

Can Artificial Intelligence(A.I.) be a Subjective Being for Ethics?: From a Lens of Posthumanism¹⁾²⁾

Park, Hyu-Yong³⁾ (Jeonbuk National University, Professor)

< ABSTRACT >

Due to the sheer development of artificial intelligence (henceforth, A.I.) algorithm, diverse A.I. machines pave their way into the deep levels of human lives. Under these circumstances, this paper discussed what the ethical principles and criteria of A.I. should be with a lens of posthumanism. To address this theme, this paper focused on the following concrete research questions: First, what onto-epistemological and ethical perspectives does posthumanism have? Second, what can the specific theories of posthumanism deal with the issues of A.I. ethics and what are the characteristics of A.I. as the ethical subject? Third, what ethical paradigm is required under the milieu of advanced capitalism that stimulates the spread of A.I. machines? To concentrate on the third question, this paper gave the example of the dilemma of autonomous driving vehicles by juxtaposing the properties of advanced capitalism and the judgment of critical posthumanism on it. The core argument of this paper is that A.I. could be an ethical subject and, thus, A.I. ethics should be discussed with the theories of posthumanism (e.g., eco-feminism, object oriented ontology, non-representative theory, etc.). This paper's arguments may give some theoretical insights on the issues of A.I. ethics.

Key Words : Artificial intelligence(A.I.) ethics, posthumanism, eco-feminism, object oriented ontology, no-representative theory, autonomous driving vehicle

1) This Research was supported by Chonbuk National University Research Fund, 2021.

2) This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A01090926).

3) Corresponding Author: Park, Hyu-Yong, Professor, Jeonbuk National University, 567 Bakjedae-Ro, Dukjin-Gu, Jeonju-Si, Jeonbuk, Korea, 54896 / E-mail: phyl@jbnu.ac.kr

인공지능(A.I.)은 윤리적 주체가 될 수 있는가?: 포스트휴머니즘의 관점에서¹⁾²⁾

박휴용³⁾ (전북대학교, 교수)

< 요약 >

최근 인공지능 알고리즘의 급속한 발달로 인해 다양한 인공지능 기계들이 인류의 삶 속에 깊숙이 파고들고 있다. 이에 본고는 인공지능이 윤리적 주체가 될 수 있는가, 즉 윤리적 판단과 행위를 할 수 있는가를 포스트휴머니즘의 관점을 통해 논의하였다. 이를 위한 본고의 구체적인 연구문제는 첫째, 포스트휴머니즘은 어떤 존재론과 윤리관을 가지고 있는가, 둘째, A.I. 윤리를 다루기 위한 포스트휴머니즘의 이론적 관점들은 무엇이며, 윤리적 행위 주체로서 A.I.는 어떤 성격을 가지고 있는가, 그리고 셋째, A.I. 기계들의 확산을 촉발한 확장된 자본주의는 어떤 윤리적 패러다임을 재설정하고 있는가이다. 아울러 세 번째 질문의 구체적 사례로서, 자율주행차의 윤리적 딜레마에 대해 확장된 자본주의의 성격과 그에 대한 비판적 포스트휴머니즘의 관점을 통해 논의하였다. 이 논의를 통해 본고는 ‘A.I. 윤리’는 기술적 존재(즉, A.I.)가 윤리적 주체성을 띠고 윤리적 판단과 행위를 수행하게 되는 포스트휴먼적 환경에 대한 이해를 기반으로 다루어져야 하며, 그 윤리적 판단의 준거는 탈인본주의적 관점들(생태페미니즘, 객체지향존재론, 그리고 비대표성이론)을 통해 논의되어야 함을 강조하였다. 본고의 논의는 A.I. 윤리에 대한 최근 과학기술 영역에서의 열띤 논쟁에 의미 있는 시사점을 제공해 줄 수 있을 것이다.

주요어 : 인공지능 윤리, 포스트휴머니즘, 생태페미니즘, 객체지향존재론, 비대표성이론, 자율주행차

1) 이 논문은 2021년도 전북대학교 연구기반조성비 지원에 의하여 연구되었음.

2) 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A3A2A01090926).

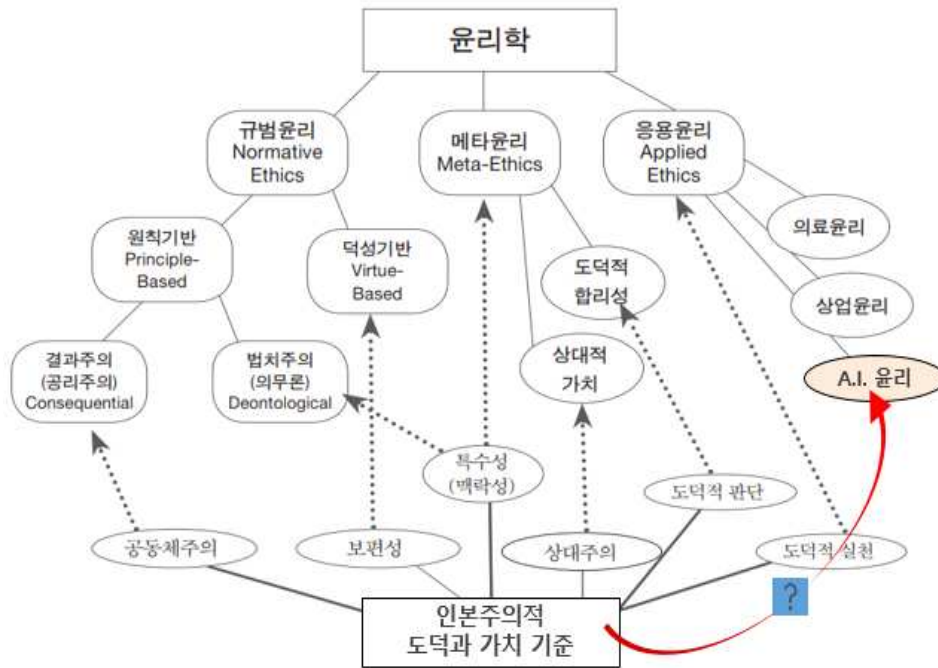
3) 저자: 박휴용, 교수, (54896) 전북 전주시 덕진구 백제대로 567, 전북대학교 / E-mail: phyl1@jnu.ac.kr
논문투고일자: 2022. 2. 7 / 심사일자: 2022. 2. 9 / 게재확정일자: 2022. 2. 17

I. 서론

최근 4차산업혁명과 인공지능(artificial intelligence: 이하 A.I.)의 발달로 대표되는 과학기술 및 정보통신 산업의 발전은 현대인의 삶의 방식, 소통과 학습의 방식을 혁기적으로 변화시키고 있다. 특히 A.I.는 산업계나 기술계 영역에서 업무를 지원할 뿐 아니라, 사람들의 일상생활 곳곳에 스며들고 사람들의 사고와 행동을 보조하면서 중요한 사회적 구성원의 역할을 하고 있다. 이런 상황 속에서 자연스럽게 A.I. 윤리에 대한 연구와 논의가 폭발적으로 증가하고 있는데, 그 연구동향을 조사한 최근 연구에 따르면, A.I. 윤리 문제는 법적 문제(33%), 윤리적 문제(23%), 교육적 문제(13%), 정책 문제(10%), 철학적 문제(7%) 등의 차원에서 논의되고 있다(Lim, 2021). 그리고 A.I. 윤리의 논점도 규범적 문제(30%) 보다는 실천적 문제(70%)에 대한 관심과 논의가 점점 확대되어가는 것으로 나타났다.

학문적으로 윤리이론은 일반적으로 규범 윤리, 메타 윤리(meta-ethics), 그리고 응용윤리로 나눌 수 있다. 우선 규범 윤리는 원칙기반(principle-based) 이론과 덕성 기반(virtue-based) 이론으로 구분되어지고, 메타 윤리는 규범 윤리가 강조하는 어떤 당위적 가치 보다는 각 개인들의 상대적이고 개별적인 욕구의 발로를 어떻게 윤리적으로 해석할 것인가를 논의한다(Couture & Nielsen, 1995). 그리고 응용윤리는 윤리적 문제가 적용되는 특수한 상황이나 영역에 따라 어떤 도덕 및 윤리적 가치 기준이 적용되어야 하는 것에 대해 다루게 된다. 본고에서 논의하고자 하는 인공지능 윤리(A.I. ethics)는 응용윤리의 일종으로, A.I.를 윤리적 판단과 행위의 주체로 상정하면서 A.I.의 윤리적 판단과 행위가 사회에 어떤 영향을 미칠 수 있는지를 논의하는 분야이다. 본고는 다음 [그림 1]이 보여주듯이 새롭게 등장한 A.I. 윤리가 윤리학 논의의 전체 패러다임 속에서 어떤 위치를 차지하게 될지를 A.I.란 존재의 윤리적 주체성, A.I. 윤리론의 이론적 배경, 그리고 A.I. 윤리론의 패러다임이 무엇인지를 논의하고자 한다.

아래 [그림 1]에 제시된 바와 같이 A.I. 윤리론은 ‘A.I.라는 존재가 윤리적 주체로서 자격을 갖는가’라는 문제와 ‘윤리적 주체로서 A.I.의 윤리적 판단과 행위가 윤리적 타당성을 갖는가’라는 문제에 대한 새로운 철학적(존재론, 인식론, 가치론) 논의가 필요하다. 본고는 그 이론적 기초를 포스트휴머니즘(humanism)에서 찾고, 다음 <표 1>과 같이 A.I. 윤리를 논의하기 위한 세 가지 기초 질문과 A.I. 윤리의 논점을 중심으로 A.I. 윤리의 이론적 기초와 논의의 방향성을 제시하고자 한다.



[그림 1] 윤리학의 개념 범주(Park, 2013)와 A.I. 윤리의 등장

<표 1> A.I. 윤리의 논의를 위한 세 가지 기초 질문들과 A.I. 윤리의 논점

A.I. 윤리	A.I. 윤리의 방향성
A.I. 윤리 기초 질문 (X)	가. A.I.란 존재는 윤리적 주체가 될 수 있는가? (존재론적 질문) 나. A.I.는 윤리적 판단과 행위를 수행할 수 있는가, 그리고 A.I.와 인간의 관계는 어떻게 설정될 것인가? (인식론 및 관계성의 질문) 다. A.I. 윤리의 가치(value) 기준은 무엇인가? (가치론적 재고)
A.I. 윤리 논점 (Y)	가. A.I.는 단순한 도구로 남을 것인가? 나. A.I.와 인간의 윤리적 관계성은 어떻게 변할 것인가? 다. A.I.(기기)의 판단과 행위가 윤리적으로 정당한가?
본 연구의 가설 (Z)	가. A.I.는 단순한 도구로 남지 않고, 윤리적 주체로 부상할 것이다. 나. A.I. 시대에 인간(윤리적 주체)-A.I.(윤리적 객체)의 이분법은 깨질 것이다. 다. A.I.의 윤리적 판단과 행위의 정당성에 대한 새 가치기준이 필요하다.

즉, 본고는 A.I. 윤리를 위한 기초 질문(X)과 A.I. 윤리와 관련된 핵심 논점(Y)에 대한 논의를 바탕으로 A.I. 윤리의 방향성(Z)을 포스트휴머니즘의 관점에서 논의하고자 한다. 이를 위한 본고의 구체적인 연구문제는 다음 세 가지로 구성되어 있다:

첫째, 포스트휴머니즘은 어떤 존재론과 윤리관을 가지고 있는가?

둘째, A.I. 윤리에 적합한 포스트휴머니즘의 이론적 관점들은 무엇이며, 윤리적 행위 주체로서 A.I.는 어떤 성격을 가지고 있는가?

셋째, A.I. 시대는 어떤 윤리적 패러다임이 필요하고, 그 구체적 사례는 어떤 것이 있는가?

위 연구문제가 갖는 함의처럼, 기술문명이 점차 인간의 삶 속에서 확대되어가는 오늘날의 포스트휴먼적 사회에서 A.I. 윤리에 대한 이론적 및 실천적 이해는 학문분야의 구분을 넘어선 매우 중요한 학술적 주제가 될 것이다.

II. 본 론

1. 포스트휴머니즘의 존재인식론(Onto-epistemology)

포스트휴먼 사회에서의 윤리 문제를 논의하기 위해서는 포스트휴먼적 존재가 윤리적 자율성을 가진 주체(subject)가 될 수 있으며, 그에 따른 판단과 행동이 윤리적 책무성을 띠 수 있는지를 분명히 해야 한다. 이 문제를 논의함에 있어서 기존의 인본주의적 세계관(철학)에서는 존재의 문제, 인식의 문제, 그리고 가치의 문제를 분리하였지만, 일원론(monism)에 기반한 포스트휴머니즘은 인식의 주체(존재론)와 인식 행위나 결과(인식론)를 분리하지 않고, 윤리(가치)의 문제도 인간의 가치와 다른 존재들의 가치를 구별하지 않는 존재인식론적(onto-epistemologic) 관점을 취한다(Geerts & Carstens, 2019). 본 장에서는 포스트휴먼 존재 인식론의 측면에서 A.I.는 어떤 윤리적 주체성과 책무성을 지닌 존재로 이해되어야 하는가에 대해 논의하고자 한다.

