

## 인공지능로봇의 의료의 도덕형이상학적 모색: 의료적 전자인간의 책임가능성\*

정창록\*\*

### 요약

본 논문의 목적은 인공지능(artificial intelligence, AI)로봇의 의료에서 전자인간 의사가 과연 책임질 수 있는지에 관해 2017년 유럽연합(European Union, EU)의 AI로봇에게 전자인간으로서의 법적 지위를 부여한 논의와 관련하여 다루어 보는 것이다. 과연 전자인간으로서 AI로봇 의사는 얼마만큼 책임질 수 있을 것인가? 동시에 권리를 주장하는 것은 가능할 것인가? 이와 관련하여, 필자는 동물과 인간의 책임가능성을 둘러싼 기존의 논의를 통해 AI로봇의 도덕형이상학적 지위 문제를 정리해 보고 있다. 그리고 본 논문은 책임의 전제조건으로서 칸트적 의미의 자유 개념에 관해서는 EU의 소위 '전자인간'에 관한 논의에 있어 '킬 스위치(kill switch)'와 관련하여 다룬다. 또한 의료에서 왓슨의 등장으로 인해 열린 새로운 지평에 관해 정리해 본다. 필자는 전자인간 의사의 등장으로 의료에서 '자가치료(self medical care)' 개념이 등장하였고, '보편적 자가의료(universal self medical care)'의 가능성이 열렸다고 본다. 그리고 이러한 상황에서 필자는 인간 의료인의 새로운 임무를 '킬 스위치(kill switcher)'로 전망한다.

### 색인어

인공지능의료, 전자인간 의사, 도덕적 책임, 킬 스위치, 자가치료, 보편적 자가의료

교신저자: 정창록, 경북대학교 사범대학 윤리교육과 · 가톨릭대학교 성심교정 법학과 외래교수. Tel: 053-950-5868, Fax: 053-950-5947, e-mail: canthos@hanmail.net

ORCID: <https://qa.orcid.org/0000-0001-6028-5595>

\* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5B5A07061510). 가천 길병원의 왓슨을 견학하도록 허락하여 본 논문을 작성하는 데 도움을 준 가천 길병원의 이 언 교수와 김영보 교수에게 감사드린다. 더불어 본 주제에 지속적인 관심을 가지고 논문 작업에 많은 도움을 주려 노력한 익명의 의사(Q)에게도 감사드린다.

\*\* 경북대학교 사범대학 윤리교육과 · 가톨릭대학교 성심교정 법학과 외래교수

www.kci.go.kr

## I. 시작하며

본 논문은 인공지능을 활용한 의료의 윤리와 법을 아우르는 규범체계에 대한 검토, 즉 인공지능(artificial intelligence, AI)로봇의료에 대한 ‘도덕형이상학적 모색’을 시도해 보려한다. 필자가 여기서 ‘도덕형이상학적 모색’이라고 하는 것은 칸트의 ‘도덕형이상학’에서 가져온 것이다. 칸트는 도덕형이상학에서 제1부 ‘덕론’과 제2부 ‘법론’을 나누고 있고 ‘덕론’은 자율적 특성을 가지는 도덕과 윤리에 관한 것, ‘법론’은 강제적 성격을 갖는 법에 관한 것이다. 이러한 모색의 시도에는 다음과 같은 상황이 영향을 미쳤다. 즉, 2017년 1월 12일 유럽연합(European Union, EU) 의회는 “로봇의 법적 지위는 ‘전자인간(electronic person)’”이라 천명하였다. 신문기사는 다음과 같이 전하고 있다. “유럽연합(EU)이 현지시간 2017년 1월 12일, 로봇의 법적 지위를 ‘전자인간’으로 인정하는 결의안을 통과시켰다. EU 의회는 이날 인공지능(AI) 프로그램에 따라 스스로 작동하는 로봇에 대해 ‘전자인간’이라는 법적 지위를 부여하는 안을 놓고 찬성 17표, 반대 2표, 기권 2표로 통과시켰다” [1]. 위 기사에 따르면 의회 조사위원인 매디 델보는 “우리 일상 속에서 로봇이 영향을 미치는 영역이 점차 증가하고 있다”며 “이러한 현실에 발맞춰 로봇이 인류에게 도움을 주는 존재임을 보다 명확히하기 위해 EU 내 탄탄한 법적 프레임을 만들어야 할 필요가 있다”고 밝혔다. 위 기사에 따르면 EU 의회는 조직 내 ‘EU 로봇국’을 신설하여 AI에 대한 기술적·윤리적 기준을 적용하는 역할을 수행할 수 있도록 할 것을 EU 집행위에 권고했으며, EU 의회는 로봇이 법률이나 안전, 윤리적 기준에 따라 작동할 수 있도록 제한하는 법률이 필요하다고 덧붙였다. 즉, 로봇 제작자들이 비상 상황에서 로봇의 작동을 멈출

수 있도록 하는 ‘킬 스위치(kill switch)’ 등을 법적으로 의무화해야 한다는 것이다.

이제 현실의 일상 생활에서 로봇을 보는 시기는 점점 더 가까워지고 있는 것 같다. 이렇게 인간 세상에서 로봇이 어떤 특정 행위를 하게 된다면, 책임질 수 있는 상황에서 로봇도 인간처럼 책임을 질 수 있을 것인가? 특히 인간의 생명을 다루기에 어떤 행위의 결과로 책임공방이 오가는 의료계에서는 AI로봇은 어떤 방식으로 책임 문제의 상황을 맞이할 것인가? 이러한 질문들에 답해보기 위해 필자는 인공지능에 관한 한국학계의 기존연구를 검토하고 책임의 개념을 간략히 정리해 볼 것이다. 이후 EU의 법적지위 부여한 소위 ‘전자인간’의 지위와 ‘킬 스위치’를 둘러싼 논란에 관해 살펴볼 것이다. 그리고, 의료에서 인공지능으로 말미암아 ‘자가치료(self medical care)’ 개념의 등장과 ‘보편적 자가의료(universal self medical care)’의 가능성에 관해 정리해 본 후에, 완전자율 의료 가능한 상황을 상정한 후 이러한 가정된 상황에서 과연 닥터 전자인간은 얼마나 책임질 수 있을지에 관해 고찰해 보고자 한다.

## II. 한국학계의 기존연구검토

KISS (Koreanstudies Information Service System)와 교보 스킨에서 키워드 ‘AI로봇윤리’로 검색한 논문들 중 의료계를 대상으로 한 연구는 아니지만, 필자는 다음과 같은 네 논문에 주목하고 싶다. 1) 변순용과 송성연[2]은 로봇윤리의 이론적 기초를 마련하기 위한 연구를 수행하였고, 2) 양종모[3]는 인공지능의 위험과 그 규제를 위한 일반적인 법적 규제방안을 마련하였다. 3) 이중기[4]는 AI로봇이 활용되는 자율주행차에 관해 연구하였으나 이를 의료계의 문제와 연관시키지는 않고 있다. 4) 류화진[5]은 지능형 로봇이 과

연 범죄주체성을 가질 수 있고 이때 형사책임을 어떻게 될 것인지에 관해 다루고 있다. 다음에 조금 자세히 내용을 정리해 보겠다.

