



인공지능 시대의 새로운 윤리학 모색 루치아노 플로리디의 정보윤리학을 중심으로*

유용민 인제대학교 미디어커뮤니케이션학과 조교수**

이 글은 새로운 기술적 존재들이 진정한 지능적 존재인지 아닌지 여부에 집착하는 실체론적 관심에서 탈피하고 대신 관계론적 관점을 채택하는 입장에서 인공적인 잠재적-자율 행위자들이 인간과 관계를 맺어가는 오늘날 요청되는 윤리학적 대안을 고찰한다. 이를 위해 이 글은 인공지능 시대의 보편 윤리학의 대안 틀로서 루치아노 플로리디의 정보윤리학에 주목하여 그 가능성과 한계를 고찰해 보고자 했다. 그의 정보윤리학은 피동자-존재 중심의 윤리학으로서 인간과 비-인간 존재들 사이의 엄격한 위계성을 전제하는 인간 중심주의적인 윤리학적 사고방식이 가진 한계를 넘어서 구성주의 관점에서 정보권을 어떻게 가꿔갈 것인가에 관한 윤리학적 청사진을 제시한다. 존재 중심의 정보윤리학은 인간을 위한 인공지능이라거나 좋은 사회를 위한 인공지능 같은 대중적인 인공지능 윤리 내러티브 속 휴머니즘 전략의 맹점에 대한 비판적 인식을 열어주면서 기술적 존재들의 도덕적 지위 문제와 관련하여 실체론적 접근의 한계를 깨뜨룬다는 점에서 인간-기계 소통 시대의 실천 윤리적 전망을 제공한다.

KEYWORDS 인공지능, 인공지능윤리, 정보윤리, 행위자, 인간 중심주의, 휴머니즘, 플로리디

* 이 연구는 2021년 방송문화진흥회 연구지원사업의 지원을 받아 수행하였습니다. 아울러 본 논문의 완성도를 제고할 수 있도록 귀중한 심사의견을 주신 익명의 세분 심사위원 그리고 연구를 보조하며 수고를 아끼지 않은 연세대학교 커뮤니케이션대학원 서도원 박사과정생에게 다시 한번 감사 말씀을 드립니다.

** eudemonia38@naver.com

1. 서론

2016년 알파고가 전세계 대중 앞에 선보인 이후 인공지능은 더 이상 낯설지 않은 존재로 자리매김 중이다. 인공지능 기술은 현재 사법, 치안, 행정, 의료보건의, 산업 그리고 일상생활과 소비의 영역에 이르기까지 우리 곁에서 공존의 범위와 영역을 넓히고 있다. 이러한 변화를 아우르는 거대 서사로서 4차산업혁명의 진전과 함께 맥루한(Marshall McLuhan)적인 의미에서 인간 능력의 확장이라 할 미디어는 이제 인간 존재의 정수에 해당되는 지능적 기능까지 모사해 인간 행위에 개입하는 존재로 진화해 감으로써, 미디어 테크놀로지가 갖는 의미를 둘러싼 본질적인 변화의 가능성을 연출하고 있다. 이는 기존의 정보사회와는 결이 다른 또 다른 성격의 정보사회의 가능성과도 연관된다. 다니엘 벨(D. Bell)이 말했던 정보사회에서 미디어 테크놀로지는 기껏해야 인간을 보조하는 도구에 불과했지만 앞으로는 도구 이상의 의미와 효과로 인간과 관계를 맺어갈 것으로 예측된다.¹⁾ 이러한 변곡점의 가능성과 관련하여 커즈와일이 말한 인공지능이 언젠가 인간 지능을 압도할 것이란 특이점(singularity) 시나리오(Kurzweil, 2005)²⁾나 인공지능이 스스로 사고하고 행동함으로써 발생하는 다양한 현실적 문제들에 대해 우려하는 시각들, 인공지능에 의해 인간이 사회의 주변부로 내몰림으로써 인간의 가치와 존재 이유를 스스로 증명해야 하는 디스토피아적인 전망들은 인류 앞에 펼쳐지고 있는 인공지능 시대를 어떻게 바라볼 것인가와 관련하여 기존 지식이 앞으로도 여전히 유효할 것인가를 질문하고 있다.

1) 이러한 전망은 아직까지 심각하게 논쟁적인 주제다. 따라서 이 글에서 인공지능의 발전 전망에 관한 거의 모든 주장들은 잠정적인 전망일 뿐이다.

2) 특이점이란 용어는 레이 커즈와일에 의해 대중적으로 확산되었기는 하지만, 실제 이 용어는 수학자이면서 공상과학 소설가 베르너 빈지(Vinge, 1993)가 포스트 휴머니즘 시대 기술적 특이점의 도래에 관한 자신의 에세이에서 언급한 용어이다.

이런 파괴적 질문이 대표적으로 쟁점화되는 영역 중 하나는 인공지능 윤리(AI ethics) 영역이다. 인공지능³⁾이 대량의 수행 능력을 발휘하여 인간 삶에 개입하는 자동화 역량과 스스로 계산하여 판단하는 자율성 역량을 기반으로 인간 삶의 전반에 개입하면서 다양한 도덕적 윤리적 문제들이 발생하고 있다. 그에 따라 전세계적으로 인공지능 윤리 문제를 어떻게 다룰 것인가에 관한 학술적 논의나 제도적 해법 모색은 활발히 진행 중이며 인공지능 윤리에 관한 가이드라인과 윤리 헌장들도 계속 쏟아져 나오고 있다. 2016년 영국의 공학-물리학연구협회는 로봇 원칙을, 2017년 전세계 인공지능 관련 전문가들이 모인 〈유익한 인공지능 회의〉에서는 인공지능 연구개발에 관한 23가지 원칙(아실로마 원칙)을, 유럽연합은 2019년 인공지능 윤리 가이드라인을, 같은 해 경제협력개발기구(OECD)는 인간 중심 인공지능 원칙을, 그리고 유네스코(UNESCO)는 2020년 인공지능 윤리 권고 초안(First Draft of the Recommendation on the Ethics of Artificial Intelligence)을 각각 발표하였다. 국내에서도 인공지능 윤리에 대한 모색은 발 빠르게 진행되었다. 2018년 카카오톡의 인공지능 윤리 가이드라인, 2019년 삼성전자의 인공지능 윤리 가이드라인, 같은 해 방송통신위원회의 인공지능 윤리 원칙 등이 발표되었으며 이런 흐름은 계속될 전망이다.

인공지능 윤리에 대한 각국 정부와 기업 그리고 산업계들의 놀라울 만큼 발 빠른 대응들을 보면 흥미로운 지점이 발견된다. 그것은 전통적으로 도덕과 윤리를 과학기술 진보나 기술혁신의 장애물로 간주하여 윤리를 부차적 문제로 바라보고 기술에 의한 윤리적 문제를 해결하는 것을 기

3) 이 글은 인공지능에 대해 경험적으로 매우 엄격한 정의는 시도하지 않는다. 이 글의 목적상 인간의 지능적 기능을 대체하기 위해 개발, 보급되는 현실의 인공지능 테크놀로지들을 광범위하고, 그 기술들의 집합을 느슨하게 지칭할 따름이다. 이 글은 인공지능이 진정으로 지능적인가, 지능이란 무엇인가, 인공지능이 인간과 같은 지능을 가질 수 있는가와 같은 기술-실체론적 질문에는 덜 관심을 둔다. 그 이유는 이 글의 서론 후반에서 상세히 언급한다.