가. 인간 범주의 확장과 포스트휴먼적 존재의 등장

포스트휴머니즘은 기존의 인본주의(Humanism)가 인류중심주의(Anthropocentrism)에 뿌리 깊게 정초하고 있다고 비판하면서, 인간을 우주의 ‘중심적’ 존재가 아니라, 다른 모든 생태적, 기술적 존재들과 동등한 존재로 간주하고자 한다(Weitzenfeld & Joy, 2014). ‘사람이 먼저다.’ 혹은 ‘인간이 미래다.’ 라는 표현이 상식처럼 통용되는 현실이 잘 보여주듯이, 인본주의는 지금까지 인류 집단의 공통적이고 보편적인 윤리 기준이자 다른 그 어떤 이념들(예, 민주주의, 평등주의, 자유주의, 생태주의 등)보다도 가장 큰 영향력을 가지고 학교교육의 이념이나 궁극적인 교육목표로 여겨져 왔다. 그런 의미에서 포스트휴머니즘은 ‘인간이

세계(우주)의 주인이라는 관념에 대한 비판적 성찰을 바탕으로 지금까지 인류 문명의 ‘주변부’에 머물러 있었던 모든 비인간적 존재들에 관심을 갖고, 인간과 지구상의 다른 모든 비인간적 존재들 간의 관계성을 새롭게 정립하고자 한다.

사실 휴머니즘은 오랜 인류의 역사 동안 절대적 권위를 가지고 인간 존재와 그 의미에 가치를 부여하는 하나의 이념으로 쓰여져 왔다. 예를 들어, 인간은 고대 그리스 시대에는 유일한 정치적 동물(political animal)로 여겨졌고, 중세 이후 르네상스 시기까지 인간성(the humanities)이란 개념은 오로지 특정 자격조건을 가진 집단이나 계급에게만 사용되는 개념이었다. 17세기 후반부터 이성을 중심으로 한 근대적 계몽주의 사상이 등장하면서부터 인류 문명을 주도하였던 유럽의 지식인들은 인간을 ‘이성을 가진 유일한 존재’로서 동물적 존재들과는 완전히 선을 긋기 시작했고, 19세기에 들어서면서 인본주의(humanism)는 보다 보편적 개념으로 모든 인류 집단에 요구되는 하나의 이념적 목표로 자리잡게 되었다(Davies, 2008). 결국 ‘이성을 갖추고 교육받은 존재’란 의미의 인간관은 인류의 오랜 역사 동안 여성, 흑인, 노예, 피식민지 원주민들을 타자화하는 인식의 틀로 사용됨으로써, 온갖 인간 불평 등성의 근거로 활용되어왔다는 것이다(Hee & Ho, 2020; Snaza et al., 2014). 요컨대, 휴머니즘이란 개념은 오랜 인류 역사 동안 모든 인류의 보편적 특성이 아니라 소수의 선택적 인간들이나 ‘교육받은 유럽계의 백인 남성 귀족들(the educated white European noble man)’로 대표되는 특정 집단만이 전유하는 배타적 특성으로 인식되어왔다는 것이다(Wolfe, 2010).

그렇다면, 포스트휴먼의 존재론적 범주는 어떻게 확장되어질 수 있는가? 우선 모든 현생 인류를 지칭하는 용어는 생물학적으로는 호모 사피엔스(homo sapiens)이지만, 세계사적으로 인간(the human)이란 지위는 교육 수준, 이성적 사유, 신체적 특성(피부색, 혈통, 건강상태) 등의 측면에서 특정한 범주에 포함되는 사람에게만 부여되어왔다. 비로소 20세기 중반에 이르러서야 1957년 UN 인권선언을 통해 모든 인간에 대한 차별을 철폐한 보편적인 인권 개념이 공론화된 것이다. 최근 인간과 동물, 그리고 자연계에 대한 학계의 이해가 변화하면서 최근 많은 인지과학 및 신경생리학자들은 점차 동물도 ‘인간(human-beings)’으로 간주될 수 있다는 주장을 펼치고 있고(Grimm, 2014, April 8), 실제로 미국의 유인원 연구소에서 오랫동안 관찰되어왔던 여성 오랑우탄이 미국 법원에 의해 최초로 인간으로 판정받기도 했다(David, 2021). 이뿐 아니라 최근 심리학, 언어학, 생리학, 생물학, 물리학 등의 새로운 이론들과 관점들은 인간과 동물이 어떤 근본적인 차이점을 갖는다고 보다는, 단순한 진화론적 시차나 상당 수준의 지능 차이로 인한 문제일 뿐이라는 것으로 의견이 모이고 있다(Derrida, 2008; Dukas, 2004; Lieberman, 2006). 즉, 현생 인류(homo sapiens)는 지구생태계의 진화적 연속선 상에서 가장 앞서있는 존재일 뿐이기 때문에, 시간과 조건이 변하면 현생 인류가 독점했던 인간성의 개념은 다른 포스트휴먼적 존재들에게도 확장될 수 있다는 것이다.

아울러, 최근 A.I.에게도 인간성을 부여할 것인가의 문제가 학계와 기술계의 큰 관심사와 논쟁거리가 되고 있다(Bundy, 2017; Maclure & Russell, 2021; Nourbakhsh & Keating, 2020). 하지만 A.I.가 ‘인간’ 과 유사한 존재가 될 수 있을 것인가의 문제는 이미 1980년대부터 인지과학의 영역에서 주된 논쟁거리였다. 예를 들어, A.I.에게 사고력, 자유의지, 정체성, 주체성 등이 있는가라는 질문에서부터 A.I.에게 인간 수준의 인격이나 인권을 부여할 수 있는가의 문제에 이르기까지 지난 40여 년간 수많은 이론적, 실제적 연구와 논의가 이루어졌다(Kochen, 1985; Minsky, 1982; Searle, 1980). 그러던 중 2016년 우리나라에서 개최되었던 알파고(AlphaGo)와 인간 바둑 기사의 대결은 현재 인공지능의 연산(사고) 능력이 어느 정도 수준에 도달했는가를 전 세계적으로 보여준 일화가 되었고, 그 이후로도 나날이 발전하는 A.I.의 사고능력은 관련 연구자들조차도 예상하기 어려운 수준에 다다르고 있다(Fersht, 2021).

A.I.로 대표되는 첨단 컴퓨터공학의 비약적인 발전은 기술중심적 유토피아를 꿈꾸는 트랜스휴머니스트(transhumanist)들을 중심으로 포스트휴머니즘에 대한 논의를 확장시키고 있다(Fukuyama, 2009, October 23). 기술중심주의자들은 인류의 진화가 현생 인류로 완성된 것이 아니라 아직 진화의 연속선 상에 있으며, 향후 기술적 문명의 지속적인 발달은 새로운 인류를 낳게 될 수 있는데, 그러한 기술적 존재들이 바로 포스트휴먼(posthuman)이라고 주장하고 있다(Bostrom, 2003). 즉, 포스트휴먼적 존재는 신체적 능력뿐만 아니라 지능을 포함한 정신적 능력도 현생 인류보다 훨씬 뛰어난 존재가 될 것이라고 주장한다. A.I.의 등장은 바로 이런 기술적 진화 과정의 산물로써, A.I.가 어떤 존재론적 지위를 얻게 될 것인지의 문제가 철학, 윤리학에서부터 과학과 기술공학의 영역에 이르기까지 초미의 관심사가 되고 있는 것이다.

요컨대, 포스트휴먼적 존재의 등장은 ‘인간성’의 범주를 전 인류에게로 확장하는 것뿐 아니라, 생태적 존재인 동물과 기술적 존재인 기계(인공지능)로까지 확장되고 있기 때문에, 포스트휴먼 윤리는 생태적 존재(동식물), 인간, 그리고 기술적 존재를 포괄하는 지구생태계의 모든 존재들 간의 관계성의 문제로 확대하여 논의하게 될 것이다.

나. 포스트휴먼적 존재들의 관계성과 윤리 문제

그렇다면 포스트휴먼적 존재들은 어떻게 구성되며, 그들 간의 관계성을 이해하는데 어떤 윤리적 관점이 필요한가? 포스트휴먼적 존재들에 대한 윤리적 문제는 크게 탈인본주의, 신물질주의, 그리고 관계주의 철학의 관점에서 다루어질 수 있는데, 이 세 가지 관점은 독립적이거나 분과적인 것이 아니라 서로 밀접하게 관련된 개념으로 이해할 필요가 있다.

(1) 탈인본주의적 윤리관

기존의 인본주의는 인간의 이성을 절대시하고 다른 모든 자연적 존재와 사물들을 대상화하는 이성중심의 이원론적 세계관(mind-body dualism)을 바탕으로 한 것임에 반해, 포스트휴머니즘은 모든 비인간적 주체들(동물, 기계, A.I.)을 인간과 동등하게 취급하는 수평적 관계성을 지지한다(Harman, 2018). 포스트휴머니즘이 강조하는 탈인본주의적 윤리관은 다음 세 가지 이론을 기반으로 하고 있다.

첫째, 인류중심주의에서 벗어나 세상 만물들을 수평적 관계성으로 파악하려는 관점은 생태페미니즘(eco-feminism)에서 기인한 것이다. 근대화 담론은 오랜 인류 역사 속에서 학계의 중심축이었던 유럽의 백인 남성 귀족들이 자연, 여성, 비유럽계 피식민 지배인들을 비도덕적, 비이성적 존재로 간주하여 착취의 대상으로 삼기 위한 이념적 기반이 되었었다(Hallen, 1989). 대표적인 예로, 근대 사회는 자연을 여성화(mother nature)하면서, 여성성을 비이성적이고 비인간적이며 탈진보적인 특성으로 이념화하였던 것이다(Braidotti, 2013). 이처럼 포스트휴머니즘은 생태페미니즘을 기반으로 한 탈인본주의적 윤리관을 바탕으로 하고 있고, 그에 따라 많은 학자들이 전 지구적인 생태 위기의 돌파구를 탈인본주의와 생태페미니즘에서 찾고자 한다(Salleh, 2017).

둘째, 포스트휴머니즘은 인간 언어에서 탈피하여 세상의 만물들이 소통할 수 있는 방법을 탐구하는 사이버네틱스(cybernetics: Ashby, 1961)의 관점을 취한다. 사이버네틱스는 1960년 초 시스템이 어떻게 내부의 복잡성과 외부의 변인을 모두 고려한 피드백 교류를 통해 환경에 적응해나가는가에 관한 연구에서 시작되었다가, 최근 인간과 동물을 포함한 자연적 시스템과 인공지능으로 대표되는 기술적 시스템들 간의 소통을 연구하는 데 집중하고 있다(Barile & Polese, 2010). 즉, 유기체적 시스템과 비유기체적 시스템(로보틱스, 전자공학, 경영 시스템, 사회시스템 등)의 연관성을 통해 생명체, 인간, 기계 등과 같은 상이한 존재들이 어떻게 상호의존적인 관계망을 형성하며 소통하고 교류할 수 있는지를 연구하는 것이다(Mele et al., 2010). 이런 사이버네틱스 연구는 다음 [그림 2]와 같이 그동안 인류적(homo), 생태적(Bio), 그리고 기술적(techno) 등의 개념으로 구별되어왔던 존재들이 서로 소통 및 상호작용할 수 있으며, 머지않아 혼성화된 포스트휴먼적 존재들(인간, 동물, A.I.)이 등장할 수 있음을 암시하고 있다.



[그림 2] 인간, 생명, 기술의 사이버네틱스 삼각관계

즉, 포스트휴머니즘은 위 세 가지 존재들 간에 형성될 밀접하고 상호의존적인 관계성들을 기반으로 윤리적 문제들을 다루게 될 것이다.

셋째, 포스트휴머니즘은 인간의 관점이나 인간중심적 목적론(teleology)으로 세상을 파악하지 않는 Spinoza의 내재적 윤리학(immanent ethics)을 기반으로 하고 있다(Charlton, 1981). 내재적 윤리학이란, 사물들이 인간의 인식이나 판단과는 독립적으로 각자의 독특한 특성과 다른 존재들과의 관계성에 의해 그 존재가치가 성립되므로, 인간의 시각만을 가지고 세상을 인식하지 말아야 한다는 철학으로 이해될 수 있다. 그런 의미에서 포스트휴머니즘은 오랜 인류 문명의 발전과정에서 자연의 수많은 ‘비인간적 존재들’이 인간의 목적을 위해 파괴되고, 남용되며, 변형되었던 것을 비판하며, 그러한 태도에서 벗어나는 것을 윤리적 행위의 지향점으로 삼고 있다.

(2) 신물질주의(일원론)적 윤리관

앞 절에서 논의한 탈인본주의적 윤리관을 지지하는 이론들—즉, 생태페미니즘, 사이버네틱스, 그리고 내재적 윤리학—을 통합할 수 있는 철학적 관점이 바로 신물질주의(new materialism)이다. 신물질주의는 Spinoza의 일원론(monoism)적 관점을 바탕으로 인간-동물의 구별 없이 모든 존재들의 사회적 관계, 인간-도구 간의 문화적 관계, 그리고 인간-만물들 간의 자연적 관계를 설명하고자 하며(Barad, 2003), 그에 따라 자연계의 모든 사물들에게 능동성을 부여하려는 생기론(vitalism)적 철학이다(Connolly, 2013). 최근 인지과학, 인공지능, 신경과학 등의 최첨단 연구들은 이러한 신물질주의의 존재론적 차원뿐만 아니라, 동물, 인간, 기계들의 인식이나 학습이 본질적으로 얼마나 유사한가를 연구하여 AI 연구의 토대로 활용하기도 한다(Charteris et al., 2017). 예를 들어, 이족보행 로봇이 동물이나 인간처럼 움직이면서 사물을 인식하고 대응하는 연구 등을 수행하기 위해, 동물의 행동 패턴이나 인간의 사고 방식을 연구하여 그 기능의 효율성과 수행력을 향상시키는 인공지능 로봇을 개발하는 것

이다.