1) 변순용과 송선영[2]은 ‘로봇 윤리의 이론적 기초를 위한 근본 과제연구’에서 이 연구의 목적은 로봇 제작 및 사용에 따라 인간 삶에 미치는 다양한 영향들에 관한 윤리적 성찰을 탐구하는 데 있다고 밝히고 있다. 현재 로봇 공학의 기술 발달에 따라 분야별로 전개되는 로봇의 제작사용의 양상은 매우 다양하게 나타나고 있고, 이로 인해 현재 서양 중심으로 전개되고 있는 로봇윤리 연구는 각 분야별 사례에 따라 개별적인 사례들 위주로 진행되고 있는 실정하기에, 이 연구에서 이러한 로봇윤리의 기본적 특징과 한계를 밝히고, 로봇윤리의 이론적 기초를 모색해보고, 향후 미래에 대한 책임윤리로서 로봇윤리의 본질을 전망하고 있다.

다음으로 2) 양종모[3]는 ‘인공지능의 위협의 특성과 법적 규제방안’에서 인간 수준의 로봇을 전제로 하는 로봇 윤리학적 측면의 실존 문제를 다루는 논의의 시급성을 다루고 있다. 이 논문에서는 현재 활용되고 있는 약한 의미의 인공지능의 특질을 제대로 분석하고, 그 개발과정이 기존의 기술에 기초한 개발과 어떠한 차이를 보이는지를 들여다본 후, 그런 기초 위에서 그 위협의 특성에 맞는 적절한 법적 규율 방안을 논하고 있다. 이 논문에서 양종모는 약한 인공지능에 국한하여 그 위협의 특성을 다양한 각도에서 들여다보고 있다. 그는 인공지능이나 그 개발과정의 특성을 고려하면 민·형사법적 책임 귀속을 통한 사후적 규율만으로 부족하다고 본다. 특히 그는 로봇 개발과 관련하여 사후적 규율 외에 개발 단계의 사전 규율이 필요하다고 보고 있다.

3) 2016년 이중기[4]는 ‘자율주행차의 발전단계로 본 운전자와 인공지능의 주의의무의 변화와

규범적 판단능력의 사전 프로그래밍 필요성’이라는 논문에서 자율주행차를 둘러싼 책임문제를 다루고 있다. 자율주행차가 2단계에서 3단계로 발전함에 따라 운행을 담당하는 운행자의 주의의무가 자율주행기능에 의해 일정 부분 대체되고 운행자의 역할이 감소하는 경향을 보이는데, 이 경우 인간운전자의 주의의무를 대체하는 자율주행차의 자율주행능력이 어느 정도가 되어야하고 이러한 자율주행능력을 확보하기 위한 자율주행차의 주의의무를 어떻게 파악해야 하는지가 문제된다. 이 논문은 인간 운전자의 주의의무가 감소하는 만큼 자율주행차의 역할이 증대되고, 역할이 증대되는 자율주행차의 주의의무는 결국 자동차 관리법상 제조자의 자기인증의 기준으로서 반영되어야 하며, 자율주행자동차의 자기인증의 기준은 인간 운전자의 주의의무 수준이 그 기준이 되어야 한다는 점을 살펴보고 있다. 또 자율주행차가 3단계에서 4단계로 발전하는 경우, 인간의 관여가 완전히 배제되기 때문에 운전대와 브레이크 페달이 없어지는 등 내부디자인이 완전히 바뀌게 되는데, 이러한 자율주행자동차에서 목표지점을 지시하고 주행시작을 입력하는 행위 외에는 주행에 관한 한 탑승자의 관여가 일어나지 않기 때문에 주행관련 운행자의 주의의무는 배제된다고 할 수 있다. 그리고 이 논문은 이런 완전 자율주행단계에서는 인간의 관여가 배제되기에 자율주행차의 주의의무의 내용에는 인간운전자의 윤리적 판단능력 및 융통성을 확보하는 조치가 필요하게 되는데, 이러한 인간운전자의 윤리적 판단능력 및 융통성을 확보하기 위해서는 사전적으로 윤리적 판단 및 융통성이 필요한 상황을 검토해야 하고 그 상황에 필요한 윤리적 판단능력과 융통성을 미리 인공지능의 주의의무 알고리즘에 반영할 필요가 있다고 강조하고 있다.

4) 2016년 류화진[5]은 ‘지능형 로봇의 범죄주

체성과 형사책임'이라는 논문에서 로봇범죄에 관하여 다루고 있다. 그는 이 논문에서 지능형 로봇의 범죄의 가능성에 관한 문제를 다루기 위해서는 기본적으로 정리되어야 할 용어들과 전제요건들로 세 가지를 다룬다. 첫째, 형사법에서 로봇범죄를 논하기 위해서 일단 법학적인 '로봇'의 개념을 확정할 수 있는가 하는 점, 둘째, 로봇범죄를 범죄로 인정하기 위해서는 로봇이 범죄의 주체가 되어야 하는데, 과연 로봇의 범죄주체성을 인정할 수 있겠는가 하는 점, 셋째, 로봇의 범죄주체성을 인정할 수 없다면 로봇이 사람에게 가하는 법익침해에 대해서는 누가 어떤 책임을 질 수 있겠는가 하는 점이다.

지금까지 살펴본 AI로봇 관련 논문들에서 공통적으로 다루고 있는 개념 중에는 '책임'이 있다. 과연 책임이란 무엇인가? 필자는 책임개념에 관해 생명윤리학자 피터 싱어의 입장과 가치윤리학자 막스 셸러를 통해 정리해 보려 한다.

### III. 책임이란 무엇인가? [6]<sup>1)</sup>

AI로봇의료에서 책임문제를 다루기 전에 동물과 인간의 가치에 대한 것을 다루는 것은 의미가 있을까? 필자는 동물과 AI로봇이 모두 동일하게 인간보다 뛰어난 면을 가진다고 여긴다. 동물은 인간보다 뛰어난 시력과 청력과 후각과 힘을 지녔다. AI로봇은 인간보다 뛰어난 계산력을 지녔으며 특정 행위 역시 동물과 같이 인간의 감각영역을 벗어날 수 있다. 필자는 인간보다 뛰어난 감각능력과 물리적 힘을 지닌 동물이 책임질 수 없듯이, AI로봇도 책임질 수 없다고 생각한다. 물론 책임질 수 없는 존재라고 하여 우리가 고려하지 말고 마음대로 다루어야 할 존재라는 것은 아니

다. 다만 어떤 행위의 결과를 책임질 만한 능력이 동물과 AI로봇에게는 없다는 것이다. 그 능력이란 필자가 생각하기에는 칸트가 자신의 철학에서 책임의 전제로 삼고 있는 '자유(Freiheit)'이다.

