46) 왜냐하면 우리의 권력 남용으로 인해 파괴되는 다른 생명이나 자연은 역으로 인간세계의 황폐화를 초래하기 때문이다.

다섯째, 한스 요나스는 현대기술의 발전이 형이상학적 물음을 제기한다고 보았다. 현대기술의 발전으로 인해 오늘날의 인류가 편리하고 풍요로운 삶을 살아가게 되었다는 사실은 부정할 수 없을 것이다. 그러나 한스 요나스는 핵, 환경파괴로 인한 인류 존속 자체의 위협 가능성과 유전공학으로 인한 인류의 유전적 불가침성(不可侵性)의 손상, 그리고 지구상에 있는 생명체들의 생존 조건을 파괴할 수도 있는 기술의 위험성을 경고하였다. 이러한 경고로부터 한스 요나스는 인간에 대한 형이상학적 물음을 제기하였다.(Hans Jonas, 1987/2014, p. 47) 인류가 과연 존재해야 하는지, 왜 존재해야 하는지, 문명의 초래가 필요했는지, 생명 자체의 존재에 관한 물음이 그것이다. 이러한 물음들은 진부하게 보일지라도 기술의 사용에 있어서 반드시 고민해야 할 물음이다. 왜냐하면 이들 물음에 대답은 인간이 기술을 사용하는 데 있어 얼마만큼의 위험요소를 감수해야 하는지, 반대로 결코 감수해서는 안 될 위험요소는 무엇인지 판단하는 데 필요하기 때문이다. 만약 이러한 물음에 대해 진지하게 고민하지 않고 개발자가 오로지 기술개발에 몰두한다면 상상하기도 꺼림직한 일들이 발생하게 될 가능성이 증가하게 될 것이라고 한스 요나스는 우려하였다.(Hans Jonas, 1987/2014, p. 47)

그렇다면 선행연구에서 밝혔던 것처럼 한스 요나스의 책임이론과 최근의 과학기술 발전 사이에 어떤 연관성이 있는 것일까? 특히 필자가 주목하는 인공지능기술을 한스 요나스의 관점을 중심으로 분석해본다면 과연 우리에게 어떤 시사점을 제시할 수 있을까? 이를 위해 먼저 인공지능이란 무엇이고 인공지능기술과 관련된 윤리적 쟁점을 살펴봄으로써 인공지능시대에 요구되는 인간의 책임의식 및 자세를 탐색해 보고자 한다.

Ⅲ. 인공지능기술 및 윤리적 쟁점

1. 로봇과 인공지능의 구분

가. ‘로봇’과 ‘인공지능’이란?

1920년에 체코슬로바키아 극작가 카렐 차페크(Karel Čapek)의 희곡 R.U.R.(Rosuum's Universal Robots)에서 로봇(Robot)이라는 용어가 처음 사용되었다고 전해진다.³⁸⁾ 이 희곡에서 로봇은 원래 사람의 노동을 대신하는 로봇이었으나 점차 개량되면서 사람처럼 학습하고 감정을 갖게 되는 로봇이 되었다.³⁹⁾ 이처럼 로봇이라는 용어가 등장했을 때는 ‘인간이 스스로 인간과 유사한 기계를 만들면 어떤 일이 생길까?’ 하는 문학적 상상력에 의해 쓰였다고 볼 수 있다. 한편 1956년에 존 매카시(John McCarthy) 등으로 구성된 다트머스 학회에서 인공지능(Artificial Intelligence)이라는 용어가 창안되었다.⁴⁰⁾ 김효은(2019)에 의하면 여기서 인공지능은 인간과 비슷한 수준의 기계나 컴퓨터를 의미하기보다는 “형식화 가능한 지능 요소의 개발이라는 ‘방향성’을 지시하는 용어”⁴¹⁾라는 의미에 더 가까웠다. ‘컴퓨터의 추론과 탐색’이라는 주제로 지능이라고 부를 수 있는 인간의 요소를 어떻게 알고리즘 등의 절차로 구현해낼지에 대한 학계의 논의차원에서 인공지능이라는 용어가 처음 쓰였다고 볼 수 있다.

이와 같이 처음 용어가 등장했을 때는 문학적 의미의 로봇과 컴퓨터과학적 의미의 인공지능으로서 그 쓰임에 차이가 있는 것처럼 보였다. 하지만 과학기술이 발전함에 따라 문학적 상상력 속에서만 존재한 로봇은 실제로 인간의 노동력을 대체하고 인간과 유사한 형태로 발달하고 있으며, 인공지능기술도 정밀해지고 전문화되고 있다. 최근에는 그 구분이 무색할 정도로 서로 밀접한 관계 속에서 개발되고 있다고 봐도 이상하지 않을 정도이다. 그러나 글에서 로봇과

38) <https://ko.wikipedia.org>. (검색일: 2020. 3. 23.)

39) <http://www.yes24.com>. (검색일: 2020. 3. 23.)

40) <https://ko.wikipedia.org>. (검색일:2020. 3. 23.)

41) 김효은. (2019). **인공지능과 윤리**. 커뮤니케이션북스(주). p. xiv

인공지능이라는 용어를 사용할 때 그 의미의 쓰임과 범위를 명확히 설정해야 말하고자 하는 내용을 명확하게 전달할 수 있기 때문에, 먼저 용어의 사전적 의미를 살펴본다.

국립국어원에 따르면, 로봇(Robot)이란 “① 『기계』 인간과 비슷한 형태를 가지고 걷기도 하고 말도 하는 기계 장치. 즉인조인간, ② 『기계』 어떤 작업이나 조작을 자동적으로 하는 기계 장치, ③ 남의 지시대로 움직이는 사람을 비유적으로 이르는 말.”⁴²⁾이다. 로봇을 줄인 말로 최근 들어 IT분야에서 자주 쓰이는 봇(Bot)이란 한국정보통신기술협회에 따르면 “① 로봇의 줄임말로 데이터를 찾아주는 소프트웨어 도구. 인터넷 웹 사이트를 방문하고 요청한 정보를 검색, 저장, 관리하는 에이전트의 역할을 한다, ② 보안이 취약한 컴퓨터를 스스로 찾아 침입해 보이지 않는 곳에서 조용히 작동하면서 컴퓨터 사용자도 모르게 시스템에게 명령을 내릴 수 있는 원거리 해킹 툴. 봇(bot)은 채팅 서버와 P2P 네트워크를 통해 컴퓨터를 감염시켜, 해커들이 마음대로 지시를 내릴 수 있으며, 다른 컴퓨터를 공격하도록 명령하거나 감염된 시스템에서 정보를 빼낼 수 있어 보안 문제 가운데 하나가 되고 있다. 좀비피시(PC)”⁴³⁾로 정의되고 있다. 봇은 의미상 긍정적인 의미와 부정적인 의미로 둘 다 쓰이고 있다고 볼 수 있다. 그리고 인공지능(人工知能, Artificial Intelligence)이란 “『정보·통신』 인간의 지능이 가지는 학습, 추리, 적응, 논증 따위의 기능을 갖춘 컴퓨터 시스템. 전문가 시스템, 자연 언어의 이해, 음성 번역, 로봇공학, 인공 시각, 문제 해결, 학습과 지식 획득, 인지 과학 따위에 응용한다. 즉에이아이.”⁴⁴⁾로 국립국어원에서 정리되었다. 사전적 의미로만 보았을 때 로봇은 유형의 기계적 특성을 가지고 인공지능은 무형의 시스템적 특성을 가진다고 볼 수 있다. 그렇지만 봇이 사전적 의미의 로봇과 인공지능의 특성을 모두 갖듯이 로봇과 인공지능을 단지 형태로서만 구분하기에는 한계가 있다.

다음으로, 선행연구를 통해 용어의 쓰임을 분석해보자. 우선 고인석(2014)⁴⁵⁾은 로봇윤리의 기본 원칙을 논하기 위해 로봇의 범주를 설정하였다. 그에

42) 로봇. <https://stdict.korean.go.kr>. (검색일: 2020. 3. 23.)

43) 봇. <http://terms.tta.or.kr>. (검색일: 2020. 3. 23.)

44) 인공지능. <https://stdict.korean.go.kr>. (검색일: 2020. 3. 23.)

의하면 로봇은 어디까지나 인공물(人工物)에 불과하지만 인간의 능동적 정신을 그대로 모방한 특성을 지니고 있다.(고인석, 2014, p. 403) 그래서 칼, 망치 등의 인공물과 달리 로봇 자체로는 행동의 능동성을 지닌 존재로 설정하였다. 이에 따르면 칼, 망치 등은 인간에 의해 다루어지지만 로봇은 시스템적 명령에 따라 최소한 스스로 움직이는 것처럼 보이는 능동적 존재라고 할 수 있다. 또한 고인석은 로봇이라는 범주를 인간이 창조한 범주인 동시에 앞으로 계속 변화될 범주로 보고, 로봇의 정의적(定義的) 특성을 확정짓기는 어렵다고 보았다.(고인석, 2014, p. 404) 그럼에도 불구하고 그는 로봇의 핵심 특성을 다음과 같이 정리하였다.

- ① 인공물임
- ② 외부 세계를 지각하는 능력을 지님
- ③ (지각된) 자료를 처리하는 계산의 능력을 지님
- ④ (전형적으로는 운동의 방식으로 나타나는) 출력 (고인석, 2014, p. 404)

이 특성에 따르면, 로봇은 지각-계산-출력의 과정을 처리하는 인공물로 볼 수 있다. 다만 다른 기계와의 차이점이라고 한다면 “로봇 자체의 체계 안에서 결정되고 조정”(고인석, 2014, p. 404)되는 ‘계산’ 과정에 있다. 그리고 ‘출력’ 과정에서는 로봇의 움직임과 같은 외적 운동과 로봇시스템의 범위 안에서 이루어지는 내적 운동을 모두 포함하며 고인석(2014)은 데이터를 단순히 수집하는 시스템(봇)처럼 운동 출력의 과정이 없는 경우도 로봇으로 보았다.(고인석, 2014, p. 404)

한편, 송선영(2017)⁴⁶⁾은 인공지능과 로봇의 개념에 대해 다음과 같이 설명하였다.

인공지능과 로봇의 개념을 간단히 구분해 보면, 인공지능은 인간 지능의 학습, 추론, 논증의 기능을 갖춘 컴퓨터 시스템으로서 이를 필요한 분야에

45) 고인석. (2014). 로봇윤리의 기본 원칙: 로봇 존재론으로부터. **범한철학**, 75(4), 401-426.

46) 송선영. (2017). 인공지능 로봇과 인간의 관계에 대한 윤리적 성찰과 전망. **영어권문화연구**, 10(3), 61-83.

다양하게 적용된다. 반면, 로봇은 간단히 말해, 인간의 노동을 대신하기 위해 프로그램에 따라 움직이는 기계이다. 이미 컴퓨터 시스템에서는 연산에 충실한 분석을 인공지능이 수행할 수 있고, 생산공정에서는 자동화시스템에 따라 로봇이 노동자의 과업을 대체하였다. 그런데 점차 과학기술의 발달에 따라 각 개념에 충실한 사물이나 기계를 보기는 어렵다. 인간의 움직임, 상호작용 및 교류는 방대한 지식과 데이터를 생산하고, 이를 특정 목적에 따라 자율적으로 활용하는 인공지능은 프로그램화된 로봇의 활동을 통해 구현된다. 이미 네트워크 중심의 지식 기반 사회가 전개가 되면서, 인공지능과 로봇의 경계는 점차 융합되고 있다. (송선영, 2017, p. 67)

이와 같은 흐름에서 송선영(2017)은 개념상 인공지능을 컴퓨터 프로그램으로, 로봇은 그 프로그램에 따라 움직이는 기계로 보았다. 하지만 송선영은 시대가 변하고 과학기술이 발전함에 따라 인공지능과 로봇의 경계는 점차 희미해지는 추세로 보고, 인공지능과 로봇을 구분하지 않고 인공지능 로봇으로 혼용하여 논의를 전개하였다.(송선영, 2017, p. 67)

변순용(2018)도 인공지능과 로봇의 구분이 과연 필요한지 의문을 제기하였다.⁴⁷⁾ 인공지능을 탑재한 자율주행자동차를 예로 들어 인공지능에 의한 결정을 수행하는 차량의 구현체로 보고 이를 'Carbot'이라고 지칭하였다.(변순용, 2018, p. 236) 만약 인공지능기술로 데이터가 처리되어 어떤 결과가 나와도 그 결과를 구현할 대상이 없다면 현실에서는 아무런 의미가 없을 것이다. 이처럼 변순용은 인공지능의 결정을 수행하기 위해서는 유·무형의 구현체가 필요하다고 보았으며, 로봇의 발전과정을 고려해볼 때 로봇만의 윤리나 인공지능만의 윤리보다는 인공지능 로봇에 대한 윤리를 고려할 필요가 있다고 밝혔다.(변순용, 2018, p. 236)

김효은(2019)은 로봇을 인공지능을 장착하지 않은 로봇(자동기계로봇), 인공지능을 장착한 로봇, 그리고 인터넷에서 소프트웨어(인공지능)로만 존재하는 무형의 로봇으로 나누어 보았다.(김효은, 2019, p. viii) 로봇윤리가 처음 등장했을 때는 로봇의 행동 제어 측면에서 주로 논의되었지만 인공지능기술이 발달하면서 인간의 개입 없이 인공지능이 스스로 결정하는 것에 대한 윤리적 논의가 필요하게 되었다. 결국 김효은은 시대적 요구와 과학기술의 발달로 인해

47) 변순용. (2018). 인공지능로봇을 위한 윤리 가이드라인 연구: 인공지능로봇윤리의 4원칙을 중심으로. **윤리교육연구**, 47(47), 233-252.

인공지능윤리의 쟁점은 로봇윤리의 쟁점을 일부 포함하게 되었으며 최근에는 구분 없이 인공지능윤리로 사용되고 있다고 한다.(김효은, 2019, p. viii)

로봇과 인공지능의 사전적 의미와 선행연구를 살펴보았을 때 로봇이란 내·외부의 제어 장치로부터 출력된 결과를 수행하는 유·무형의 인공물이며, 인공지능은 인간의 지능이 가지는 학습, 추리, 적응, 논증 따위의 기능을 데이터 기반으로 구현해내는 기술이라 할 수 있다. 로봇의 지각-계산-출력이라는 관점에서 살펴보았을 때, ‘지각’과 ‘계산’ 과정에는 인공지능기술이 주로 관여하며 ‘출력’ 과정에는 로봇의 수행이 주가 된다고 하겠다. 이러한 관점에서 ‘인공지능윤리’라고 하였을 때 크게 알고리즘, 시스템 과정에 대한 인간의 윤리적 접근이라 할 수 있고, ‘로봇윤리’라 하였을 때 로봇의 행위, 산출물에 대한 인간의 윤리적 접근이라 할 수 있다. 이러한 접근은 마치 로봇과 인공지능을 구분해야만 할 것처럼 보이지만 인공지능기술이 발전함에 따라 로봇개발에 인공지능기술이 적용되고 있으며 인공지능기술이 로봇개발뿐만 아니라 사회, 경제, 문화, 삶에 주목할만한 변화를 이끄는 점도 고려해야 할 것이다. 이로 따라 로봇과 인공지능이 이분법적으로 구분할 대상이라고 보기에 현실적으로 한계가 있지만 본 연구에서는 인공지능을 ‘판단 과정’의 관점에서, 로봇을 ‘행위 결과’의 관점에서 용어를 사용하고자 한다.

나. 선행연구로서 로봇윤리

인공지능기술의 사용에 대한 윤리적 접근은 비교적 최근에 대두되었기 때문에, 인공지능 분야의 연구자들은 선행연구로서 기존의 로봇윤리 연구를 먼저 살펴보곤 한다. 하지만 로봇윤리는 그 범위가 매우 광범위하여 로봇의 정의, 로봇을 연구하는 관점, 로봇이 활용되는 분야에 따라 달라지는 윤리적 쟁점 등 로봇에 대한 총체적인 연구물을 종합적으로 고찰할 필요가 있다. 따라서 로봇윤리와 관련된 연구내용을 먼저 살펴봄으로써 연구자가 로봇윤리를 어떠한 관점에서 연구할 수 있는지 살펴보자.

고인석(2014)은 로봇윤리를 “로봇공학의 실행과 광범위하게 연관된 윤리적 물음들을 다루는 분석과 평가와 토론의 체계”라고 하였다. 이를 설명하기 위해 공학자의 관점과 윤리학자의 관점으로 나누어 각 주체의 관점에서 바라본 로봇

윤리를 논하였다. 요약해보자면 공학자의 관점에서 로봇윤리는 로봇의 제작, 설계, 관리에서 담당자가 염두 해야 할 측면을 다루고 있으며, 윤리학자의 관점에서 로봇윤리는 로봇의 행위와 판단에 대해서 인간은 어떻게 받아들이고 평가할지에 대한 논의를 다룬다. 특히 윤리학자의 관점에서 로봇을 도덕적 주체 또는 지위로서 어떻게 보느냐에 따라 연구방향이 달라지기 때문에 고인석은 이러한 관점에 대한 설정에서부터 신중한 접근이 필요하다고 하였다.

그리고 변순용·송선영(2015)은 로봇에 대해 기존의 윤리이론을 적용하여 접근하였다. 이러한 접근은 로봇과 관련된 윤리적 물음들을 세부화하고 윤리적 탐구의 기초자료로 활용하는 데 도움을 준다. 로봇과 관련된 윤리적 물음들을 정리해보면 <표 III-1>과 같다.

<표 III-1> 로봇과의 공존에서 우리가 마주해야 하는 새로운 윤리적 질문들

영역 이론	로봇 설계자 및 제작자	로봇 관리자	로봇 사용에 따른 피드백	로봇 사용자
의무론	- 로봇에게 도덕적 행위자로서 지위를 부여할 것인가?	- 로봇을 인간과 유사한 도덕적 행위자로서 간주할 것인가?	- 로봇에게 선의지에 따른 프로그램 명령을 어느 범위에까지 입력할 것인가?	- 로봇을 나와 상대방을 위해 사용할 것인가? - 로봇을 제한된 범위에서 도덕적 행위자로 대우할 것인가?
공리주의	- 이윤을 많이 남길 수 있는 로봇을 만들 것인가? - 전체 인류의 행복에 기여할 수 있는 로봇을 만들 것인가?	- 다양한 수요에서 생산된 로봇들을 어떻게 관리할 것인가?	- 로봇기술을 수요에 맞게 발전하는 것이 과연 질적 행복에 기여하는 것인가?	- 반복된 로봇 사용이 결과적으로 그리고 지속적으로 사용자들의 질적 행복에 기여하는 것인가?