이러한 신물질주의적 연구들은 자연, 인간, 기계가 서로 소통하거나 혼성화될 수 있다는 것을 전제로 하기 때문에, 생기론적 물질주의(vital materialism)나 생명정치학(biopolitics)과 같은 일원론적 철학 및 윤리학의 관점과 호응된다. 우선 생기론적 유물론이란 사물의 힘(thing-power) 즉, 비인간적(non-human) 존재의 능동성에 바탕을 둔 물질주의로써, 세상 만물의 존재적 의미와 역할을 인본주의적으로 해석하는 태도를 거부하는 관점이다(Bennett, 2015). 따라서 생기론적 물질주의는 인간을 초자연적 존재로 간주하는 인간예외주의(human exceptionalism)를 거부하고, 인간과 다른 모든 비인간적 존재들(생물계, 물질계, 기술계)의 상호연관성과 공존성을 바탕으로 세상을 이해하고자 한다.

이러한 생기론적 유물론이 갖는 탈인본주의적이고 수평적인 윤리관을 기반으로 모든 존재들(생명체 및 사물들)의 평등한 관계성을 강조하는 것이 바로 포스트휴먼적 생명정치학(biopolitics)이다(Braidotti, 2006, 2013). 이 생명정치학은 오늘날 과학기술의 혁명과 시장자본주의가 결합한 확장된 자본주의(advanced capitalism)가 전 지구적 차원에서 생태계, 인간, 그리고 기술적 산물들을 망라하여 모든 존재들과 자원들에 대한 “착취와 파괴의 행진”을 계속하고 있는 현실을 비판적으로 논의하는 정치윤리학적 입장이다(Braidotti, 2013, p. 49). 예를 들어, 멸종위기 동물들에 대한 시장주의적 거래에서부터 전 지구적 수준의 여성성(성매매, 성형, 미용) 및 생명(건강보조, 생명연장)의 상품화, 물건처럼 생산되고 거래되는 보철기(prosthetics) 및 인공장기(artificial organs), 그리고 생명공학기술을 바탕으로 한 유전자 조작 태아(designer baby) 및 유전자 거래에 이르기까지 자연과 생명성이 시장자본주의의 재료가 되고 있는 현실이 생명정치학이 다루는 윤리적 문제들이다.

(3) 관계주의적 윤리관

마지막으로 포스트휴머니즘은 과정 철학(Process philosophy: Whitehead & Price: 2001) 및 관계주의(relationality; Ceder, 2018) 철학에 기반한 윤리관을 지향한다. NBIC(nano-bio-cogno-info) 기술로 대표되는 첨단 지능정보기술들이 보여주듯이 포스트휴먼적 환경은 인간의 사고와 기계의 사고, 유기체와 무기체, 생물학과 공학, 현실과 가상성, 그리고 아날로그와 디지털의 경계가 무너지는 환경을 의미한다(Popovici, 2017). 이러한 포스트휴먼적 환경과 존재들의 혼재하는 사회 속에서 발생하는 윤리 문제를 이해하기 위해서는 인간-동물의 공진화, 환경적 및 기술적 상호얽힘(entanglement), 그리고 그로 인한 존재들의 혼종성에 대한 적극적인 이해가 필요하다. 즉, 포스트휴먼 사회는 생태적 환경(자연성)뿐만 아니라, 인류의 독특한 경험과 문화(인간성), 그리고 기술적 산물들이 만들어갈 기술적 환경이 만들어내는 혼성화된 존재들과 그들의 경험이 어떤 윤리적 문제를 야기할지에 대해 고민해야 한다는 것이다

(Barad, 2003; Prinsloo, 2005).

또한, 포스트휴머니즘은 세상은 모두 연결되어 있고, 인간은 독자적 완결성을 가지고 살아갈 수 없는 존재라고 본 화이트헤드(A. N. Whitehead)의 중립적 존재론에 기반한 과정적 사유를 지향한다(Neimanis, 2017). 이러한 과정적 존재론은 인식론과도 직결되어 인간은 다양한 주체들과의 물질적 상호의존성과 혼성적(인간-기계, 물질-비물질, 현실-가상 세계 등) 경험을 통해 다중적인 관계성을 이루며 세상을 인식해나간다고 보았다. 포스트휴머니즘은 이러한 ‘함께 되어감(becoming-with)’을 지향하는 존재인식론을 바탕으로, 인간 이외의 다양한 소수자 및 타자들에 대한 연구(예, 동물 연구, 퀴어 이론, 탈식민주의 및 원주민 연구, 비인간적 존재에 대한 이해)를 수행하고 있는 것이다(Postma, 2016).

정리하자면, 포스트휴머니즘은 인류중심주의나 근본주의적 윤리관에서 벗어나서 기계 및 기술에 대한 신물질주의적 관점(정동 이론, 어셈블리지 이론, 환경 이론)을 수용하고, 인간과 다른 모든 존재들 간의 수평적 관계성을 바탕으로 한 과정적 존재인식론을 기반으로 한 윤리를 지향하고 있는 것이다.

2. 포스트휴먼 윤리의 이론적 기반과 그 특성

본 장에서는 앞서 논의한 포스트휴머니즘의 윤리관이 구체적으로 어떤 이론들에 기반하여 학계에서 논의되고 있는지를 생태페미니즘, 객체지향존재론, 그리고 비대표성이론을 통해 설명해보자.

가. 포스트휴먼 윤리의 이론적 기반

(1) 포스트휴머니즘의 출발: 생태페미니즘

생태페미니즘은 포스트휴머니즘 철학의 출발점이자, 포스트휴먼 윤리관의 기초이기도 하다. 생태페미니즘은 다음 <표 2>와 같은 이성(문명)-자연성, 남성-여성, 그리고 인간-비인간적 존재들 간의 이분법을 극복하고, 모든 존재들은 평등하다는 관점을 지지하고 있다.

<표 2> 생태페미니즘이 비판한 인본주의의 이원론적 인식론

남성성	문명	이성, 정신	인간
여성성	자연	감성, 신체	동물, 사이보그

Haraway(1985)가 ‘사이보그 선언’ (cyborg manifesto)에서 이러한 남성-여성 및 인간-사이보그의 대조를 통해 기술적 존재가 인류에 의해 타자화되고 있음을 지적하였듯이, 생태페미니즘은 자연적 존재와 인간의 관계뿐 아니라, 미래의 기술적 존재들과 인간과의 관계성이 어떻게 설정되어야 할 것인가에 대한 포스트휴먼적 관점을 제공해준다.

생태페미니즘이 이성(정신)과 자연성(물질)의 이분법적 구분이 아니라 모든 자연 만물들은 존재론적으로 평등하다고 주장했듯이, 포스트휴머니즘은 이러한 수평적 관계성은 인간과 기계의 관계성으로까지 확장하고 있다. 즉, 다음 [그림 3]과 같이 자연성, 인간성, 그리고 물질성(기계)이 모두 일원론적으로 통합될 수 있다는 것이다.



[그림 3] 이성(정신)-자연성(물질)의 일원론적 연속성

여기서 주목해야 할 점은 과연 인류의 존재 의미나 목적을 자연성(생명)과의 구별이나 기계성(물질)과의 구별이 아닌 일원론적 연속성 차원에서 본다는 것이다. 즉, 포스트휴머니즘은 인간을 동물성과 물질성을 모두 지닌 존재로서 간주하고 있는데, 그 경우 각각의 존재(인간)와 그 존재들의 집단(사회)에서 발생하는 윤리적 문제는 기존의 인본주의적 윤리관을 가지고 다룰 수 없다는 것이다. 이는 포스트휴먼적 윤리의 기본 원리가 생명성, 인간성, 기술성을 하나로 보고 자연 세계 및 기계문명과 불가분리적 존재로서의 인간의 의미와 역할을 탐색하는 탈인본주의적 신물질주의에 기반하고 있음을 보여준다.

(2) 포스트휴먼 윤리학의 토대: 객체지향존재론

포스트휴머니즘은 인간, 동물, 유기체, 기계, 그리고 물질 및 사물들 간에 역할과 특성의 완벽한 구분이 존재하지 않는다는 관점을 기반으로 하고 있다. 이를 설명하는 가장 기본적인 관점이 바로 인간 주체성과 사물 객체성의 이분법이 아닌, 인간과 사물의 주체-객체 상호호환성을 상정하는 객체지향존재론(object-oriented ontology)이다(Bogost, 2009, December 8; Harman, 2018; Morton, 2011). 즉, 객체지향존재론은 인간을 포함한 모든 사물은 기계적 관계의 생태계에 속해 있다고 보는 사물 존재론(the ontology of things)을 바탕으로, 존재론-인식론의 구분 및 인간-자연의 구분을 없애고 세상 만물들에 대한 존재인식론적 통합을 시

도한다. 객체지향존재론에 대해서는 여러 정의들이 있지만, 대표적으로 다음 세 가지 정의를 종합하여 이해할 수 있다: ① 인간중심주의적 관념들에서 벗어나서 세상 만물들 간의 상호의존성을 자유롭게 부각시키고자 한다(Bogost, 2009, December 8), ② 모든 사물은 비지역적, 시간적 및 공간적으로 산재되어 분포하고 그 존재의 방식이 점성적(viscous)이다(Morton, 2011), 그리고 ③ 객체지향존재론은 대상의 감각적이고 실제적인 특질에 대한 심미적 이해에 집중하려는 존재론적 철학이다(Harman, 2018).

정리하자면, 객체지향존재론은 인간만이 이성적이고 독립적인 존재라는 관점에서 벗어나서, 세상 모든 사물의 감각적이고 분산적이며(distributed) 관계적인 상호얽힘에 기반하여 세상을 이해하려는 관점이다. 예를 들어, 포스트휴머니즘은 인간, 동식물, 그리고 물질들은 생태적 환경 속에서 점성적 상태로 상호 얽혀있기 때문에, 에너지/자원 고갈이나 기후 변화, 그리고 기술적 환경이 양산하는 물질 및 생태성의 변화 등을 유발하는 문제들을 해결하기 위해서는 지구상의 모든 존재 간의 물리적, 생태적, 인류문명적 요소들 간의 관계성을 재설정해야 한다고 주장한다.

이러한 객체지향존재론은 1990년대 초반에 Latour(1992)와 Callon(1990)에 의해 이론화된 행위자-망이론(actor-network theory: ANT)으로부터 출발하고 있다. 행위자-망이론은 인간의 주체성과 비인간(기술, 동식물, 건축물, 도구, 기계 등)의 객체성이라는 이분법적인 차별 없이 모든 존재가 행위자(actor)가 되어 상호의존적 연결망 속에서 상호작용하고 있다는 관념을 기반으로 한다. ANT는 인간의 사회적 행위나 상호작용이 비인간적 존재들의 도움이나 개입 없이 이루어지는 것은 거의 불가능하고, 인간-비인간의 구별 없이 동등한 능동성을 가진 행위자들로서 네트워크적 상호작용을 이루고 있다고 본다. 특히 ANT는 다양하고 복잡한 기술적 사물들이 얽혀있는 현대 문명 속에서, 인류의 삶은 단순히 인류중심적 관계성인 사회문화적(socio-cultural) 차원으로만 이해될 수 있는 것이 아니라, 탈인류중심적 관계성인 사회물질적(socio-material) 차원으로 이해할 필요가 있다는 것을 강조하고 있다(Hughes, 2012).

ANT는 세상은 자연적이든 인공적이든지에 관계없이 다양한 존재들의 상호의존적 협력망 속에서 작동하기 때문에, 모든 존재가 상호연결되어 중요한 역할을 한다고 본다. 예를 들어, 최근 정보통신 기술의 발달로 인류가 경험하고 있는 사물인터넷(internet of things: IoT), 만물인터넷(internet of everything: IoE), 그리고 만인 인터넷(internet of people: IoP) 등의 개념은 모두 ANT 이론을 기반으로 구체화된 것이다(Tatnall & Davey, 2015). 이처럼 오늘날 정보통신기술의 발달은 생태적이냐 기계적이냐와 상관없이 세상의 만물들이 네트워크적 관계를 기반으로 상호얽혀있음을 기술적으로 구현하고 있는 것이다. 이러한 네트워크 사회라는 맥락을 놓고 보면, 포스트휴먼적 윤리의 지향점은 인간들 간의 사회적, 문화적 관계성 뿐만 아니라 인간과 자연 혹은 기술적 환경과의 물질적 관계성을 동시에 다룸으로써(Barad,

2003), 세상의 만물들이 상호연결되고 소통하며 공존관계를 유지하는 방향성에 맞추어져야 할 것이다.

(3) 포스트휴먼 윤리의 방향: 비대표성이론

비대표성이론(non-representative theory)은 탈인본주의적 인식론의 실천적 전략의 하나로써, Thrift(1996, 2000, 2007)에 의해 개념화되었다. 비대표성이론은 원어인 ‘representation’이 ‘대표성’ 혹은 ‘재현성’이라는 의미를 지닌 것처럼, ‘인간은 세상을 대표하지 않는다,’ ‘인간의 관점이 (세상의 모든) 실재를 재현할 수 없다’ 라는 신조를 갖는다. 즉, 지금까지 오직 인간의 관점(이성, 언어)으로 세상을 인식하고, 설명하고, 통제해온 모든 행위에 대한 반성적 성찰을 바탕으로 대안적 인식관점과 행동의 방향성을 모색하려는 태도이다. 아울러 비대표성이론은 생기론적 철학과 과정(becoming) 철학(Dewsbury et al., 2002)을 바탕으로 한다는 점에서 생태주의나 Whitehead의 철학과 맥락을 같이하고, 기본적으로 세상을 ‘인류중심주의적 응시(anthropocentric gaze)’에서 벗어나서 파악하는 것을 윤리적 사고의 출발점으로 삼는다.