#### 1. 동물과 인간의 가치

의료윤리학자 피터 싱어는 인간과 일부 고등 동물은 지능이나 자의식 혹은 과거감, 미래감, 연대감을 가질 수 있다는 점에서 유사한 가치를 가진다고 평가했다[7]. 그러나 인간이 가진 가치는 동물 전체가 가진 가치와 다르다. 인간이 가진 지능은 동물이 가진 것과 다르며 인간이 세계에 부여하는 가치 또한 동물들만이 있는 세계를 상상해 볼 때 가지게 되는 가치와 다르다. 인간에게 기술 문명의 혜택이 없었을 때 동물은 우리를 위협하는 존재였고 나아가 정복해야 할 대상이었다. 동물에게도 인격이 있는가? 도대체 인격이란 무엇인가? 그런데 인간이 무엇이기에 인격이라는 것을 논하는 것일까? 만약 이 세계에 인간이 없다면 어떨까? 무엇이 가치 있는 것일까? 이 세계에 인간이 없고 동물들만 있다면 무엇이 가치 있는 것일까?

독일의 철학자 칸트[8]는 인간이 가진 가치에 관해 만일 세계가 무생물들이나 혹은 일부분은 생명이 있지만 이성이 없는 존재자들로 이루어졌다면, 그러한 세계에서는 가치에 관해서 조금이라도 이해하는 존재자는 없을 것이므로, 그와 같은 세계의 현존재는 아무런 가치가 없을 것이라고 말한다. 그에 반해서, 이성적 존재자들이 있기는 하지만, 그러나 그들의 이성이 사물들의 현존재의 가치를 자연과 그들의 관계 내에서만 정할 수가 있을 뿐이요, 그러한 가치를 근원적으로 스

1) 학위논문 "피터 싱어의 생명윤리론에 대한 비판적 고찰" [6]에서 일부 내용을 가지고 와서 이 논문의 논의에 맞추어 재구성하였다.

스로 만들 수가 없다고 한다면...실로 그러한 세계 역시 상대적 목적은 있겠지만, 절대적 궁극 목적은 없을 것이라고 말한다. “칸트는 인류가 피조계에서 특별한 위치를 차지한다고 생각했다. 물론 칸트만 이렇게 생각한 것은 아니다. 이것은 예전부터 내려오는 생각이다. 고대부터 인간은 그들 자신을 다른 피조물들과는 본질적으로 다르다고 생각해 왔다. 단지 다를 뿐만 아니라, 다른 피조물보다 더 나은 존재라고 여겨 왔던 것이다. 사실 인간은 전통적으로 그들 자신을 무척이나 대단한 존재라고 여겼다. 칸트도 분명 그러했다. 칸트가 보기에 인류는 “본질적 가치, 즉 존엄성”을 지녔으며, 이것으로 인해 그 무엇보다도 가치 있는 존재가 되는 것이다. 반면, 다른 동물들은 인간의 목적에 기여하는 한에서만 가치를 지닌다. 따라서 우리는 동물을 우리 마음대로 이용할 수 있다. 심지어 우리에게서는 동물을 학대하지 말아야 할 “직접적인 의무”조차 없다” [9]. 칸트[10]는 이 세계는 인간에 의해 궁극 목적을 갖는다고 본다. 그는 가치를 규정하는 입법자 자신은 바로 그 때문에 존엄을, 즉 무제약적이고 비교할 바 없는 가치를 가져야 한다고 본다.

여기서 칸트가 말하고 있는 인간의 가치를 문성학은 다음과 같이 말하고 있다. “도덕적 행위 주체로서 인간은 동시에 가치 평가의 주체로서 사물들에 대해 가격을 매길 수 있는 존재이다. 사물들은 유한한 값을 가지며, 따라서 규정할 수 있는 값을 가진다. 그리고 “가격을 가지는 것은 같은 가격의 다른 어떤 것과 교체될 수 있다”. 그러나 가격을 매기고 가치를 규정하는 입법적 주체인 인간 그 자신은 결코 가격 매겨질 수가 없다. 달리 말하면 인간은 유한하거나 규정할 수 있는

고정된 값을 가지는 존재가 아니다. 인간은 가격을 초월해 있다. 즉 무한한 값을 가진다” [11].

칸트[12]는 인간과 동물의 차이에 관해 동물들은 그들의 본성에 의해 그들이 이전에 될 수 있었던 모든 것이라고 본다. 또, 칸트는 최초로 어떤 낮은 이성이 동물들에게는 모든 것을 제공하였으나 인간은 그 자신의 이성을 필요로 한다고 말한다. 칸트는 인간은 본능의 결핍으로 말미암아 자신을 인도할 계획표를 스스로 만들지 않으면 안 된다고 하였다. 즉, 동물들은 자연적으로 프로그래밍된 바에 따라 살아간다는 것이다. 이 부분은 AI로봇에 관해서도 유사하다. AI로봇은 인간의 특정목적을 위해서 제작된다. 동물이 자연선택에 의해 본능을 가지듯이, AI로봇은 인간의 특정목적을 위해서 프로그래밍된다. 이에 반하여 인간은 자신을 어떤 방식으로 살아가도록 할지에 관해 스스로 프로그래밍해야 한다. 동물은 자기의 본능에 종속적인 지능을 가지고 자연 속에서 살아간다. AI로봇 역시 특정목적을 위해 어떻게 행위해야 할지에 관해 외부에서 주어진 최종명령(ultimate order)<sup>2)</sup>을 수행하기 위해서 프로그래밍된다. 그러나 인간은 동물이나 AI로봇처럼 고정된 한계 속에 갇혀 있는 존재가 아니라, 끊임없이 자신이 도달한 한계를 넘어서는 방식으로 자신의 동일성을 유지하는 존재이다[13]. 즉 인간은 자신의 최종명령은 스스로 정한다.

문성학은 인간의 이러한 본질적 측면을 ‘가능적 무한자로서의 인간’이라는 관점에서 말한다. 루소[14]도 다음처럼 말했다. “내가 알고 있는 한 그 어떤 철학자도 감히 인간에게 ‘그대는 여기까지만 가야만 하며 그 이상 가서는 안 된다.’고 말하지 못했다. 우리는 자연이 우리가 무엇이 되도

2) 최종명령이란 필자가 정해본 것이다. 이세돌과 바둑을 두었던 인공지능 알파고의 최종명령은 ‘바둑게임에서 이기는 수를 찾으라’는 것이다. 퀴즈쇼에 나온 IBM 왓슨의 최종명령은 ‘퀴즈쇼의 문제를 풀어라’는 것이다. 이러한 최종명령은 오로지 인간에 의해서만 주어진다.

록 허락한 것인지 모른다”. 문성학은 가능적 무한자로서의 인간이라는 인간관은 서양철학의 대가들이 다양한 방식으로 표현한 사상이기도 하다고 말한다. “인간은 자기의 모든 업적을 넘어서, 따라서 자기의 본래적 가능성을 넘어서 밖으로 나가려 하기 때문에, 그는 중국에 가서는 자기 자신을 넘어서 나가며 그 자신을 초월한다. 파스칼은 이를 두고 “인간은 무한히 인간을 넘어선다”고 말했다. 셸러는 인간을 “자기 자신을 영원히 넘어서는 삶을 사는 존재”라고 하면서 이 사상을 이렇게 표현했다. “자기를 넘어서 밖으로 나가는 열정”이야말로 유일한 참된 ‘인간성’이다” [15].