피하거나 정부나 시민사회의 비판이 제기된 경우에 한해 윤리 이슈에 소극적으로 대응하고 나서는 산업계조차도 인공지능 기술 개발 및 도입 과정에서부터 윤리 이슈와 쟁점을 적극 검토하고 있다는 것이다. 왜 이러한 아이러니가 관찰될까? 여러 가지 대답이 가능하겠지만 기술과 인간, 기술과 윤리, 윤리와 산업의 관계가 질적으로 변화하고 있음을 시사하는 징후로 해석해 볼 수 있다(손화철, 2018; 허유선·이연희·심지원, 2020).

물론 이러한 변화가 갖는 의미를 지나치게 강조해서는 안 된다는 목소리들도 있다. 적어도 현시점에서 인공지능은 인간의 지능 수준과 비교하기에는 현격한 차이가 있기 때문에 인공지능이란 존재의 영향력을 과장할 필요는 시기상조라는 것이다(Leslie-Kaelbling, 2021, 10월)⁴. 모든 상황에 인간처럼 대처 가능한⁵ 강한 인공지능(strong AI)/일반 인공지능(artificial general intelligence)은 실현 가능성이 극히 불투명한 것으로 전망되고 있으며 방대한 양의 데이터를 고속으로 처리하는 머신 러닝(machine learning)은 인간의 지적 역량 중에서 기껏해야 계산 능력을 반복하는 알고리즘 프로그래밍에 불과하다는 비판들(이호영 외, 2020; 이희은, 2021)도 존재한다. 실제 약한 인공지능에 해당되는 기술조차 아직은 단순한 판단도 실수하는 유아적 오류와 결함을 안고 있다. 그래서 인공지능의 등장으로 인간의 지능이 대체된다거나 인공지능이 스스로 느끼고 생각하는 상황을 전제로 인공지능 윤리 문제를 이해하려는 접근은 비과학적이며, 인간과 기계의 관계라든가 윤리적 세계를 살아가는 데 있어서 기계의 의미와 힘 그리고 그것이 인간에게 미치는 영향

4) <http://www.aitimes.com/news/articleView.html?idxno=141175>

5) 일반 인공지능에 대한 기술 개발 과정의 역사와 주요 논쟁에 관해서는 메사추세츠 공대 테크놀로지 리뷰(MIT Technology Review)의 2020년 10월호 특집 기사 〈일반 인공지능: 우리는 얼마나 가까이 왔나? 그리고 여전히 그것을 시도한다는 것의 의미는 무엇인가?(Artificial General Intelligence: Are We Close, and Does it even Make Sense to Try?)〉를 참조하라. 일반 인공지능에 대한 개발 시도는 초기부터 강력한 비판을 받아 왔지만 최근까지도 일반 인공지능의 가능성을 예언하거나 그에 대한 장밋빛 선전들을 찾아보긴 어렵지 않다.

력을 과대하게 인식하는 것은 오히려 현실을 잘못 이해하는 계기로 작용할 수 있다. 인공지능이 인간의 힘과 역량을 축소시킬 것이란 담론이 대유행하고 있지만 실제로 인공지능은 인간의 힘과 역량을 배가시키는 쪽으로 활용되고 있는 것이 더 진실에 가깝다(김진석, 2019, 11쪽). 인공지능을 누가 어떻게 개발 활용하고 이익은 누가 얻는가라는 정치경제학적 맥락에서 보면 보편적 인간 대 기계를 비-위계적으로 대립시키는 관점은 인공지능 문제를 사고하는 일반 틀이 되기도 어렵다. 근본적으로는 지능, 생각, 마음의 실체 같은 것들도 아직은 완전한 규명이 되지 않았다. 윤리라는 측면에서 보면 더더욱 그러하다. 그렇다면 인간의 윤리나 기계 윤리와 관련하여 어떤 실체적인 이해나 그에 근거한 실천은 불가능한 것일까?

이와 관련하여 기술적인 존재를 인간과 동일한 속성이나 형질을 갖는지 아닌지 여부에 따라 기술적 존재의 지위를 묻는 실체론적 관점에서 출발해 인공지능 윤리를 사고하려는 경향 또한 적지 않다. 그러나 그런 관점은 기술적 존재들이 인간이나 자연 같은 다른 존재들과 맺어가는 관계로부터 실질적으로 발생하는 문제들에 덜 주목한다는 한계를 갖는다. 콤켈버그(Coeckelbergh, 2014)는 그러한 관점을 도덕적 속성의 실체론적 접근이라고 부른다. 이에 따르면 인간과 기계 사이의 윤리적 공존을 위한 고민을 기계의 도덕적 지위가 어떤 속성이 근거하는지를 확인하는 인식론에 얽매일 필요는 없다. 인공지능이 그런 속성을 갖는지 여부와 별개로 인간이 기술적 존재들과 접촉, 소통하는 관계 속에서 창발되는 도덕적 지위의 가능성 그리고 그 윤리적 관계의 필요라는 관점도 살펴 봐야 한다. 즉 특정 속성에 대한 인식이 해결되어야만 관계의 윤리성이 현상되는 것은 아니다. 이런 관계론적 관점을 채택하면 도덕이란 무엇인가? 인공지능은 도덕적 지위를 갖는가? 같은 실체론적인 고민거리가 해결되지 못하더라도 인간이 비인간 존재들과 어떤 관계를 맺어갈 것인가라는 실천-윤리학적 질문을 보다 적극 주시할 수 있다(Coeckelbergh, 2014).

관계론적 관점에 주목한다면 인공지능 윤리 문제는 근본적으로 윤리가 처하는 변화된 조건을 실질적으로 만들어가는 구성주의 측면에 보다 주목할 필요가 있다. 이미 인공지능 기술 발달과 더불어 인간과 로봇 사이의 경계가 허물어지고 있으며 인간이 기술을 자신의 신체의 일부로 받아들이는 현상도 나타나고 있다.⁶⁾ 인공지능 기술이 접목된 로봇, 컴퓨터, 자율주행차량, 사물인터넷 또한 인간의 직권위임에 의하여 자율적인 역할을 수행함으로써 인간의 직접 통제와 개입 없이도 작동하는 잠재적인 인공적인 행위자로 우리 사회 곳곳에 배치되면서 인간이라는 존재의 일부를 이루고 있다. 그에 따라 인간이 아닌 인공지능에 의해 벌어진 도덕적 해악의 당사자는 누구인가? 인공지능이 누군가를 차별하거나 혐오하는 행위를 했을 때 그 책임을 누가 져야 하는가? 인공지능의 의도된 혹은 의도되지 않은 작동에 의하여 누군가의 사생활이 침해된다면 사생활 침해의 잘못은 누구에게 있는가? 혹은 인간은 로봇이나 인공지능을 어떻게 대하는 것이 좋은 것인가? 등과 같이 선뜻 답하기 쉽지 않은 종류의 질문들이 앞다투어 제기되고 있다.