영역 이론	로봇 설계자 및 제작자	로봇 관리자	로봇 사용에 따른 피드백	로봇 사용자
덕윤리	- 로봇 설계자 및 제작자는 어떤 덕을 갖추고 있어야 하는가?	- 로봇 설계자, 제작자, 사용자의 덕을 어떻게 지속적으로 개발할 것인가?	- 피드백에 따른 로봇기술을 발전시키는 데 있어서 로봇 자체의 반성적 능력을 적용시켜야 하는가?	- 로봇 사용이 사용자의 덕성 함양에 도움이 된다면, 일생에 걸친 인간과 로봇의 상호작용은 바람직한 것인가?
책임윤리	- 로봇 설계와 제작이 현재와 미래의 전체 인류에 미치는 영향을 고려해야 할 수 있는가?	- 누가 책임을 갖고 있고, 책임의 판단은 누가 내릴 것인가?	- 로봇의 제한적 자율성에 대한 책임의 범위를 고려할 수 있는가?	- 사용자와 로봇은 교류의 순간마다 상호적극적인 윤리적 책임을 질 수 있는가?
정서주의	- 어떤 범위까지 로봇의 정서적 소통 능력을 확대할 것인가?	- 누가 로봇에게 정서들을 입력해야 하고, 관리할 것인가? 그리고 입력된 정서들은 보편타당한 것인가?	- 수요에 따른 정서 입력의 범위를 무조건 받아들여 확대해야 하는가?	- 사용자와 로봇은 제한된 목적으로 정서교류를 하면서 자기 정체성을 확립하는 것이 가능한가?

※ 출처: 변순용·송선영. (2015). **로봇윤리란 무엇인가?**. 어문학사. pp. 50-51

굴리에모 탐블리니(Guglielmo Tamburrini)는 로봇윤리를 “현재 그리고 미래의 로봇 사용과 관련된 윤리적인 문제들을 추출하여 분석하고자 노력하는 응용윤리의 한 분야”⁴⁸⁾라고 하였다. 이를 위해 아래의 여섯 가지 물음을 제시하였는데, 이러한 물음들은 인간의 자율성, 책임, 기술자원에 대한 공정성, 사회문화적 차별, 행위자(Agent)의 윤리적 차원을 확대하여 생각하게 해준다.

48) Rafael Capurro·Michael Nagenborg. (2013). **로봇윤리: 로봇의 윤리적 문제들**(변순용·송선영 역). 어문학사. p. 57

- (1) 서비스 내지 개인용 로봇이 야기한 손해에 대해서 누가 책임을 져야 하는가?
- (2) 인간과 로봇으로 구성된 팀에서 통제를 위한 위계서열을 만들 때 윤리적인 제한이 있는가?
- (3) 인터넷과 연결되는 개인용 로봇에 의해 사생활의 권리가 위협받는가?
- (4) 인간의 언어능력과 문화가 인간보다 저급한 능력을 갖춘 로봇과의 포괄적인 상호작용으로 저하하는가?
- (5) 군사 로봇이 전장에서 살인할 수 있도록 허용해야 하는가?
- (6) 로봇을 마치 인간처럼 도덕적 행위자로 간주해서 기본적인 권리가 있다고 보아야 하는가? (Rafael Capurro·Michael Nagenborg, 2013, pp. 57-58)

이와 같이 로봇윤리에 대한 다양한 관점은 로봇윤리가 현재 발생하고 있거나, 또는 제기될 가능성이 있는 로봇에 대한 윤리적 차원의 문제를 다루는 것과 동시에 로봇윤리가 로봇을 활용하는 인간의 윤리와 밀접한 관련이 있음을 시사한다. 이와 더불어 오늘날 인공지능기술을 적용한 로봇의 행위와 그것으로 인해 발생하는 쟁점에 대해 우리는 어떤 태도를 보여야 하는지 진지하게 고민할 수 있어야 한다. 한스 요나스가 과학기술의 발달로 발생한 현대사회의 문제점에 주목했듯이 인공지능기술의 발달로 과거에는 생각할 수 없었던 위와 같은 새로운 유형의 문제와 쟁점에 대해 신중한 태도를 보일 수 있어야 한다.

그리고 로봇의 개발은 과학과 기술 분야에 이목을 끄는 주요 분야가 되고 있다. 로봇 연구는 단순히 로봇공학뿐만 아니라 다른 학문과의 연결을 요구한다. 왜냐하면 로봇으로 인해 사회적으로 미칠 영향은 특정 분야에 한정되는 것이 아니라 사회를 구성하는 거의 모든 분야에 영향을 미치기 때문이다. 따라서 로봇을 연구하는 실험실에서부터 로봇을 제작하는 제조사, 그 로봇을 사용하는 사용자까지 로봇의 개발부터 사용까지 참여하는 인간의 역할과 마음가짐에 윤리는 바람직한 방식으로 작용할 수 있어야 한다. 특히 인공지능기술은 인간의 노동환경 개선과 업무구조 변화에 이미 쓰이기 시작했으며 이로 인해 발생하는 윤리적 문제 또는 예상 가능한 윤리적 문제에 대한 진지한 고찰이 제기되고 있다. 이렇게 로봇 연구는 로봇개발의 방향과 그에 따른 활용범위가 광범위하여서 컴퓨터과학, 정보윤리, 생명윤리부터 시작하여 사회학, 경제학까지 로봇개발과 관련된 분야와의 면밀한 검토가 필요하게 되었다.

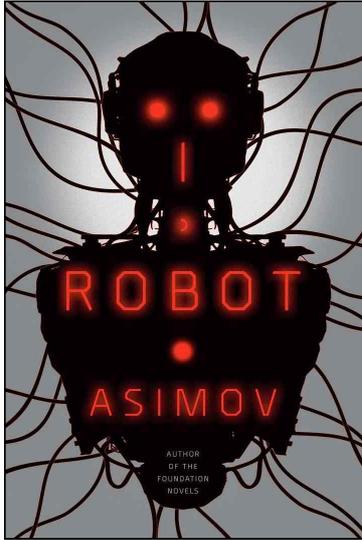
일찍이 베루지오(Gianmarco Veruggio)와 오페르토(Fiorella Operto)는 이러한 사회적 요구와 급변하는 사회현상에 관심을 두고 로봇윤리 연구의 필요성을 제기하였다.⁴⁹⁾ 그들은 로봇을 연구하는 데 로봇의 행위 결과에 초점을 맞출 것인가, 아니면 로봇의 행위 원인에 초점을 맞출 것인가에 따라서 로봇윤리 연구의 주제는 달라진다고 보았다. 로봇의 행위 결과에 초점을 맞추게 되면 로봇의 행위에 대한 옳고 그름, 로봇에 행위를 인간처럼 대우할 것인가, 아니면 단지 기계로 대할 것인가 등에 대한 인식론적 물음, 로봇을 사용하게 되는 인간의 태도 등을 연구주제로 삼을 수 있다. 로봇의 행위 원인에 초점을 맞추게 되면 지능, 자율성, 학습, 의식, 판단력, 감정, 자유의지 등에 대한 새로운 접근 또는 재검토를 요구하게 된다.(Gianmarco Veruggio·Fiorella Operto, 2006, p. 3)

결국, 로봇을 연구하는 관점은 어느 하나라고 단정 짓기 어려우며 또한 어느 것을 먼저 해결해야 할지 우선순위를 정하는 것조차 힘든 상황이다. 그렇기 때문에 연구자의 입장에서 연구자가 기존에 진행하고 있는 연구와 로봇을 접목시켜 학문 간 연계를 추구해야 할 것이다. 생명윤리와 로봇, 정보윤리와 로봇, 사회학과 로봇, 언어학과 로봇, 문화와 로봇 등 종합적인 접근으로 로봇을 연구할 수 있어야 한다. 한 가지 예로, 베루지오와 오페르토는 이러한 관점에서 문화와 로봇의 관계를 살펴보았다.(Gianmarco Veruggio·Fiorella Operto, 2006, p. 3) 로봇이 현실 속 논의대상이 되기 전에 로봇은 소설, 영화 등의 문학적 소재로 많이 쓰였다. 문학작품 속의 로봇은 인간에게 적인지 친구인지에 따라 크게 두 부류로 나눌 수 있다. 아이작 아시모프의 ‘나는 로봇이야(1950)[그림 III-1]⁵⁰⁾’와 데즈카 오사무의 ‘우주소년 아톰(1952)[그림 III-2]⁵¹⁾’에서의 로봇은 두 부류의 관점을 설명하는 데 자주 쓰이는 사례라 할 수 있다.

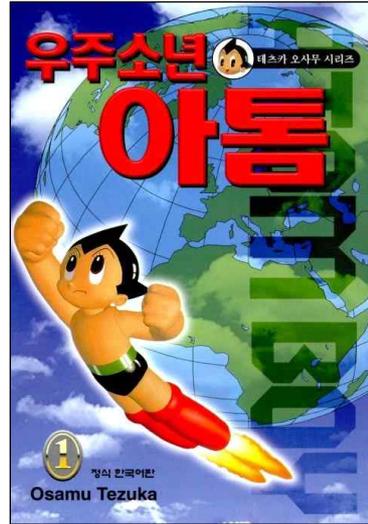
49) Gianmarco Veruggio, Fiorella Operto. (2006). Robotics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics. *International Review of Information Ethics*, 6, 2-8.

50) <https://books.google.co.kr>. (검색일: 2020. 2. 14.)

51) <https://ridibooks.com>. (검색일: 2020. 2. 14.)



[그림 III-1] 나는
로봇이야(1950)



[그림 III-2] 우주소년
아톰(1952)

이러한 비교를 통해 베루지오와 오페르토는 로봇을 바라보는 문화적 인식의 차이가 있을 것이라는 가능성을 제기하였다. 이에 그들은 이러한 문화적 인식의 차이를 참고하여 로봇공학자, 연구자, 대중이 바라보는 로봇의 관점을 네 가지로 정리하였다.(Gianmarco Veruggio·Fiorella Operto, 2006, p. 4)

첫째, 단지 기계로서의 로봇이다. 로봇이 아무리 정교하고 인간의 모습과 비슷하게 발전하더라도 로봇은 단지 기계일 뿐이다. 크고 작은 부품들을 조립하고 프로그램을 입력하여 정해진 운동을 수행하는 동적 장치로서 로봇은 인간의 노동력을 대신하고 산업현장에서 위험하거나 인간이 할 수 없는 일을 수행한다. 이러한 관점에서 주요 연구주제는 로봇의 의식, 의지, 자율성 논쟁보다는 인간의 삶의 질을 향상시키기 위한 올바른 로봇의 사용에 초점을 맞추게 된다. 한스 요나스의 관점에서 로봇을 단지 기계로 본다면, 로봇의 행위는 인간의 권력, 즉 의지가 연장되는 것으로 볼 수 있다. 따라서 로봇의 행위는 인간의 조작에 의한 행위로, 로봇의 행위에 대한 책임은 전적으로 인간에게 있다.

둘째, 윤리적 차원을 갖는 로봇이다. 인간이 동물과 다른 윤리적 차원을 갖듯이 로봇 또한 고유한 윤리적 차원을 갖는다는 것이다. 여기서 윤리적 차원이란 인간의 고유한 자유의지로 인해 나타나는 윤리적 행위, 생각, 표현 따위가

다. 현재까지 로봇은 인간을 기반으로 연구되고 인간이 설계하고 제작하기 때문에, 로봇은 인간의 언어사용방식, 감정, 운동원리, 사고 및 추론과정과 점차 닮아간다고 볼 수 있다. 따라서 로봇의 윤리적 차원 또한 인간과 닮아가거나 로봇의 고유한 영역이 있을 것이라는 생각도 해볼 수 있게 된다. 예를 들어 로봇의 자율성, 의식, 존재론적 의미 등을 연구하는 것이 있다. 한스 요나스의 관점에서 로봇을 윤리적 차원의 존재로 보았을 때 인간의 사고, 감정, 행위방식을 로봇에게 빗대어 접근할 수 있다. 그렇다면 우리는 로봇도 하나의 존재물로 인식해야 한다. 이에 따라 우리에게서 로봇이 안전하게 작동하고 업무를 효율적으로 처리할 수 있도록 로봇에 대한 유지와 관리의 책임이 발생하게 된다.

셋째, 도덕적 행위자로서 로봇이다. 이 관점에서는 로봇의 행위에 대해 알고리즘에 의한 절차적 성공(Success) 또는 오류 등의 실패(Failure)로 구분하는 것이 아니라, 선(Good)과 악(Bad)으로 구분하는 것이다. 로봇이 실행한 행위에 대해 도덕적 관점에서 따져보고 판단해봄으로써 선과 악에 대한 진지한 논의를 이끌어낸다. 예를 들어, “다섯 사람을 구하기 위해 한 사람을 죽이는 것이 도덕적으로 허용 가능한지에 관한 질문”⁵²⁾의 트롤리 딜레마 같은 도덕적 딜레마 상황에서 로봇은 어떻게 행동해야 하는지, 그리고 군사용 로봇의 도입에 대한 논의, 위급상황에서 자율주행자동차의 판단에 대한 논의는 ‘로봇’ 그 자체가 아닌 ‘인간’의 윤리적 판단을 중심으로 이루어지게 된다. 결국 이러한 관점은 로봇에 대한 도덕적 판단이 인간의 도덕적 판단으로 논의를 이끌어낸다. 이러한 관점은 윤리적 주체가 아닌 단지 윤리적 객체로서만 존재하는 윤리적 차원의 로봇에서 한 단계 더 나아가 프로그램화된 판단기준, 선택기준에 따르는 로봇이라고 할 수 있다. 이러한 관점은 인간의 윤리적 판단 과정과 유사하게 로봇이 스스로 처리하는 것처럼 보이게 한다. 그러나 로봇의 판단기준은 인간이 만들었고, 인간의 판단기준은 인간의 내면에서 자발적으로 나타난 자유의지에 의한 것이라는 차이가 있다. 한스 요나스의 관점에서 로봇을 도덕적 행위자로 삼았을 때, 책임의 주체는 여전히 윤리적 존재로서의 인간이다. 인간에 의해 설계된 로봇을 윤리적 존재라고 보기엔 어려움이 있다.

넷째, 새로운 종(種)으로서의 로봇이다. 이러한 관점에서 로봇은 자유의지와

52) 트롤리 딜레마. <https://terms.naver.com>. (검색일: 2020. 3. 1.)

의식을 갖는 것뿐만 아니라 인간의 노력으로 만들어낸 이전과 전혀 다른 종의 탄생을 이야기한다. 예를 들어 공상과학소설이나 영화, 만화에 등장하는 것처럼, 인간과 대등하거나 그 이상의 존재로 로봇을 인식하는 것이다. 이때에는 로봇이 인간의 지적 능력을 뛰어넘기 때문에, 우리가 생각해내지 못한 새로운 형태의 윤리적 논의를 이끌어낸다. 한스 요나스의 관점에서 로봇을 새로운 종의 탄생으로 보았을 때 책임의 주체는 인간뿐만 아니라 로봇으로도 볼 수 있는 가능성이 발생하게 된다.

그러나 네 가지 관점 중에서 어느 관점으로 연구를 해야 하는지는 결정하기 힘들 것이다. 현재 로봇윤리 연구는 로봇공학, 컴퓨터과학, 인공지능, 철학, 윤리, 생물학, 물리학, 인지과학, 신경과학, 법철학, 사회학 등 다양한 분야와 관련되어 연구가 진행되고 있기 때문에 어느 관점이 맞고 틀리고를 따질 수 없는 분야이다. 그러므로 베루지오와 오페르토는 로봇윤리를 연구할 때 종합적인 고려가 필요하며 다양한 사례를 분석하여 로봇윤리 자체를 민주적 토론의 장으로 이끌어야 한다고 주장하였다.(Gianmarco Veruggio·Fiorella Operto, 2006, p. 7) 또한 로봇윤리 연구를 통해 인간은 미래에 대해 비관적으로 전망할 것이 아니라 인간의 삶의 질 향상과 행복 추구를 위한 올바른 로봇 인식을 함양할 것을 강조하였다.(Gianmarco Veruggio·Fiorella Operto, 2006, p. 7)

이러한 취지에서 EURON(European Robotics Research Network)은 2006년에 로봇윤리 로드맵⁵³⁾을 발표하였다. 시간적 범위로는 현재와 예측 가능한 미래로 설정하였고, 대상은 오직 인간, 그리고 방법으로 세부 분야에 따라 요약된 형태의 로드맵을 제시하였다. 이러한 로드맵의 취지는 군사, 의료, 사회적 복지 차원에서 쓰이는 로봇 등 이미 상용화되거나 개발될 로봇에 대해 윤리적으로 접근하는 것이다. 이를 통해 로봇윤리 연구는 로봇개발에 걸림돌이 되어서는 안 되며, 동시에 사회적 논쟁을 초래한다거나 인류에게 위협이 되는 것을 초기에 예방해야 한다고 하였다. 이를 위해 로봇 설계자, 제작자, 사용자 등 로봇을 대하는 주체들뿐만 아니라 생물학, 철학, 사회학 등의 학제적 차원

53) Gianmarco Veruggio. (2006). EURON Roboethics Roadmap. *Restricted Circulation*, 1(1), 1-42.

에서 공동의 협의체 마련을 제안하였다.(Gianmarco Veruggio, 2006, p. 38)

로봇윤리와 비슷하면서도 다른 것이 인공지능윤리이다. 로봇윤리의 하위 영역으로서 인공지능윤리는 최근 들어 국가와 기업에서 주목을 받는 분야이다. 특히 한국에서는 알파고와 이세돌의 대국을 기점으로 인공지능에 대한 대중의 관심이 높아졌다.⁵⁴⁾ 국가와 기업의 인공지능윤리 현장과 미래의 핵심기술로 주목받고 있는 인공지능기술은 어떤 윤리적 쟁점과 관련이 있는지 고찰할 필요가 있다.