인간성을 이렇게 정의한다면 동물과 인간의 질적 차이를 고민할 필요가 없을 것이다. 왜냐하면 인간 이외에는 자신을 끊임없이 의식적으로 넘어서는 것을 그 본질로 하는 동물은 없기 때문이다[16]. 이는 AI로봇에게도 마찬가지이다. AI로봇 역시 동물과 같이 끊임없이 의식적으로 자신을 넘어서 수 없다. 새로운 생명 윤리의 문제 상황에서 인류는 인간이라는 존재가 무엇인지에 관계고 고민하고 있다. 인간이란 무엇인가라는 질문은 오래되었지만 항상 새로운 것이다. 현재 이 질문으로 우리는 현기증을 느끼고 있지만 곧 극복할 것이다. 왜냐하면 인간은 스스로 도달한 한계를 넘어가는 방식으로만 인간일 수 있기 때문이다. 인간은 동물들 중 유일하게 인간 종뿐만 아니라 전체 동물들을 고려할 수 있다. 인간은 자연계 속에서 유일한 진리의 위탁자로서 하나의 생물종으로서의 인간의 관점을 벗어나서 이성의 관점, 전체를 고려하는 관점, 즉 세계 개방적인 관점에서 자연을 보는데서 가장 인간답게 된다[16]. 자연계 내에서 인간이 가지는 존엄성은 다음과 같이 설명할 수 있다. “이 지구상에서 인간만이 자신의 자연종속성을 인식하며, 또한 자연의 유기적 연관성을 파악하여 지구 친화적 방식으로 지구상의

생명계에 개입할 수 있는 유일한 존재이기에 인간은 지구상에서 고귀한 존재이고 존엄한 존재일 것이다. 그러나 이 경우 인간의 존엄성은 인간이 전 자연의 궁극목적 혹은 최종목적이 되는 방식으로 확보되는 그런 칸트적 존엄성은 아니고, 오히려 전 자연 시스템의 자연스런 흐름이 자연스럽게 흘러갈 수 있도록 도울 수 있는 능력을 가진 유일한 생명체라는 의미에서 존엄하게 된다” [11].

위와 같이 인간은 ‘전 자연의 시스템의 자연스런 흐름이 자연스럽게 흘러갈 수 있도록 도울 수 있는 능력을 가진 유일한 생명체’이기에 존엄하다고 할 수 있다. 싱어의 이익 평등 고려의 원칙을 지구온난화라는 상황에서 판단해 보면, 동물과 인간 사이의 천칭은 인간에게로 기운다. 이유는 인간에게 존엄성이 있기 때문이다. 싱어가 이익 평등 고려의 원칙도 결과론적으로 접근할 때는 각 개체가 가진 고통의 정도보다는 전체적 고통의 해소 정도에 따라 고려해야 한다고 했던 것을 적용한다면, 이제 인간과 동물은 평등하지 않다. 더욱이 AI로봇은 감각을 가지지 않고 고통과 쾌락을 느끼지 않기에 이익평등고려의 원칙의 적용대상이 되지 못한다. 이러한 상황에서 싱어의 이익 평등 고려의 원칙을 적용한다 하더라도, 인간은 동물이나 AI로봇과 달리 존엄성을 가지기에 동물이나 AI로봇보다 우선적으로 고려되어야 한다. 또한 싱어가 천칭에 비유하기도 했던 개체성을 떠나 전체성을 고려하는 이익 평등 고려 원칙에 의해서도, 인간이라는 종은 전체적으로도 특별하게 고려되어야만 한다. 싱어의 천칭의 비유는 공리주의의 결과론적 입장을 대변하는 것이고, 이 입장은 언제나 개체가 아닌 전체의 이익을 우선한다. 어떠한 위기 상황에서는 고통받는 개체의 이익은 전체의 이익을 고려할 때 무시될 수도 있다. 살펴본 바와 같이 싱어는 지구 환경 위기

의 심각성을 강조하고 있다. 지구 환경 위기는 지진이 난 상황보다 심각했으면 심각했지 덜하지 않다. 지구 위의 생물체들은 앞으로의 생존자체가 문제되기 때문이다. 이러한 상황에서 지구 생태계의 전체 운명은 인간이라는 특수한 종의 역할에 의해 좌지우지 된다고 보아도 무방하다. 현재 인간이라는 종은 인간 아닌 동물에 대해 지진이 난 상황에서의 의사와 같은 입장에 있다. 그런데 싱어가 말하는 이익 평등 고려의 원칙으로는 인간 종의 이익을 우선시킬 수 없다는 데에 그 원칙의 문제점이 있다. 그러나 이러한 이익평등고려의 원칙을 AI로봇에게 적용시켜본다면 문제 양상이 매우 이상해진다. 왜냐하면 만약 결과적으로 지구온난화 문제를 AI로봇이 해결할 수 있고 그 해결점이 그 문제를 일으킨 인간이고 인간이 소비활동을 하지 않는다면 이 문제를 해결할 수 있다고 AI로봇이 결론을 내리고 시행한다면 큰 문제이기 때문이다. 여기서 우리는 인간중심적 윤리로 이 문제를 해결할 수 있다. 인간은 인간중심성에서 벗어날 수 없는 존재이다. AI로봇도 결국 인간이 만든 도구이다. 따라서 인간이라는 대전제를 벗어난 해결책은 진정한 해결이 아니다. 다음에 과연 어떤 존재가 책임질 수 있을지에 관해서 살펴보자.

## 2. 누가 책임질 수 있는가?<sup>3)</sup>

“칸트는 동물을 학대하는 것이 잘못된 일일 수 있다고 인정했지만, 그 이유는 동물이 상처받기 때문이 아니었다. 그 이유는 단지 그 결과로 인간이 간접적으로 고통을 당할 수도 있기 때문인데, 즉 ‘동물을 잔인하게 대하는 사람은 인간도 거칠게 대하게 될 것’이기 때문이다. 따라서 칸트가

보기에, 동물 그 자체에는 어떠한 도덕적 중요성도 없다” [9]. 여기서 동물과 인간의 책임문제를 한 신문 기사를 통해서 다루어보려 한다. 만약 우리가 침팬지와 같은 동물에게 인간이 갖는 권리를 주어야 한다면 동시에 권리에 대한 그 책임도 물을 수 있어야 한다. 우리가 대형 유인원인 침팬지에게 인간이 갖는 권리를 준다면 침팬지는 자신이 가진 권리에 상응하는 의무를 다할 수 있을 것인가? 다음의 경우를 살펴보자. “침팬지에게 얼굴과 손을 물어 뜯겨 중상을 당했던 미국 차를라 내시(57)씨가 지난 5월 안면 전체 이식수술을 받은 후 몰라보게 복귀된 모습을 공개했다. 미국 ABC뉴스에 의하면 오하이오 주 클리블랜드에 사는 내시 씨는 2개월 전 쯤 보스턴에 있는 브리검 여성병원에서 뇌사자에게 기증받은 안면을 전체 이식하는 페이스오프 수술을 받았다” [17]. 미국의 한 여성이 침팬지에게 상해를 입었다는 기사이다. 싱어의 주장대로라면 침팬지는 이 사건에 대해 인간이 지는 것과 같은 법적이고 도덕적인 책임을 져야 한다. 만약 우리가 침팬지에게 인간과 동일한 책임을 묻는다면 상해죄를 저지른 침팬지를 감옥에 보내야 한다. 그러나 필자가 생각할 때 침팬지가 갇힌 감옥은 감옥이 아니라 동물원이다. 왜냐하면 침팬지는 자신의 권리에 상응하는 책임을 질 수 없기 때문이다. 이는 우리가 동물에게 책임을 물을 수 없다는 사실을 알려준다. 즉 동물은 우리가 책임을 져야 할 대상이지 책임을 물을 수 있는 대상은 아니다. 이는 AI로봇에게도 유사한 것 같다. 이봉재[18]는 이를 다음과 같이 표현한다. “프로그램이 완전한 지성을 갖추지 못한다는 전제를 받아들인다면, 프로그램 자체의 책임은 미미한 듯하다. 더군다나 책임을 묻는다고 해도 프로그램 자체에 대한 처벌이 어떤 것일지