앞으로 이런 질문은 더더욱 기술과 관련한 윤리적 논쟁의 중심에 놓일 것으로 전망된다. 인공지능 시대를 맞이하여 새롭게 등장한 이러한 윤리적 질문은 그 자체로 답하기 쉽지 않다. 인공지능이나 로봇 같은 새로운 기술들의 등장은 바로 윤리적 사고의 대전제가 되는 근본 범주들의 타당성을 흔들고 있기 때문이다. 인간, 정신, 신체, 자연, 인공, 원인(인과), 행위(자), 의사결정, 책임, 생명 등등 거의 윤리학적 근본 개념과 범주들은 이제 데카르트 이래 지금까지 지속되고 있는 인간=주체, 기술=객체 등식으로 구분되는 이원론적 사고방식으로는 포섭되지 않은 현상들의 출현 속에서 패러다임 전환의 필요와 맞닥뜨리고 있다.

6) 마즐리시(Mazlish, 1995)에 따르면, 인간과 기계, 자연과 인공 사이의 근본적 구분이 불확실해지는 현상은 인간과 다른 존재들 사이에 아주 근원적인 불연속과 단절이 있다는 휴머니즘의 타당성에 도전한다.

이런 도전 앞에서 인공지능 윤리의 원칙이나 기준들을 재정립하고 이를 인공지능 개발과 보급 과정에 적용하려는 노력들이 최근 들어 큰 관심을 모으고 인공지능 윤리와 관련한 원칙과 기준들을 산업 현장과 법제도에 반영하기 위한 노력이 앞다투어 벌어지고는 있다. 하지만 아직까지 인공지능이 발생시키는 또는 인공지능을 매개로 발생하는 윤리적 문제의 책임을 어떻게 해결할 것인가와 관련하여 뚜렷한 해답이 나온 것은 아니다.⁷⁾ 인공지능을 인간과 동등한 존재로 간주해야 하는지 아니면 단순한 도구로 취급해야 하는지? 인공지능에 의해 발생한 윤리적 문제의 책임을 인공지능을 개발, 운용하는 기업이나 정부에게만 물으면 충분한지? 거꾸로 인간이 인공지능을 윤리적으로 대해야 할 책임과 의무가 있는 것인지? 등등 인공지능 윤리의 문제는 이제 막 제출된 질문의 형태일 따름이다. 여기서 인간이 인공지능과 일상적 관계를 맺는 경험은 이제 막 출현했다는 점에서 인공지능 윤리를 둘러싼 현재의 지적 혼란들은 부정적이고 세기말적인 윤리적 위기라기보다는 스스로 작동하는 기술들과 인간이 앞으로 어떤 방식으로 관계를 맺어가는 것이 좋은가라는 윤리학적 실천이 어떻게 가능한지를 재사유하는 윤리학적 전환의 계기로 이해할 필요가 있다.

이러한 문제의식에서 이 글은 이탈리아 출신으로 옥스퍼드 대학에서 정보윤리 분야에 천착하고 있는 정보철학자 루치아노 플로리디(Luciano Floridi)의 정보윤리학에 주목하여 그의 정보윤리학이란 무엇이고 그리고 그것이 인공지능 같은 전례없는 기술적 존재들과 인간의 공존에 요구

7) 더구나 최근 하나둘씩 발표된 윤리 헌장들은 - 가령, 2016년 영국과 미국 그리고 유럽연합이 발표한 보고서들 - 하나같이 '좋은 사회를 위한 인공지능 윤리라는 테제를 표방함으로써(Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018) 인공지능 윤리를 '인간을 위한'이라는 맥락에서 접근하고 있지만 여기서 말하는 인간주의가 모두를 위한 보편적인 도덕적 가치를 지향하는 것은 아니다. 어디까지나 해당 국가나 정부가 추구하는 가치나 그들이 처해 있는 상황 속에서 그 의미가 자의적으로 제한될 수밖에 없다.

되는 윤리학적 대안으로서의 가능성과 한계를 탐색해 보고자 했다. 그의 정보윤리학은 서구 휴머니즘의 오랜 인간-중심적 윤리학의 세계관으로부터 탈피해 인간과 비인간 존재들 사이의 윤리적 관계의 가능성에 주목함으로써 존재 일반의 윤리학을 위한 윤리학적 확장을 꾀한다.

그의 정보윤리학을 다각도로 검토하기 위하여 유럽 학계에서는 별도의 심포지움이 열릴 정도로 윤리학의 이론적 갱신 문제에 대한 관심이 있었지만 한국에서는 인공지능 윤리 이슈에 대한 관심은 대체로 인공지능 윤리나 로봇 윤리를 산업적으로 어떻게 대처할지 혹은 법제도적으로 어떻게 다스릴 것인가라는 실무적인 측면에서 이슈화되는 경향이 큰 것 같다. 알파고 열풍에서부터 보다 최근의 이루다 논란처럼 기술 기업이 제공하는 인공지능 기반 미디어 서비스가 윤리적 논란에 휘말리자 관련 기업들은 인공지능 윤리 현장과 가이드라인 등을 만드는 일에 앞장섰다. 하지만 정작 그곳에서 제시되는 인공지능 윤리란 인공지능을 ‘사람을 위한 인공지능’ 같은 식으로 표현하거나 ‘사회 윤리 안에서’ 인공지능 윤리를 모색한다고 선언함으로써, 인공지능 윤리를 사회 윤리, 즉 인간 중심의 윤리라는 전통적 위계 안에 위치시키고 있다.⁸⁾ 물론 현 단계 기술 발전 수준을 고려할 때 인공지능 기술에 의한 도덕적 해악이나 피해의 발생을 예방하는 현실적인 조치가 필요하다 해도 그러한 인간 중심주의 접근은 인공지능이나 로봇 같은 것들을 철저히 인간의 편리하고 윤택한 삶을 돕는 기술적인 도구라는 관점에 여전히 국한시킴으로써 인간 이외의 기술적 존재들의 행위 또는 작용을 통해 야기되는 새로운 성질의 도덕적 윤리적 문제에 대처하는 데 한계를 가진다. 특히 그 논의들은 공통적으로 ‘좋은 사회’, ‘인간을 위한’이라는 윤리학적 은유들은 제시하고 있지만 ‘인간적인 것’이나 ‘사회적인 것’이 무조건 윤리적인 것이라는 명제는 어떤 철학적

8) 박성은 (2021. 2. 17). 네이버·카카오, AI 윤리 대책 앞다퉀 발표...이루다 사건 대응 성격 <AI타임즈>. URL: <http://www.aitimes.com/news/articleView.html?idxno=136619>.