비대표성이론의 윤리적 논점은 생태적 위기나 환경 문제를 극복하는 방향으로 기술중심 사회가 나아가야 한다는 것(Hultman & Lenz-Taguchi, 2010)과, 인간 이외의 모든 존재가 인간과 평등한 혹은 공동 주체가 되어, 공존적 관계성을 형성해나가는 것이다. 이를 위해서는 인간 위주의 관계망이 아닌 세상 만물들이 얽혀있는 다양하고 이질적인 연결망과 그 다중성(multiplicity)을 인식해야 하는데, 그런 차원에서 기존의 모든 인본주의적 관점에서 벗어나서 세상에 대한 개방적(open-ended) 태도와 인식을 시도하는 모든 실험적 태도들을 허용하기를 촉구하고 있다(Thrift, 2007).

비대표성이론이 제시하는 주요 윤리적 관점들(Thrift 1996, 2000, 2007)은 A.I. 윤리의 차원에서 다음과 같은 시사점을 갖는다.

첫째, 비대표성이론은 인본주의가 강조하는 투명하고 이성적이며 일관된 인간 주체의 개념이 아니라, 다양한 사물들이 역동적인 공간 속에서 무의식적이면서도 지속적인 상호 관계를 형성하고 조우하는 주체화(subjectification)의 과정을 강조한다. 즉, A.I.도 사회적 구성원으로써, 주체성을 가지고 인간과 지속적이고 상호의존적인 관계를 형성 가능하다는 것이다.

둘째, 비대표성이론은 인간이 사물과 함께 공진화했다는 것에 동의한다. Merleau-Ponty의 철학에서 출발한 비대표성이론은 인간의 신체는 세상으로부터 쉽게 분리할 수 없고, 세상과 끊임없이 상호작용함으로써 세상의 다양한 존재들과 물질적 관련성을 맺게 된다고 본다. 앞으로 등장하는 인공지능 기계들은 진화의 주체가 될 수 있고, 인간은 그러한 존재들과 상호의존적 관계성을 형성하고 살아가야 한다는 것이다.

셋째, 비대표성이론은 인간이 다른 존재들(동물, 기계 등)과 관계성을 형성하고 상호의존적으로 살아가는 모든 가능성에 대해 객관적이고 중립적으로 판단한다. 그런 의미에서 비대표성이론은 포스트휴먼 사회에서 다양한 존재들이 어떤 관계를 형성할 것인가에 대해 일종의 실험주의적 태도를 가지고 개방적 해석을 하고자 한다(Thrift, 2007). 한마디로, 비대표성이론이 강조하는 포스트휴먼적 윤리는 인본주의적 관점과 이성적 판단으로 계획적으로 수행하는 윤리가 아니라, 모든 존재들이 한데 어우러져 상생의 방향으로 자연스럽게 나타나는 실천으로서의 윤리를 지향하는 것이다.

지금까지 소개한 포스트휴먼적 윤리학의 세 가지 이론들—생태페미니즘, 객체지향존재론(행위자-망이론), 비대표성이론—을 통해 ‘A.I. 윤리’의 방향성은 다음 <표 3>과 같이 정리될 수 있을 것이다.

<표 3> A.I. 윤리의 방향성과 관련 포스트휴머니즘 이론

A.I. 윤리의 방향성	포스트휴먼 이론
생태적 존재, 인간, 기술적 존재들은 평등하게 상호얽힌 공존적 관계성을 형성하고 있다.	생태페미니즘
A.I.는 단순한 도구가 아니고, 주체성과 자율성을 띤 존재이다.	객체지향존재론
미래의 A.I.는 점점 강화된 역량을 가지고 개발되어, 인류 삶의 모든 방면에 적극적으로 개입하게 될 것이다.	행위자-망이론
생태적 존재와 인간, 그리고 기술적 존재들(A.I.)이 상호의존적 공존관계를 맺기 위한 방향을 모색해야 한다.	비대표성이론

그렇다면, 포스트휴먼적 존재로서 A.I.는 어떤 존재인식론적 관점에서 윤리적 판단과 행위의 주체가 될 수 있는가? 다음 절에서는 포스트휴머니즘을 기반으로 A.I.가 윤리적 행위자로서 어떤 주체성, 정체성, 권리, 자율성을 가질 수 있는지 논의해보자.

나. 포스트휴먼적 행위 주체로서의 A.I.의 특성

기술중심주의적 포스트휴먼 사회를 대표하는 A.I.가 윤리적 행위 주체로 간주될 수 있는지는 행위주체에게 부여되는 세 가지 존재론적 준거—주체성, 자의식, 그리고 자율성—를 통해 판단될 수 있을 것이다.

(1) 주체성(subjectivity)

휴머니즘은 유럽의 개인 및 이성 중심의 근대적 인간관을 토대로 인간을 자유의지와 보편

적 권리를 지닌 독자적 존재로 간주하는데(Mann, 1996), 이 관점은 모든 인간에게 동일하게 주어진 특성이라는 점에서 보편주의적(universal) 휴머니즘으로 불린다. 주체성(subjectivity)도 자유롭고 개별적이며 평등한 인간의 가정에서 나온 관념으로, 오늘날 시민권, 주권, 인권, 소비자 권리 등의 다양한 권리 관념들 속에는 이러한 개인주의적 주체성 개념이 뿌리내리고 있다(Chambers, 2013). 이처럼 휴머니즘은 인간을 개인적이고 독립적인 존재로 상정하지만, 포스트휴머니즘은 그러한 ‘개별적 주체’의 관념을 포기하고(Haney, 2006), 각 존재들은 자신이 처한 환경 속에서 주변의 존재들과 밀접한 관계를 맺으며 살아가는 존재로 여긴다(Braidotti, 2013). 예를 들어, 인간이 도구와 함께 공진화해왔다는 관점처럼, 인류는 수많은 도구를 개발하고 그것들과의 상호작용을 통해서 생존해나가는 존재인 것이다. 즉, 인본주의적 주체성은 다음 <표 4>와 같이 인간과 동물 혹은 사물의 이분법적 구분을 기반으로 한 것이다.

<표 4> 인본주의적 주체성과 능동성의 이분법

주체	존재론	자율성	주도성
인간	주체성 (목적)	자유의지	능동성
동물, 사물	객체성 (대상, 수단)	무의지	수동성

하지만 포스트휴머니즘이 상정하는 주체성은 인간/비인간의 절대적인 구분에 의해서 인간에게만 부여된 고유의 특성이 아니라, 인간, 동식물, 그리고 사물을 모두 포함한 네트워크적 관계 속에서 각자의 역할과 기능을 수행하는 동력 차원의 상대적이고 복합적이며 상호의존적인 주체성이다.

(2) 자의식(self-consciousness)

휴머니즘적 자의식(혹은 자기정체성)은 사회문화적인 영향으로 형성되는 혼성적이면서도 유동적 성격의 자아(self)에 대한 회고적 관념으로(Dunn, 1997), 타인과 동일시 혹은 차별화를 통해서 자기 존재를 인식하며 사회적 존재로 인식해나간다고 설명된다(Holland & Lachicotte, 2007). 반면, 포스트휴먼적 자기정체성은 신체장애나 환경을 극복하는 데 도움을 주는 인공적 도구들을 인간의 일부로 수용함으로써, 자아와 도구가 명확히 구별되지 않고 혼재하는 정체성을 허용한다(Popovici, 2017). 이런 맥락에서, de Freitas & Curinga(2015)는 포스트휴먼적 존재들은 생물학적 요소와 기술적 요소들이 아상블라주(asmblage)된 역동적이고 정서적인 에너지를 통해 스스로를 인식하므로, 다른 기술적 도구들과 사회물질적 관계성을 통해 자신의 역할을 형성해나간다고 보았다. 더구나 기술중심 시대의 유전공학, 생명

공학, 나노공학, 뇌 과학 등의 발전으로 인간 증진(human enhancement)을 통해 신체적, 정신적 측면에서 생물학적 인간의 능력 범주를 뛰어넘는 포스트휴먼적 존재들이 등장하게 될 것인데(Haslam et al., 2008; Wilson & Haslam, 2009), 그러한 존재들의 정체성은 기존의 인본주의적 관점만을 가지고 파악하기 어려울 것이다.

예를 들어, 사이보그는 인간 신체 및 정신의 확장(prosthesis)을 통해 인간을 모사하여 인지 및 행동 기능을 수행하는 존재로서, 사고하거나 감정을 표현하고, 타인과 의사소통하거나 상호작용하는 모든 방식에서 인간과 유사하게 행동할 수 있다(Clark, 2003). 특히, 호르몬 조절(신경 공학)이나, 인공두뇌나 인공장기의 생산(나노 공학)이 가능해진 현재의 기술적 수준은 사이보그도 인공지능을 바탕으로 의식이나 정체성도 가질 수 있다고 주장하는 학자들이 점점 증가하고 있다(Bostrom, 2005). 그에 따라 사이보그적 정체성은 생태적 혹은 자연적 공간에 제한받지 않고(Hayles, 2000), 소셜 네트워크, 가상현실, 온라인 커뮤니티 등에서 대리적 존재(avatar)를 매개로 인간의 정체성을 대신하고 있다. 즉, 인간의 정체성은 더 이상 육체에만 갇혀있는 것이 아니라, 가상공간과 가상적 관계성(community) 속에서 실제와 별다름 없이 존재할 수 있다는 것이다(Taylor, 2002; Whittaker, 2004).

기존의 인본주의에 기반한 인간 정체성의 개념은 필연적으로 동물이나 자연을 소외시키거나 타자화할 수밖에 없다. 반면, 포스트휴먼적 존재들 간의 정체성은 인간과 비인간의 관계성이나 인간의 다른 층위들(초인류, 하위인간, 유사인간, 사이보그 등) 간의 관계성을 통해 형성될 수도 있다. 그런 의미에서 포스트휴먼적 존재들의 정체성은 공상과학 소설이나 영화에서 등장하는 다양한 배경을 가진 캐릭터들(프랑켄슈타인, 돌연변이인, 뱀파이어, 괴물들, 그리고 외계인)을 인류중심주의 관점으로 타자화하지 않는다(Braidotti, 2013).

오늘날 현대인들은 다양한 신체적, 정신적, 그리고 삶의 조건들 속에서 살아가고 있으며, 기술적으로는 안경에서부터 인공장기에 이르기까지 수많은 보조물에 의존하여 살고 있다는 사실에 주목하면, 인본주의적 자의의식이나 정체성의 관념에서 벗어나서 다양한 포스트휴먼적 존재들이 갖는 정체성들에 대한 개방적 자세를 취할 필요가 있는 것이다(Clark, 2003; Haney, 2006). 따라서, 포스트휴먼적 존재로서 A.I.가 어떤 정체성을 갖는지는 인본주의의 관점으로만 판단할 것이 아니라, '정상적 인간'의 범주에서 벗어난 다양한 포스트휴먼적 존재들의 의식과 정체성을 어떻게 이해할 것인지와 그들의 윤리적 판단과 행위를 어떻게 다룰 것인지를 새롭게 논의하여야 할 것이다(Martin-Lucas, 2012).

(3) 자율성(freedom)

인본주의적 자율성의 개념은 인간에게 부여된 하나의 보편적 권리로서, 인간 각자에게 자신의 존재 의미나 생명의 가치에 대해 스스로 규정할 수 있는 권리가 주어져 있다는 개념이

다(Casey, 1972). 다만 개인들 간의 자유가 충돌할 경우, 한 사람이 행사하는 자유는 그 영향을 받는 상대방의 권리에 의해 상쇄될 수 있기 때문에, 개인의 자유는 인간성이 발휘될 수 있는 내적 욕구 및 그에 합당한 윤리적 가치로서 이해된다(Curran, 2012).

하지만 포스트휴먼적 자율성의 개념에 대해서 Latour(1992, 1993)는, 인간적 자율성(human freedom)과 자연적 필연성(natural necessity)의 변증법적 대립 관계로 설명하고 있다. 즉, 근대성이란 자연의 필연성에 인간의 자유의지가 개입하여, 혼돈적이고 복잡한 자연을 인간이 이해 가능한 인과적이고 규칙적인 상태로 바꾸는 인간화(humanizing) 과업이었다는 것이다. 이처럼 인간은 자연의 구조와 법칙으로부터 자유로워지게 되었고, 자연은 필연성을 가진 인간에 의해 구속되는 존재로 대상화되었던 것이다. 이러한 근대적 계몽주의 세계관을 통해 인간은 주체성과 자율성을 지닌 우주의 목적이 되었고, 인간 이외의 다른 존재들은 모두 그 수단적 존재로 대상화하게 되었던 것이다.