3) 학위논문 “피터 싱어의 생명윤리론에 대한 비판적 고찰” [6]에서 일부 내용을 가지고 와서 참조하였다.

상상하기 어렵다. 프로그램을 구금할 것인가, 활동중지시킬 것인가?”

## IV. 전자인간의 지위와 킬 스위치

### 1. EU의 법적지위 부여: 소위 ‘전자인간’

EU 의회가 2017년 1월 12일(현지시간) 벨기에 브뤼셀에서 AI를 탑재한 로봇의 법적지위를 ‘전자인간’으로 인정하는 결의안을 의결했다<Table 1>. EU의회는 AI로봇을 제작할 때 인류에 기여할 수 있도록 알고리즘을 설계해야 한다고 규정했다. 즉 AI로봇의 일탈행위에 대비할 수 있도록 시스템 작동을 강제 종료할 수 있는 ‘킬 스위치’도 반드시 탑재할 것을 정한 것이다. EU의회는 동시에 인간 역시 AI로봇과 협력하며 살아가는 4차 산업혁명 시대에 철저히 대비해야 한다고 표명했다. 더불어 EU 의회는 “AI로봇 기술의 발전과 함께 인간의 노동영역이 급변할 수 있다”면서 새로운 고용모델을 제시할 것을 촉구했다. 인공지능 로봇 알고리즘에 이와 같이 비상상황일 때 작동을 멈추는 킬 스위치 탑재를 의무화한 것은 세계적으로 첫 번째 가이드라인이라 할 수 있겠다.

<Table 1> 유럽연합(European Union, EU)의회의 ‘AI로봇 결의안’ [19]

구분	주요내용
법적 지위	· AI 로봇을 ‘전자인간’으로 인정
‘킬스위치’	· 로봇작동을 멈추는 버튼 장착
주요 원칙	· 로봇은 인간을 위협하면 안 됨 · 늘 인간의 명령에 복종해야 함 · 로봇 역시 자신을 보호해야 함
핵심 권고	· EU 안에 AI기술 · 윤리기구 신설
최종 의결	· 다음날 본회의 투표 실시 예정

AI : artificial intelligence.

이번 결의안은 AI로봇 시대를 대비하여 프로그램 오류나 해킹 등 비상상황에 대한 막연한 불안감을 해소하기 위해 ‘킬 스위치’ 탑재를 의무화했다. 즉 인간이 위협적이라고 느낄 경우에는 AI로봇의 작동을 강제로 멈출 수 있도록 한 것이다. 또한 AI로봇을 전쟁이나 테러에 악용할 소지를 없애고, 한편으로는 무차별적인 규제도 예방하기 위한 조치로 보인다. 그리고 이와 함께 AI로봇 개발자들은 AI로봇을 인간의 명령에 무조건 복종하면서 위협을 가하지 않도록 설계해야 한다. 더불어 이번 EU의회 가이드라인에 의하면 로봇 개발자들은 해당 국가기관에 로봇을 등록해야 하고 로봇의 작동으로 인하여 사고가 일어나면 당국의 시스템 코드에 접근할 수 있는 접근권한을 제공할 의무도 규정하였다.

더 나아가 EU 의회는 결의안을 통해 EU 집행조직 안에 ‘로봇AI 기구’를 신설하여 AI로봇을 둘러싼 기술적이고 윤리적 기준을 논의하도록 EU 집행위원회에 권고했다. 또 AI로봇이 인간사회의 법률이나 도덕기준에 따라 작동할 수 있도록 제한하는 법률이 필요하다는 의견도 덧붙였다. 또 EU 의회는 AI로봇 기술의 발전으로 인간의 노동영역이 급변할 수 있다고 보아 새로운 고용모델을 제시할 것을 촉구하며, 동시에 조세시스템도 개편해야 한다고 권고하였다. 즉, AI로봇이 실용화되면 대규모 실직이 발생할 수 있는데, AI로봇을 활용하는 이들에게 일명 ‘로봇세’를 물려 이에 대비하겠다는 것이다. 이 결의안은 2017년 2월 본회의 의결을 비롯하여 EU 회원국들의 논의를 거쳐 법으로 제정된다. 이른바 ‘로봇시민법’으로, 다른 나라가 EU 회원국에 AI로봇을 수출할 때도 이 규정을 따라야 한다고 한다[19].

## 2. ‘킬 스위치’를 둘러싼 논란

EU는 AI로봇이 인류 생존을 위협할 수 있기에, 필요할 때 로봇의 기능을 멈추는 ‘킬 스위치’가 필요하다는 결의안을 채택했다고 한다[20]. 그리고, 연합뉴스[20]에 따르면 2017년 세계경제포럼(World Economic Forum, WEF)은 연례 글로벌 리스크 보고서에서 인공지능과 로봇공학 기술이 인간의 통제를 벗어날 위험이 있다고 지적했다고 파이낸셜타임스가 2017년 1월 12일(현지시간) 전했다고 한다. 이 보고서는 “AI 시스템의 의사 결정 방식에 대해 완전히 이해하지 못한 채 이 시스템에 전적으로 의존하는 것의 위험성”을 경고하며, 특히 군사용 로봇 개발 경쟁이 일어나고 있다고 지적한 데 이어 “초지능(superintelligence)이 인류의 생존을 위협할 수 있다”는 우려를 나타냈다고 한다. 또한 앞의 기사에 따르면, 마이크로소프트의 창업자 빌 게이츠는 2016년 “나도 초지능에 대해 걱정하는 쪽”이라면서 “먼저 기계는 우리의 일을 많이 대신하고 초지능은 아닐 것이다. 이는 우리가 잘 통제한다면 긍정적이다. 하지만 수십 년이 더 지나면 인공지능은 우려하기에 충분할 만큼 강해질 것”이라고 말했다고 한다. WEF 보고서는 AI가 많은 분야의 생산성과 의사 결정을 향상시킬 가능성이 있다면서도 충분한 통제가 이뤄지지 않을 우려를 부각했는데, 보고서의 저자 가운데 한 사람인 존 더직 마쉬 글로벌 리스크&스페셜티스 사장은 “인공지능은 어떻게 해서든 결정을 내린다. 프로그램되는 대신 패턴 인식에서 학습할 수 있을 것”이라고 말했다고 한다[20]. 이러한 우려들을 사전에 방지하기 위하여 EU는 제작되는 AI로봇에 킬 스위치를 탑재하도록 한 것이다.