이론적 근거를 갖는다고 보기 어렵다.)⁹⁾

이 글은 이런 문제의식을 플로리디의 정보윤리학을 중심으로 검토해 봄으로써 앞으로 다가올 본격적인 인공지능 시대에 필요한 윤리학의 갱신 필요성과 방향을 탐색해 보고자 한다. 그는 실체론적 접근보다는 구성주의 관점에서 정보사회의 새로운 도약 앞에서 고조되는 인간과 비인간 존재들 사이의 관계적 측면에 주목한 거시적이고 일반적인 정보윤리학의 확장을 꾀한다는 점에서 인공지능 시대의 윤리학을 둘러싼 논의에 의미 있는 논의의 틀을 제공한다. 이를 위해 이 글은 크게 다음과 같은 세 단계의 논의로 구성하였다. 첫째, 플로리디 정보 윤리학이 이론적으로 넘어서고자 하는 기존 윤리학의 한계의 주된 성격에 해당되는 인간 중심주의적 사고를 비판적으로 검토하고, 이러한 비판적 검토의 함의를 포스트 휴머니즘 논의나 최근 인공지능 윤리 논의에 적용하여 살펴보았다. 이를 통해 필자는 플로리디의 정보윤리학 같이 인간 중심주의에서 탈피하여 보다 관계론적이고 구성적인 정보윤리학을 적극적으로 주목해야 할 필요와 연결지었다. 다만 이 글에서 근대 이후 사상의 보편적이고 지배적 사고방식으로써 인간 중심주의에 관한 모든 논의를 조망한다는 것은 불가능한 일이다. 따라서 여기서는 인간을 행위하는 유일 존재, 즉 주체로 간주하는 입장에 특히 비판적 초점을 두었다. 둘째, 인간 중심적 윤리학과 차별화되는 정보윤리에 관한 플로리디의 핵심 논의들을 탐색적으로 살펴보고, 그의 정보윤리학이 인공지능 시대의 윤리학적 대안으로서 갖는 가능성을 고찰한다.¹⁰⁾ 셋째, 끝으로 이 글은 플로리디의 정보윤리학이 갖는 맹점

9) 우리 언론학 분야 주요 학술저널에서 인공지능 윤리를 주제로 하는 학술 논의들 또한 대부분 인공지능 챗봇, 건강/보건 영역에서의 인공지능 활용, 인공지능 도입에 따른 방송정책 변화 등 인공지능 기술의 산업적 도입에 따른 윤리적 이슈에 대한 응용윤리적 접근에 머물거나 인공지능 윤리 문제를 다룬다 해도 기존의 정보격차, 사회적 약자 계층의 기술-소외 문제 등 전통적인 인간 중심적 관점에서의 미디어 윤리적 관점에서 출발해 이를 인공지능 기술 문제에 대입하는 논의들에 머물러 있다(가령, 김용찬, 2020; 박남기, 2020; 박남기·김윤정·장은채, 2021; 손승혜, 2020; 송하연·김지은·김태년, 2020; 홍대식, 2020).

과 한계에 대해서도 아울러 검토함으로써 인공지능 시대 정보윤리학에 남겨진 숙제를 확인하고, 이를 인공지능 시대 윤리 이론의 나아갈 바와 연관지어 전망해 보았다.

2. 인간 중심주의와 인공지능 윤리에 대한 검토

인간 중심주의는 인간의 경험, 가치를 근간으로 세계를 해석하는 입장을 아우르는 개념이다. 이러한 개념에서 세계는 인간이 행위하는 무대로 간주된다. 이러한 관점 안에서 인간은 행위의 유일한 주체가 되며 나머지 존재들은 모두 객체로 상정된다. 이러한 이원론은 많은 비판과 반론을 받았음에도 데카르트 이후 지금까지 지속되고 있지만, 인공지능 같은 지능적 테크놀로지들이 등장하면서 인간 중심적 사고에 대한 기술철학적 반론과 논쟁으로부터 다시 한번 큰 도전을 받고 있다. 그 중심적 이유는 크게 두 가지 측면에서 살펴볼 수 있다. 하나는 데카르트적인 인간 중심주의의 타당성과 관련하여, 인간만이 정신과 의식을 소유하고 있는 반면, 인간을 제외한 모든 존재는 정신과 의식을 소유하고 있지 않다는 데카르트

10) 본론에 해당하는 논의는 다시 다음 세 가지 맥락으로 구성하였다. 첫째, 플로리디의 정보윤리학은 기존의 윤리학을 ICT나 디지털 영역에 대입하는 응용윤리보다 윤리학적 세계관을 재정의하는 거시적 보편적 윤리학의 성격을 갖는다는 점을 고려할 때 윤리학이 투영되는 세계 그 자체에 대한 그의 이해에 우선 주목할 필요가 있다. 이에 이 글은 우선 초역사(hyper-history) 시대와 인포스피어(infosphere)라는 그의 테제를 살펴보고자 한다. 둘째, 인포스피어 시대에 요구되는 윤리학으로 플로리디가 제시하는 피동자 중심(patient-oriented) 그리고 존재 중심적 윤리학에 대해 검토하고 그것이 다가오는 인공지능 시대 전통적인 인간 중심적 그리고 행위자 중심의 윤리학이 갖는 한계와 오류를 어떻게 극복할 수 있는지에 관한 윤리학적 대안으로서의 잠재력을 중심으로 고찰해 보고자 한다. 이를 위해 피동자 중심의 윤리학 그리고 분산된 도덕성 모델에 관한 그의 논의를 중점적으로 살펴볼 것이다. 셋째, 플로리디의 정보 윤리학에서 윤리적 존재로서 인간의 상은 어떻게 재형성 되는지와 관련한 플로리디의 정보 생태주의의 관점을 그가 제시한 호모 포이에티쿠스(homo poieticus) 인간관 개념에 대해 검토하고, 그것이 플로리디의 정보윤리학의 성격에 어떤 함의를 제공하는지를 필자 나름의 시선을 통해 살펴보고자 한다.

트적인 인간 중심주의의 대전제(Candea, 2010)의 타당성이 흔들리고 있다는 점이다. 둘째, 인간이 주인이 되어 만들어낸 현대 문명의 위기와 관련해서, 인간 중심주의가 타당하다 해도 그것이 야기한 결과가 현대문명의 위기를 해결하기보다 오히려 그 위기를 심화시키거나 지속시키는 원인이 되고 있다는 문제의식에서 인간 중심주의의 한계를 극복해야 할 필요하다. 보다 최근에는 인류세(anthropocene)에 대해 주목하면서 급변하는 현대 문명의 위기 속에서 요구되는 인간 존재에 대한 근본적 성찰의 필요성과 맞물려 인간 중심주의에 대한 비판적 논의가 활발히 모색되고 있기도 하다.