2. 인공지능기술과 관련된 윤리적 쟁점

가. 좁은 인공지능(ANI)과 범용 인공지능(AGI)

인공지능에 대한 윤리적 논의는 인공지능을 어떻게 보느냐에 따라 논의의 주제와 대상이 달라진다. 대체로 인공지능을 연구하는 학자들은 인공지능을 세분화하여 약한 인공지능(Weak Artificial Intelligence, Weak AI)과 강한 인공지능(Strong Artificial Intelligence, Strong AI)으로 구분하여 인공지능을 연구하거나 또는 좁은 인공지능(Artificial Narrow Intelligence, ANI), 범용 인공지능(Artificial General Intelligence, AGI), 초인공지능(Artificial Super Intelligence, ASI)으로 나누어서 인공지능에 대한 철학적 논의를 시도한다. 이러한 인공지능의 구분은 어느 한 사람이 정의했다고 보기는 힘들고, 인공지능학자, 컴퓨터공학자, 사설기관, 개인에 의해 정의를 다양하게 하고 있다. 인공지능의 구분과 관련하여 현재 사용되고 있는 의미를 정리해보면 <표 III-2>와 같다.

54) 인간과 인공지능의 대결, 문화와 과학의 대결 등으로 수많은 화제 불러일으키며 역사적인 세기의 대결로 주목받은 첫 대국은 한국어와 영어로 공식해설이 제공된 유튜브(YouTube) 생중계를 통해 전 세계 8천만 명 이상이 시청하였으며, 대한민국에서는 한국방송공사(KBS2)에서 생중계되었다.(<https://ko.wikipedia.org>. 검색일: 2020. 3. 2.)

<표 III-2> 현재 사용되고 있는 인공지능 구분

구분	의미
약한 인공지능 (Weak Artificial Intelligence, Weak AI)	<ul style="list-style-type: none"> - “특정 문제를 해결하는 지능적 행동, 사람의 지능적 행동을 흉내 낼 수 있는 수준, 대부분의 인공지능 접근 방향.”⁵⁵⁾ - “특정 영역의 문제를 푸는 인공지능기술로써 문제 해결, 이상적 업무의 연구의 처리에 최적화된 인공지능, 알고리즘과 기초 데이터, 규칙을 입력해야 함.”⁵⁶⁾ - ““특정한 문제를 인간처럼 풀기’를 목표로 하는 인공지능 연구 분야.”⁵⁷⁾
강한 인공지능 (Strong Artificial Intelligence, Strong AI)	<ul style="list-style-type: none"> - “사람과 같은 지능, 마음을 가지고 사람처럼 느끼면서 지능적으로 행동하는 기계.”⁵⁸⁾ - “인간의 지능을 가지고 생각을 할 수 있는 컴퓨터, 명령을 받지 않아도 스스로 일을 할 수 있으며, 인공지능이 스스로 생각했을 때 불합리하다면 명령을 거부할 수도 있음.”⁵⁹⁾ - “학습, 추리, 적응, 논증과 같은 인간의 지각적 기능과 독립적 자아를 스스로 갖춘 컴퓨터 시스템.”⁶⁰⁾
좁은 인공지능 (Artificial Narrow Intelligence, ANI)	<ul style="list-style-type: none"> - “특정 영역의 문제를 푸는 인공지능”(김효은, 2019, p. x) - “소셜 미디어의 이미지 분류 서비스나 얼굴 인식 기능 등과 같이 특정 작업을 인간 이상의 능력으로 해내는”⁶¹⁾ 인공지능 - “오늘날 우리가 흔히 만나볼 수 있는 기술들을 아우른다. 즉 특정 임무에 한하여 기능을 발휘하도록 한 인공지능기술.”⁶²⁾ - “음성 인식, 바둑 등 특정한 문제를 해결하는 인공지능”⁶³⁾
범용 인공지능 (Artificial General Intelligence, AGI)	<ul style="list-style-type: none"> - “인간 수준의 인공지능”(김효은, 2019, p. x) - “인간의 감각, 사고력을 지닌 채 인간처럼 생각하는 인공지능”⁶⁴⁾ - “보다 광범위한 일을, 보다 사람처럼, 할 수 있게 해주는 기술”⁶⁵⁾ - “특정 문제뿐만 아니라 주어진 모든 상황에서 생각, 학습하고 창작하는 인공지능”⁶⁶⁾
초인공지능 (Artificial Super Intelligence, ASI)	<ul style="list-style-type: none"> - “모든 분야에서 사람보다 뛰어난 인공지능”(김효은, 2019, p. x) - “인간보다 1,000배 이상 뛰어난 지능을 가진 AI. AI는 효율, 자기보존, 자원획득, 창의성 등의 원초적 욕구를 기반으로 끊임없이 자가발전한다는 특징을 갖고 있음.”⁶⁷⁾

55) <https://blog.naver.com>. (검색일: 2020. 3. 7.)

56) <https://www.samsungsds.com>. (검색일: 2020. 3. 7.)

57) <http://scienceon.hani.co.kr>. (검색일: 2020. 3. 7.)

위의 표에서 알 수 있듯이, 약한 인공지능(Weak AI)과 좁은 인공지능(ANI)은 비슷한 의미로서 특정분야의 문제를 효과적으로 해결하는 인공지능기술을 말한다. 이러한 의미에서 인공지능기술은 하나의 인공지능 시스템이 분야 또는 영역을 자유롭게 이동할 수 없는 것이 주요 특징이다. 예를 들어 하나의 인공지능 프로그램이 수학 분야에서 언어 분야로, 체스 분야에서 바둑 분야로 자유롭게 이동할 수 없다. 한 가지 분야에서 인간보다 월등한 수준의 문제해결 과정을 보일지라도 분야 간 전이는 일어나지 않아 수단으로서의 기술적 특징을 확인할 수 있다. 이러한 인공지능기술에서 조금 더 발전한 의미로서 강한 인공지능(Strong AI)과 범용 인공지능(AGI)이 대등하게 사용된다. 이는 간단히 말해 인간 수준의 인공지능이라고 볼 수 있다. 여기서 말하는 인간 수준을 어떻게 정의할지 학자마다 의견이 다르지만, 대체로 인공지능기술이 인간의 지능 수준과 비슷해지거나 분야 간 전이가 자유롭게 일어날 경우, 이 시점을 범용 인공지능의 출현이라고 보고 있다. 마지막으로 초인공지능(ASI)은 인간의 수준을 뛰어넘는 인공지능으로, 공상과학소설이나 영화에서 등장하는 인공지능 로봇의 경우를 예로 들 수 있다. 초인공지능이 현실적으로 가능한지에 대한 여부부터 관점에 따라 다양한 의견이 나오지만, 현재 인공지능기술을 개발하고 사용하는 인간의 처지에서 미래에 대한 낙관론과 비관론을 이야기할 때 초인공지능의 모습이 거론되고 있다.

김효은(2019)은 좁은 인공지능(ANI)과 범용 인공지능(AGI)을 구분할 필요가 있다고 파악하고 있다. 예를 들어 로봇의 '인격'에 대한 논의에는 좁은 인공

58) <https://blog.naver.com>. (검색일: 2020. 3. 7.)

59) <https://www.samsungsds.com>. (검색일: 2020. 3. 7.)

60) <https://opendict.korean.go.kr>. (검색일: 2020. 3. 7.)

61) <https://sdc-james.gitbook.io>. (검색일: 2020. 3. 7.)

62) <https://www.boannews.com>. (검색일: 2020. 3. 7.)

63) <https://m.etnews.com>. (검색일: 2020. 3. 7.)

64) <https://sdc-james.gitbook.io>. (검색일: 2020. 3. 7.)

65) <https://www.boannews.com>. (검색일: 2020. 3. 7.)

66) <https://m.etnews.com>. (검색일: 2020. 3. 7.)

67) <http://www.dt.co.kr>. (검색일: 2020. 3. 7.)

지능에 대한 법적 논의와 범용 인공지능에 대한 철학적 논의가 섞여 있다고 하였다.(김효은, 2019. p. xi) 법적 논의에서의 ‘인격’은 주로 회사법인 형태로서 전자 인격의 가능성을 살펴보고 있으며, 철학적 논의에서 ‘인격’은 사람으로서의 인격을 지칭한다. 이를 참고하였을 때 인공지능 연구는 그 관점을 세분화함으로써 혼란을 최소화하고 윤리적 쟁점의 관련 영역과 범위를 명확히 할 필요가 있다. 따라서 필자는 인공지능기술을 언급할 때 사전적 의미의 ‘범위(範圍)’에 초점을 맞춰 ‘좁은’ 인공지능(ANI)과 ‘범용’ 인공지능(AGI)으로 구분하여 사용하고자 한다. 이 범위에는 ‘시간적’ 범위와 ‘공간적’ 범위가 모두 포함된다. 이를 정리해보면 <표Ⅲ-3>과 같다.

<표 Ⅲ-3> ‘시간적’ 범위와 ‘공간적’ 범위로서 인공지능

범위 구분	‘시간적’ 범위	‘공간적’ 범위
좁은 인공지능 (Artificial Narrow Intelligence, ANI)	- 현재 - 현재 사용되고 있는 인공지능 기술과 관련된 기술적 쟁점	- 특정 분야 - 기술, 국방, 의료, 교육, 오락 등 특정 분야별로 발생하는 응용윤리적 쟁점
범용 인공지능 (Artificial General Intelligence, AGI)	- 가까운 미래 - 인간과 유사한 인공지능기술 의 개발과 출현에 관한 철학 적 쟁점	- 복합 분야 - 인간과 유사한 인공지능의 출현으로 발생할 수 있는 범 지구적 차원의 모든 변화와 철학적 쟁점

나. 좁은 인공지능(ANI)에 관한 윤리적 쟁점

‘시간적’ 범위에 대한 논의로, 김효은(2019)은 현재 시점에서 발생하고 있는 좁은 인공지능(ANI)과 관련된 윤리적 쟁점으로 ‘알고리즘 편향’과 ‘투명성’, 그리고 ‘책임’으로 정리하였다. 알고리즘 편향이란 “기계학습에서 쓰이는 데이터를 선택·수집·분류·사용할 때, 그리고 알고리즘을 구성할 때 불공평한 기준이 개입되는 것을 말한다.”(김효은, 2019, p. 28) 여기서 알고리즘(Algorithm)이란 수학과 컴퓨터과학, 언어학 또는 관련 분야에서 어떠한 문제를 해결하기 위해

정해진 일련의 절차나 방법을 공식화한 형태로 표현한 것으로, 계산을 실행하기 위한 단계적 절차를 의미한다.⁶⁸⁾ 컴퓨터과학 분야에서 설계자는 프로그램에 알고리즘을 입력하고 프로그램을 실행한다. 인공지능 알고리즘은 그 특성상 인간이 몇 가지의 코드만 입력하여 프로그램 자체에서 특수한 패턴을 찾아낸다. 하지만 이 알고리즘으로 출력할 수 있는 경우의 수는 직접 헤아릴 수 없을 정도로 많기 때문에 이러한 방식의 학습과 훈련을 거쳐 나온 결과에 대해서 설계자는 왜 그러한 결과가 나오게 되었는지 이해하기 어렵게 된다. 이러한 인공지능 알고리즘의 모습에 대해 김효은은 안의 상황을 들여다볼 수 없다는 의미인 “블랙박스(Black Box)”로 표현하였다.(김효은, 2019, p. 29) 이때 발생할 수 있는 문제점으로 첫째, 인공지능의 최종 의사 결정이 잘못되는 경우로 데이터가 편향되게 수집되는 경우와 둘째, 불공정한 기준으로 알고리즘이 만들어지거나 학습되는 경우를 제시하였다.(김효은, 2019, p. 30)

첫 번째의 경우, 2015년 구글 포토 서비스를 이용하던 흑인 프로그래머가 흑인 여성 친구와 함께 찍은 사진을 올렸는데 흑인 여성 친구 얼굴 아래에 ‘고릴라’라는 태그가 달렸다고 한다. 입력된 데이터가 고릴라라고 판단되도록 구글 포토 서비스에 의해 분석에 필요한 데이터가 비정상적으로 수집된 것이다. 이에 구글은 ‘프로그램 오류’라고 사과했으며 개선을 약속하였다.(김효은, 2019, pp. 30-31) 이와 같은 구글 포토 서비스의 인식 오류처럼 인공지능 알고리즘에 쓰이는 데이터의 양 또는 질적인 차이로 인해 편향된 결과도출이 발생할 수 있다. 두 번째의 경우, 콤파스(COMPAS) 사례를 예로 들 수 있다. 2016년 미국의 독립언론 ‘프로퍼블리카(ProPublica)’는 탐사보도를 통해, 미국의 재범 위험 평가 소프트웨어인 콤파스가 흑인을 차별한다는 사실을 밝혀냈다. 콤파스는 피고의 범죄 참여, 생활 방식, 성격과 태도, 가족과 사회적 배제 등을 점수로 환산해 재범 가능성을 계산하여 판사에게 구속 여부를 추천하는 알고리즘이다. 콤파스는 인종을 변수로 포함하고 있지 않으나 흑인의 재범 가능성을 백인보다 2배 위험하다고 판단했다.⁶⁹⁾ 평가항목으로만 따졌을 때, 흑인이 백인보다 높은 평가를 받기에는 현실적으로 어려웠던 것이다. 이처럼 알고리즘의 설계부

68) 알고리즘. <https://ko.wikipedia.org>. (검색일: 2020. 3. 8.)

69) 편견·차별 부르는 AI 알고리즘…해법은 있을까. <http://www.hani.co.kr>. (검색일: 2020. 3. 8.)

터 특정 집단이나 계층이 유리하게 적용되거나 불리하게 설계된 경우가 발생할 수 있다. 알고리즘 편향은 인공지능이 스스로 편향된 결과를 도출한 것이 아니라 데이터 수집 과정에서 인종, 문화, 경제, 사회, 종교 등 전반의 분야에 대해 데이터가 고르게 수집되지 못함에 따라 발생한 것이다. 이에 대해 김효은(2019)은 이러한 데이터는 인간의 활동, 기록, 영상, 검색 등에서 비롯된 자료이기 때문에, 인간의 무의식적 성향이 인공지능 알고리즘의 처리 과정에 반영될 가능성이 있다고 하였다.(김효은, 2019, p. 32) 이러한 알고리즘 편향에 대한 기술적 쟁점은 편견, 차별, 불평등이라는 윤리적 쟁점으로 발전하게 된다. 이러한 알고리즘 편향을 최소화하려는 방안이 현재 논의 중에 있다.

닉 보스트롬(Nick Bostrom)과 엘리저 유고스키(Eliezer Yudkowsky)도 인공지능기술의 발전에 따라 점차 복잡하고 집적된 인공지능 알고리즘을 개발할 때 인간의 복잡한 신경망을 모방한다면, 알고리즘이 어떠한 과정으로 데이터를 판단하는지 이해하는 것은 거의 불가능할 것으로 전망하였다.⁷⁰⁾ 이렇게 복잡한 과정의 알고리즘 추론과정에 대한 설명의 요구가 '투명성'의 요구다.(김효은, 2019, p. 42) 알고리즘의 처리 결과에 대해 잘못되었다고 인간이 판단하였을 때, 어디서부터 어떻게 잘못되었는지 그 원인을 따져야 한다. 하지만 인공지능 알고리즘이 정밀하고 복잡해질수록 그 지점이 어디인지를 정확히 파악하는 것이 어려워지게 된다. 인공지능 개발자가 의도한 대로, 목적에 알맞게 결과가 나온다는 보장이 없게 되면, 이는 곧 그 기술을 탑재한 제품을 이용한 사용자 안전의 보장과 신뢰 확보에도 심각한 문제를 일으킬 수 있다. 그렇다고 해서 인공지능 알고리즘의 원리를 모두 공개하는 것은 정보를 민감하게 관리해야 하는 공공기관의 보안 문제로 이어질 수 있다. 예를 들어, 투명성을 확보하고자 하는 긍정적인 의도로 출발했으나 해킹으로 알고리즘 기록에 침투하는 등의 투명성을 악용하는 일이 발생할 수 있다.⁷¹⁾ 그뿐만 아니라 알고리즘 체계는 일반인이 이해하기에 어려운 용어와 절차로 이루어져 있는데, 이 복잡하고 어려운 컴퓨터 공학 용어를 어떻게 대중에게 알기 쉽게 설명할 것인지에 대해 문

70) Nick Bostrom·Eliezer Yudkowsky. (2011). The Ethics of Artificial Intelligence. Cambridge University Press, 1, 1-20.

71) 투명성이 만병통치약?... 블랙박스 AI의 위험을 완화하는 방법. <http://www.ciokorea.com>. (검색일: 2020. 3. 8.)

제를 제기할 수 있다는 것이다. 따라서 ‘투명성’ 확보의 범위를 어디까지 설정해야 하는지에 대한 구체적인 검토가 필요하다.

‘투명성’ 확보의 문제는 곧 법적 책임의 문제로 확대된다. 인공지능기술로 인해 발생한 문제에 대해서 누가 책임을 져야 하는지는 현재까지 논의되고 있는 물음이다. 미국의 공유차량 업체 우버(Uber)의 자율주행자동차가 시험 운행 중 자전거 이용자를 치어 숨지게 한 사고, 테슬라의 ‘모델 3’ 차량이 추돌사고를 내어 운전자가 사망한 사고 등⁷²⁾의 사례로 보았을 때 인공지능기술이 인간의 생명과 안전을 전적으로 보장한다고 보기에 현재로서는 힘들다. 그러므로 인간의 생명과 직접적으로 관련된 군사, 자율주행자동차, 의료 분야는 물론이고 인간의 삶 속에서 활용되고 있는 인공지능기술에 대한 법적 책임의 주체가 누구인지 신중하게 따질 필요가 있다. 이때의 책임은 법적 책임으로서 처벌과 보상 측면을 이야기한다. 닉 보스트롬과 엘리저 유고스키에 따르면 시민들에게 예측 가능한 사회적 환경을 제공하고 그들의 권리와 의무를 보장하는 법의 기능상, 오늘날 인공지능시대는 인공지능기술과 관련된 법률의 구축이 필요한 시점이라고 하였다.(Nick Bostrom·Eliezer Yudkowsky, 2011, p. 2) 그러나 사회가 전문화되고 인공지능기술 개발과 관련된 주체들이 다양해짐에 따라, 이러한 시도는 인공지능 과학자, 윤리학자, 법학자, 사회학자 등 필요 주체들이 모여서 논의되어야 할 거대한 영역이 되었다. 이때 김효은(2019)은 인공지능 알고리즘의 판단 측면에서 인공지능에게 맡길 수 있는 판단과 맡길 수 없는 판단을 구분하는 것이 인공지능에 대한 법적 책임을 논의하는 시작점이 될 것으로 전망하였다.(김효은, 2019, pp. 56-57)

‘공간적’ 범위로서 좁은 인공지능(ANI)의 윤리적 쟁점은 인공지능기술이 적용된 다양한 분야의 로봇과 관련된다. 여기서는 제조·산업용 로봇, 메디컬·헬스케어 로봇, 서비스·엔터테인먼트 로봇, 가정교육용 로봇, 탐사 로봇, 군사 로봇, 유희 로봇, 사회적 네트워크형 로봇 등 분야별로 투입되는 인공지능 로봇을 개발하거나 사용하면서 발생할 수 있는 윤리적 쟁점들을 다룬다. 분야별 로봇은 목적에 따른 용도(用途)가 설정되어 있기 때문에 분야 간 교차수행은 원

72) 자율주행차 과연 안전할까 - 인간의 돌발 행동에도 충분히 대처해야. <https://news.joins.com>. (검색일: 2020. 3. 8.)