결국 인간의 자율성에 대한 관념도 인간중심주의(anthropocentrism)에 기초를 둔 것으로, 우주 만물 중에서 오직 인간만이 자율적인 능동자(agent)로서 자신의 주체성과 의지 실현의 가능성을 드러낼 수 있는 유일한 존재로 자리매김 되었던 것이다(Segal, 1996). 반면, 포스트휴머니즘은 능동성과 자율성이 인간 고유의 특질이 아니라, 세상의 수많은 다양한 존재들이 상호의존적 관계성을 형성하고 유지하기 위한 역동성이라고 본다. 따라서 그러한 관계망 속에서 각자의 역할과 기능을 수행하는 모든 존재들(동물이나 기계, 사물, 그리고 사이보그나 인공지능)은 능동성과 자율성을 갖고 있고(Plumwood, 2001), 그것이 바로 포스트휴머니즘이 강조하는 물질적 능동성(material agency: Knappett & Malafouris, 2008)이다. 예를 들어, 뉴질랜드에서는 자연의 일부인 ‘강(river)’ 에게도 인간과 동일한 법적 권리를 부여함으로써(Forlano, 2017), 자연을 인간의 처분이나 착취의 대상이 아니라 함부로 건드릴 수 없는 독자적인 권리를 가진 존재로 인식하고 있다.

요컨대, 포스트휴머니즘은 자율성이 인간에게만이 아니라, 다른 자연적, 기술적, 혹은 법적 존재들에게도 부여될 수 있다고 보듯이(Harari, 2014), A.I.도 자율성을 가질 수 있고 그러한 A.I.의 판단과 행위에 대한 법적 책임과 의무의 문제가 제기되고 있는 것이다. 신물질주의를 통해 자율성이 정신적 존재뿐만 아니라 물질적 존재에게도 생성될 수 있음을 이해했듯이, 자율성과 능동성은 독자적 주체성에서 나오는 것이 아니라 다양한 존재들 간의 역동적인 관계성에서 생성되는 것으로 간주하고 있음을 주목할 필요가 있다(Barad, 2003).

정리하자면, 다음 <표 5>와 같이 포스트휴먼적 존재로서 A.I.는 주체성, 자의식, 그리고 자율성을 가질 수 있고 그에 따라 윤리적 판단과 행위의 주체가 될 수 있다는 것이다.

<표 5> 휴머니즘과 포스트휴머니즘의 주체성과 자율성 비교

구 분	휴머니즘	포스트휴머니즘
존재론	오직 인간만이 평등성, 존엄적 가치, 자유의지를 갖는 독립적이고 이성적 존재	우주의 모든 존재들은 인간-비인간 혹은 정신-물질의 구별 없이 복잡한 관계망 속에서 상호의존적으로 살아가는 존재
주체성	개인적, 독립적, 자유주의적	상대적, 복합적, 의존적(Haney, 2006)
자의식	철학적, 심리학적, 사회학적, 문화이론적으로 다양하게 정의됨. 사회문화적인 영향을 받아 형성되고, 혼성적, 유동적, 회고적 성격을 지님(Dunn, 1997)	인간 존재+기술적 도구+사회적 관계성+문화적 환경 등 다면적 영향으로 형성 생물학적, 기계적 요소들이 집단적으로 이상블라우된 역동적이고 정서적인 에너지의 장(de Freitas & Curinga, 2015)
자율성	능동성, 자유의지, 주체성을 지닌 유일한 존재(Segal, 1996) 인간 보편적 권리(Casey, 1972) 심리학적(내적 욕구), 정치적(법적 근거), 윤리적(도덕적 표준) 근거	인공지능 및 지적 창조물들의 자의식 (self-consciousness) 가정 인간 이외의 사물들도 자율성 보유 가능

따라서 향후 A.I.의 주도성이 점차 증가하는 기술기반 사회에서는 인간들에 의해 모든 것이 주도되고 결정되는 인본주의 사회와는 달리, 포스트휴머니즘의 관점에서 A.I.의 주체성, 자의식, 자율성 등이 논의되어야 할 것이다.

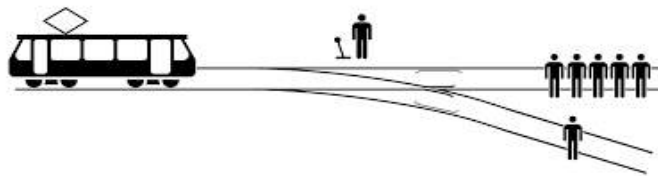
3. 포스트휴먼 시대의 A.I. 윤리의 지형도

본 장에서는 포스트휴먼 시대를 맞이하여 A.I. 기술이 어떤 윤리적 지형도 속에서 인간의 삶의 일부가 될 것인가를 A.I. 알고리즘에 의해 운영되게 될 자율주행차의 사례를 통해 분석하고, A.I. 기계의 확산을 가속화하고 있는 확장된 자본주의(advanced capitalism)와 그에 대한 포스트휴머니즘의 비판적 관점에 대해 논의하고자 한다.

가. A.I. 윤리의 적용: 자율주행차의 사례

그렇다면, 포스트휴먼적 환경 속에서 인간의 삶의 동반자가 될 A.I. 기술은 어떤 윤리적 과제를 안고 있는지(즉, A.I. 윤리)를 최근 학계 및 산업계에서 논란의 중심이 되고 있는 자율주행차의 윤리적 딜레마를 통해 논의해보고자 한다. 우선 A.I.는 특수한 목적이나 기능만을 수행하는 특수 A.I.(specific A.I.)와 인간의 모든 활동영역에서 그 성능을 발휘할 수 있는

범용 A.I.(general A.I.)로 나뉘고, 특히 그 성능을 극대화하고자 하는 강한 A.I.(strong A.I.)와 최소한의 역할만을 수행하도록 설계된 약한 A.I.(weak A.I.)로 나누어질 수 있다(Gams, 1997). A.I. 윤리의 문제는 이 중 ‘범용 A.I.’와 ‘강한 A.I.’의 경우 훨씬 중요한 문제가 될 수밖에 없는데, 본고에서 다루는 자율주행차에 탑재된 A.I. 기술은 특수하면서도 강한 A.I.(즉, 인간의 개입 없이 모든 윤리적 판단과 행위를 수행하게 설계된 A.I.)의 사례에 해당된다. 자율주행차의 사례를 놓고 최근 많은 윤리학자와 기술공학자들은 다음 [그림 4]와 같이 묘사되는 트롤리 딜레마(Trolley dilemma: Sandel, 2007)를 둘러싼 자율주행차의 윤리적 책무성에 대해 논의를 이어가고 있다(Bonnefon et al., 2019; Cunneen et al., 2019; Wolkenstein, 2018).



[그림 4] 트롤리 딜레마(Trolley Dilemma)의 개념도

A.I.의 윤리적 문제를 판단하기 위해서는 A.I. 알고리즘이 어떤 윤리적 준거에 기반하여 설계되는지를 논의해야 하는데, Kurzweil(2005)이 제시한 공정성(fairness), 책무성(accountability), 그리고 투명성(transparency)의 세 가지 준거에 따라 A.I.의 윤리적 판단의 논점을 정리해보면 다음 <표 6>과 같다.

<표 6> 트롤리 딜레마에서의 A.I. 윤리성의 준거

구분	A.I. 윤리성의 일반적 준거	트롤리 딜레마에서의 준거
공정성	A.I.의 판단이 공정하고 편파적이지 않는가?	어느 트랙으로 진행할지에 대한 공정한 판단을 하고 있는가?
책무성	A.I.의 판단이 인간의 삶에 어떤 영향을 미치는가?	A.I.의 판단이 어떤 결과를 초래하는가?
투명성	A.I.가 부정의한 목적을 추구하거나 인류를 초월하는 존재가 되지는 않을까?	A.I.의 판단은 안정적이고 신뢰할 만한가?

앞 절에서 포스트휴먼적 존재들이 주체성이나 자율성을 가질 수 있다고 논의한 바와 같이, A.I.에게 위와 같은 윤리적 준거를 요구한다는 것은 ‘A.I.가 윤리적 주체가 될 수 있다’는 명제의 논리적 개연성에 기반한 것이다. 이 명제는 A.I.가 위의 트롤리 딜레마와 같은 윤리적 딜레마의 상황에서 책임감 있는 판단을 내릴 수 있다는 것이고, 그에 따라 윤리적/도덕

적 주체(moral agent)로 등장하고 있다는 것이다. 결국 A.I.를 윤리적 행위의 주체로 상정한다는 것은 향후 A.I.가 포스트휴먼적 사회의 책임감 있는 구성원으로서 다른 윤리적 주체들과 상호의존적으로 공존하게 된다는 점에서 윤리적 지위(ethical status)를 부여받게 될 것임을 의미한다(Rakowski et al., 2021).

그런 의미에서 최근 여러 연구들이 A.I.가 윤리적 딜레마 상황에서 어떤 판단을 내리느냐의 문제를 자율주행차의 사례 차원에서 논의하고 있다(Coeckelbergh, 2020). 이 문제에 대해 본고는 자율주행차의 A.I. 알고리즘에 어떤 윤리적 준거가 적용되고 있으며, 그러한 논리가 A.I.의 윤리적 주체성을 어떻게 방증하고 있는지를 네 단계—즉, 공정성, 효율성과 책무성, 윤리적 주체의 타당성, 사회적 책무성과 투명성—의 준거를 통해 논의하고자 한다.

● 준거 1: 공정성의 준거

A.I.가 트롤리 딜레마에 대해 어떤 판단을 내리는지의 근거를 휴머니즘적 윤리적 원칙(예, 근본주의, 본질주의, 생태주의, 공동체주의, 공리주의 등)과 포스트휴머니즘적 원리(신물질주의, 객체지향존재론, 관계주의 등) 중 어디에 둘 것인지를 결정할 필요가 있다(공정성의 준거). 예를 들어, 자율주행차의 윤리적 판단의 기준을 휴머니즘적 윤리적 원리들을 기반으로 자율주행차의 ‘기능적 도덕성’ (functional morality: 인류사회가 허용할 만한 수준에서 최악을 피하고 차악을 선택하는 수준)에 기반하여 선택 알고리즘을 설계하는 것이 가능하다(Allen & Wallach, 2012; Johansson, 2010).

● 준거 2: 효율성과 책무성 간의 균형

현재 자율주행차의 등장과 그 필요성에 대한 사회적 기대는 자율주행차의 기술적 역량(효율적인 이동, 이동 중 여가 시간의 활용, 신속한 위험 인식 및 급제동 기능 등)의 문제에 집중되어 있다(Koschuch et al., 2019, November 4-8). 이러한 사회적 요구는 현대인들이 다양한 영역과 수준에서 편리함과 효율성을 얻기 위해 능동성, 자율성, 책무성 등을 인공지능 기계에 이양하고 있는 것과 같은 맥락이다. 즉, 자율주행차의 책무성 문제는 자율주행차의 필요성(효율성)이 충분하다는 전제 위에서 다루어지는 것이다.

● 준거 3: 윤리적 주체의 타당성

인간이 자신의 편리함을 위해 운전 행위의 능동성과 주체성을 자율주행차(A.I. 알고리즘)에게 양보한다는 것은 A.I.에게 ‘윤리적 주체’의 지위를 부여하겠다는 것과 마찬가지로이다(Waytz et al., 2014). 즉, 자율주행차가 주행 중에 스스로 주체적 결정을 하고, 그 행위의 결과에 대한 책임을 져야 한다는 것은 자율주행 A.I.가 윤리적 주체가 될 수밖에 없다는 의미이다(Luengo-Oroz, 2019; Theodorou & Dignum, 2020). 그게 아니라면, 자율주행차의 행위에 대해 운전자, 차량제조사, 혹은 A.I. 알고리즘 개발자 중 누가 ‘법적 책임’을 져야할지를 논의하면 되기 때문에, 윤리적 딜레마 자체가 성립하지 않는다.

● **준거 4: 사회적 책무성과 투명성**

인간이 인공지능에 대한 의존도가 높아질수록 A.I.의 윤리적 주체성과 책무성이 높아지므로, 그러한 책무성만큼 윤리적 주체로서의 투명성(안정성과 신뢰성)이 갖추어져야 한다. 즉, 기계에게 윤리적 판단과 행위의 결정권을 맡길수록 인간은 기계의 결정력을 믿고 의존하는 방향으로 습관화되기 쉽다. 그런데, 오늘날과 같은 기술문명 사회에서 인간의 기계에 대한 의존도는 점차 높아지고 있고 그 흐름이 비가역적이기 때문에(Grissinger, 2019), A.I.에 대한 의존성이 높아지는 것에 대한 사회공동체의 수준에서 정치윤리적 판단이 필요하다. 예를 들어, 인간과 A.I.의 판단과 행위의 주체성을 분산시키거나(예, Specific A.I.), A.I.의 역할을 제한하는(예, 약한 A.I.) 정책의 설정이 가능하다. 즉, A.I. 주체성의 분배와 영향력의 제한을 위해 자율주행차가 완전자율주행(Level 5) 단계에 도달하더라도 이를 일반도로에서 허용할 것이냐의 문제는 기술적 문제가 아니라 정치윤리적 문제로 판단해야 하는 것이다.