이와 관련하여 소위 4차산업혁명으로 담론화된 테크놀로지의 급진적 발전으로 인간과 비인간, 자연적인 것과 인공적인 것, 주체와 객체 사이에 설정되는 데카르트적인 이분법적 영역으로 기술 문명의 변화가 갖는 의미를 이해하려는 방식의 타당성이 도전받는 일들이 늘어나고 있다. 인간과 기술의 경계가 점점 더 불분명해지는 하이브리드적 세계 안에서 인간만을 유일한 행위 주체로 상정하는 인간 중심적 사고는 인공지능 윤리 논쟁을 포함하여 현재 기술 변화와 더불어 발생하는 현대사회의 다양한 위기들을 해결하는 데 적절한 해법이 되고 있지도 못하다. 인공지능이 인간과 유사하게 지능적 존재로 발전해 갈 것이란 기술-낙관적 전망에 따라 로봇, 인공지능 같은 기술적 도구들의 존재론적 지위를 객체가 아닌 주체로 재조정해야 할 필요도 활발히 논쟁되고 있다. 자율성, 판단 능력, 창조성을 지닌 인공지능 기술이 발달하면서 인간의 본질은 사유에 있고 다른 모든 사물과 구별되는 인간 고유의 특징은 의식이며 의식을 갖춘 인간만이 자율적 존재라는 데카르트적 인간 중심주의로는 더 이상 인간의 개념을 특징짓는 데 한계가 발생하고 있다(강영안 이상현, 2013, 154-155쪽). 인간만을 주체로 상정하는 인간 중심주의의 또 다른 중요한 문제점은 인간 이외의 비인간적 존재들을 수동화하고 주변화한다는 것이다. 인간 중심주의 담론은 인간이라는 주체가 갖는 가치를 향상시키는 지향성을 갖

고있는 것이며, 따라서 그러한 인간성의 향상을 위한 실천 과정에서 비인간 존재들을 수동적 존재의 지위에 머무르는 일은 정당화된다. 그러나 문제는 인간이 저지르는 많은 도덕적 해악이나 인간이 자연이나 사회에 미치는 악한 영향력이 바로 이러한 인간의 의도와 목적으로 인해 주도된다는 점이다.

인간 중심주의에 대한 비판은 기술과 인간을 구분하는 실체론적 차원에서 인간 중심주의에 제기되는 도전뿐만 아니라 관계론적 측면에서도 불가피하다. 인공지능이나 로봇 같은 기술적 존재들이 인간과 소통하고 정서적 관계를 맺거나 인간 대 인간이 나누던 상호작용을 인간과 기계가 시도해가는 것이 가능해지면서 인간과 비인간 사이에서 발생하는 윤리적 문제를 고려하기 위해서는 인간과 비인간 사이에 설정된 위계성을 넘어설 필요가 있다. 이를 위해서는 인간 주체에만 초점을 맞추는 대신 인간과 사물 사이의 관계적 연결망과 그것들의 연합이 주고받는 영향에 주목해야 하며(Delanda, 2006; Latour, 2005), 이는 자연스럽게 자연, 사회, 기술의 나머지 부분과 인간이 관계를 맺을 때 인간의 지위를 배타적으로 특권화하는 이론적 입장에 물음표를 던진다. 가령 라투르의 행위자-연결망 이론은 관계론과 실체론을 종합하면서 인간과 비인간 사이의 관계를 기반으로 사회를 바라보아야 한다고 말한다. 이런 관점에서 사회에 대한 이해를 인간과 비인간이란 위계적 이분법에 기반하는 사고에서 벗어나 인간과 비인간 존재들의 관계성을 포괄적으로 포용하는 관점으로 이행할 필요가 강조되고 있기도 하다(김왕배, 2021). 인류 문명을 비판적으로 사유하는 최근의 논의들 또한 인간이 만들어가는 역사 시대를 더 좋게 만들기 위해 보다 '강력한 인간'이 필요하다는 논리는 오히려 비판적 검토의 대상이 되어야 한다는 인간 중심주의에 대한 실천적 문제제기도 들리고 있다(이에 대해서는 김환석, 2018을 보라).

인간 중심주의를 넘어서고자 하는 일련의 사상적 시도는 현재 포스트 휴머니즘 영역에서 활발히 논의되고 있다. 가령 페퍼렐(Pepperell,

2009)은 포스트 휴머니즘의 조건에 관한 논의에서 인간을 유일 주체로 상징하는 역사 발전 단계는 끝났으며 인류는 휴머니즘 이후의 시대로 이미 진입했음을 주장하면서 인간이란 과거와 동일한 방식으로 존재할 수 있다는 생각은 불가능하다고 주장한다. 이러한 주장의 근거는 인간은 기계를 자신의 일부로 받아들이는 한편, 기계는 점점 더 인간을 닮아가는 현상에 있다. 이러한 현상이 심화되어 간다면 인간과 기계 사이의 관계는 생명과 비생명 사이에 놓여 있던 위계적 경계를 넘어 상호 수렴하며 그것들 간의 범주적 구분의 타당성이 점차 모호해지거나 약해지는 상황, 즉 생명 이후의 시대적 조건에 위치해 갈 것으로 전망된다.

물론 인간 중심주의라는 사고방식은 인류 역사에서 긴 역사를 갖고 다양한 층위와 맥락에서 논의된 것이기 때문에, 인간 중심주의는 물론이거니와 그것을 넘어서고자 하는 새로운 사고의 방식들 간의 논쟁 관계를 인간중심 대 탈/비 인간중심이라는 이분법으로 간단히 종합하기는 어렵다. 포스트 휴머니즘 논의만 하더라도 이념, 민족, 종교, 생태, 동물 윤리, 페미니즘, 탈식민 이론 등 다양한 영역에서 담론화되고 있다. 하지만 그러한 담론들의 공통점으로 인간, 기계, 생명, 존재, 행위 등의 근본 개념들에 대해 새로운 이해를 적극적으로 요청하고 특히 그러한 요청에 있어서 인간과 비인간 존재의 조화와 공존을 모색하려고 시도한다는 점에 주목해 볼 수 있다. 이를 위해서 인간을 유일한 역사 주체이자 이성 행위자로 간주하고 다른 생명체나 자연 그리고 기술적 존재들을 주체의 자리에서 배제하는 관점은 인간의 사유와 행위만이 도덕적 세계를 만들 수 있다는 독단적이고 임의적인 인간 중심주의는 인간 개념에 내재된 위계에 수반되는 차별과 불평등 그리고 폭력 같은 문제들을 어떻게 극복할 것인가라는 도전 앞에서 비판과 극복의 대상이 된다. 물론 그러한 제각각의 주목들이 논리적 귀결이 인간중심주의로부터의 완전한 탈피인지 아니면 인간중심주의의 재구성인지는 간단히 대답할 수 있는 문제는 아니고, 이 글에서 상세히 종합하기도 어렵다. 하지만 어떤 경우를 지향하든 기존의

인간중심주의에 대한 비판적 시선을 도입하는 일은 불가피해 보인다.

인공지능 윤리 논쟁 또한 이러한 인간중심주의의 도전을 받을 부분이 많지만, 인공지능 윤리 논의는 여전히 다각도의 의미에서의 인간 중심주의 안에 안락하게 머물러 있는 듯하다. 1956년 다투머스 회의¹¹⁾에서 인공지능 개념이 공식적으로 처음 검토된 이후 1960년대를 넘어서면서 인공지능 윤리에 관한 연구들이 본격적으로 등장하기 시작했다(Samuel, 1960). 하지만 인공지능 윤리가 본격적으로 활성화된 시기는 2010년대 중반부터로 이후 인공지능에 관한 철학적, 정책적 논의들이 활발히 검토되고 있다. 이들 논의들은 이미 야기되었거나 앞으로 다가올 인공지능 관련 윤리 문제를 확인하고 그 대응 방안을 다루고 있는데 대부분 '인간성'이라는 가치를 공통적으로 중시한다는 공통성을 갖는다(Hagendorff, 2019). 일일이 열거하기 어려운 최근의 인공지능 윤리 논의들은 거의 대부분 인간적 가치의 지속과 증진의 문제를 다루고 도덕적 선을 위해 요구되는 기술적 문제들을 살펴보는 방식들을 취하고 있다.