칙적으로 제한한다. 하지만 로봇의 종류에 따른 윤리적 쟁점은 응용윤리의 특성상 문제의 유형이 인간의 삶과 밀접하게 관련이 있으며, 윤리학적 관점에서 문제를 해결하고자 한다. 따라서 분야별로 쟁점이 중복될 수 있으며, 문제 해결의 근거로 쓰이는 윤리학적 이론이 겹치는 경우가 발생할 수 있다. 이 점을 참고하여 변순용(2015)과 송선영(2015)은 로봇 분야별 미래 시나리오에 대한 윤리적 쟁점을 제시하고 그 대응 방안을 <표 III-4>로 정리하였다.

<표 III-4> 분야별 로봇으로부터 발생 가능한 윤리적 쟁점

분야	윤리적 쟁점
<p style="text-align: center;">제조용· 산업용 로봇</p>	<ul style="list-style-type: none"> - 미래 사회에서 제조 및 산업용 로봇으로 인해 인간 노동의 질은 향상될 수 있는가? - 미래 사회에서 인간 노동자들은 제조 및 산업용 로봇으로부터 소외를 당할 가능성은 전혀 없는가? - 미래 사회에서 제조 및 산업용 로봇의 편리함에 다른 육체노동 감소로 인해 인간 노동자들에게는 새로운 직업병이 발생하지 않을까? - 인간 노동자들은 제조 및 산업용 로봇의 공정에서 주로 어느 역할을 담당하게 될까?
<p style="text-align: center;">메디컬· 헬스케어 로봇</p>	<ul style="list-style-type: none"> - 케어 로봇의 인공지능과 프로그램은 누가 어떤 내용으로 만들고 누가 관리해야 하는가? - 정서심리학적 치료도 가능할 것인가? - 케어 대상자가 케어 로봇에 대한 특정 '성(gender) 정체성'을 요구한다면, 부여할 것인가? - 가족 구성원들은 서로의 안부를 직접 챙길 필요가 있겠는가?
<p style="text-align: center;">서비스· 엔터테인먼트 로봇</p>	<ul style="list-style-type: none"> - 더 자극적인 서비스 및 엔터테인먼트 프로그램에 점차 중독될 가능성은 전혀 없는가? - 감시 기간 중 외부 방문자에 대한 오류 인식의 가능성은 전혀 없는가? 있다면, 어떤 문제가 발생할 수 있을까? - 독신자가 사랑하는 사람, 죽은 가족 구성원과 동일한 로봇의 외형을 요구한다면, 이를 허용해야 하는가? - 독신자가 자신의 욕구 충족을 위해 로봇의 외형을 자신의 이상형의 모습으로 요구한다면, 이를 허용해야 하는가?

분야	윤리적 쟁점
가정교육용 로봇	<ul style="list-style-type: none"> - 단지 지식 습득의 수준을 넘어 친구 간의 관계를 돈독하게 형성하는 데 기여할 수 있을까? - 가정교육용 로봇에 몰입하여 굳이 학교에 갈 필요가 없지 않을까? - 로봇 성능의 차별에 따라 가격이 결정되면, 이는 또 다시 빈부 격차를 심화시킬 수 있는 것은 아닌가?
탐사 로봇	<ul style="list-style-type: none"> - 탐사 로봇이 오작동을 일으켜 고장이 나면, 환경에 대한 심각한 파괴가 발생하지 않는가? 인공위성 우주 쓰레기는 누구의 책임인가? - 어떤 탐사용 로봇을 통해 한정된 자원을 빠른 속도로 편리하게 개발하는 것은 생태계의 혼란을 야기하고, 결국 인류의 파멸을 더욱더 앞당기는 것은 아닌가?
군사 로봇	<ul style="list-style-type: none"> - 무고한 민간인이 다칠 수 있음에도 불구하고, 전투 로봇에 대해 공격 명령을 내리는 것은 정당화할 수 있는가? - 전투 로봇의 즉각적인 현장 대처(공격) 능력을 인정해야 하는가?
유희 로봇	<ul style="list-style-type: none"> - 유희 로봇이 담당하고 있는 성관계는 인간적인 것인가? 아니면 단순한 도구적인가? - 인간적인 성관계의 실현을 위해 선천적이거나 후천적인 성적 장애를 가진 사람들을 위해서만 유희 로봇 사용은 정당화될 수 있는가? - 유희로봇의 활용은 어떻게 통제될 수 있을까?
사회적 네트워크형 로봇	<ul style="list-style-type: none"> - 사회의 안정을 위해 우리의 모든 것에 대한 감시는 정당화될 수 있는가? - 교도소 및 유치장과 같은 장소에서 경계 로봇에 의한 전방위 감시는 범죄자에 대한 인권침해인가? 범죄자를 보호하는 것인가? - 직접적인 인간의 확인을 거치지 않고서도, 네트워크를 통해서 모든 시설물에 대한 안전을 확보할 수 있는가?

※ 출처: 변순용·송선영. (2015). **로봇윤리란 무엇인가?**. 어문학사. pp. 125-170

한편 국내 정부기관에 의해 설립된 ‘대통령직속 4차산업혁명위원회’⁷³⁾는 인공지능윤리이슈의 대상을 제조, 의료, 금융, 국방 등 4개 산업 분야로 선정하였다. 이는 국내외 인공지능 제품/서비스별 산업 집중도를 분석한 결과로서 그 비율이 높은 제조, 의료, 금융의 3개 분야를 우선 선정하고, 국제적인 이슈와 동시에 국가안보적 차원의 법적·윤리적 쟁점을 야기하는 국방 분야를 검토대상으로 선정하였다. 이때 대통령직속 4차산업혁명위원회는 기술의 생산에서부터 소비에 이르는 ‘기술의 생애단계’와 그 단계와 관련된 ‘행위자’를 중심으로 윤리적 쟁점을 분석하였다.⁷⁴⁾ 대통령직속 4차산업혁명위원회에서 분석한 결과는 다음과 같다.

<표 III-5> 제조 분야 인공지능윤리이슈(자율주행자동차를 중심으로)

단계	행위주체	윤리이슈
개발 단계	개발자	- 인간에 대한 안전보호 의무 및 사회윤리적 공리를 반영하는 알고리즘 개발(하향식의 윤리적 프로그래밍 강조)
	제조사	- 알고리즘 개발 과정의 투명성 제고
	공적 주체 (중앙 및 지방 행정정부)	- 자율주행자동차의 개발 및 상용화를 위한 포지티브(positive) 규제 방식 채택 ⁷⁵⁾ (금지조항의 최소화 및 디폴트적 허용)
	소비자 (운행자 포함)	- 윤리적 기준의 자율적 선택권 보장에 대한 요청
	시민단체 (보행자 포함)	- 사회적 수용성 판단을 위한 공론의 의견 형성 및 반영 요청
인허가 단계	개발자 및 제조사	- 시범운행 단계에서의 윤리 기준 충족 및 운행결과에 대한 정보공개의 책무성
	공적 주체	- 인허가를 위한 윤리적 검증 기준의 정립 및 적용
	소비자 및 시민단체	- 인허가 과정에서의 윤리문제에 대한 모니터링 및 의견개진 절차 요청

73) <https://www.4th-ir.go.kr/>. (검색일: 2020. 3. 14.)

74) 대통령직속 4차산업혁명위원회. (2018). 4차산업혁명시대 산업별 인공지능윤리의 이슈 분석 및 정책적 대응방안 연구. 정보통신정책연구원. pp. 113-114

단계	행위주체	윤리이슈
상용 단계	개발자	- 알고리즘에서 사회윤리 및 사회적 수용성의 지속적인 변화 추 이 반영(상향식의 윤리적 프로그래밍 강조)
	제조사	- 자율주행자동차의 제조물책임보험을 통한 사후적 차원에서의 윤리문제 고려
		- 법인 내 기관윤리위원회 조직 및 운영을 통한 윤리문제의 절 차주의적 해결
	공적 주체	- 국가·사회적 차원에서 공적 위험분배를 위한 사회보험 등 공 적 기금 확보
		- 인허가단계의 윤리적 기준 충족 여부에 대한 사후적 검증 도 구 마련
		- 스마트 시티 및 지능형 도로교통체계와의 연계에 따른 보안 문제 해소 방안
소비자	- 윤리적 프로그래밍에 대한 선택권 및 자기결정권 보장의 지속 적 유지(프로그래밍 변경가능성 포함)	
시민단체	- 사회윤리 및 사회적 수용성 관련 지속적인 의견 형성을 위한 공론에서의 촉매 기능	
	- 도로교통체계의 지능화로 인한 프라이버시 및 개인정보 침해 예방 및 모니터링	

※ 출처: 대통령직속 4차산업혁명위원회. (2018). **4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구**. 정보통신정책연구원. pp. 135-136

75) 법률·정책상으로 허용하는 것을 구체적으로 나열한 뒤 나머지는 모두 금지하는 방식의 규제를 말한다. 법률·정책상으로 금지한 행위가 아니면 모든 것을 허용하는 네거티브(negative) 방식보다 규제 강도가 훨씬 세다. (포지티브 규제. <https://terms.naver.com>. 검색일: 2020. 3. 21.)

<표 III-6> 금융 분야 인공지능윤리이슈(로보어드바이저를 중심으로)

단계	행위주체	윤리이슈
개발 단계 (디자인, 프로그램밍)	개발자	<ul style="list-style-type: none"> - 알고리즘 개발단계에서의 윤리 특수성, 금융규제의 목적 투입 - 투자전문가가 투입(인정된 투자이론에 기초하여 설계)·상시 감독
	자율 규제기관	- AI 개발 촉진정책과 규제의 균형
	공적 규제기관	- AI 개발 촉진정책과 규제의 균형
	금융종사자	- AI 이용과 관련한 윤리교육, 금융규제 준수 교육
허가 단계 (판매·광고·이용)	개발자	- 허가를 위한 금융윤리적 자가검증도구 개발 및 적용
	공적 규제기관	- 허가과정에서 금융윤리적 검증도구 개발 및 역량 강화(알고리즘과 데이터 평가능력, 이용자에게 제시 및 의사결정하게 제시되는 선택구조, 정보기술인프라, 인공지능이용이 대규모화될 경우 시장에 미치는 영향 등)
이용 단계 (판매·광고·이용)	금융자문인	<ul style="list-style-type: none"> - AI 로보어드바이저⁷⁶⁾의 특수성 및 위험 교육 - 이용자 보호 프로세스 준수 - 이해상충 방지 프로세스 준수 - AI 위험을 쉽게 알 수 있는 방법으로 설명·공시기법 선택 - 개인정보 보호
	금융종사자 (은행·증권회사·보험회사·자산운용회사·선물회사)	<ul style="list-style-type: none"> - 이용자 보호를 위한 최적화된 알고리즘 이용 및 검증 담보 - 이해상충 방지 프로세스 검증 - AI 위험을 쉽게 알 수 있는 방법으로 설명·공시기법 선택 - 알고리즘 투명성 담보 - 개인정보 보호 - 해킹방지 등 침해방지를 위한 인적·물적 시설 적정수준 담보 - 금융시장 안정성 담보 - 이용자 피해 발생 시 책임범위

단계	행위주체	윤리이슈
이용 단계 (판매·광고·이용)	자율 규제기관 (거래소, 금융협회 등)	<ul style="list-style-type: none"> - 금융회사 이용자 보호와 규제의 균형 - 금융시장 보호(금융시장의 공정한 가격발전기능과 안정성 확보) - 금융윤리의 규범화 수준과 범위 - AI 기술에 관한 정보 및 대응 공유 - AI 기술의 수혜 격차 문제
	공적 규제기관	<ul style="list-style-type: none"> - 금융회사 이용자 보호와 규제의 균형 - 금융시장 보호와 규제의 균형 - 금융윤리의 규범화 수준과 범위 - AI기술에 관한 정보 및 대응 공유(감독기관간, 전문기관간, 전문가간) - AI 기술의 수혜 격차 문제
	이용자	<ul style="list-style-type: none"> - 결정과 선택권 문제 - 윤리적 투자 - 금융정보의 활용과 프라이버시 보호 - 해킹에의 안전
	기타 (기관투자자)	<ul style="list-style-type: none"> - 결정과 선택권 문제 - 윤리적 투자

※ 출처: 대통령직속 4차산업혁명위원회. (2018). 4차산업혁명시대 산업별 인공지능 윤리의 이슈 분석 및 정책적 대응방안 연구. 정보통신정책연구원. pp. 153-154

76) “‘로보어드바이저’(robo-advisor)는 로봇(robot)과 투자전문가(advisor)의 합성어다. 고도화된 알고리즘과 빅데이터를 통해 인간 프라이빗 뱅커(PB) 대신 모바일 기기나 PC를 통해 포트폴리오 관리를 수행하는 온라인 자산관리 서비스를 일컫는다. 직접 사람을 마주하고 상담하지 않고도 온라인 환경에서 자산 배분 전략을 짜주기 때문에 개인 맞춤형 서비스를 제공할 수 있고, 수수료가 저렴하며, 낮은 투자금 하한선을 설정할 수 있는 것이 특징이다.”(로보어드바이저. <https://terms.naver.com>. 검색일: 2020. 3. 14.)

<표 III-7> 의료 분야 인공지능윤리이슈

단계	행위주체	윤리이슈
개발 단계	개발자	- 알고리즘 개발 단계에서 국가별/지역별 윤리의 특수성 고려
	공적 주체	- AI 개발을 위한 촉진 정책과 규제의 적절한 균형
	의사	- AI 이용과 관련한 윤리 교육(의과대학에서의 윤리 교육 포함)
인허가 단계	개발자 및 제조사	- 허가를 위한 윤리적 자가 검증 도구 개발 및 적용
	공적 주체	- 허가 과정에서의 윤리적 검증 도구 개발
	소비자 및 시민단체	- 인허가 과정에서의 윤리문제에 대한 모니터링 및 의견개진 절차 요청
상용 단계	의사	- 수가(酬價) 책정의 문제 - 의료사고 시 책임 면제 범위의 문제
	병원 경영자	- 병원의 수익 창출에 최적화된 알고리즘 적용의 문제
	공적 주체	- 의료사고 시의 손해배상 체계와 사전 행정적 규제의 검토 - 윤리적 내용의 규범화 범위 - 의료 기술의 수혜 격차 문제
	소비자	- 데이터 활용과 프라이버시 보호의 문제 - 소비자의 선택권 문제 - 자가진료 내지 무허가 의료행위의 문제

※ 출처: 대통령직속 4차산업혁명위원회. (2018). 4차산업혁명시대 산업별 인공지능윤리의 이슈 분석 및 정책적 대응방안 연구. 정보통신정책연구원. p. 168

<표 III-8> 국방 분야 인공지능윤리이슈

단계	행위주체	윤리이슈
정책 결정 단계	군	<ul style="list-style-type: none"> - 치명적 자율무기⁷⁷⁾ 개발의 필요성 및 허용여부 - 방어목적의 AI를 채택할 때 얻을 수 있는 고유한 장점과 취약점은 무엇인가 - AI가 미래 전쟁을 어떻게 변화시킬 것인가에 대한 연구
	최고통수권자	<ul style="list-style-type: none"> - LAWs는 경쟁국들과의 군사적 균형에 어떤 영향을 미칠 것인가 - 국가안보 최고수호자로서의 윤리의식
	윤리/법률가	<ul style="list-style-type: none"> - LAWs의 사고책임에 대한 법적·윤리적 쟁점에 대한 신속한 대응책이 강구되었는가 - 쟁점들에 대한 국제법 준수 및 대응 입법은 준비되었는가
개발 단계	개발자	<ul style="list-style-type: none"> - 알고리즘에 어떤 윤리기준을 주입시킬 것인가 - 전쟁법에서 요구하는 의미 있는 인간통제를 어떻게 구현할 것인가
	윤리/법률가	<ul style="list-style-type: none"> - 법적/윤리적 기준을 제시할 수 있는가
	운용책임자	<ul style="list-style-type: none"> - 운용자가 필요로 하는 별도의 기준은 없는가
실험 단계	개발자 및 제조사	<ul style="list-style-type: none"> - LAWs에 대한 안정성은 담보되었는가
	공적 주체	<ul style="list-style-type: none"> - 지향하는 LAWs는 극대의 살상이라는 효용성은 보장되는가
	소비자 및 시민단체	<ul style="list-style-type: none"> - 테러리스트 등에 대한 보안수준
국제 협력 단계	국방/외교	<ul style="list-style-type: none"> - LAWs에 대한 국제 보편의 기능적, 윤리적 기준은 무엇인가
	윤리/법학자	<ul style="list-style-type: none"> - 기존 국제법질서와의 관계는 어떻게 되는가
	운용책임자	<ul style="list-style-type: none"> - 보편과 특수의 조화를 어떻게 도모할 것인가

77) “치명적 자율무기(Lethal Autonomous Weapons: LAWs), 일명 살인 로봇은 인간의 개입 없이 시스템 스스로가 목표물을 선택하고 공격하도록 설계된 자율무기를 의미한다. LAWs는 LAWS(lethal autonomous weapon systems, 치명적 자율무기 시스템), LAR(lethal autonomous robots, 치명적 자율로봇), 로봇 무기(robotic weapons), 살상용 로봇(killer robots)이라고도 불린다.”(Krishnan, Armin. 2009; Anderson & Waxman. 2012; 대통령직속 4차산업혁명위원회. 2018)