킬 스위치[21]란, 휴대전화 같은 정보기기를 분실했을 때 원격으로 조작해 개인 데이터를 삭제

하고 도용을 막도록 하는 스위치다. 프로그램 오류나 해킹 등 비상 상황에서 로봇을 통제할 수 없을 때 즉각 멈출 수 있도록 하는 기능으로 일종의 ‘자폭 기능’이다. 하지만 최근 AI가 인류 생존을 위협할 수 있다는 우려가 나오면서 이와 같이 ‘킬 스위치’의 필요성이 강조되고 있다. 앞서 언급한 바와 같이 영국 일간지 가디언은 EU 의회가 AI 로봇의 법적 지위를 ‘전자 인간(electronic personhood)’으로 지정하는 결의안을 통과(벨기에 브뤼셀에서 열린 EU 의회 법사위원회에서 찬성 17표, 반대 2표, 기권 2표로 통과)시켰다고 전했다. 이는 국가 차원에서 AI로봇의 지위·개발·활용에 대한 기술적·윤리적 가이드라인을 제시한 것이어서 전 세계에서 비상한 관심을 불러일으키고 있다.

필자는 여기서 잠시 칸트의 도덕적 책임으로서의 전제조건으로서 자유에 관해 알아보려고 한다. 문성학[22]은 칸트윤리학에서 책임가능성의 전제조건은 바로 자유임을 다음과 같은 사례를 통해서 설명하고 있다. “칸트가 ‘너는 해야 하기 때문에 너는 할 수 있다’라고 말할 때, 칸트는 그런 의무의식 만능주의를 주장하려는 것이 아니고 당위는 그 실천가능성이 전제되어야 한다는 것 달리 말하면 만약 당위의 실천가능성이 애초부터 전무하다고 한다면 도덕이란 것은 환상에 불과하다는 것, 그리고 당위의 실천(자유)은 인간의 신체를 통해 표현될 수밖에 없다는 것을 주장하고 있는 것이다. 예컨대, 스스로도 주체할 수 없는 정욕에 불타서 지나가는 처녀를 성폭행한 청년이 있다고 하자. 그 청년은 생리적인 인과법칙에 따라 행위한 것이 된다. 그런데 생리적인 인과법칙도 인과법칙인 한, 필연의 법칙이고 거역할 수 없는 법칙이다. 그럼에도 불구하고 그의 내면에서는 ‘나는 그런 나쁜 짓을 해서는 안 된다’는 소리가 미약하게 울려 퍼진다. 이 경우 그 내면의

소리, 즉 당위의 명령이 의미가 있으려면 그 청년에게는 자신을 지배하려는 생리적인 인과법칙의 강제에도 불구하고 처녀를 폭행하지 않을 수 있는 가능성이 전제되어 있지 않으면 안 된다. 만약에 이 가능성이 전제되어 있지 않다면, ‘나는 그런 나쁜 짓을 해서는 안 된다’는 그 내면의 명령은, 마치 ‘나는 빛보다 더 빨리 달려야만 한다’와 같이 아무런 의미가 없는 말이 될 것이다. 칸트는 도덕적 명령은 인과법칙의 관점에서 볼 때는 불가능한 명령이지만, 그렇다고 그 명령이 자연법칙을 무시하는 기적과 같은 일 즉 인과법칙을 파괴하는 일을 일으키는 명령이라고 생각하지는 않았다”. 우리가 도덕적으로 마땅히 해야 할 일은 우리에게 당위의 실천가능성이 있기 때문이다.

과연 AI로봇에게는 이러한 의미의 실천가능성이 존재하는가? 필자는 AI로봇에게 탑재될 킬 스위치는 AI로봇의 실천가능성을 제한한다고 생각한다. 따라서 AI로봇에게는 두 가지 의미에서 자유는 부재하다. 첫째, AI로봇은 알고리즘에 의해 최종명령에 따르도록 설계되어 있다는 점에서 자유의 가능성은 매우 부족하다. 둘째, AI로봇에게 탑재될 킬 스위치의 존재는 설령 AI로봇의 유사 자유행위에 대한 유사 자유가능성 조차 없애고 있다. 따라서 AI로봇에게 자유란 근본적으로 불가능하다.

## V. 의료에서 전자인간 의사

### 1. 의료에서 자가치료 개념의 등장과 보편적 자가의료의 가능성

우리는 앞서 AI로봇에게 자유가 불가능함에 대해 칸트윤리학적 설명방식으로 살펴보았다. 의료에서 AI로봇은 기존의 의사-환자 관계를 넘어 환자 스스로 자가치료를 이론적으로 가능하게 하

고 있으며 앞으로 등장할 수많은 AI로봇과 그 발전은 이를 실천적으로도 가능하게 할 것으로 전망된다. 양우진 등[23]은 ‘가속화되는 고령화 시대 의료 인공지능 융합에 관한 연구’라는 2017년 한국통신학회 하계종합학술발표회 발표문에서 다음과 같이 말하고 있다. “만성·중증 질환자의 치료를 위해 딥러닝(deep learning) 등 인공지능 기술이 필수이며 예방·관리 중심으로 의료서비스가 변화하고 있다. 사물인터넷 기술을 활용해 수집한 고혈압, 당뇨 등 만성질환자의 생활데이터, 진단·치료과정에서 발생된 의료데이터가 폭발적으로 증가하고 있으며 방대하게 수집된 개인 의료빅데이터를 토대로 유전자 분석 및 질병예측을 위해서는 딥러닝 등 인공지능 알고리즘이 필수적이다. 이는 데이터의 다양성, 비구조화, 복잡성 등 의료데이터가 가지고 있는 특성 때문이다. 인공지능 기술은 방대하고 복잡한 의료데이터 분석을 통해 치료 중심의 의료 서비스가 예방과 관리 중심으로 변화될 것이 예상된다. 또한 사전진단에서 처방, 진료 및 환자 상담 등 다양한 영역에서 인공지능 기술이 융합될 것이다”. 우리는 심심찮게 TV 등에서 미래세계 사람들이 아침에 일어나면 인공지능 로봇이 건강상태를 체크하고 실시간으로 치료를 받는 모습들을 볼 수 있다. 현재 기술은 이론적으로 그러한 미래를 가능하게 하고 있다. 지금은 병원에 있는 왓슨이 스마트폰으로 들어올 수도 있을 것이다. 그리고 스마트폰으로 들어온 왓슨은 TV나 VR 등 여타의 형태로 된 디지털 존재로 현실세계에 등장하여 우리에게 건강과 관련된 조언이나 정보를 제공할 수 있을 것이다. 이러한 기술은 점점 더 확장일로에 있으며 이러한 세계를 살아가는 사람에게 병원이라는 특수공간은 점점 더 그 필요성을 잃어갈 수도 있다. 의료는 병원이라는 공간적 한계를 넘어 보편적인 생활세계로 들어오게 될 것이다. 즉 의료에서 AI

로봇의 등장으로 ‘자가치료’ 개념이 등장함과 ‘동시에 보편적 자가의료’의 가능성이 열리게 되는 것이다. 병원이라는 특수장소의 필요성이 저하되면서 동시에 의료인이라는 전문인들의 필요성 역시 저하될 가능성이 이론적으로 가정된다. 과연 이렇게 AI로봇이 의료에 등장함으로써 열릴 수 있는 가정적 미래에서 의료인들은 어떤 역할을 할 수 있을 것인가?