이와 관련하여 인공지능 윤리 논의에서 간과되고 있는 지점은 개별적인 기술적 이슈들에 대한 검토를 넘어서 상이한 이슈들, 주제들을 포괄적으로 사고할 수 있는 철학적 논의가 상당히 부재하다는 점이다(Höffe, 1993/2013). 가령 허유선 등(2020)은 인공지능 기술 산업이 앞으로 계속해서 발전해가는 상황에서 인공지능 윤리를 아우를 수 있는 전체 조망과 비전에 대한 논의가 더욱 요청된다는 점을 지적한다. 이들은 인공지능 윤리에 대한 문제의식은 양적으로 상당한 성장을 거두었지만 그에 비해 인공지능 윤리의 근본적인 정체와 목적에 대한 관심과 그에 기초한 합의나 공감대가 여전히 부재하다는 점을 지적한다. 실제로 이들은 인공지능 윤리를 다루는 국내외 주요 문헌들을 분석하면서 인공지능 윤리에 대한

11) 인공지능이라는 고유한 과학적 분야가 확립되는 데 초석을 닦은 학술회의를 지칭한다. 1956년 다투머스대학에서 개최되었고, 존 매카시, 마빈 민스키, 너대니얼 로체스터, 끌로드 웨넌 등이 참여했다.

문헌들은 다양함에도 인공지능 윤리가 왜 요청되는지에 대한 논의가 심도 깊게 전개된 경우를 찾아보기 어렵다는 점을 지적한다. 이러한 경향 속에서 인공지능 윤리는 상당 부분 기술의 신속한 상품화와 그 과정에서 통과의례적으로 다뤄지는 식으로 논의되고 있을 가능성도 크다. 이는 곧 인공지능 윤리 논쟁에서 인공지능 윤리 개념의 외연과 함축이 협소하게 설정되어 있음을 시사한다.

상술한 바와 같이 인공지능 윤리 논쟁은 너무나 풍성한 것처럼 보이지만 실상 인공지능 윤리가 왜 요구되고 그것이 무엇을 지향해야 하는가에 관한 전체적인 의미를 펼쳐보고 그것을 조망해가는 논의는 역설적으로 부족하다. 이와 관련하여 다양한 요인들을 고려할 수 있지만, 그 중에서 이 글의 문제의식과 관련한 쟁점은 인공지능을 논의하는 주체도, 인공지능의 윤리가 적용되는 주체도 어디까지나 인간이라는 개념에 국한되어 있다는 점일 수 있다. 가령 하켄도르프(Hagendorff, 2019, p. 3)는 인공지능 윤리는 여전히 서구 백인들에 의해 논의되고 있으며, 그에 따라 인공지능 윤리 담론이 다양화 다면화되고 있지 못하다고 지적한다. 이러한 상황에서 인공지능 윤리에 대한 전지구적 합의를 추구하려는 시도는 다양한 지정학적, 인종적 차이나 다양성을 인공지능 윤리 문제에 반영하지 못하는 인공지능 윤리의 정당성, 대표성의 결핍을 불러온다(Crawford et al., 2019; Jobin, Marcello, & Effic, 2019). 더구나 인공지능 윤리라는 개념이 인공지능이 일으키는 문제를 해결하는 '수단'으로 간주되는 경향이 우세하다 보니, 인공지능 윤리 논쟁이 단지 기술적 문제로 축소됨으로써, 기술을 우선시하는 논리로부터 인공지능 윤리를 구성해가는 논의에 치우칠 가능성도 커지고 있다.

문헌검토를 통해 살펴본 이러한 인공지능 윤리 논의의 지배적 특징들 안에는 인간 중심주의가 그 중심에 일정하게 자리잡고 있다고 볼 수 있다. 기존의 사회적 논의 구도나 지형 자체가 인공지능 윤리 논의 구도를 한정하면서 그 내용 또한 인간이 주체가 되어 풀어나가는 윤리 문제로

다뤄가는 방식들이 그렇다. 그러나 이러한 접근은 인공지능 기술을 이해하는 데에도 그리고 기술과 인간이 관계를 맺어가는 과정 상에도 적절한 윤리적 관점이 되긴 어려워 보인다. 인공지능 기술은 전적으로 통제 예측될 수 있는 기술로 보이기도 어려울 뿐만 아니라, 지금의 기술 문명은 인간과 기술이 상호 영향을 미치며 공-구성해가는 단계로 나아가고 있다는 점을 보다 적극적으로 고려할 필요가 있다. 인공지능 시대의 윤리학적 모색은 기술과 윤리에 설정된 기존의 인간 중심주의적 논리들의 제약과 전통을 확장하고 진화할 필요가 있다. 그러한 관점이 인간 중심주의를 완전히 포기하는 것이어야 할지는 확실치 않지만, 그러한 관점보다는 인간 중심성에 보다 개방적이고 포용적인, 그리고 그 결과 성찰적인 윤리적 동력을 제공할 수 있는 관점이 필요하다는 문제의식은 충분히 검토해 볼 수 있다. 이 글이 플로리디의 정보 윤리학에 주목해 그것의 가능성과 한계를 탐색해 보려는 이유도 여기에 있다. 다음 장에서는 플로리디의 정보윤리학의 배경과 그 열개를 상세히 고찰해 보고자 한다.

3. 루치아노 플로리디의 정보윤리학

1) 정보윤리학의 토대

최근의 기술-사회 담론은 4차산업혁명 시대, 인공지능 사회, 지능정보사회(배영 외, 2021) 등 다양한 표현으로 불리고 있지만 그 초점은 디지털 테크놀로지의 발전에 힘입어 가속화되는 정보 혁명이라 할 수 있다. 플로리디에게 정보 혁명이란 표현은 기존의 디지털 담론에서 말하는 정보 혁명과는 다른 의미로 이해된다. 기존의 정보혁명이 인터넷 혁명, 디지털 혁명 같이 새로운 기술에 의해 나타나는 사회 변화를 설명하는 데 초점이 주어진다면 플로리디의 정보혁명은 우리가 누구인가에 관한 우리 스스로에 대한 이해의 문제로 확장된다. 이를 위해 플로리디는 소위 말하는 인

터넷 혁명, 디지털 혁명 같은 것들을 네 번째 혁명(the fourth revolution)이라는 표현으로 재정의한다(Floridi, 2013a, pp. 13-14). 첫 번째 혁명은 코페르니쿠스에 의한 혁명으로 인간이 우주의 중심이 아니라 주변의 일원이라는 인식의 등장으로 야기된 혁명이다. 두 번째 혁명은 찰스 다윈(C. Darwin)에 의한 것으로 인간이 생명의 중심이라는 발상에서 인간 또한 많은 생명과 같은 진화와 적응의 산물이라는 사고로부터 파급된 혁명이다. 세 번째 혁명은 인간 무의식을 발견한 지그문트 프로이트(S. Freud)나 포스트 모더니즘 사조가 던진 테제, 즉 인간 이성은 합리적 이성적 성격으로 보기 어렵다는 자기 인식 혁명이다. 그리고 이다음 등장한 혁명이 제4의 혁명으로 정보통신기술의 비약적 발전으로 인하여 자연과 인공의 관계를 둘러싼 변화가 급진전되고 그 과정에서 인간과 우리의 외부 사이의 관계를 둘러싸고 강력한 변화가 일어났다고 플로리디는 말한다(Floridi, 2013a, p. 1, pp. 13-14).