단계	행위주체	윤리이슈
실전 배치 단계	부대책임자	- LAWs의 살상력은 충분한가 - 아군과 적군, 민간인과 무장군인을 정확히 구분할 수 있는 여부
	정책담당자	- 상대세력의 LAWs에 필적하는지 여부 - 정책부서와 정보부서의 유기적인 협조가 이루어지는가 - 사이버공간에서도 구동이 가능한 수준의 LAWs인지 여부 - 프라이버시 유출에 대한 대비가 적절히 이루어졌는지 여부 - 현장 배치된 LAWs의 프로그램 해킹에 대한 대비가 충분한지 여부
	윤리/법률가	- 어떤 유형의 피해를 인정할 것인지 여부

※ 출처: 대통령직속 4차산업혁명위원회. (2018). 4차산업혁명시대 산업별 인공지능윤리의 이슈 분석 및 정책적 대응방안 연구. 정보통신정책연구원. p. 184

다. 범용 인공지능(AGI)에 관한 윤리적 쟁점

인공지능기술 개발은 인간의 지능 모방을 목표로 하고 있으며, 이를 위해 인공지능 연구자는 지능의 주요 속성인 합리성과 효율성에 주목하고 있다.⁷⁸⁾ 이렇게 인공지능기술이 인간의 정신활동과 신체활동을 모방하는 방향으로 개발되고 있다면, 자연스럽게 인간 수준의 인공지능, 즉 범용 인공지능에 대한 논의로 이어지게 된다. 범용 인공지능에 대한 논의는 기술공학적 관점의 좁은 인공지능과는 다른 관점으로 접근하게 된다. 김효은(2019)에 의하면 크게 존재론적 관점의 접근으로 ‘인공지능이 인간의 지능을 실제로 가질 수 있는가’, 인식론적 관점의 접근으로 ‘인공지능이 인간의 지능을 가졌다는 것을 인간이 어떻게 알 수 있는가’로 나누어 볼 수 있다.(김효은, 2019, p. xii) 인식론적 관점의 접근으로 튜링 테스트⁷⁹⁾, 중국어방 논변⁸⁰⁾, 러브레이스 테스트⁸¹⁾ 등 다양한

78) 변순용 편. (2019). 윤리적 AI로봇 프로젝트. 어문학사. p. 164

79) “앨런 튜링은 1950년에 철학 저널 ‘Mind’에 발표한 Computing Machinery and Intelligence

시도들이 있었다. 하지만 인식론적 관점의 경우 우리가 인간과 인공지능을 구분할 수 없다는 특징으로 인공지능이 인간의 지능을 가졌다고 볼 수 있는가에 대한 의문을 제기하게 만든다. 예를 들어 2014년 튜링 테스트를 통과한 인공지능 ‘유진 구스트만(Eugene Goostman)’에 대해 인간의 지능을 가졌다고 선불리 받아들이기 어려운 경우를 고려하면 그렇다.⁸²⁾ 왜냐하면 인식론적 관점으로 인공지능이 인간의 지능을 가졌다고 말하기 위해서는 우선 인간의 지능을 어떻게 정의해야 하는지에 대한 근원적인 접근이 필요하기 때문이다. 따라서 인공지능에 대해 인식론적으로 접근하기 위해서는 인공지능에 대한 존재론적 물음을 먼저 고찰해야 한다.

인간 수준의 인공지능이라고 명명하기 위해서는 ‘인간 수준’이 무엇인지 그 조건을 고려해야 한다. ‘인간 수준’이란 무엇인가? 어떠한 특징이 나타났을 때 우리는 ‘인간’이라고 할 수 있는가? ‘인간’이란 무엇인가? 인간 수준의 인공지능을 탐구하기 위한 이와 같은 존재론적 물음은 다시 ‘인간’에 대한 근원적인 물음으로 되돌아가게 한다. 국내외적으로 이성, 자유, 의식, 의지, 감정에 대한 철학적인 고찰은 범용 인공지능 개발에 주요한 연구 과제로 자리 잡았다. 닉 보스트롬과 엘리저 유고스키에 따르면 사실 인간에 대한 철학적 접근은 고대에서부터 시작되었으며, 시대가 변하고 기술이 발전하면서 인간을 연구하는 분야도 윤리 철학에 국한된 것이 아니라 언어학, 신경과학, 생물학, 인류학, 뇌과

에서 기계가 지능적이라고 간주할 수 있는 조건을 언급했다. "기계는 생각할 수 있는가?"라는 질문에 대해 그는 긍정적이라고 답변하면서, "컴퓨터가 생각할 수 있다면 그것을 어떻게 표현해야 하는가?"라는 핵심 질문에 대해 그는 "컴퓨터로부터의 반응을 인간과 구별할 수 없다면 컴퓨터는 생각(사고, thinking)할 수 있는 것"이라고 주장하였다. 만일 지성 있는 사람이 관찰하여 기계가 진짜 인간처럼 보이게 하는 데 성공한다면 확실히 그것은 지능적이라고 간주해야 한다는 주장이다." (<https://ko.wikipedia.org/wiki>. 검색일: 2020. 3. 11.)

80) “중국어 방 혹은 중국인 방(영어: Chinese room)은 존 설(John Searle)이 튜링 테스트로 기계의 인공지능 여부를 판정할 수 없다는 것을 논증하기 위해 고안한 사고실험이다.” (<https://ko.wikipedia.org>. 검색일: 2020. 3. 11.)

81) 인공지능에게 예술 작품을 만들라고 한 다음 그 결과가 어떻게 나왔는지 사람이 설명할 수 없는 경우, 또는 인공지능의 작품과 인간의 작품을 구분할 수 없는 경우에 테스트를 통과한 것으로 보자는 제안이다. (<http://www.ciokorea.com/news>. 검색일: 2020. 7. 4.)

82) ‘유진 구스트만’, 단순한 채팅 로봇인가, 아니면 ‘생각하는’ 인공지능인가?. <http://www.itworld.co.kr>. (검색일: 2020. 3. 11.)

학, 심리학으로 확대되었다고 한다.(Nick Bostrom·Eliezer Yudkowsky, 2011, p. 3) 이러한 변화의 모습으로 비추어보았을 때 인간 수준에 대한 논의가 활성화되기 위해서는 단일 학문의 깊이 있는 연구뿐만 아니라 연구 분야 간 긴밀한 협력이 필요하고, 범용 인공지능의 출현 가능성을 염두에 두어 인공지능 분야와의 연계 방안을 모색해야 할 것이다.

특히 인간의 '직관'에 대한 연구는 좁은 인공지능(ANI)이 범용 인공지능(AGI)으로 인정받기 위한 핵심적인 부분이라 할 수 있다. 여기서 직관(直觀)은 “감각, 경험, 연상, 판단, 추리 따위의 사유 작용을 거치지 아니하고 대상을 직접적으로 파악하는 작용.”⁸³⁾으로서 인간이 대상을 파악하는 하나의 방법이다. “인간에게 쉬운 것은 컴퓨터에게 어렵고 반대로 인간에게 어려운 것은 컴퓨터에게 쉽다.”라는 모라벡의 역설⁸⁴⁾은 인간의 직관적 사고 능력과 인공지능 알고리즘의 형식화된 처리 능력의 대비를 보여준다. 예를 들어 어린아이에게 물체 집기는 몇 번의 연습을 통해서 쉽게 가능하지만, 로봇이 인간처럼 물체를 자유자재로 집기란 매우 어렵다. 반대로 복잡한 수학적 계산을 할 때 인간에게는 어렵지만 로봇은 몇 가지의 코드만 주어진다면 순식간에 해결한다. 추형석(2018)에 의하면 인간과 인공지능 사이의 격차는 인간의 직관적인 사고의 원리는 무엇이며, 그것을 어떻게 알고리즘으로 구현해내는가에 달려있다고 보았다.⁸⁵⁾

만약 인간의 도덕적 판단 기준을 알고리즘으로 구성하여 인공지능기술에 적용한다면 어떨까? 웬델 월러치(Wendell Wallach)와 콜린 알렌(Colin Allen)은 도덕 행위자의 범위가 인공지능 시스템으로 확대될 경우, 인공적 도덕 행위자(Artificial Moral Agent, AMA)⁸⁶⁾로 명명하여 인공지능 시스템이 인공적 도덕 행위자가 되기 위한 몇 가지 접근법을 제시하였다. 크게 하향식(Top-down) 접근법, 상향식(Bottom-up) 접근법, 그리고 혼합식(Hybrid) 접근법

83) 직관. <https://stdict.korean.go.kr>. (검색일: 2020. 3. 11.)

84) 모라벡의 역설. <https://terms.naver.com>. (검색일: 2020. 3. 11.)

85) 추형석. (2018). 범용 인공지능의 개념과 연구 현황. **소프트웨어정책연구소 이슈리포트**, (2), 1-16.

86) Wendell Wallach·Colin Allen. (2016). **왜 로봇의 도덕인가**(노태복 역). 메디치. (원저 2009 출판)

이 있다. 하향식 접근법이란 “어떤 구체적인 윤리 이론을 택한 다음 컴퓨팅적 요건을 분석함으로써 그 이론을 구현할 수 있는 알고리즘과 서버 시스템의 설계를 이끌어내는 방식을 가리킨다.”(Wendell Wallach·Colin Allen, 2009/2016, p. 139) 이 접근법의 경우에 어떤 윤리 이론을 로봇에 적용할지, 윤리 이론을 알고리즘으로 어떻게 구현할지에 대한 고민이 필요하다. 다음으로 상향식 접근법은 “행위자가 행동 과정들을 탐구하고 배우며 도덕적으로 칭찬받을 만한 행동에 대해 보상을 받는 환경을 창조하는 데 초점이 놓인다.”(Wendell Wallach·Colin Allen, 2009/2016, p. 139) 콜버그의 도덕 발달 단계 이론을 적용한 아동의 도덕성 발달 과정이 상향식 접근법의 한 가지 예라고 볼 수 있다. 이때 인공지능 학습의 경우 데이터를 기반으로 이루어지기 때문에 좁은 인공지능의 윤리적 쟁점에서 다루었던 데이터 편향 가능성이 다시 제기될 수 있다. 그리고 학습을 통해 프로그램화된 인공지능의 도덕적 판단을 우리가 받아들이 수 있는가에 대한 논란이 제기될 수 있다. 마지막으로 혼합식 접근법은 이분법적인 하향식 접근법과 상향식 접근법의 혼합 형태로 “함양된 내재적 가치에 의해 제공된 하향식 가치를 포함하는 다양한 입력 및 영향과 더불어 환경에 대한 풍부한 이해를 통합시킬 수 있는 컴퓨팅 시스템”(Wendell Wallach·Colin Allen, 2009/2016, p. 204)을 가리킨다. 이를 위해 웬델 윌러치와 콜린 알렌은 아리스토텔레스의 덕윤리를 기반으로 연결주의 네트워크를 활용한 컴퓨터 시스템의 개발 가능성을 제시하기도 하였다. 여기서 연결주의는 “언어와 규칙이 제공하는 추상적 이론보다는 경험과 사례를 통한 신경망의 개발과 훈련에 초점을 둔다.”(Wendell Wallach·Colin Allen, 2009/2016, p. 206) 이를 위해서는 하향식 접근법과 상향식 접근법의 철저한 이해가 바탕이 되어야 하기 때문에 실현 상의 난제가 발생한다. 그리고 연결주의 학습 시스템으로 내린 판단에 대한 근거와 설명이 충분히 보장되어야 하며, 사회, 정치, 문화적 성향을 반영한 하향식 구조를 어떻게 구현해낼지에 대한 사항들은 여전히 연구 과제로 남아있다.

이처럼 인공지능 연구자 중에서 범용 인공지능(AGI)이 인간의 사고방식과 행동을 기반으로 개발하고자 다양한 가능성을 열어두어 연구가 이루어지고 있다. 그렇다면 인간이 도덕적 지위(Moral Status)를 갖듯이 기계의 도덕적 지

위에 대해 생각하지 않을 수 없게 된다. 닉 보스트롬과 엘리저 유고스키는 인간에게는 고유한 도덕적 지위가 있다고 하지만, 그 도덕적 지위란 무엇인지 단일한 기준으로 규정할 수 없다고 하였다.(Nick Bostrom·Eliezer Yudkowsky, 2011, p. 6) 그럼에도 불구하고 인간의 고유한 도덕적 지위를 인정한다면, 인간 수준의 범용 인공지능에 대한 도덕적 지위도 논의가 필요한 주제가 된다. 범용 인공지능의 도덕적 지위를 인정한다면 어떠한 근거로 뒷받침하는지, 또는 도덕적 지위를 인정할 수 없다면 인정할 수 없는 근거가 무엇인지에 대한 논의가 활성화되어야 한다. 닉 보스트롬과 엘리저 유고스키는 이와 같은 범용 인공지능의 도덕적 지위는 생명윤리, 동물윤리, 의료윤리 등의 다른 응용윤리학에서 다루어지는 도덕적 지위 논쟁을 중심으로 설명하였다.(Nick Bostrom·Eliezer Yudkowsky, 2011, p. 6) 예컨대 태아의 도덕적 지위, 시험관 아기, 낙태, 대리모 문제, 유전자 조작, 인체실험, 동물실험, 동물복제 등 응용윤리 분야에서 윤리적 검토가 필요한 쟁점사항들은 그 대상의 도덕적 지위를 어떻게 설정하느냐에 따라서 논의 관점이 달라진다는 것이다. 이는 결국 생명의 근원과 존재에 대한 형이상학적 접근을 요구하게 된다.

그러나 현재 시점에서는 범용 인공지능에 대한 윤리적 쟁점은 좁은 인공지능에 대한 윤리적 쟁점과 비교해서 광범위하고 추상적이며 비현실적인 쟁점처럼 보인다. 실제로 인공지능기술이 상용화되면서 이미 발생했거나 앞으로 발생할 수 있는 문제점과 갈등양상에 대해서 사람들은 법 또는 윤리적 규범에서 그 해결방안과 현실적인 절충점을 찾으려 한다. 또한 인공지능기술이 적용되는 분야가 다른 과학기술에 비해 그 범위가 넓고 다양하며 그 영향력 또한 우리의 일상뿐만 아니라 사회구조적으로도 변화를 불러일으키고 있기 때문에, 인공지능기술을 취급할 때 지켜야 할 윤리적 지침은 기업, 국가, 협력체 단위로 발표되고 있다. 이렇게 발표된 윤리적 지침에는 현실적 쟁점과 더불어 인공지능기술과 인간에 대한 본질적인 물음이 밀바탕으로 되어있다. 그렇기 때문에 인간의 존엄성, 자율성, 책임의 주체, 인공지능의 특성에 대한 고찰이 이루어져야 인공지능에 대한 윤리적 지침이 일회성의 지침으로 남지 않을 것이다. 이 점이 범용 인공지능에 대한 윤리적 쟁점을 고려해야 할 이유이기도 하다.

IV. 인공지능시대를 대비한 한스 요나스의 책임이론 적용

1. 한스 요나스 기술관에 근거한 인공지능기술 분석

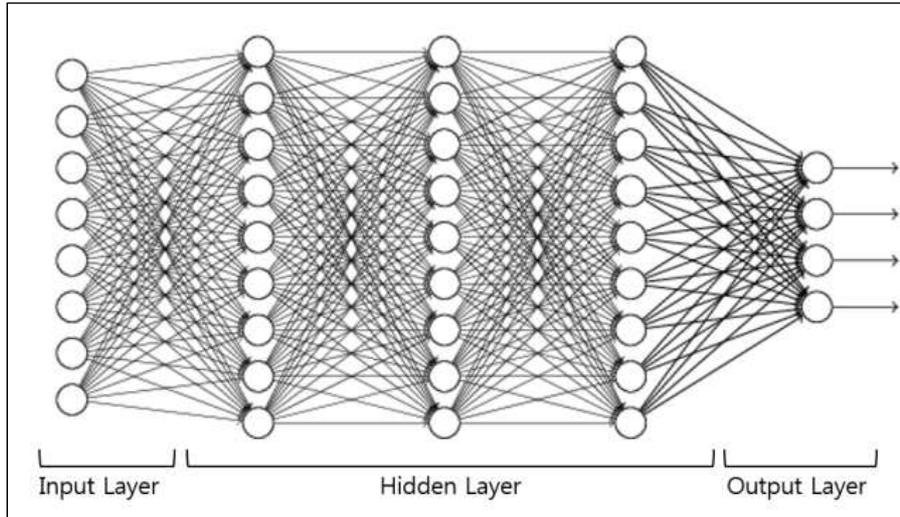
앞서 살펴본 바와 같이, 인공지능기술이 발달함에 따라 인공지능기술 자체로 발생하는 윤리적 쟁점, 인공지능기술과 로봇이 결합하면서 로봇의 행위, 또는 존재와 관련된 윤리적 쟁점이 현대사회에서 새롭게 대두되고 있다. 이러한 인공지능기술과 관련된 윤리적 쟁점을 직시하는 것이 현재 살아가고 있는, 그리고 앞으로 살아갈 인류에게 필요한 윤리적 지혜이다. 이를 위해 한스 요나스의 기술관과 책임관을 살펴보았다. 한스 요나스의 기술관을 바탕으로 인공지능기술의 특성을 분석하고, 분석한 내용에 대응하기 위한 윤리적 책임의 형태를 한스 요나스의 관점에서 제시해보고자 한다.

요컨대 한스 요나스는 현대기술의 특징을 ‘결과의 모호성’, ‘적용의 강제성’, ‘시공간적 광역성’, ‘인간중심주의의 파괴’, ‘형이상학적 물음의 제기’로 나누어 설명하였다. 한스 요나스는 이러한 특징을 가지는 현대기술 중에서 의학 분야, 인체실험, 유전공학, 인간복제, 죽음의 지연술을 중심으로 자신의 주장을 뒷받침하였는데 인간의 생물학적인 특성과 직접적으로 연관된 현대기술들을 예로 하여 자신의 이론에 적용하였다. 이와 비슷한 흐름으로 본 연구에서 다루고 있는 인공지능기술을 한스 요나스의 관점에서 살펴본다면 다음과 같은 내용으로 이어지게 된다.