## 2. 전자인간 의사는 얼마나 책임질 수 있을까?

앞서 필자는 동물과 AI로봇과 인간의 가치에 관해 살펴보았다. 동물과 AI로봇은 이미 주어진 목적이 있다. 동물에게는 자연의 목적이 주어지고, AI로봇에게는 인간이 부여한 인간세계의 목적이 주어지고 있다. 동물에게 자연의 목적이 있다는 것은 생물학적으로 동물은 자연선택에 의해 존재한다는 의미이고, AI로봇에게 인간의 목적이 있다는 것은 사회적으로 AI로봇은 인간의 특정 역할을 수행하기 위해 존재한다는 의미일 것이다. 즉, 동물은 본능이라는 자연의 프로그램에 따라, AI로봇은 알고리즘이라는 인간의 프로그램에 따라 움직이게 되어 있다. 이 두 존재 즉 동물과 AI로봇에게는 앞서 언급한 칸트적 의미의 ‘자유의 가능성’은 없다고 할 수 있다. 특히 AI로봇에게 킨스위치가 탑재될 경우에는 더더욱 그러하다. 이제 AI로봇은 의료의 영역으로 들어오게 되어 있다. 우리 의료사회에 이미 닥터 왓슨은 들어와 인간의 사와 공존하고 있다. 이러한 상황에서 과연 전자인간으로서의 인공지능 의사는 얼마나 책임질 수 있을까? 이러한 질문에 대해 지금까지 살펴본 바로 간단히 AI로봇에게 자유의 가능성이 없으므로 책임가능성이 없다는 결론에 도달할 수 있을 것이다. 하지만, 이에 대해 조금 더 살펴보도록 하자.

의료에서 “인간 의사와 인공지능의 협업 단계에서 ‘산술적’인 오진 가능성은 전자 쪽이 더 높다고 할 수 있다. 그러나 인공지능시스템의 특정한 정보처리연산 능력이 산술적으로 인간 의사보다 월등하고 신뢰할 만한 것이라 할지라도, 진료 행위에는 해석과 판단력이 요구되며 이러한 측면에서 인공지능 시스템은 인간과 같은 자연지능의 판단과 커다란 차이를 보인다. 첫째, 인공지능 시스템은 데이터화 가능하게 처리된 정보 이외의 것을 판단 근거로 활용할 수 없다[23]. 그러나 유능한 인간 의사는 환자의 상태를 데이터화 불가능한 것을 포함하여 총체적으로 파악하여 진료에 반영한다. 둘째, 인공지능 시스템은 환자의 병증을 환자의 포괄적인 건강이나 행복의 관점에서 해석하지 못한다. 동일한 병원균의 감염이라는 병증에 대해서도 환자의 심리적 기질이나 성향, 가치관에 따라 처방이 달라질 수 있다. 이런 심리학적·가치론적 해석의 부분 역시 알고리즘화 된 프로그램으로서 감당하기 어려운 부분이다. 전문가 프로그램이 그것에게 제공되는 정보영역 밖에 존재하는 환자의 상태나 기질, 가치관 등을 고려할 수는 없기 때문이다[18]” 따라서 의료현장에서 인공지능 시스템의 도입은 의료인의 책임 경계를 다시 생각하게 한다. 필자는 의료에서 인간 의료인의 임무는 킨스위치라고 생각한다.

## VI. 마치며: ‘킨 스위치’로서의 인간 의료인의 요청

앞서 한국학계의 기존연구검토를 통해 로봇윤리의 이론적 기초와 인공지능의 위험과 그 규제를 위한 일반적인 법적 규제방안 및 AI로봇이 활용되는 자율주행차와 지능형 로봇이 과연 범죄주체성을 가질 수 있는지 여부와 형사책임을 어떻게 될 것인지 등에 관해 간략히 살펴보았다. 이러

한 논의의 핵심의제는 ‘책임’이라 할 수 있다. 책임 논의에서 칸트는 인간과 동물의 차이에 관해 동물들은 그들의 본성에 의해 그들이 이전에 될 수 있었던 모든 것이라고 보았는데, 최초로 어떤 낯선 이성이 동물들에게는 모든 것을 제공하였으나 인간은 그 자신의 이성을 필요로 하는 존재로서 본능의 결핍으로 말미암아 자신을 인도할 계획표를 스스로 만들어야 하는 존재였다[12]. 즉, 동물들은 자연적으로 프로그래밍된 바에 따라 살아가고, AI로봇은 인간의 특정목적을 위해서 제작된다. 동물이 자연선택에 의해 본능을 가지듯이, AI로봇은 인간의 특정목적을 위해서 프로그래밍된다. 이에 반하여 인간은 자신을 어떤 방식으로 살아가도록 할지에 관해 스스로 프로그래밍해야 한다. 이러한 인간의 능력을 피터스는 인격의 ‘지적 입법적 기능’으로 보았다.

피터스는 인격에는 세 기능이 필요한데, 지적 입법적 기능과 의지적 실천적 기능 및 반성적 기능이 그것이다. 이 세 기능은 자유를 전제로 한다. 그런데 이 기능 중 가장 중요한 기능이 바로 지적 입법적 기능이다. 동물은 자기의 본능에 종속적인 지능을 가지고 자연 속에서 살아가고, AI로봇 역시 특정목적을 위해 어떻게 행위해야 할지에 관해 외부에서 주어진 최종명령을 수행하기 위해서 프로그래밍되기에 동물이나 AI로봇은 지적 입법적 기능을 가지고 있지 않다. 오로지 인간에게서만 지적 입법적 기능이 발견된다. 물론 인간에게서 지적 입법적 기능이 언제나 발견된다거나 어떤 상황에서도 발현된다는 의미는 아니다. 왜냐하면 지적 입법적 기능은 인류라는 종에게 주어진 특수한 기능이며 개별적 인간에게서는 적게 주어지거나 아직 발현되지 않았을 수도 있다. 이럴 경우 인간의 의무는 그러한 취약한 환경에 처한 인간을 대신하여 지적 입법적 자유를 보장해야 한다. 이것이 의료윤리에서 강조되는 자율성

존중의 원리인데 취약한 동의자의 보호 원칙을 통해 살펴볼 수 있다.