이러한 정보혁명의 의미를 이해하기 위해 플로리디는 역사(history) 시대와 초 역사(hyper-history) 시대를 구분하는 기준을 도입한다(Floridi, 2014). 플로리디에게 역사 시대와 초역사 시대는 모두 정보의 관점으로 범주화 가능한 시대적 단계다. 미디어 분야에서 정보혁명은 대체로 다니엘 벨에 의해 주창된 후기산업사회/탈산업사회(post-industrial society)의 도래(Bell, 1976)와 더불어 등장한 정보사회라는 맥락을 지칭한다. 그러나 플로리디는 이보다 범위를 더 넓혀 문자체계가 등장한 시점부터를 정보 시대로 명명한다. 문자 정보를 도구로 활용하여 쓰고 기록할 수 있는 기술적 조건의 출현이야말로 그 이전의 선사 시대와는 질적으로 다른 성격을 사회에 부여했다는 것이다.

초역사 시대는 디지털 테크놀로지에 의해 촉발된 새로운 역사적 단계로 디지털 테크놀로지에 대한 인간 삶의 의존도가 비약적으로 확장된 시대다. 이러한 차이는 플로리디의 역사 구분에 약간의 의문을 제기하는데, 역사 시대와 초역사 시대 모두 '정보'의 시대라는 일련의 연속성과 연

관성을 갖는다면, 역사 시대와 초 역사 시대라는 구분은 단순히 정도나 강도의 문제일 뿐, 본질적으로 서로 다른 시대로 규정되기 어렵다고 볼 수 있다. 플로리디는 이에 대해 초역사 시대의 본질적 특성은 정보가 도구/자원의 범주에서 벗어나 우리가 살아가는 환경, 인간의 행위를 가능케 하면서 동시에 제약하는 조건 그 자체로 작용한다는 점을 근거로 역사/초역사 구분의 타당성을 찾는다. 역사 시대가 정보 저장과 전달의 측면에서 혁명적이었다면 초역사 시대는 정보 전달과 저장을 넘어서 인간이 정보 환경 그 자체를 존재의 존립 조건으로 삼는 혁명적 변화를 의미한다 (Floridi, 2014, pp. 1-24).

이런 맥락에서 역사 시대와 초역사 시대는 서로 다르면서 동시에 정보라는 층위에서 연속성을 갖는다. 플로리디는 정보라는 것이 기존에는 존재하지 않았지만 기술 발전에 의해서 출현한 것으로 이해하기보다는 초역사 시대에 이르러 다시 인간의 시야에 포착된 것, 즉 항상적으로 존재하고 있었던 것으로 간주하는 관점을 도입한다. 바이오스피어(biosphere)와 인포스피어(inforsphere)라는 대조적 비유가 그 대목인데, 플로리디에 따르면 초역사 시대를 맞이하여 물질, 지구, 자연 환경 같은 바이오적 장에 의해 가려져 있던 정보적 장을 인식할 수 있게 됐다는 점을 강조함으로써, “정보적 존재론”¹²⁾의 가능성을 개방한다. 인포스피어는 인간이 자연과 물리적 조건을 환경 삼아 살아가는 바이오스피어와는 다른 인식이 가능한 곳으로 모든 정보적인 행위자들이 서로 상호작용하는 정보적 환경 그 자체를 의미한다(Floridi, 2013a, 2014, pp. 25-48).

이러한 인포스피어 개념은 외견상 온라인과 디지털 공간 같은 가상

12) 이러한 존재론적 입장은 현재 서구 학계에서 논쟁의 지점이다. 대표적으로 정보존재론이나 디지털존재론이나 하는 입장 차이는 기술철학자들 사이에서 논란이 되고 있는데, 가령 플로리디의 경우는 정보존재론을 이야기하지만 기술철학자 중 한 명인 욱허이의 경우는 디지털존재론을 이야기한다. 이 논쟁이 갖는 의미나 필요에 대해서는 다음 기회에 본격적으로 소개하고자 한다.

세계로 환원되는 개념인 것처럼 보인다. 그러나 플로리디의 인포스피어 개념은 오프라인과 온라인을 모두 포괄하는 공간이자 정보를 주고 받으며 상호 작용하는 모든 존재들이 공존하는 곳이다. 이런 개념 정의는 인간이나 동물 같은 유기체적 생명체를 중심으로 하는 기존의 생태학적 환경 개념을 넘어 인공지능, 로봇, 사물인터넷 같은 기술적인 존재들까지 포괄하는 보다 더 큰 혹은 상위의 존재론적 개념이다. 이런 개념을 통해 플로리디가 의도하는 바는 우리가 자연물은 물론 기술적인 존재들과 공존하는 세계란 오프라인이나 온라인이라는 이분법 또는 정보란 오직 인간의 도구 활용 목적을 돕는 기술적 도구들이라는 관점으로 환원되기 어려운 것임을 강조하는 데 있다. 현재 온라인과 오프라인의 경계가 약화되고 상호작용하면서 상호-구성해가는 현실은 인포스피어가 이미 도래 중인 현실임을 시사한다. 플로리디는 이제 우리가 말하는 현실이란 정보적으로 이해되는 세계에 점점 더 다가서고 있다고 말한다(Floridi, 2014, p. 7).

인포스피어 안에서 주체는 본질적으로 정보적이다. 플로리디는 그러한 주체를 인포그(inforg)라고 부른다(Floridi, 2014, pp. 94-96).¹³⁾ 인포그는 정보적인 유기체들이다. 그러나 유기체라고 하면 기계가 배제되는 것은 아니라 정보적으로 이해 가능한 모든 존재자들을 말한다. 인간과 기술의 상호작용은 근현대에 이르러 촉발된 것이 아니라 오래전부터 시작된 것이다.¹⁴⁾ 인간은 과학기술을 신체나 정신 기능의 일부로 수용한 포스트 휴먼이나 트랜스 휴먼적 존재가 아니더라도 오래전부터 인포그적

13) 이러한 인포그 개념은 아마도 사이보그(cyborg) 개념과 유사해 보인다. 다만 기계와 인간의 결합체를 지칭하는 사이보그 개념에는 여전히 인간 중심성, 생명과 분리되지 않은 존재론적 사유가 깃들어 있는 반면, 인포그는 그보다 존재의 개념을 더 확장한다는 의미가 담겨 있다.

14) 이에 관한 대표적인 논의로는 행위자-네트워크 이론(대표적으로 Latour, 1991) 그리고 태초부터 기술의 도움 없이는 삶의 영위하기 어려웠던 보철적 존재로서의 인간에 관한 기술철학적 논의로는 스티글러(Stiegler, 2015)의 논의를 참고.