먼저 ‘결과의 모호성’ 측면에서는 인공지능기술의 특징을 나타내는 말 중에서 ‘블랙박스(Black Box)⁸⁷⁾’라는 말의 유래를 살펴보면 이 특징을 파악할 수 있다. 여기서 블랙박스라는 용어는 비행기에 비치하는 자동 기록 장치라기보다는 “기능은 알지만 작동 원리를 이해할 수 없는 복잡한 기계 장치⁸⁸⁾”를 뜻한다. [그림 IV-1]과 같이 딥러닝(Deep Learning)의 주요 특징인 인공 신경망은 그 층위가 여러 겹이고 일정한 패턴으로 인해 복잡한 처리과정을 거치게 된다.⁸⁹⁾

87) James Barrat. (2016). **파이널 인벤션**(정지훈 역). 동아시아. (원저 2013 출판) p. 180

88) black box. <https://en.dict.naver.com>. (검색일: 2020. 2. 14.)



[그림 IV-1] 딥러닝 네트워크

이러한 처리과정은 인공 신경망을 구성하는 층위가 많아질수록 더욱 복잡해 지는데 오늘날 개발되는 인공지능기술의 처리과정은 전문가들도 해석하기 어렵거나 아예 접근조차 못 하는 실정이다. 이러한 딥러닝 네트워크의 특징을 비유적으로 표현하는 말이 ‘블랙박스’이다. 이러한 특징으로 인해 인공지능기술로 출력된 결과물은 왜 그렇게 나왔는지 인간이 예측할 수 없고 다만 프로그램에 정확한 데이터를 많이 입력함으로써 출력물의 정확도를 높일 뿐이다. 이렇게 인공지능기술의 처리과정이 복잡해지면, 인공지능기술로 인해 사고가 났을 때 그 원인을 규명하는 일이 힘들게 된다. 더 나아가 이러한 문제가 반복적으로 발생하게 되면 인공지능기술 개발에 걸림돌이 될 것이다. 이를 방지하기 위해서는 앞에서 언급한 투명성이 확보되어야 한다.

둘째, ‘적용의 강제성’ 측면에서 인공지능기술은 기술의 핵심 원리를 다양한 분야에 접목시킴으로써 소위 ‘인공지능시대’의 도래를 촉진하게 해주었다.⁹⁰⁾ 인공지능기술은 현재 일상생활 속에 다양한 형태로 쓰이고 있다. 개인비서, 또는 자신의 일정을 관리해주는 스마트폰, 자신이 선호하는 채널 또는 시청 시간을 파악하여 프로그램을 추천해주는 스마트 TV, 음악 감상이나 라디오 청취, 블

89) <https://namkugkim.wordpress.com>. (검색일: 2020. 2. 14.)

90) 뉴 사이언티스트 외. (2018). **기계는 어떻게 생각하고 학습하는가**(김정민 역). 한빛미디어(주)

루투스 연결에 활용되던 스피커를 음성인식 기술과 접목한 AI 스피커⁹¹⁾, 세탁기, 그 외의 가전제품 등 가정에서 쓰이는 인공지능기술을 쉽게 찾아볼 수 있다. 더 나아가 자율주행자동차, 드론, 의료, 수학, 예술, 무역, 사회 등 거의 모든 분야에서 인공지능기술이 쓰이고 있다. 인공지능기술 개발자들은 자신들이 개발한 인공지능기술을 오픈소스⁹²⁾로 제공하여 그 활용범위를 자연스럽게 확대시키고 있다. 우리가 알아채지 못한 사이 이미 인공지능기술은 사회 곳곳에 쓰이고 있으며 그 유용성과 효율성으로 인해 인공지능기술 개발을 중단하거나 기술 자체를 거부하는 것은 현재로서는 어려워 보인다.

셋째, ‘시공간적 광역성’의 측면에서 인공지능기술은 미래의 핵심기술로 자리 잡혔으며 인공지능기술을 연구하기 위한 투자비중도 다른 기술들에 비해 커지고 있는 실정이다. 이러한 흐름은 미래의 일자리부터 시작하여 인간의 생활양식, 더 나아가 우주산업, 그리고 우리가 상상할 수 없는 영역으로의 확장까지 나아가게 해준다. 그리고 인공지능기술은 사회구조의 변화, 산업구조의 변화, 직무구조의 변화 등 일상의 변화를 일으키는 요소 중에 하나로 작용하고 있다. 하지만 부정적인 시각도 있다. 직업의 형태 변화에 따라 기술을 다룰 수 있는 일반적인 개인의 능력과 고도로 전문화된 기술 사이에 차이가 있는 기술 불일치 문제⁹³⁾, 금전적으로 여유가 있는 사람과 없는 사람 사이의 기술적 격차인 기술 불평등 문제, 무분별한 인공지능 개발에 대한 미래 비관론자들의 주장들까지 현재 인공지능기술에 대해서 긍정적인 시각과 부정적인 시각이 동시에 나타나고 있는 실정이다. 이에 따라 미래의 후손들을 생각하고 인공지능기술의 파급력을 고려하여 윤리적 판단을 포함한 신중한 접근을 요구한다.

넷째, 인공지능기술은 전통윤리학적인 인간중심주의를 넘어 범지구적인 차원의 접근을 요구한다. 마틴 포드(Martin Ford)에 의하면, 데미스 허사비스(Demis Hassabis)는 인공지능기술이 국가 간 경쟁을 부추길 위험을 경고하

91) 인공지능(AI) 스피커. <https://terms.naver.com>. (검색일: 2020. 2. 14.)

92) 오픈소스 소프트웨어, OSS라고도 한다. 소프트웨어의 설계도에 해당하는 소스코드를 인터넷 등을 통하여 무상으로 공개하여 누구나 그 소프트웨어를 개량하고, 이것을 재배포할 수 있도록 하는 것 또는 그런 소프트웨어를 말한다. (오픈소스. <https://terms.naver.com>. 검색일: 2020. 2. 14.)

93) Martin Ford. (2019). **AI 마인드**(김대영 외 역). 터닝 포인트. p. 151

면서 국제적인 협약이나 규제, 협력을 강조하였다. 이를 통해 데미스 허사비스는 인공지능기술이 인공지능에만 국한된 문제를 넘어 불치병 치료, 수질 정화, 기후 변화 등의 환경문제의 개선에 도움이 되는 기술을 개발할 필요가 있다고 주장하였다.(Martin Ford, 2019, pp. 50-52) 또한 조슈아 테넨바움(Joshua Tenenbaum)은 인공지능기술 열풍이 불고 있는 사회 분위기를 경계하면서 상대적으로 신경 쓰지 못한 기후변화 문제에 관심을 가질 필요가 있다고 하였다.(Martin Ford, 2019, p. 572) 이처럼 인공지능기술은 인간의 뇌 구조와 정신활동을 모방하는 것으로 시작했으나 인공지능기술의 적용 대상은 인간을 넘어 사회문제, 자연으로 확대되고 있다. 따라서 인공지능기술에 대한 윤리적 접근은 인간의 문제를 넘어 사회, 생태계, 범지구적 차원을 함께 고려할 필요가 있게 되었다.

다섯째, 인공지능기술은 형이상학적 물음을 제기하기도 한다. 현재의 인공지능기술은 특정 분야와 특정 학습 방법으로만 프로그램화되어 있어 인간과 똑같이 생각하고 행동할 수는 없다. 하지만 전문가들은 머지않은 미래에 인간과 비슷한 인공지능이 출현하게 될 것이라고 예측하고 있다.⁹⁴⁾ 그렇게 된다면 인간과 인공지능을 구분하는 것부터 시작해 존재 자체에 대한 물음이 자연스럽게 발생하게 될 것이다. 그뿐만 아니라 인공지능기술이 개발되고 있는 현재에도 인간과 기계 사이의 형이상학적 물음을 제기한다. 인간의 정신은 어떻게 구현되는지, 의식이 무엇인지, 감정이 무엇인지, 인간의 학습능력은 어떻게 발달했는지, 인간의 복합적 사고가 어떻게 가능한지 등 인간의 의식과 정신, 자아와 관련된 물음과 함께 인공지능기술은 연구되고 있다. 따라서 인공지능기술은 수학자, 컴퓨터 공학자, 생물학자, 뇌과학자, 윤리학자, 천문학자, 바이오 공학자, 사회학자 등 거의 모든 전문분야 전문가들이 한데 모여 연구가 이루어져야 할 총체적 연구 분야가 되었다.

94) 초지능 인공지능 대안이론 ‘리프레이밍 슈퍼인텔리전스’. <http://scimonitors.com>. (검색일: 2020. 2. 17.)

2. 한스 요나스 책임관에 근거한 인공지능기술 행위주체별 책임

인공지능기술에 대한 기술적 특성을 한스 요나스의 기술관에 비춰 분석해보았다. 이러한 인공지능기술을 다루는 것은 결국에는 우리 인간이다. 따라서 우리에게 요구되는 책임의 자세를 살펴볼 필요가 있다. 본 연구의 제3장 2절 ‘인공지능기술과 관련된 윤리적 쟁점들’에서 정리하였듯이, 기술의 생애단계는 기술을 개발하고 연구하는 개발단계, 사회적 기준을 근거로 개발된 기술의 타당성과 적절성을 평가하는 인허가단계, 그리고 인허가된 기술을 대중에게 보급하는 상용 또는 이용단계로 나누어 볼 수 있다. 이러한 과정에서 다양한 행위주체들이 참여하게 되는데 본 연구에서는 기술의 생애단계에서 참여하게 되는 개발자·제조사, 공적주체, 사용자로 나누어 한스 요나스의 책임관에 근거한 그들의 책임을 제시해보고자 한다. 여기서 개발자·제조사는 인공지능 과학자, 연구자, 그리고 인공지능기술과 관련된 프로그램 제작 및 탑재, 인공지능기술이 적용된 제품을 생산하는 제조사를 의미한다. 그리고 공적주체는 지방정부, 정치인, 국제기구로서 인공지능과 관련된 법령을 정하고 인공지능기술에 대한 규제 또는 완화 정책을 논의하고 시행하는 주체이다. 마지막으로 사용자는 인공지능기술을 현장에서 사용하고 평가하는 모든 주체이다. 이들은 의료, 제조, 금융, 교통, 농업, 국방, 관광 등 인공지능기술을 활용하는 모든 산업분야에 종사하는 사람들부터 그 서비스를 제공하는 수요자이자, 민간차원에서 인공지능기술에 대한 적절성을 평가하는 평가자의 역할을 한다. 다만 이러한 행위주체들은 기술의 생애단계 중 반드시 어느 한 단계 속에서만 활동하는 것이 아니라, 모든 단계를 거쳐 지속적으로 다른 행위주체와 교류하는 특성을 지니고 있다. 그렇기 때문에 각 주체들은 분절적이기 보다는 서로 영향을 주고받는 관계라 할 수 있다.

그리고 기술의 생애단계에서 이끌어낸 행위주체 외에도 교육주체를 추가하여 그들의 책임을 제시해보고자 한다. 인공지능기술의 적용 범위가 대부분의 산업분야로 확대되면서 그에 따르는 문제점이 발생하고 있다. 주로 인공지능기술을 어떻게 다루고 취급하는가에 대한 기술사용문제, 인공지능기술과 관련된 인력 부족문제, 인공지능기술의 보급에 있어 사회적 계층 간의 차이가 발생하는 것

을 예로 들 수 있다. 국내에서도 이러한 문제의식을 바탕으로 인공지능 교육과 관련된 정책을 추진하고 있다.⁹⁵⁾ 초등, 중등, 고등, 대학, 그리고 평생교육까지 국내의 교육단계별로 어떤 인공지능교육을 할지의 계획을 구상하고 있다고 한다. 이렇게 인공지능교육에 대한 교육주체의 역할이 중요해지고 있다. 이러한 방향에서 교사, 산업체 연수기관, 교육기관 등의 교육주체들이 지녀야 될 책임을 한스 요나스의 관점에서 제시해보고자 한다.

가. 개발자·제조사의 책임

국내에서 인공지능기술을 적용한 산업 분야 중에서 비중이 크면서 논란이 되는 주요 분야는 자동차 분야라고 할 수 있다. 자동차를 구매할 때 차주들은 자율주행 옵션을 추가하여 자신의 안전과 편의를 보장하기도 하지만 자율주행 기능으로 인해 크고 작은 사고가 발생할 때마다 안전성에 대한 논란이 제기되고 있다. 현재 국내 법률상으로는 자동차를 주행할 때 발생하는 모든 책임은 운전자에게 있다.⁹⁶⁾ 현재까지 자동차 제조사는 자율주행 기능을 ‘운전자 보조 시스템’으로 명시하고 그 세부내용을 지침으로 제공하고 있기 때문에 조작과 취급에 대한 책임은 운전자에게 있는 것이다. 이때의 책임은 법률상의 책임으로 어떤 문제가 발생했을 때 법적인 처벌과 보상 차원의 책임을 의미한다. 이러한 법률적 책임의 논쟁은 자동차 분야에만 국한되지는 않는다. 인공지능기술을 탑재한 의료기기와 관련된 사고, 국방 분야 등 인간의 생명과 직접적으로 관련된 분야에서 법률적 책임의 논쟁은 꾸준히 제기되고 있다. 이렇듯 인공지능기술이 인간의 생활 곳곳에 들어오면서 그에 따르는 사회적 논란이 발생하고 있다. 이러한 상황 속에서 인공지능 개발자와 제조사는 인공지능기술을 개발할 때 기술의 사회적 영향력을 고려하여 신중한 태도를 갖춰야 한다.

과학자나 연구자가 기술의 사회적 영향력을 고려해야 하는 이유는 무엇일까? 연구에 있어 연구자의 자유와 공익이 충돌하는 경우가 현대에 들어 많이 발생하고 있다. 한스 요나스는 이러한 이유를 산업혁명 이후 순수 과학영역이 기술

95) '인공지능시대' 초등부터 단계별 'AI교육 기본계획' 만든다. <https://www.yna.co.kr>. (검색일: 2020.6.10.)

96) 4자율주행차 상용화... “사고는 ‘너’ 책임”. <https://shindonga.donga.com>. (검색일: 2020. 6. 10.)

영역으로 확대되었기 때문이라고 보았다.(Hans Jonas, 1987/2014, p. 91) 과거 순수 이론으로서 자연의 법칙이나 진리 탐구의 경우에는 실험 또는 논리적 사고 과정을 통해 과학적 흥미와 가치중립적 태도를 바탕으로 연구에 몰입해도 그 연구 결과가 곧 사회적 변화를 일으킬 만큼 거대하지 않았다. 하지만 자연과학이 기술과 융합되면서 과학의 영향력은 점차 일상으로 확대되고 그로 인해 사회적 변화와 경제성장이 일어나면서 그에 따른 새로운 문제점이 발생하게 되었다. 물론 과학기술의 발전으로 인간의 삶의 질이 향상되고 사회적 발전을 도모하게 되었으나 환경문제, 인간소외문제, 계층 간 소득격차문제 등 과거에는 크게 문제가 되지 않았던 주제들이 쟁점화되었다. 이는 곧 오늘날 과학기술과 사회 사이에 밀접한 관련이 있음을 시사하고 개발자·제조사는 연구의 자유와 사회적 공익 사이에서 신중한 태도를 보여야 할 것을 강조한다.

그리고 인공지능 개발자와 제조사는 예방과 보호, 보존의 가치를 중심으로 인공지능기술을 다루어야 한다. 한스 요나스는 과학기술 발전을 과학적 진보로 보고 과학적 진보의 이면에 주목하였다.(Hans Jonas, 1984/1994, p. 276) 한스 요나스가 말하는 진보란, 좀 더 나은 것, 적어도 더 바람직한 것으로 도약하는 것으로서 한스 요나스는 진보를 개인의 도덕적 진보와 사회의 과학적 진보로 나누었다.(Hans Jonas, 1984/1994, p. 277) 사회의 과학적 진보는 개인의 도덕적 진보보다 그 영향력이나 사회적 변화가 뚜렷하게 나타난다. 하지만 이러한 과학적 진보에는 언제나 포기하게 되는 비용이 발생한다. 이때의 비용은 단순히 돈을 의미하는 것이 아니라 과학적 진보로 발생하게 되는 환경 파괴, 생명경시풍조 등 과학기술의 발달로 인한 여러 사회적 문제를 의미한다. 그렇다고 과학적 진보 자체를 거부하기에는 이미 생활 속에 자리 잡은 기술력의 크기가 상당해서 과거의 생활로 돌아가기는 불가능하다. 따라서 이러한 사회적 문제를 최소화하기 위해 개발자와 제조사는 인공지능과 관련된 사고를 방지하기 위한 예방 측면에 더욱 관심을 기울여야 할 것이다. 그리고 개발자와 제조사는 인공지능의 판단은 어디까지나 인간의 이성적인 판단을 보조하는 수단으로서 수행하고, 최종 결정권은 인간에게 있다는 것을 유념해야 한다. 또한 인공지능기술을 인간의 편의제공을 위해서만 사용하는 것이 아니라, 수질 관리, 멸종위기종 관리, 공해유발 원인을 차단하는 등 다른 생명과 환경을 보존

하고 관리하는 데에도 유용하게 쓰일 수 있도록 인공지능기술의 개발 방향을 다른 분야로 확대할 필요가 있다.

나. 공적주체의 책임

공적주체의 책임은 한스 요나스가 말하는 정치인의 책임과 밀접한 관련이 있다. 정치인의 책임은 미래를 고려한 책임이라고 할 수 있다. 오늘날의 공적주체는 정책을 수립하는 데 과거에 비해 많은 요소를 고려해야 한다. 이러한 요소들을 분석하기 위해 자료(Data)가 필요한데, 이를 바탕으로 경제성장률, 인구변화, 일자리 전망, 주요 정책수립을 통해 미래를 대비한 국가 운영에 힘쓰고 있다. 그러나 현재 시점에서 미래를 예측한다는 것은 꽤나 조심스러운 일이다. 과학기술이 발전하고 사회 모습이 다양해짐에 따라 앞으로의 모습을 단순히 예측할 수 없게 된 것이다. 이렇게 ‘변화’ 자체가 일상이 된 현재의 모습 속에서, 정책의 방향은 현재 발생하고 있는 문제의 원인을 분석하고 해결하는데 초점을 맞추는 것처럼 보인다. 이러한 정책의 방향은 오늘날 살아가고 있는 국민의 요구와 사회적 여론을 적극적으로 반영한다고 볼 수 있다. 인공지능관련 정책 또한 이러한 상황 속에서 제안되는 정책이라고 할 수 있다. 데이터 수집·활용 기반 구축, 5G 등 네트워크 고도화, 인공지능 인프라 확충 및 융합 확산 등의 ‘디지털인프라 구축’, 그리고 ‘비대면 산업 육성’, ‘SOC 디지털화’의 영역을 중심으로 한 국내의 정책⁹⁷⁾ 추진 배경이 코로나19의 확산에 따른 경제활동 감소, 변화된 일상의 모습과 관련이 있다는 것을 예로 들 수 있다.