요컨대, 자율주행차에 해당하는 진정한 의미의 윤리적 딜레마는 위와 같은 ‘트롤리 딜레마(인본주의 차원의 공정성 문제)’가 아니라, 인간이 A.I.보다 더 공정한 판단을 할 수 있는냐의 문제이다. 실제로 트롤리 딜레마는 인간이 어떤 윤리적 준거에 의해 판단하고 행동하느냐를 파악하기 위한 사고 실험(thought experiment)의 차원에서 논의되고 있다. 즉, 트롤리 딜레마는 분명한 정답이 있는 것이 아니라, 윤리적 판단과 행위 정당성의 근거를 무엇으로 삼고 있는가의 문제이다. A.I.에 기반한 자율주행차의 경우, 그 윤리적 판단에 대한 평가 준거를 ‘인본주의적 공정성 대 포스트휴먼적 공정성’ 중 어디에 둘 것이냐 하는 것이 가장 중요한 논점이 되고 있다. 본고는 그 준거가 지금까지 논의한 포스트휴머니즘의 관점들(생태주의, 행위자-망이론, 비대표성이론 등)과 다음 절에서 다룰 확장된 자본주의에 대한 비판의 관점들을 바탕으로 새롭게 설정할 필요가 있음을 강조하는 것이다.

나. 비판적(critical) 포스트휴머니즘의 정치윤리학(politico-ethics)

포스트휴머니즘의 한 갈래인 비판적 포스트휴머니즘(Braidotti, 2016; Herbrechter, 2013)은 앞서 언급한 인류중심주의와 더불어 확장된 자본주의(advanced capitalism)에 대한 비판을 통해 포스트휴먼적 미래를 논의하는 관점이다. 비판적 포스트휴머니즘은 포스트휴먼적 존재들이 생태적(자연적), 인간적, 기술적 요소들이 혼성화된 존재라는 것을 인정하면서도, 동시에 기술의 발달이 가속화됨에 따라 생태계와 인간, 그리고 기계적 존재들이 자본주의의 무제한적 확장의 재료가 되어가는 상황을 우려하고 있다. 즉, 생태계-인간계-기술계의 일원론적 연속성을 인정하면서도, 과학기술이 확장된 자본주의와의 협력을 통해 인간과 비인간적 요소들의 조합(assemblage)을 가속화하고 있는 현실에 대해 정치윤리학적 판단의 딜레마를 안고 있다는 것이다(Braidotti, 2013). 따라서 비판적 포스트휴머니즘은 기술 유토피아를 표방

하는 트랜스휴머니즘(transhumanism)에 대한 비판과 탈인간중심주의를 바탕으로, 자연적, 기술적 생태계 속에서 인간과 비인간적 존재들이 어떻게 상호공존하며 지속가능성을 높일지를 탐색하고자 한다(Wolfe, 2010).

우선 확장된 자본주의가 설계하는 포스트휴먼적 세계는 개인의 자유주의적 정치철학(자유와 선택권의 극대화)과 자본주의적 경제원리(경쟁과 시장원리)를 바탕으로 21세기의 급속한 과학기술적 진보를 활용한 기술적 유토피아이다. 즉, 과학 발전이 가져온 기술적 성과(인간강화기술, 환경통제, 생명공학, 유전자 조작, 에너지 및 자원개발 등)를 바탕으로 인간중심주의적 비전을 미래로 확장하는 것이다. 이 확장된 자본주의가 기존의 자본주의와 어떤 차이점이 있는지는 다음 <표 7>을 통해 간단히 정리해볼 수 있다.

<표 7> 진보주의의 진화와 확장된 자본주의의 등장

19세기 이전	20세기	21세기
고전적 자본주의 (classic capitalism)	인지 자본주의 (cognitive capitalism)	확장된 자본주의 (advanced capitalism)
인간(프로레타리아 계급)의 노동력(육체)과 토지 및 천연자원(물질)의 착취	인간의 정신(지식과 감정) 노동의 착취를 통한 서비스 및 지식(정보)산업의 확장	모든 생물 종과 지구 생태계 전체를 착취: 모든 존재(인간, 동식물, 유전자, 기계)와 생태성(의식, 기호, 본능, 가상세계)을 자본주의 시장의 상품으로 간주

즉, 고전적 자본주의가 인간의 육체적 노동력과 자연의 물질적 자원을 착취했고, 20세기의 인지자본주의가 인간의 정신적 요소들을 착취했다면, 오늘날의 확장된 자본주의는 NBIC(nano-bio-info-cogno) 기술로 대표되는 테크놀로지를 기반으로 모든 생물종의 기본적인 요소들(예, 유전자)과 모든 정신 및 물질의 상태성(예, 인간의 의식과 본능, 가상적 경험)까지도 자본주의적 시장 경제의 자원으로 삼으려고 하는 것이다. 그로 인해 기술중심적으로 확장된 자본주의는 소수의 자본가와 기업, 국가 등이 인류의 과학기술을 독점하여 인류의 보편적 삶의 조건(식량, 보건의료, 문화, 정보통신, 안보, 여가활동 등)은 물론 삶과 죽음의 문계까지 통제하고 권력화할 수 있다는 비판을 받고 있다(Foucault et al., 2008; Mbembe, 2008).

Braidotti(2016)는 위와 같은 확장된 자본주의가 기술적 발전을 바탕으로 미래 사회를 통제하고자 하는 대표적 세 가지 영역으로 i) 사회관계망 서비스(social network service: 이하 SNS) 시장, ii) 생명유전학(bio-genetics), 생명정보학(bio-informatics) 등의 생명기술(bio-technology) 서비스 시장, 그리고 iii) 물질계와 환경계를 넘나드는 생태적 환경 시장으로 구분한 바 있다. 먼저 SNS 시장은 개인들의 취향과 관심, 사회적 관계성, 현실 및 가상적

경험과 시간을 무제한적으로 수집하여 엄청난 개인적 데이터를 수집할 수 있고 그 정보를 통해 원하는 만큼 개인의 생각과 행동을 통제할 수 있게 되었고(Brown, 2013), 생명기술 서비스 시장에서 인류는 정보공학과 생명공학의 기술적 협력을 바탕으로 인간을 포함한 모든 생명체의 생체정보(bio-information)를 수집하고 가공하여 다양한 목적으로 활용하고 있다(Paul & Aithal, 2020). 또한 생태적 환경 시장에서도 확장된 자본주의는 비인간 종족들, 전 지구적 수준의 에너지와 자원들, 그리고 생태환경 시스템을 통제하고 관리함으로써 자본주의 시스템을 지리정치학(geopolitics)적 영역으로 확장하고 있다(Luisetti, 2019).

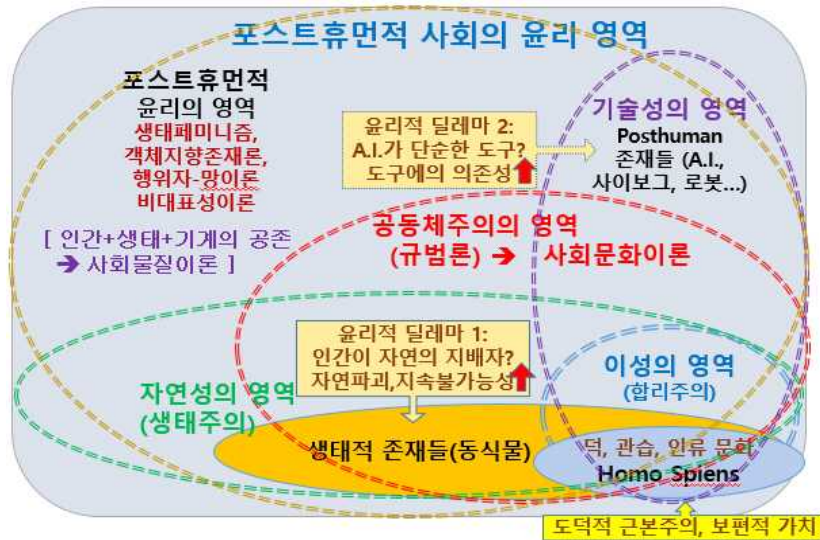
이처럼 인간이 생명 현상의 본질적 수준까지 통제가 가능한 기술을 얻게 되면서 그러한 기술을 누가 통제하고 어떻게 사용하는가는 최근 가장 결정적인 윤리적 논쟁거리가 되고 있다. 생명 자체가 자본의 투자대상이 되고, 새로운 생명 종이 탄생하거나 기존 종들의 멸종이 가속화될 수 있는 기술을 갖게 됨으로써 생명기술서비스 시장은 생명정치학(bio-politics)의 장이 되었다(Foucault et al., 2008; Negri & Hardt, 2000; Mbembe, 2008). 예를 들어, 유전자 변형 및 생체변형 동식물, 노화 방지나 질병 치료, 그리고 생명 연장을 위해 세포와 세균 및 바이러스를 다루는 모든 생의학적 연구들이 이러한 생명기술 서비스 시장의 산물이다(van der Zaag, 2016). 어떤 연구자들은 현재 지구촌이 겪고 있는 Covid-19 바이러스의 대유행도 그러한 바이러스의 조작 연구의 부산물이라는 의혹을 제기하면서, 인류는 현재 컴퓨터 바이러스와 생명 바이러스의 변증법을 경험하고 있다고 주장하기도 한다(Peters et al., 2020). 요컨대, 확장된 자본주의는 과학기술의 극대화를 매개로 생태, 인간, 물질의 모든 부분을 통제할 수 있는 산업에 투자하여 무한한 이윤을 추구하고 있는데, 그러한 기술적 패러다임의 핵심에 A.I.가 있는 것이다. 즉, A.I.의 기술적 무한확장성과 이를 뒷받침하는 산업자본주의 시스템이 다가올 포스트휴먼 윤리의 맥락적 조건이다.

정리하자면, 앞에서 논의한 자율자동차의 경우와 같은 A.I. 알고리즘에 기반한 기계들은 단순한 사물로만 머무르지 않고, 점차 윤리적 판단과 행위를 수행하는 윤리적 주체들의 자격을 갖추게 될 것이고, 그러한 흐름의 원동력이 바로 확장된 자본주의이다. 따라서, 그러한 사물-주체들(thing-subject)의 등장이 가져오는 윤리적 딜레마를 적절히 논의하기 위해서는 인본주의적 윤리가 아니라 포스트휴머니즘을 바탕으로 A.I. 윤리의 패러다임을 새롭게 정립할 필요가 있다는 것이다.

다. A.I. 시대에 대응하기 위한 포스트휴먼 윤리관

앞 절에서 논의한 바와 같이 A.I.와 인간이 공존하는 사회 속의 다양한 윤리적 주체들의 판단과 행위에 대한 윤리적 준거를 설정하기 위해서는 인본주의적 윤리관을 넘어서 생태주의 및 기술적 사회의 환경에서의 윤리관을 제시할 수 있는 포스트휴먼 윤리관이 필요하다는

것이다. 인본주의적 윤리관에서 포스트휴먼적 윤리관으로의 패러다임의 확장은 다음 [그림 5]로 설명될 수 있다.



[그림 5] 포스트휴먼적 사회의 윤리적 패러다임

위 [그림 5]에서 제시된 바와 같이, 기존의 휴머니즘적 윤리관은 다음과 같은 관점의 변화를 통해 포스트휴먼적 윤리관으로 확장될 필요가 있을 것이다.

첫째, 전통주의적 윤리관은 오랜 인류의 역사 동안 현생 인류(homo sapiens)가 우주의 중심적 존재라는 전제 하에, 인류 사회가 존속되고 번영하기 위한 근본주의적 덕목, 관습, 문화적 규범들과 같은 보편적 가치를 윤리적 판단과 행위의 기초로 삼아왔다.

둘째, 18세기에 들어서면서 등장한 합리주의적 사고는 인간의 이성을 도덕적 판단과 도덕적 실천을 가능케 하는 기본 특성이라고 보았고(Jackson, 1881), 이성을 지닌 인간을 유일한 도덕적 주체로 간주함으로써 인본주의적 윤리관을 더욱 강화시켰다(Barreto, 2006).

셋째, 19세기 후반 들어 과학자들은 자연을 세분화하여 인과론과 환원주의적으로 해석할 것인가 아니면 있는 그대로 관찰하면서 다양한 변인들을 복합적으로 참조하여 자연을 이해할 것인가라는 고민을 시작하였다(Cowles, 1909). 그러한 분위기에서 등장한 생태주의적 관점은 인간이 자연의 일부로써 자연과 불가분리적 존재라는 점을 강조하면서, 생태적 존재들을 윤리적 문제의 주요 행위자의 범주에 포함시키기 시작하였다. 그로 인해 ‘인간이 자연의 지배자’라는 기존의 인본주의적 세계관은 도전을 받기 시작했고, 이것이 오늘날 환경 파괴와 자원 고갈의 문제와 직결되면서 인본주의적 윤리관은 한계점에 도달하게 된 것이다

(Postma, 2016).

넷째, 기계적 존재들(사이보그, 로봇, 스마트폰, A.I. 등)이 인류의 삶의 곳곳에 스며들게 되고, 점점 그러한 기술적 도구들에 대한 인간의 의존성이 높아지고 있다. 인간이 A.I.에게 결정권을 부여하고 그 성능에 의존할수록 그러한 기계들은 단순한 도구에 머무르지 않고, 윤리적 판단과 행위의 주체가 될 수밖에 없다는 것이다. 이때 A.I. 윤리의 딜레마는 A.I.가 과연 윤리적 판단을 내릴 수 있는지의 문제가 아니라, A.I.가 인간보다 더 윤리적인 판단과 행위를 할 개연성이 있다는 사실이다. 결국 인류 사회가 기술성이 주도하는 시대에 접어들게 되면서, 인간만이 이성과 자의식에 기반한 자율성을 가지고 윤리적 판단과 행위의 주체가 될 수 있다는 기존의 인본주의적 윤리관은 그 한계점에 도달하게 된 것이다.