인간은 동물이나 AI로봇처럼 고정된 한계 속에 갇혀 있는 존재가 아니라, 끊임없이 자신도 달한 한계를 넘어서는 방식으로 자신의 동일성을 유지하는 존재이다[24]. 문성학은 인간의 이러한 본질적 측면을 ‘가능적 무한자로서의 인간’이라는 관점에서 말하였다. 그런데 가능적 무한은 동물과 AI로봇에게는 본질적으로 불가능하다. 특히 킬 스위치가 탑재된 AI로봇에게는 이론적으로 불가능하다. 그리고 이론적으로 불가능하다함은 현상적으로 AI로봇이 자유롭게 보일 수는 있으나 자유라는 개념과 지적 입법적 기능을 지니는 인격체의 특징에서 살펴보면 불가능하다는 의미이다.

그러나 기술은 사회를 끌고가는 힘이 있어 의료에서 AI로봇의 도입은 거스를 수 없는 대세이다. 이제 의료는 이러한 전자인간 의사의 등장으로 확대되었다. 넓어진 의료의 지평에서 ‘자가치료’개념의 등장하였고, 이론 말미암아 ‘보편적 자가의료’가 가능해졌다. 앞서 우리는 AI로봇은 동물과 마찬가지로 책임질 수 없음에 관해 도덕형이상학적으로 고찰해 보았다. 의료행위에는 책임이 뒤따른다. 즉 보편적 자가의료는 AI로봇에 의해 가능하게 되었다고 하더라도 이를 전체적으로 통제하며 감시할 인간 의료인의 존재는 필요한 것이다. 왜냐하면 전자인간 의사는 책임질 수 없기 때문이다. 이것은 마치 동물이 책임질 수 없는 것과 마찬가지로이다. 따라서 이제 AI로봇의료로 말미암아 자가치료와 보편적 자가의료의 가능해진 시대에도 ‘킬 스위치’로서의 인간 의료인의 존재는 요청된다고 할 수 있겠다. ㉞

## REFERENCES

- 1) 중앙일보. EU 의회, “로봇의 법적 지위는 ‘전자 인간’”. 2017. 1. 12. Available from: <http://news.joins.com/article/21116071> [cited 2018 May 5]
- 2) 변순용, 송선영. 로봇윤리의 이론적 기초를 위한 근본 과제 연구. 윤리연구 2013 ; 88 : 1-26.
- 3) 양종모. 인공지능의 위험의 특성과 법적 규제 방안. 홍익법학 2016 ; 17(4) : 537-565.
- 4) 이중기. 자율주행차의 발전단계로 본 운전자와 인공지능의 주의의무의 변화와 규범적 판단능력의 사전 프로그래밍 필요성. 홍익법학 2016 ; 17(4) : 443-472.
- 5) 류화진. 지능형 로봇의 범죄주체성과 형사책임. 과학기술과 법 2016 ; 7(2) : 213-242.
- 6) 정창록. 피터싱어의 생명윤리론에 대한 비판적 고찰. 박사학위 논문. 경북대학교 대학원, 2012.
- 7) Singer P. Practical Ethics, 2nd ed. New York : Cambridge University Press, 1993 : 72, 74-75.
- 8) Kant I. Kritik der Urteilskraft, Hamburg : Felix Meiner Verlag, 1974 : 320-321.
- 9) Rachels J. The Elements of Moral Philosophy. 노혜련, 김기덕, 박소영 역. 도덕철학의 기초. 서울 : 나눔의집, 2015 : 239-240.
- 10) Kant I. Grundlegung zur Metaphysik der Sitten, ed by Kant I. Werkausgabe VII, hrsg., W. Weischedel. Frankfurt : Suhrkamp Taschenbuch Verlag, 1974.
- 11) 문성학. 칸트 윤리학과 형식주의. 대구 : 경북대학교 출판부, 2006 : 222, 306.
- 12) Kant I. über Paedagogik, ed by Kant I. Werkausgabe XII, Frankfurt : Suhrkamp Taschenbuch Verlag, 1974 : 697.
- 13) 문성학. 인간의 길, 자유의 길. 철학논총 2003 ; 34 : 283-285.
- 14) Rousseau JJ. Emile. New York : Everyman's Library, 1974 : 29.
- 15) Hessen J. Lehrbuch der Philosophie Vol. II. Muenchen : Ernst Reinhardt Verlag, 1959 : 125.
- 16) 문성학. 삶의 의미와 윤리. 대구 : 형설출판사, 2007 : 172, 252.
- 17) 데일리중앙. 칩팬지 공격으로 얼굴 잃은 여성, 안면이식수술로 얼굴 찾아. 2011. 8. 12. Available from: <http://www.dailiang.co.kr/news/articleView.html?idxno=31695> [cited 2018 May 5]
- 18) 이봉재. 인공지능과 책임의 문제. 대동철학 2006 ; (37) : 73-92.
- 19) 파이낸셜뉴스. “AI로봇은 전자인간” EU, 법적 지위 부여. 2017. 1. 13. Available from: <http://www.fnnews.com/news/201701131750426924> [cited 2018 May 5]
- 20) 연합뉴스. 로봇이 인류 위협하나... EU ‘킬 스위치’ 검토. 2017. 1. 13. Available from: <http://www.yonhapnews.co.kr/bulletin/2017/01/13/0200000000AKR201701130808000009.HTML> [cited 2018 May 5]
- 21) 시사용어: 킬 스위치(Kill Switch). Available from: <http://ptnews.myvilpt.com/board/contentsView.php?idx=1186074&bid=17492> [cited 2018 May 5]
- 22) 문성학. 칸트철학의 인간학적 비밀. 울산 : 울산대학교 출판부, 1996 : 278.
- 23) 양우진, 변상익, 신현국. 가속화되는 고령화 시대 의료 인공지능 융합에 관한 연구. 한국통신학회 하계종합학술발표회 발표문 2017 : 201-261.
- 24) 문성학. 인간의 길, 자유의 길. 철학논총 2003 ; 34 : 283-285.

# The Search for a Moral Metaphysics in Artificial Intelligence Robot Medicine: The Responsibility of an Electronic Human in Medicine\*

JEONG Chang Rok\*\*

## Abstract

The European Parliament is currently considering a proposal to grant artificial intelligence (AI) robots the legal status of “electronic persons”. This article examines the metaphysical status of AI robots and considers some of the ethical questions that arise from the use of AI robots in the practice of medicine. In particular, would electronic human doctors have rights? And to what extent could they be held responsible for their behavior? In answering these questions this article considers the concept of a “kill switch” and how it relates to the notion of freedom that is presupposed by a Kantian understanding of responsibility. The article also outlines some of the new horizons opened up by the arrival of Watson in medicine, including the emergence of “self-care medicine” and the possibility of “universal self-care.” This article argues that in light of the advent of electronic human doctors, the new role for human health care personnel is that of a “kill switcher.”

## Keywords

artificial intelligence medicine, an electronic human in medicine, moral responsibility, kill switch, self-care medicine, universal self-care medicine

---

\* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5B5A07061510). I would like to thank Professor Lee Yun and Professor Kim Young-boo of Gachon Gil Hospital for helping me to write this paper with permission to visit Watson at Gachon Gil Hospital. I would also thank anonymous doctor who has been interested in this subject and tried to help me.

\*\* Department of Ethics Education, Teacher's College, Kyungpook National University · School of Law, The Catholic University of Korea: **Corresponding Author**