인공지능기술의 특성상, 그 발전 방향을 예측할 수 없으며 지속적인 변화가 발생할 것이라는 점에서 공적주체의 역할은 더욱 중요해지고 있다. 그러나 현재 나타나는 문제점과 여론에 초점을 맞추어 변화하는 시대에 공적주체가 서둘러 따라가면 안 될 것이다. 오히려 인공지능기술을 활용하여 사회적 변화에 대한 공적주체의 깊이 있는 분석과 통찰력을 통해 변화하는 시대를 이끌어야 한다. 한스 요나스는 현재 우리가 전현대적 선조들보다 미래에 관해 더 많이 알고 있으면서도, 동시에 조금밖에 모른다고 하였다.(Hans Jonas, 1984/1994, p. 210) 과거에 비해 엄청난 양의 분석자료와 지식을 토대로 미래의 모습

97) <https://terms.naver.com>. (검색일: 2020. 6. 10.)

을 구체적으로 전망하는 것이 통계적으로 가능하지만, 동시에 예기치 못한 일 또는 새로운 현안이 언제 어디서 발생할지 모르기 때문에 ‘변화’ 그 자체를 하나의 사회적 현상으로 받아들여야 한다는 것이다. 그렇다고 해서 변화를 거부해서는 안 되고 정책을 수립하면서 새로운 변화 요소를 언제나 고려해야 한다. 여기서 새로운 것은 그전에 없었던 것뿐만 아니라 그동안 정책 수립과정에서 놓쳤던 부분들, 예를 들어 사회적 약자의 목소리, 시민단체의 요구, 환경문제, 범지구적 차원의 캠페인 등을 포함한 세부적인 요소들이다.

변화하는 시대에 걸맞은 정책수립뿐만 아니라 인공지능기술과 관련된 법률의 수정 또는 보완도 공적주체가 책임져야 할 사항이다. 일반적으로 인공지능기술과 관련하여 발생한 사고에 관해서는 법률에 근거하여 처벌 또는 보상이 이루어지고 있다. 이때의 인공지능기술은 좁은 인공지능으로 인간의 판단을 보조하는 역할에 불과하다. 그렇지만 인공지능기술의 쓰임의 형태가 다양해지고 복잡해짐에 따라 선불리 누구의 책임이라고 판단하기가 어렵게 되었다. 게다가 머지않은 미래에 인간 수준의 인공지능은 출현할 것이라고 대부분의 미래학자는 전망하고 있다.⁹⁸⁾ 그렇기 때문에 인공지능, 로봇의 법적 지위와 관련된 법률에 대한 논의는 불가피하다. 이러한 상황 속에서 법률을 제정하는 공적주체의 책임은 더욱더 무거워진다. 공적주체의 입장에서 인공지능기술 개발을 저해하지 않으면서 존엄성, 자유, 권리, 의무 등 인류의 가치가 훼손되지 않는 방향에서 법률의 보완은 필요하다고 본다. 한스 요나스라면 인공지능기술로 인해 발생할 수 있는 쟁점사항들을 분석하여 법률로서 제조사와 사용자의 권리를 보호하고 개인 또는 기업의 이익보장보다는 인공지능기술에 대한 합리적인 규제를 주장할 것이다.

다. 사용자의 책임

사용자는 인공지능기술을 직접 사용하고 평가하는 주체이다. 인공지능기술에 대한 이들의 수요가 있기 때문에 인공지능기술 발전은 멈추지 않고 계속 진행된다. 좀 더 정확하고, 안전하고, 믿을 수 있는 인공지능기술을 개발하기 위한 원동력이 바로 사용자라고 할 수 있다. 이처럼 인공지능기술 개발에 있어 사용

98) Yuval Noah Harari et al. (2020). **초예측**(오노 가즈모토 역). 웅진지식하우스, p. 95

자의 안전과 편의를 보장하는 방향으로 진행되고 있다면 그에 따라 사용자도 인공지능기술을 안전하고 건전하게 사용해야 할 것이다. 하지만 그렇지 못한 사례가 종종 언론을 통해 보도된다. 자율주행자동차의 경우, 자율주행 중에 발생한 사고 대부분은 사람의 과실로 밝혀졌다.⁹⁹⁾ 그리고 인공지능기술을 악용하여 데이터를 조작하거나 가짜뉴스를 퍼트리고, 여론을 조작하는 등의 부정적인 사례들도 등장한다. 더불어 인공지능기술에 사용되는 데이터가 편향하거나 부적절하여, 개인의 이익을 위해 인공지능의 판단 자체가 편향성을 띠거나 어느 집단에 유리하게 조작하는 예도 확인할 수 있다.

이러한 인공지능기술 사용자의 잘못된 사용으로 비춰보았을 때, 일상생활에서 인공지능기술을 사용하는 주체로서 절제의 미덕이 필요하다. 현대 사회는 인공지능기술을 사용하도록 부추기고 있다고 봐도 무방하다. 이미 대부분의 가전제품에 인공지능기술이 탑재되었으며, 국가차원에서 금융, 제조, 의료, 국방 등 분야 대부분에 인공지능기술을 적용한 새로운 인프라 구축을 제시하고 있다. 그뿐만 아니라 심층학습, 사물인식 등 인공지능의 주요 기술이 대중에게 공개되어 인공지능기술을 취급하고 활용하는 것에 대한 진입장벽이 낮아졌다. 이러한 분위기 속에서 단지 개인의 이익을 위해 인공지능기술을 무비판적으로 쓰는 것은 지양해야 한다. 한스 요나스는 일찍이 대규모의 기술 활용에 있어 어떤 일을 일단 감행해 놓고 보자는 식은 절대로 용납할 수 없다고 하였다.(Hans Jonas, 1987/2014, p. 68) 안일한 생각으로 어떤 일을 저질러놓고 부적절한 결과가 초래하게 되는 것은 과거의 사례를 통해 확인할 수 있다. 환경오염, 핵무기개발, 생태계파괴 등 인간의 욕심과 태만으로 인해 인간뿐만 아니라 지구적 차원의 위험을 초래하였다. 그렇다고 이러한 일이 기술 개발자, 특정 집단, 대규모의 연구에 의해서만 발생하는 것은 아니다. 현재 시점에서 사용자에게 의해 언제 어디서 인공지능기술로 인한 문제가 발생해도 전혀 이상하지 않다. 그렇기 때문에 사용자로서 절제가 필요하다. 인공지능기술이 탑재된 제품을 사용할 때 알맞은 방식과 주의사항을 지켜 현명한 소비자로 거듭나야 한다.

99) "자율주행차 사고 38건 중 37건은 사람 과실". <https://www.nocutnews.co.kr>. (검색일: 2020. 6. 10.)

사용자를 평가자로 인식하여 바라볼 때 사용자는 인공지능기술 개발과 사용이 적절하게 이루어지고 있는지 꾸준한 관심을 두고, 정부의 인공지능 정책수립에 의견을 제시해야 한다. 한스 요나스는 시대가 암울하기 때문에 인간은 더욱더 도덕 이상의 것을 요구하고 있다고 하였다.(Hans Jonas, 1987/2014, p. 58) 인간성의 상실과 이기주의의 만연, 물질만능주의적 사고방식은 인공지능기술을 활용하는 데 있어 특히 우려되는 부분이다. 한스 요나스의 주장은 사용자뿐만 아니라 개발자, 기업, 정당에도 해당되는 부분으로, 인공지능기술을 사용하는 데 있어서 특정한 이익과 목적으로 쓰이는 것, 특히 생명의 존엄성과 인류에게 해가 되는 경우를 경계해야 함을 강조한다. 이를 위해 사용자는 소외되는 계층 없이 두루 사용하고 인공지능기술이 안전하면서도 미래지향적인 기술로 자리매김할 수 있도록 지속적인 관심과 주의를 기울여야 한다. 오늘날 변화의 주체는 어느 한 집단이 아니라 대중이기 때문에 변화의 주체로서 사용자는 그 역할을 성실하게 수행할 필요가 있다.

라. 교육주체의 책임

디지털 네이티브(Digital Native) 라는 말이 있다. 이 용어는 1979년 이후에 태어난 사람들을 가리키는 용어로, 개인용 컴퓨터, 태블릿, 휴대전화로 이루어진 디지털 세상에서 성장한 첫 세대를 말한다.¹⁰⁰⁾ 이들은 현대기술이 적용된 제품들을 특별한 것으로 여기지 않고, 그저 자신이 살아가는 환경의 일부로 여긴다. 그들에게 과학기술은 이미 삶의 깊숙이 들어와서 과학기술이 없는 세상은 상상할 수 없다. 머지않아 인공지능기술도 삶의 일부로 자리 잡아 인공지능 기술 없이는 어떤 일을 수행할 수 없을지도 모른다. 또한 오늘날의 세상은 끊임없이 변화하고 있다. 이러한 변화는 기술을 시작으로 집, 학교, 사회, 가치관의 변화를 초래하는데 이렇게 변화하는 시대 속에서 교육은 어떤 가치에 주목해야 하는가?

한스 요나스는 ‘변화’ 자체를 시대적 특징으로 인정하고 예측 불가능한 미래에 대해서도 인류는 책임을 져야 한다고 강조하였다. 미래의 모습은 어떨지 그 누구도 정확하게 예견할 수 없지만, 여기에서 한 가지 확실한 것은 “미래는 불

100) John Couch·Jason Towne. (2020). **교실이 없는 시대가 온다**. 어크로스. p. 33

확실하다.”라는 사실이다. 한스 요나스는 이를 “불확실성의 확실성”이라고 표현하였다.(Hans Jonas, 1984/1994, p. 206) 인공지능기술뿐만 아니라 생명공학, 유전공학, 나노기술 등의 현대 과학기술로 인해 앞으로 어떤 일이 일어나고 인류에 어떤 영향을 미칠지 아무도 알 수 없다고 볼 수 있다. 과학기술의 발달이 인류의 번영과 평화를 이끌기도 하지만 그로 인한 문제점과 갈등이 세계 곳곳에서 새롭게 발생하고 있다는 점을 고려하면 그렇다. 때문에 교육주체는 이러한 변화 속에서도 변하지 않는 가치를 추구할 필요가 있다. 이러한 가치로는 인류애, 생명존중, 존엄성 등의 도덕적 가치로, ‘인류’를 위한 감정을 일깨우고 길러주는 것이 교육적·지성적 과제라 할 수 있다. 인공지능기술의 발달과 교육현장의 디지털화로 인해 교육의 형태가 달라진다고 하더라도 ‘가치형성의 조력자’로서의 교육주체 역할이 변해서는 안 될 것이다.

인류의 문제는 위로부터가 아니라 아래로부터의 실천을 통해 해결되어야 한다고 한스 요나스는 주장하였다.(Hans Jonas, 1987/2014, p. 72) 데이터 문제, 안전성 문제, 보안 문제, 딥페이크, 범용 인공지능의 출현 등 인공지능기술의 발달 자체로 논란이 되고 있음에도 불구하고 사회는 인공지능기술의 확대 적용을 요구하고 있다. 인공지능기술을 개발하고 보급하는 데 정부와 기업체의 역할이 크지만, 그들은 대체로 사회적 요구와 대중의 인식을 바탕으로 움직인다. 인공지능에 대한 대중의 올바른 인식과 가치관을 확립하기 위해서는 무엇보다도 교육주체의 역할이 중요하다. 이는 현재 존재하는 세대만을 위해서가 아닌 미래의 후손과 생명을 위한 첫걸음이라 할 수 있다. 이를 위해 교육주체는 지식의 전수만으로는 개인의 욕구와 이상을 모두 충족할 수는 없다는 것을 인정해야 한다. 저마다 하고 싶은 것과 원하는 것이 다르고, 이를 이루려는 방법 또한 다양함을 인정해야 한다는 것이다. 또한 교육적으로 최고를 추구하는 것에 대해서도 다시 생각해봐야 한다. 최고를 추구하기보다는 저마다의 삶의 가치를 발견하고 이를 가꾸어 나가는 데 주력해야 한다. 이는 평생교육의 관점에서도 유효하다고 볼 수 있다. 인간의 일생과정에서 어떤 일이 일어나고 무엇을 할지는 정해지지 않았을 뿐만 아니라 아무도 모르기 때문에 교육주체는 인간의 가치를 지속적으로 탐구하고 가꾸는 데 노력해야 할 것이다.

V. 결론 및 제언

한스 요나스의 책임이론은 이미 발생한 사고나 죄에 대한 처벌 또는 보상 차원의 법률적 책임을 이야기하는 것이 아니다. 한스 요나스는 사고가 발생하지 않도록 미리 방지하는 차원에서 우리가 마땅히 해야 하는 일에 대한 책임을 논하는 것이다. 여기서 우리는 누군가의 부모, 아들, 딸처럼 가족의 구성원일 수 있다. 혹은 교사, 회사원, 의사, 정치인처럼 사회의 구성원일 수 있다. 이때 개인의 위치와 역할에 따라 책임져야 할 대상이 다를 수도 있지만 각자 그 대상을 대하는 마음가짐은 동일하게 작용한다. 필자는 이러한 책임의 감정을 인공지능기술로 확대해보았다. 인공지능기술이 사회의 제반 영역으로 확대되면서 나타나는 사회적 변화와 윤리적 쟁점은 우리가 변화 속에서 할 수 있는 일, 또는 우리가 해야 하는 일이 무엇인지 고민하게 한다. 고민의 과정을 거쳐 어떤 결정을 내리게 되었을 때, 우리는 우리 스스로 그 결정에 대해 책임을 질 수 있어야 한다. 이러한 책임의 대상은 인류뿐만 아니라 자연환경, 생태계, 미래 등의 모든 영역이라고 할 수 있다. 따라서 한스 요나스가 인간의 책임을 강조한 내용을 참고하여 필자는 인공지능기술에 대한 인간의 윤리적 책임을 아래와 같이 몇 가지로 정리해보았다.

첫째, 인공지능기술 개발자는 기술의 영향력이 사회의 제반 분야로 확대되었음을 인지하고 신중한 태도로 인공지능기술을 개발해야 한다. 이는 로봇윤리를 논할 때부터 제기되어 온 근본적이면서 핵심적인 주장이다. 기술의 발전은 일반적으로 생활의 편리함과 효율성을 추구하는 것을 목적으로 하지만, 인간이 그것을 과도하게 추구하거나 기술을 악용해서 환경오염, 생태계 파괴, 군사적 충돌 같은 문제점이 발생하기도 한다. 특히 인공지능기술은 군사용 로봇, 수송용 로봇, 의료형 로봇 등 사람의 생명과 직·간접적으로 관련된 분야에서 활용되기 시작하였다. 이에 따라 로봇의 사용과 관련된 새로운 윤리적 쟁점이 발생하게 되었다. 이 윤리적 쟁점 중에서도 특히 사람의 생명과 관련된 윤리적 쟁점은 당분간 사회적 이슈로 자리매김하여 인공지능 연구자들의 신중한 접근을 요구할 것이다. 그렇다고 해서 인공지능기술에 대한 윤리적 접근이 인공지능기술 개발 과정에 걸림돌이 되어서는 안 된다. 이미 생활 깊숙이 자리 잡은 인공지능기술이기 때문에 개발을

중단하는 것은 현실적으로 불가능하다. 따라서 인공지능 기술 개발과 규제 사이의 적절한 균형을 추구해야 할 것이다. 이러한 접근이 한스 요나스가 말하는 연구자의 책임 있는 자세이고 연구자라면 마땅히 갖추어야 할 덕목이라고 할 수 있다.

둘째, 입법자를 포함한 정치인은 사회적 약자, 취약계층을 포함한 사회 구성원 모두가 인공지능기술을 쉽게 접하고 사용할 수 있도록 인프라를 구축해야 한다. 신기술의 도입 초기에는 기술의 혜택을 받는 사람과 받지 못하는 사람 사이의 격차가 발생하곤 한다. 19세기 초반 노동자 계급의 저항운동인 러다이트 운동(Luddite)은 공장의 기계 도입에 저항하는 대표적인 사례라 할 수 있다.¹⁰¹⁾ 이러한 유형의 사회적 운동은 ‘불평등’과 ‘불공정한 대우’에 대한 시민의 목소리라 할 수 있다. 인공지능기술이 사회 곳곳에 도입된 지금, 직접적인 표출이 아니더라도 러다이트 운동과 비슷한 현상이 곳곳에서 발생하고 있다. 인공지능기술로 인해 사라지는 일자리가 생기고 있으며 새로운 역량을 요구하는 신직업이 발생함에 따라 취업을 준비하는 데 있어 새로운 준비와 경험을 필요로 하는 것이 그 예라 할 수 있다. 이를 해소하기 위해 정부차원에서 인공지능기술 관련 교육을 지원하고 세대 간 기술 격차를 해소하기 위한 구체적인 방안을 제시할 필요가 있다.

셋째, 인공지능기술 사용자는 인공지능기술을 오용 또는 악용하지 않도록 절제의 미덕을 갖추어야 한다. 오늘날 발생하고 있는 인공지능기술 관련 사고 대부분은 인간의 과오나 그릇된 사용, 부주의로 인한 사고이다. 기술의 사용 주체는 최종적으로 사용자인기 때문에 기술 자체보다는 사용자에게 윤리적 책임을 강조해야 한다. 사용자는 인공지능기술이 탑재된 제품을 사용할 때 알맞은 방식과 주의사항을 지켜 현명한 소비자로 거듭날 수 있어야 한다. 또한 인공지능기술을 개발하고 사용하는 것이 적절하게 이루어지고 있는지 비판적으로 살펴보고 정부와 제조사에 사용자 의견을 제시해야 한다. 인공지능기술에 무턱대고 의존하게 되면 우리가 힘들다고 여기거나 노동력이 필요한 자리를 인공지능 로봇으로 대체할 가능성이 있다. 그렇게 되면 인간적 교류의 기회마저 잃을 수 있고 잘못된 사용으로 인해 사고가 발생할 수 있는 위험이 크다. 한스 요나스의 관점에서 인공지능

101) 러다이트 운동. <https://ko.wikipedia.org>. (검색일: 2020. 4. 4.)