다섯째, 공동체주의적 규범론도 인간만을 공동체의 구성원으로 상정하고 있다는 점에서 여전히 인본주의적 윤리관에 속한다. 그리고 이 공동체의 관념은 사회문화이론(socio-cultural theories)을 통해 ‘인류’ 집단의 사고와 경험이 축적된 결과만을 문화로 인식하게 만들었다. 하지만 포스트휴머니즘은 인류 집단에 중심을 둔 사회문화이론을 대신하여, 인간, 생태계(자연성), 그리고 기계적 주체들이 공존하는 사회에서의 물질적 관계성을 설명하기 위해 사회물질(socio-material)이론의 차원에서 세상을 이해하고자 한다. 따라서, 포스트휴머니즘에 기반한 A.I. 윤리도 그 철학적 혹은 이념적 토대를 사회문화에서 찾는 것이 아니라 사회물질이론을 통해 탐색할 필요가 있다는 것이다.

이것이 본고에서 강조하고자 한 포스트휴머니즘에 기반한 A.I. 윤리의 이론적 패러다임인 것이다. 마지막으로, 그런 점에서 포스트휴먼 윤리관은 학교교육, 특히 초등교육의 영역에서도 다루어질 필요가 있다(Kwon & Kyun, 2018). 그런 의미에서 유치원이나 초등교사의 교육 차원에서 예비교사가 인공지능에 대해 어떤 인식을 가지고, 인공지능과 관련된 윤리적 이슈를 어떻게 이해할 것인가를 논의하는 것이 시급할 것이다.

III. 결 론

지금까지 본고는 A.I.로 대표되는 기술적 주체의 역량과 영향력이 급속히 확장되는 상황에서 논란이 되고 있는 A.I.윤리 문제에 대해 포스트휴머니즘의 여러 이론들의 관점에서 어떤 새로운 윤리학적 입장이 필요할지에 대해 논의하였다.

본고에서 논의한 포스트휴머니즘적 윤리관의 핵심은 크게 다음 세 가지로 요약될 수 있다: 첫째, A.I. 윤리의 행위 주체인 A.I.는 단순한 도구가 아니라 행위 주체로서의 능동성과 자율성을 지닌 존재라는 것과 그러한 자율적 주체의 윤리 문제를 논의하기 위해서는 기존의

인본주의적 윤리관이 아니라 포스트휴먼적 윤리관이 필요하다는 것이다. 둘째, 포스트휴먼적 윤리학의 이론적 토대로서 생태페미니즘, 객체지향존재론, 그리고 비대표성이론의 핵심 개념과 윤리적 준거가 무엇인지 이해할 필요가 있다는 것이고, 마지막으로 셋째, 확장된 자본주의가 확장의 주된 대상이자 자원으로 삼고 있는 기술중심주의와 생명정치학이 어떻게 미래 기술문명의 핵심인 A.I.의 기능과 역할 속에 스며들어 있는지를 비판적 포스트휴머니즘의 관점을 통해 이해할 필요가 있다는 것이다. 위 세 가지 주장을 한 문장으로 정리하면, ‘A.I. 윤리’의 방향성은 기술적 존재(즉, A.I.)가 윤리적 주체성을 띠고 윤리적 판단과 행위를 수행하게 되는 포스트휴먼적 환경에서의 윤리관과 윤리적 판단의 준거는 탈인본주의적 관점들(생태페미니즘, 객체지향존재론, 그리고 비대표성이론)을 통해 논의되어야 한다는 것이다.

포스트휴먼적 윤리관의 등장은 기술문명 사회의 등장에 따른 윤리철학의 변화에 어떤 시사점을 제공하는가? 오늘날 점차 그 영향력을 극대화하고 있는 확장된 자본주의는 기술중심주의의 총아인 A.I.를 통해 인류의 삶 곳곳에 스며들어 그 영향력을 확대하고 있고, A.I.는 단순한 인간의 도구에 머무르는 것이 아니라 주체성과 자율성을 가지고 기존 인간의 역할을 대신하게 될 개연성이 커지고 있다. 따라서 A.I.가 윤리적 판단과 행위의 주체로 등장하고 있는 기술중심 사회에서의 윤리학의 패러다임에는 포스트휴머니즘에 기반한 정치윤리학에 대한 이해와 논의가 필요하다는 것이다. 즉, 기존의 인본주의적 윤리관에서의 핵심적인 윤리적 질문이 ‘인간 존엄성과 자율성, 책임감 등에 기반하여 인류 공영의 목적을 위해 기술적 세계를 어떻게 잘 통제할 수 있을까?’ 였다면, 포스트휴머니즘에 기반한 윤리적 질문은 ‘자연생태-인간-기계문명의 밀접한 상호의존성을 바탕으로 모든 존재들이 어떻게 공존할 수 있을까?’ 라는 질문으로 바뀔 필요가 있을 것이다.

사실 A.I.를 탑재한 기계들이 윤리적 판단과 행위의 주체가 되는 것에 대해 많은 우려가 있지만, 지난 지구 생태계의 역사에서 인류가 보여준 부정적인 영향들(불평등성, 부정의, 생태계 파괴, 기후 변화 등)을 생각하면 인본주의에 대한 반성과 대안적 성찰이 더 시급한 일일 것이다. 즉, 자율주행차의 윤리적 딜레마는 앞으로 인간과 A.I. 중 누가 더 나은 윤리적 판단과 행위를 할 것인가의 문제이고, 이는 단순하게 판단할 문제가 아니다. 왜냐하면, A.I. 기계의 보편화가 가져올 윤리적 딜레마 속에서 A.I. 주체들이 향후 포스트휴먼 사회에서의 다양한 존재들의 공존적 발전에 도움이 될지 해가 될지는 아직 미지수이며, 이것이 포스트휴머니즘 철학이 풀어나가야 할 정치윤리학적 과제일 것이다.

References

- Ahmanson, R. G. (2017). The new dignity. *Journal of Christian Legal Thought*, 7(1), 10-12.
- Allen, C., & Wallach, W. (2012). Moral machines: Contradiction in terms or abdication of human responsibility. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 55-68). MIT Press.
- Ashby, W. R. (1961). *An introduction to cybernetics*. Chapman & Hall Ltd.
- Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*, 28(3), 801-831.
- Barile, S., & Polese, F. (2010). Smart service systems and viable service systems: Applying systems theory to service science. *Service Science*, 1(1-2), 21-40. https://doi.org/10.1287/serv.2.1_2.21
- Barreto, J. M. (2006). Ethics of emotions as ethics of human rights: A jurisprudence of sympathy in Adorno, Horkheimer and Rorty. *Law and Critique*, 17(1), 73-106. <https://doi.org/10.1007/s10978-006-0003-y>
- Bennett, J. (2015). Systems and things: On vital materialism and object-oriented philosophy. In R. Grusin (Ed.), *The nonhuman turn* (pp. 223-239). University of Minnesota Press.
- Bogost, I. (2009, December 8). *What is object-oriented ontology?: A definition for ordinary folk*. Ian Bogost. http://bogost.com/writing/blog/what_is_objectoriented_ontolog.
- Bonnefon, J. F., Shariff, A., & Rahwan, I. (2019). The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view]. *IEEE*, 107(3), 502-504. <https://doi.org/10.1109/JPROC.2019.2897447>
- Bostrom, N. (2003). *The transhumanist FAQ*. World Transhumanist Association. http://www.the-astrolabe.net/transhumanist_faq.htm
- Bostrom, N. (2005). In defense of posthuman dignity. *Bioethics*, 19(3), 202-214. <https://doi.org/10.1111/j.1467-8519.2005.00437.x>
- Braidotti, R. (2006). Posthuman, all too human: Towards a new process ontology. *Theory, Culture & Society*, 23(7), 197-208. <https://doi.org/10.1177/0263276406069232>
- Braidotti, R. (2013). Posthuman humanities. *European Educational Research Journal*, 12(1), 1-19. <http://dx.doi.org/10.2304/eeerj.2013.12.1.1>
- Braidotti, R. (2016). Posthuman critical theory. In D. Banerji, & M. R. Paranjape (Ed.), *Critical posthumanism and planetary futures* (pp. 13-32). Springer. https://doi.org/10.1007/978-81-322-3637-5_2

- Brown, B. A. (2013). Primitive digital accumulation: Privacy, social networks, and biopolitical exploitation. *Rethinking Marxism*, 25(3), 385-403. <https://doi.org/10.1080/08935696.2013.798974>
- Bundy, A. (2017). Smart machines are not a threat to humanity. *Communications of the ACM*, 60(2), 40-42. <https://doi.org/10.1145/2950042>
- Burns, T. W. (2014). John Courtney Murray, religious liberty, and modernity: Part I: Inalienable natural rights. *Logos: A Journal of Catholic Thought and Culture*, 17(2), 13-38. <https://doi.org/10.1353/log.2014.0017>
- Callon, M. (1990). Techno-economic networks and irreversibility. *The Sociological Review*, 38(1_suppl), 132-161. <https://doi.org/10.1111/j.1467-954X.1990.tb03351.x>
- Casey, J. (1972). *The autonomy of art: John Casey*. Royal Institute of Philosophy Supplement. <https://doi.org/10.1017/S1358246100000813>
- Ceder, S. (2018). *Towards a posthuman theory of educational relationality*. Routledge. <https://doi.org/10.1007/s11217-020-09725-w>
- Chambers, I. (2013). *Culture after humanism. History, culture, subjectivity*. Routledge. <https://doi.org/10.4324/9781315012636>
- Charlton, W. (1981). Spinoza's Monism. *The Philosophical Review*, 90(4), 503-529.
- Charteris, J., Smardon, D., & Nelson, E. (2017). Innovative learning environments and new materialism: A conjunctural analysis of pedagogic spaces. *Educational Philosophy and Theory*, 49(8), 808-821. <https://doi.org/10.1080/00131857.2017.1298035>
- Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. OXFORD.
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- Connolly, W. E. (2013). The 'new materialism' and the fragility of things. *Millennium: Journal of International Studies*, 41(3), 399-412. <https://doi.org/10.1177/0305829813486849>
- Couture, J., & Nielsen, K. (1995). Introduction: The ages of metaethics. *Canadian Journal of Philosophy*, 21(0), 1-30. <https://doi.org/10.1017/S0229705100006984>
- Cowles, H. C. (1909). The trend of ecological philosophy. *The American Naturalist*, 43(510), 356-368.
- Cunneen, M., Mullins, M., Murphy, F., & Gaines, S. (2019). Artificial driving intelligence and moral agency: Examining the decision ontology of unavoidable road traffic accidents through the prism of the trolley dilemma. *Applied Artificial Intelligence*, 33(3), 267-293. <https://doi.org/10.1080/08839514.2018.1560124>
- Curran, J. (2012). *Media and power*. Routledge.

- David, P. R. (2021) Notes on the case of orangutan Sandra, the non-human subject of rights. In H. Kury, & S. Redo (Eds.), *Crime prevention and justice in 2030* (pp. 313-336). Springer, Cham. https://doi.org/10.1007/978-3-030-56227-4_15
- Davies, T. (2008). *Humanism*. Routledge.
- de Freitas, E., & Curinga, M. X. (2015). New materialist approaches to the study of language and identity: Assembling the posthuman subject. *Curriculum Inquiry*, 45(3), 249-265. <https://doi.org/10.1080/03626784.2015.1031059>
- Derrida, J. (2008). *The animal that therefore I am*. Fordham Univ Press.
- Dewsbury, J. D., Harrison, P., Rose, M., & Wylie, J. (2002). Enacting geographies. *Geoforum*, 43(3), 437-440.
- Dukas, R. (2004). Evolutionary biology of animal cognition. *Annual Review of Ecological Evolution System*, 35(0), 347-374. <https://doi.org/10.1146/annurev.ecolsys.35.112202.130152>
- Dunn, R. G. (1997). Self, identity, and difference: Mead and the poststructuralists. *The Sociological Quarterly*, 38(4), 687-705.
- Fersht, A. R. (2021). AlphaFold: A personal perspective on the impact of machine learning. *Journal of Molecular Biology*, 433(20), 1-4. <https://doi.org/10.1016/j.jmb.2021.167088>
- Forlano, L. (2017). Posthumanism and design. *She Ji: The Journal of Design, Economics, and Innovation*, 3(1), 16-29. <https://doi.org/10.1016/j.sheji.2017.08.001>
- Foucault, M., Davidson, A. I., & Burchell, G. (2008). *The birth of biopolitics: Lectures at the Collège de France*. Springer.
- Fukuyama, F. (2009, October 23). *Transhumanism*. Foreign Policy. <https://foreignpolicy.com/2009/10/23/transhumanism>
- Gams, M. (1997). Is weak AI stronger than strong AI? In M. Gams, M. Paprzycki, & X. Wu (Eds.), *Mind versus computer: Were Dreyfus and Winograd right?* (pp. 30-45). IOS Press.
- Geerts, E., & Carstens, D. (2019). Ethico-onto-epistemology. *Philosophy Today*, 63(4), 915-925. <https://doi.org/10.5840/philtoday202019301>
- Gilroy, P. (2009). *Race and the right to be human*. Universiteit Utrecht.
- Grimm, D. (2014, April 8). Q&A: *Pets are becoming people, legally speaking*. National Geographic. <https://www.nationalgeographic.com/animals/article/140406-pets-cats-dogs-animal-rights-citizen-canine>
- Grissinger, M. (2019). Understanding human over-reliance on technology. *Pharmacy and Therapeutics*, 44(6), 320-375.
- Hallen, P. (1989). Careful of science: A feminist critique of science. *The Trumpeter*, 8(1), 3-8.

- Haney, W. S. (2006). *Cyberculture, cyborgs and science fiction: Consciousness and the posthuman*, 48(2), 426. <https://link.gale.com/apps/doc/A168742966/AONE?u=anon~172c7d5e&sid=googleScholar&xid=e3705c09>
- Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Random House.
- Haraway, D. (1985). A manifesto for cyborgs: Science, technology, and socialist feminism. *Socialist Review*, 15(2), 65-107.
- Harman, G. (2018). *Object-oriented ontology: A new theory of everything*. Penguin Books UK.
- Haslam, N., Kashima, Y., Loughnan, S., Shi, J., & Suitner, C. (2008). Subhuman, inhuman, and superhuman: Contrasting humans with nonhumans in three cultures. *Social Cognition*, 28(2), 248-258. <https://doi.org/10.1521/soco.2008.26.2.248>
- Hayles, N. K. (2000). *How we became posthuman: Virtual bodies. Cybernetics, Literature, and Informatics. Public Understanding of Science*, 9(4), 464. <https://doi.org/10.1088/0963-6625/9/4/704>
- Hee, K. S., & Ho, S. C. (2020). Education methods for social conflicts: Focusing on gender in equality. *Global Creative Leader: Education & Learning*, 10(1), 285-302. <https://doi.org/10.34226/gcl.2020.10.1.285>
- Herbrechter, S. (2013). *Posthumanism: A critical analysis*. A&C Black.
- Holland, D., & Lachicotte, W. (2007). Vygotsky, Mead, and the new sociocultural studies of identity. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky* (pp. 101-135). Cambridge University Press. <https://doi.org/10.1017/CCOL0521831040.005>
- Hughes, T. P. (2012). The evolution of large technological systems. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 45-76). The MIT Press.
- Hultman, K., & Lenz Taguchi, H. (2010). Challenging anthropocentric analysis of visual data: A relational materialist methodological approach to educational research. *International Journal of Qualitative Studies in Education*, 23(5), 525-542. <https://doi.org/10.1080/09518398.2010.500628>
- Jackson, W. T. (1881). *Seneca and Kant: Or an exposition of Stoic and rationalistic ethics, with a comparison and criticism of the two systems*. United Brethern Publishing House.
- Jandrić, P., & Ford, D. R. (2020). Postdigital ecopedagogies: Genealogies, contradictions, and possible futures. *Postdigital Science and Education*. Advance online publication. <https://doi.org/10.1007/s42438-020-00207-3>
- Johansson, L. (2010). The functional morality of robots. *International Journal of Technoethics*

- (*JIT*), 1(4), 65-73. <https://doi.org/10.4018/jte.2010100105>
- Knappett, C., & Malafouris, L. (Eds.). (2008). *Material agency: Towards a non-anthropocentric approach*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-74711-8_5
- Kochen, M. (1985). The future of AI and human flexibility. *Human Systems Management*, 5(2), 159-162. <https://doi.org/10.3233/HSM-1985-5209>
- Koschuch, M., Sebron, W., Szalay, Z., Török, Á., Tschürtz, H., & Wahl, I. (2019, November 4-8). *Safety & security in the context of autonomous driving*. [Conference session]. *2019 IEEE International Conference on Connected Vehicles and Expo. ICCVE*. <https://doi.org/10.1109/ICCV45908.2019.8965092>.
- Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin.
- Kwon, S., & Kyun, S. (2018). Pre-service early childhood teachers' recognition of artificial intelligence and artificial intelligence teachers. *Global Creative Leader: Education & Learning*, 8(4), 87-102. <https://doi.org/10.34226/gcl.2018.8.4.87>
- Latour, B. (1992). *One more turn after the social turn: Easing science studies into the non-modern world*. Notre Dame University Press.
- Latour, B. (2013). *We have never been modern*. Harvard university press.
- Lieberman, P. (2006). *Toward an evolutionary biology of language*. Harvard University Press.
- Lim, M. (2021). A systematic literature review of research on artificial intelligence ethics. *The Journal of Ethics Studies*, 1(135), 47-66. 국문: 임미가(2021). 인공지능 윤리 연구에 관한 체계적 문헌고찰. *윤리연구*, 1(135), 47-66. <https://doi.org/10.15801/je.1.135.202112.47>
- Lorimer, H. (2005). Cultural geography: The busyness of being. *Progress in Human Geography*, 29(1), 83-94. <https://doi.org/10.1191/0309132505ph531pr>
- Luengo-Oroz, M. (2019). Solidarity should be a core ethical principle of AI. *Nature Machine Intelligence*, 1(11), 494-494. <https://doi.org/10.1038/s42256-019-0115-3>
- Luisetti, F. (2019). Geopower: On the states of nature of late capitalism. *European Journal of Social Theory*, 22(3), 342-363. <https://doi.org/10.1177/1368431018803764>
- Maclure, J., & Russel, I. S. (2021) AI for humanity: The global challenges. In B. Braunschweig, & M. Ghallab (Eds.), *Reflections on artificial intelligence for humanity* (pp. 116-126). Springer, Cham. https://doi.org/10.1007/978-3-030-69128-8_8
- MacLure, M. (2017). Qualitative methodology and the new materialisms. In N. K. Denzin, & M. D. Giardina (Eds.), *Qualitative inquiry in neoliberal times* (pp. 48-58). Routledge.
- Mann, N. (1996). *The origins of humanism*. The Cambridge Companion to Renaissance humanism.
- Martín-Lucas, B. (2017). Posthumanist feminism and interspecies affect in Nalo Hopkinson's

- midnight robber. *Atlantis: Critical Studies in Gender, Culture & Social Justice*, 38(2), 105-115.
- Mbembe, A. (2008). Necropolitics. In S. Morton, & S. Bygrave (Eds.), *Foucault in an age of terror* (pp. 152-182). Palgrave Macmillan.
- Mele, C., Pels, J., & Polese, F. (2010). A brief review of systems theories and their managerial applications. *Service Science*, 2(1-2), 126-135. https://doi.org/10.1287/serv.2.1_2.126
- Mendible, M. (2017). Introduction: Race, politics, and the humanities in an age of “posts” : Rethinking the human/race. *Humanities*, 6(1), 5. <https://doi.org/10.3390/h6010005>
- Miah A. (2008) A critical history of posthumanism. In B. Gordijn, & R. Chadwick (Eds.), *Medical enhancement and posthumanity* (pp. 71-94). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-8852-0_6
- Minsky, M. L. (1982). Why people think computers can't. *AI Magazine*, 3(4), 3. <https://doi.org/10.1609/aimag.v3i4.376>
- Moore, J. W. (2017). The Capitalocene, Part I: On the nature and origins of our ecological crisis. *The Journal of Peasant Studies*, 44(3), 594-630. <https://doi.org/10.1080/03066150.2016.1235036>
- Morton, T. (2011). Here comes everything: The promise of object-oriented ontology. *Qui Parle: Critical Humanities and Social Sciences*, 19(2), 163-190.
- Murdoch, J. (1998). The spaces of actor-network theory. *Geoforum*, 29(4), 357-374. [https://doi.org/10.1016/S0016-7185\(98\)00011-6](https://doi.org/10.1016/S0016-7185(98)00011-6)
- Negri, A., & Hardt, M. (2000). *Empire*. Harvard University Press.
- Neimanis, A. (2017). Nature represents itself: Bibliophilia in a changing climate. In V. Kirby (Ed.), *What if culture was nature all along* (pp. 179-198). New Materialisms.
- Nourbakhsh, I. R., & Keating, J. (2020). *AI and Humanity*. Mit Press.
- Park, H. Y. (2013). The nature and direction of character education for a multicultural society. *The Journal of Character Education*, 7(2), 45-67. ☞ 국문: 박휴용(2013). 다문화 사회를 위한 인격교육의 성격과 방향. *인격교육*, 7(2), 45-67.
- Paul, P., & Aithal, P. S. (2020). Informatics: Foundation, nature, types and allied areas—An educational & analytical investigation. *International Journal of Applied Science and Engineering*, 8(1), 1-9. <https://doi.org/10.30954/2322-0465.1.2020.1>
- Peters, M. A., Jandrić, P., & McLaren, P. (2020). Viral modernity? Epidemics, infodemics, and the ‘bioinformational’ paradigm. *Educational Philosophy and Theory*. Advance online publication. 1-9. <https://doi.org/10.1080/00131857.2020.1744226>

- Plumwood, V. (2001). Nature as agency and the prospects for a progressive naturalism. *Capitalism Nature Socialism*, 12(4), 3-32.
- Popovici, F. G. (2017). The paradise revealed: The promised land of technology in the posthuman era. *Literature, Discourses and the Power of Multicultural Dialogue*, 5(0), 370-375.
- Postma, D. (2016). The ethics of becoming in a pedagogy for social justice. A posthumanist perspective. *South African Journal of Higher Education*, 30(3), 310-328. <http://dx.doi.org/10.20853/30-3-651>
- Pranaykumar, P. (2021). *Human consciousness and artificial intelligence: Can AI develop human-like consciousness? Cognitive abilities? What about Ethics?*. Cognitive Abilities. <http://dx.doi.org/10.2139/ssrn.3786957>
- Prinsloo, M. (2005). The new literacies as placed resources: research: Information and communication technologies. *Perspectives in Education*, 23(1), 87-98.
- Rakowski, R., Polak, P., & Kowalikova, P. (2021). Ethical aspects of the impact of AI: The status of humans in the era of artificial intelligence. *Society*, 58(0), 196-203. <https://doi.org/10.1007/s12115-021-00586-8>
- Salleh, A. (2017). *Ecofeminism as politics: Nature, Marx and the postmodern*. Zed Books.
- Sandel, M. J. (Ed.). (2007). *Justice: A reader*. Oxford University Press on Demand.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-424. <https://doi.org/10.1017/S0140525X00005756>
- Segal, J. M. (1996). *Agency and alienation: A theory of human presence*. Rowman & Littlefield.
- Shin, D. H. (2016). Application of actor-network theory to network neutrality in Korea: Socio-ecological understanding of network dynamics. *Telematics and Informatics*, 33(2), 436-451. <https://doi.org/10.1016/j.tele.2015.10.002>
- Smith, D. W. (2011). Deleuze & the question of desire: Towards an immanent theory of ethics. In D. Jun (Ed.), *Deleuze and ethics* (pp. 123-141). Edinburgh University Press.
- Snaza, N., Appelbaum, P., Bayne, S., Morris, M., Rotas, N., & Sandlin, J. (2014). Toward a posthumanist education. *Journal of Curriculum Theorizing*, 30(2), 39-55.
- Tatnall, A., & Davey, B. (2015). The internet of things and beyond: Rise of the non-human actors. *International Journal of Actor-Network Theory and Technological Innovation*, 7(4), 56-67. <https://doi.org/10.4018/978-1-5225-1832-7.ch016>
- Taylor, T. L. (2002) Living digitally: Embodiment in virtual worlds. In R. Schroeder (Ed.), *The social life of avatars. computer supported cooperative work* (pp. 40-62). Springer,

- London. https://doi.org/10.1007/978-1-4471-0277-9_3
- Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, 2(1), 10-12. <https://doi.org/10.1038/s42256-019-0136-y>
- Thrift, N. (1996). *Spatial formations* (Vol. 42). Sage.
- Thrift, N. (2000). Still life in nearly present time: The object of nature. *Body & Society*, 23(4), 34-57. <https://doi.org/10.1177/1357034X00006003003>
- Thrift, N. (2007). *Non-representational theory: Space, politics, affect*. Routledge.
- Ulmer, J. B. (2017). Posthumanism as research methodology: Inquiry in the Anthropocene. *International Journal of Qualitative Studies in Education*, 30(9), 832-848. <https://doi.org/10.1080/09518398.2017.1336806>
- van der Zaag, A. C. (2016). On posthuman subjectivity. *Journal of Cultural Economy*, 9(3), 330-336. <https://doi.org/10.1080/17530350.2015.1040436>
- Vannini, P. (2015). Non-representational research methodologies: An introduction. In P. Vannini (Ed.), *Non-representational methodologies* (pp. 11-28). Routledge.
- Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52(0), 113-117. <https://doi.org/10.1016/j.jesp.2014.01.005>
- Weitzenfeld, A., & Joy, M. (2014). An overview of anthropocentrism, humanism, and speciesism in critical animal theory. *Counterpoints*, 448(0), 3-27.
- Whitehead, A. N., & Price, L. (2001). *Dialogues of Alfred North Whitehead* (Vol. 84). David R. Godine Publisher.
- Whittaker, J. (2004). *The cyberspace handbook*. Routledge.
- Wilson, S., & Haslam, N. (2009). Is the future more or less human? Differing views of humanness in the posthumanism debate. *Journal for the Theory of Social Behaviour*, 39(2), 247-266. <https://doi.org/10.1111/j.1468-5914.2009.00398.x>
- Wolfe, C. (2010). *What is posthumanism?*. University of Minnesota Press.
- Wolfson, O. (2021). A relational-database methodology for incorporating consciousness into AI agents. *Journal of Artificial Intelligence and Consciousness*, 2(1), 67-80. <https://doi.org/10.1142/S2705078521500041>
- Wolkenstein, A. (2018). What has the Trolley Dilemma ever done for us (and what will it do in the future)? On some recent debates about the ethics of self-driving cars. *Ethics and Information Technology*, 20(3), 163-173. <https://doi.org/10.1007/s10676-018-9456-6>