기술의 확대는 우리 인간이 영향을 미치는 범위가 확대되는 것과 동일한 것으로 본다. 이러한 관점에서 사용자는 기술의 영향이 미치는 범위에 대해 소외되는 계층이 없이 올바른 방식으로 인공지능기술이 사용되는지 면밀히 따져봐야 한다.

넷째, 교육주체는 윤리적 차원의 인공지능 교육을 강화해야 한다. 현재 국내에서는 국가경쟁력 확보 및 미래 사회의 핵심역량 구축을 목표로 소프트웨어 교육을 강화하고 있다.¹⁰²⁾ 이러한 취지에서 코딩(Coding)이 정규 교육과정으로 편성되어 초·중·고 단위로 교육이 이루어지고 있다. 코딩 교육은 인공지능기술을 프로그램 형태로 다루는 데 필요한 작업능력이나 언어를 이해하는 데 필요한 교육이다. 코딩 작업을 하면서 학생들에게 논리력, 창의력, 문제해결력, 정보처리능력 등의 역량향상을 기대할 수 있다. 이와 더불어 디지털 리터러시(Digital Literacy) 교육이 필요하다. 디지털 리터러시 교육은 정의가 다양하지만 대체로 데이터를 다루는 단순한 기능과 능력을 넘어 데이터를 취급하는 태도 및 가치관 등의 역량을 기르는 관점의 교육이라 알려져 있다.¹⁰³⁾ 인공지능기술을 다루는 능력도 중요하지만, 디지털 리터러시 교육을 통해 인공지능기술로 인해 발생하는 수많은 윤리적 쟁점에 대해서 인공지능기술을 다루는 주체의 입장이 되어 생각을 표현할 기회를 제공할 필요가 있다. 이를 바탕으로 학생들이 인공지능기술에 대해 무조건적인 수용보다 비판적으로 바라보며 현재와 미래를 아우르는 미래지향적 사고를 갖추는 데 도움이 되어야 한다. 윤리적 판단력을 기초로 한 미래지향적 사고는 인공지능기술이 미래에도 안정적으로 정착될 수 있게 해주는 책임의 한 방식이라 할 수 있다.

본 연구에서 다룬 책임은 거시적 관점의 책임으로 전반적인 사회적 변화 추세에 대한 대응이라 할 수 있다. 그렇기 때문에 포괄적이고 추상적인 주장처럼 보일 수 있다. 이를 보완하기 위해 현실 속에서 인공지능기술을 활용한 사례와 쟁점들을 접목하고자 하였다. 그럼에도 불구하고 인공지능기술의 다양한 적용 범위를 고려한다면 책임윤리 또한 산업 분야별로, 인공지능기술을 다루는 주체별로 접근하여 책임윤리의 확립방안을 연구할 필요가 있다.

102) ‘소프트웨어’ 교육은 ‘코딩’교육일까?. <https://www.eduinnews.co.kr>. (검색일: 2020. 4. 5.)

103) 디지털 리터러시. <https://ko.wikipedia.org>. (검색일: 2020. 6. 26.)

참 고 문 헌

단행본

- 이희승 (편). (1994). **국어대사전**(3판). 민중서림.
- 편집부 (편). (1998). **한한대사전**(증보판). 민중서림.
- 변순용·송선영. (2015). **로봇윤리란 무엇인가?**. 어문학사.
- 김효은. (2019). **인공지능과 윤리**. 커뮤니케이션북스(주).
- 변순용 편. (2019). **윤리적 AI로봇 프로젝트**. 어문학사.
- 뉴 사이언티스트 외. (2018). **기계는 어떻게 생각하고 학습하는가**(김정민 역). 한빛미디어(주)
- Hans Jonas. (1994). **책임의 원칙:기술 시대의 생태학적 윤리**(*Das Prinzip Verantwortung. Versuch einer Ethik Für die technologische Zivilisation*)(이진우 역). 서광사. (원저 1984 출판)
- Rafael Capurro·Michael Nagenborg. (2013). **로봇윤리: 로봇의 윤리적 문제들**(변순용·송선영 역). 어문학사.
- Hans Jonas. (2014). **기술 의학 윤리**(*Technik, Medizin und Ethik: Praxis des Prinzips Verantwortung*)(이유태 역). 솔출판사. (원저 1987 출판)
- Wendell Wallach·Colin Allen. (2016). **왜 로봇의 도덕인가**(노태복 역). 메디치. (원저 2009 출판)
- James Barrat. (2016). **파이널 인벤션**(정지훈 역). 동아시아. (원저 2013 출판)
- Martin Ford. (2019). **AI 마인드**(김대영 외 역). 터닝 포인트.
- Yuval Noah Harari et al. (2020). **초예측**(오노 가즈모토 역). 웅진지식하우스.
- John Couch·Jason Towne. (2020). **교실이 없는 시대가 온다**. 어크로스.

논문

- 양해림. (2000). 생태계의 위기와 책임윤리의 도전-한스 요나스의 책임개념을 중심으로-. **철학**, 65, 241-267.
- 이윤택. (2005). 요나스의 미래윤리와 책임. **동서철학연구**, 36(36), 113-137.
- 김종국. (2005). 미래와 힘, 한스 요나스의 '책임의 원칙' 그 이후. **사회와 철학**, (10), 89-104.
- 양천수. (2009). 책임구상의 다양한 지평-한스 요나스의 책임원칙을 예로 하여 -. **안암법학**, 30(30), 147-181.
- 김은철·송성수. (2012). 과학기술시대의 책임윤리를 찾아서: 한스 요나스의 '책임의 원칙'을 중심으로. **공학교육연구**, 15(1), 72-78.
- 고인석. (2014). 로봇윤리의 기본 원칙: 로봇 존재론으로부터. **범한철학**, 75(4), 401-426.
- 신현우. (2015). 도덕과 교육에서 로봇윤리의 교육적 함의. **윤리연구**, 105(105), 269-286.
- 손화철. (2016). 인공지능시대 인간의 자리. **윤리학**, 5(2), 55-80.
- 송선영. (2017). 인공지능 로봇과 인간의 관계에 대한 윤리적 성찰과 전망. **영어권문화연구**, 10(3), 61-83.
- 변순용. (2018). 인공지능로봇을 위한 윤리 가이드라인 연구: 인공지능로봇윤리의 4원칙을 중심으로. **윤리교육연구**, 47(47), 233-252.
- 유은순·조미라. (2018). 포스트휴먼 시대의 로봇과 인간의 윤리. **한국콘텐츠학회논문지**, 18(3), 592-600.
- 정결. (2018). 과학기술시대의 새로운 윤리학의 필요성에 대한 연구 -거시 윤리학의 관점을 중심으로-. **인문학연구**, 112(3), 87-113.
- 정채연. (2018). 지능정보사회에서 지능로봇의 윤리화 과제와 전망-근대적 윤리 담론에 대한 대안적 접근을 중심으로-. **전북대학교 동북아법연구**, 12(1), 87-121.
- 이영록. (2019). 책임의 의미와 성격에 관한 역사적 탐색. **일감법학**, (42), 195-221.

- Gianmarco Veruggio, Fiorella Operto. (2006). Robotics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics. *International Review of Information Ethics*, 6, 2-8.
- Nick Bostrom·Eliezer Yudkowsky. (2011). The Ethics of Artificial Intelligence. *Cambridge University Press*, 1, 1-20.
- Anna Jobin·Marcello Ienca·Effy Vayena. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399.

보고서

- 조영임. (2016). 인공지능기술 동향 및 발전 방향. **정보통신기술진흥센터**. 13-26.
- 대통령직속 4차산업혁명위원회. (2018). 4차산업혁명시대 산업별 인공지능윤리의 이슈 분석 및 정책적 대응방안 연구. **정보통신정책연구원**.
- 추형석. (2018). 범용 인공지능의 개념과 연구 현황. **소프트웨어정책연구소 이슈리포트**, (2), 1-16.
- Gianmarco Veruggio. (2006). EURON Roboethics Roadmap. *Restricted Circulation*, 1(1), 1-42.

기타자료

- 유신. (2012. 8. 14). 강한 인공지능, 약한 인공지능. **사이언스온**. 2020. 3. 7, <http://scienceon.hani.co.kr/media/49829>
- 류현정. (2017. 1. 16.) EU, AI 로봇에 '전자인간' 지위 부여... "무조건 인간에 복종하도록 설계해야". **조선비즈**. 2020. 5. 3. https://biz.chosun.com/site/data/html_dir/2017/01/16/2017011601845.html.
- 김수연. (2017. 4. 17). [알아봅시다] 인공지능(AI)의 종류. **디지털타임스**. 2020. 3. 7, http://www.dt.co.kr/contents.html?article_no=2017041802101831102001

- 엄민재. (2018. 7. 8). [취재파일] 가짜뉴스보다 더 무서운 가짜 동영상 '딥페이크'. **SBS NEWS**. 2020. 4. 4, https://news.sbs.co.kr/news/endPage.do?news_id=N1004837041
- 김지혜. (2018. 10. 25). 인공지능 혁명 핵심은 '데이터 과학'. **전자신문**. 2020. 3. 7, <https://m.etnews.com/20181025000097?obj=Tzo4OiJzdGRDbGFzcyI6Mjp7czo3OiJyZWZlcmVyIjtOO3M6NzoiZm9yd2FyZCI7czoxMzoid2ViIHRvIG1vYmlsZSI7fQ%3D%3D>
- 김미림. (2018. 11. 6). '웰다잉 산업'이 뜬다. **넥스트이코노미**. 2020. 3. 20, <http://www.nexteconomy.co.kr/news/articleView.html?idxno=11802>
- 정하늘. (2019. 5. 31). '소프트웨어'교육은 '코딩'교육일까?. **에듀인뉴스**. 2020. 4. 5, <https://www.eduinnews.co.kr/news/articleView.html?idxno=14884>
- 문가용. (2019. 7. 6). [주말판] 인공지능, 고민을 멈추고 직접 해봐야 할 때. **보안뉴스**. 2020. 3. 7, <https://www.boannews.com/media/view.asp?idx=81202>
- 김선영. (2019. 8. 29). [약 이야기]집중력 높이고 성적 올리려 ADHD약 먹는다?. **중앙일보**. 2020. 2. 13, https://jhealthmedia.joins.com/article/article_view.asp?pno=20810
- 구본권. (2019. 10. 3). 딥페이크 탐지기술, 가짜뉴스 해결사 될까?. **한겨레**. 2020. 4. 4, <http://www.hani.co.kr/arti/science/future/911878.html>
- 이창균. (2019. 10. 5). [이코노미스트] 자율주행차 과연 안전할까 - 인간의 행동에도 충분히 대처해야. **중앙일보**. 2020. 3. 8, <https://news.joins.com/article/23595546>
- 구본권. (2019. 10. 27). 편견·차별 부르는 AI 알고리즘...해법은 있을까. **미래&과학**. 2020. 3. 8, <http://www.hani.co.kr/arti/PRINT/914757.html>
- 이재영. (2020. 1. 5.). '인공지능시대' 초등부터 단계별 'AI교육 기본계획' 만든다. **연합뉴스**. 2020. 6. 10. <https://www.yna.co.kr/view/AKR20200103105500004>
- 조규희. (2020. 2. 2.). 쉼자율주행차 상용화... "사고는 '너' 책임". **신동아**. 2020. 6. 10. <https://shindonga.donga.com/3/home/13/1965704/1>.

- 이성규. (2020. 3. 19). 코로나와 AI 스타트업. **AI타임스**. 2020. 4. 4, <http://www.aitimes.com/news/articleView.html?idxno=126621>
- 김민수. (2018. 8. 30.). "자율주행차 사고 38건 중 37건은 사람 과실". **노컷뉴스**. 2020. 6. 10. <https://www.nocutnews.co.kr/news/5023954>.
- Kim Namkug. (2017. 5. 15). 인공지능이 블랙박스에 대한 기술적 답변. **Kim, Namkug's Ideas**. 2020. 2. 14, <https://namkugkim.wordpress.com/2017/05/15>
- Thomas Macaulay. (2018. 5. 14.). 컴퓨터의 창의력, 인류에게 어떤 의미일까?. **CIO Korea**. 2020. 7. 4. <http://www.ciokorea.com/news/38232>
- David. (2019. 6. 3). 초지능 인공지능 대안이론 '리프레이밍 슈퍼인텔리전스'. **The Science Monitor**. 2020. 2. 17, <http://scimonitors.com>
- James Kobielus. (2020. 1. 6). 투명성이 만병통치약?... 블랙박스 AI의 위험을 완화하는 방법. **CIO Korea**. 2020. 3. 8, <http://www.ciokorea.com/news/140790>
- '유진 구스트만', 단순한 채팅 로봇인가, 아니면 '생각하는' 인공지능인가?. (n.d.). 2020. 3. 11, <http://www.itworld.co.kr/print/87899>
- [AI] '무어의 법칙'보다 7배 빠르다...인공지능의 질주. (n.d.). 2020. 4. 4, <https://blog.naver.com/nopil02/221755474364>

A B S T R A C T *

A Study on Artificial Intelligence Technique and Human Responsibility

- focused on Hans Jonas's Perspective -

Kim, Do Yun

Major in Elementary Ethics Education
Graduate School of Education
Jeju National University

Supervised by Professor Jang, Seung Hee

For us today, artificial intelligence technique is no longer an unfamiliar technique that is out of touch with our daily lives. The artificial intelligence secretary, which can be accessed through a smartphone, not only informs users of information such as weather, news, video and schedule according to their tastes, but also

* A thesis submitted to the committee of Graduate School of Education, Jeju National University in partial fulfillment of the requirement for the degree of Master of Education conferred in August, 2020.

communicates with users in a natural response to questions. And self-driving systems that are installed in cars also provide drivers with an optimized vehicle environment. Like this, artificial intelligence technique is being used in various forms in our daily lives.

However, artificial intelligence has not always brought everyday convenience. Artificial intelligence technique is causing new problems that have not existed in the past, such as safety problems raised by accidents involving self-driving cars, damage cases caused by unauthorized use of personal information or manipulation of information by exploiting artificial intelligence technique, and cases in which programs applying artificial intelligence technique produced biased results. What we should note here is that these problems are often caused by human error, wrong use, or carelessness rather than by flaws in artificial intelligence itself. Therefore, today's society demands the right responsibility for humans using artificial intelligence technique.

In response, I look at the contents focusing on ethical issues related to artificial intelligence technique, especially those related to "responsibility," and compare the responsibilities of legal and ethical meanings. After that, we will summarize several responsibilities for each actor who deals with artificial intelligence technique. To that end, he will analyze the theory of responsibility, focusing on Hans Jonas, and examine the issues caused by artificial intelligence technique from his point of view.

What we should pay attention to Hans Jonas is that Hans Jonas analyzed the characteristics of the problems caused by the development of science and technique and presented the technical and responsible officers to respond to them. He first compared the

characteristics of pre-modern science and technique, which had relatively no effect on human life, with the characteristics of post-modern science and technique that led to fundamental changes in the way of life. And the post-modern science and technique were summarized into two main characteristics. First, humans themselves have become targets of technique. With the development of genetics, biotechnology, and brain science, human physiology, psychology, and behavior have become subjects of science and technique, a change of perception of human beings has occurred. Second, modern technique has become the subject of ethics as humans have become the subject of technique. People have begun to pay attention not only to human ethical problems but also to the ethical use and influence of technique.

Through this change of perception, Hans Jonas demanded a corresponding responsibility, which differs from the responsibilities used in the law. In the legal sense, responsibility refers to causative responsibility resulting from the point of view of damages or punishment by carefully examining the causes, actors and motivations for any wrongdoing. But Hans Jonas insisted on the responsibilities required for an uncertain future rather than emphasizing the responsibilities of legal meaning. In other words, Hans Jonas's responsibilities are human responsibility for the existence itself, responsibility for what human beings deserve to do, and responsibility for future generations.

Meanwhile, ethical issues related to artificial intelligence technique are becoming complicated and subdivided by the development of artificial intelligence technique in the industrial sectors of medical, military, manufacturing and finance, and related actors. In particular, as the process of processing information in artificial

intelligence technique has become more sophisticated, recognizing things and solving problems are becoming similar to humans. This artificial intelligence technique is interpreted in light of Hans Jonas' point of view as follows.

First, among the major areas of artificial intelligence technique, the information processing process of artificial neural networks is becoming more complicated. Artificial neural networks have the characteristic that the more layers that make up them, the more accurate results they produce. This has made it difficult to analyze the information processing process of artificial intelligence technique. It has also become difficult for us to predict how one input, like AlphaGo, will be output through an artificial neural network.

Second, changes in industrial structure, changes in job structure, changes in legal system, changes in living environment are emerging due to artificial intelligence technique. Just as technique produces technique, it is inevitable to develop more efficient artificial intelligence technique to cope with these changes. Therefore, the development of artificial intelligence technique has become inevitable.

Third, scholars studying artificial intelligence technique argue that artificial intelligence technique should be used to clean up water quality, improve environmental problems and manage ecosystems, rather than just for human convenience and benefit. This shows that artificial intelligence technique should go beyond human life and be used on a global level.

Fourth, with the development of artificial intelligence technique, artificial intelligence robots are gradually becoming more and more similar to the human thought process. This raises metaphysical questions related to human consciousness, existence, from academia.

The subject that deals with these features of artificial intelligence technique is human at the moment. Therefore, in reference to Hans Jonas' emphasis on human responsibility, I have presented several ethical responsibilities for artificial intelligence technique.

First, the developer of artificial intelligence technique should develop artificial intelligence technique with a cautious attitude, recognizing that the influence of technique has expanded to various sectors of society. At this time, developers should set goals to preserve the environment, protect the ecosystem, prevent accidents and secure safety.

Second, politicians, including legislators, should build infrastructure so that all members of society, including the socially disadvantaged and the vulnerable, can easily access and use artificial intelligence technique. Also, laws and systems should be reorganized to reflect the changing social landscape of artificial intelligence technique.

Third, users should have the virtue of moderation to prevent misuse or abuse of artificial intelligence technique. It should also critically examine whether the development and use of artificial intelligence technique is being done properly, and present user opinions to governments and manufacturers.

Fourth, educational subjects such as educators and training institutions should strengthen artificial intelligence education on an ethical level as well as software education. And the subject of education should explore the unchangeable moral values amid the global changes caused by artificial intelligence technique, awakening and cultivating the moral sentiments of the subject being educated.

Keyword: Hans Jonas, Responsibility, Robot, Artificial Intelligence, Robot Ethics, Artificial Intelligence Ethics