인 존재라 할 수 있다. 디지털 네트워크가 멈추면 인간 활동이 정지되는 사회, 스마트 미디어가 없이는 사회적 관계를 추구하기 힘들어진 인간의 모습은 우리가 인포스피어 속 인포그로 체현된 삶을 영위 중이라는 존재론적 상황을 보여준다. 이러한 존재론적 상황은 인간과 비인간 사이의 유사성과 연속성에 주목하면서 인간 이외의 존재들 또한 정보적인 존재로 바라보는 존재론적 패러다임으로 제시한다. 그 속에서 생명과 비생명, 인간과 기계, 자연과 인공, 주체와 객체 같은 기존의 존재론적 범주들은 정보적 패턴들로 환원되어 이해될 수 있는 존재들이기에, 인간과 기계, 자연과 인공의 관계에는 불확정적이고 우연적이며 임의적인 그리고 연속적인 관계로 재정의된다. 이는 인간을 정점에 놓고 자연, 사물, 기계 등을 그 아래에 놓는 위계적인 인간 중심적 존재론으로부터의 일정한 초월을 의미하는 것으로, 인간과 비인간이 정보적으로 상호작용하면서 세계를 공-구성해가는 정보권 시대의 존재론적 대안으로서의 의미, 즉 가상의 인공적인 존재들이 인간과 소통하면서 상호작용하는 새로운 행위 주체로 출현하는 변화를 반영한 새로운 존재론이다. 이러한 존재론은 많은 디지털 담론에 내재된 인간에 대한 전통적인 휴머니즘적 이해를 넘어서 보다 포스트 휴먼적인 주체로 인간의 의미를 재개념화하려는 시도이기도 하다.

물론 그의 논의에서 정보란 무엇인가에 대해서는 논란의 여지가 남는다. 정보가 우리 삶에서 지배적 영향을 미치는 개념이 된 것 자체는 부정할 수 없지만 과연 정보가 무엇인가에는 확실한 답이 아직까지도 마련되어 있지 않다(김재희, 2021). 플로리디의 정보철학은 세계에 존재하는 문제를 정보적인 문제로 환원하여 이해할 수 있다는 것이지, 정보 그 자체에 대한 궁극적 원리나 설명을 제공하지 못한다. 다만 정보철학 관점에서 정보는 인간, 기계, 자연, 생명 같은 세계 내 존재들을 사유할 수 있는 존재론적인 토대이며 이 토대에 대한 관심을 통해 우리는 기술에 관한 정보, 생물에 관한 정보, 기호에 관한 정보, 사회에 관한 정보, 자연에 관한 정보 등 정보가 세계를 이해하는 보편형식 중 하나임을 확인하고 이를 통

해 세계 자체에 대한 이해를 확장해갈 수 있다는 점에서 의미가 없진 않다. 우리가 거주하는 세계가 정보에 대한 의존도가 높아지고 있다면 현실에 대한 정보적 해석이 요구되며 이는 곧 세계 이해의 방법으로서의 철학의 정보(론)적 전환이 필요하단 뜻이다. 이런 전환적 세계관 속에서 인간이나 다른 존재들 모두 인포스피어 내에서 살아가는 존재자¹⁵⁾들이며 그 존재자들 모두는 정보적 시스템으로 이해될 대상으로 인식할 수 있다. 그래서 플로리디의 정보철학을 정보 자체의 근원을 규명하는 철학으로 오해하는 것은 곤란하다. 플로리디 정보철학은 정보에 대한 모든 문제를 해결하는 궁극의 원리를 제공하는 것이 아니라 인포스피어 같은 정보적 세계에서 발생하는 문제들을 다루기 위한 개념적 토대를 제공하는 기획에 가깝다(Floridi, 2013a, pp. 313-314)는 점을 유념할 필요가 있다.

플로리디의 하이퍼 역사 시대는 정보나 정보를 다루는 디지털 기술들이 인간의 목적 달성을 위한 도구의 범주로부터 벗어나 인간의 삶의 본질적 조건으로 작용하는 시대이다. 하이퍼 역사 시대는 정보를 처리하는 기계 장치들이 양적으로 늘어나면서 기계나 기계가 다루는 정보량 자체가 폭발적으로 커지는 특징을 보인다. 하지만 이는 다니엘 벨이 말했던 후기산업사회의 맥락과도 크게 다르지 않다. 기계장치의 수와 정보량이 증대되는 현상 자체는 일련의 연속성을 갖기 때문이다. 그러나 하이퍼 역사 시대가 과거의 정보사회와 구별되는 중요한 이유는 그 기계들이 인간의 직접 개입 없이도 상호 협력과 조정을 통해서 정보 환경의 구성에 영향을 미친다는 점에서(Floridi, 2014, pp. 4-11) 초역사 시대라는 은유는 인공지능 시대가 기존 시대와 구분되는 불연속적이고 단절적인 기술적 맥락과 의미를 전망해가는 데 의미가 있다.

이러한 기계들의 자율적인 활동은 이미 전지구적인 범위에서 일어

15) 물론 기술적 존재가 존재자가 될 수 있는냐는 확실히 답하기 어려우며, 하이데거적 관점에서 기술은 존재자가 될 수 없다는 견해들이 아직은 우세하다(이에 대해서는 김병선, 2007을 보라).

나는 현상으로 하이퍼 역사 시대의 지배적인 기술의 성격을 보여준다. 플로리디는 이러한 변화 양상과 그 의미를 기술에 대한 3단계 구분론을 통해서 설명하는데, 1차 기술과 2차 기술 그리고 3차 기술의 개념들이 그것이다(Floridi, 2014, pp. 25-32). 플로리디는 기술의 사용자(user), 촉진자(prompter) 그리고 사이성(in-betweenness) 개념을 적용하여, 이들 1차, 2차, 3차 기술을 구분한다.¹⁶⁾ 여기서 기존의 기술 개념과는 구별되게 제시되는 개념이 3차 기술 개념이다. 하이퍼 역사 시대로 넘어가는 과정에서 새롭게 상정되는 지배적인 기술로서 3차 기술이다. 3차 기술이 기존의 기술들과 갖는 차이는 3차 기술의 경우 사용자와 촉진자 모두가 자연이나 인간이 아닌 기술이나 기계라는 점이다. 기계 장치들 사이에 존재하는 이 3차 기술은 기술들 사이에서 그 기술들을 매개한다. 알고리즘은 바로 대표적인 3차 기술이다. 3차 기술이 출현한 결과 인간은 수평적인 기술과의 사이성을 맺는 관계적 존재가 아니라 사이성을 갖는 기술들의 경계 외부에서 기술로부터 무언가를 얻는 존재가 된다. 즉 기술의 이용자에서 기술의 소비자로 탈바꿈되는 것이다(신상규, 2016). 예를 들어 사람이 라디오를 이용하기 위해서는 다양한 의사결정이 요구되었지만 이제는 알고리즘이 선호하는 노래를 추천하고 자동으로 플레이시켜줌으로써 인간의 사용 행위는 대폭 감소되고 있다. 스스로 학습하는 알고리즘이 내장된 사물인터넷이나 인공지능이 접목된 스마트 미디어 장치들은 이런 기술들을 매개하는 사이성을 갖는 기술에 해당된다. 이런 제3의 기술은 3차산업혁명 시기까지 개발된 기술들과 달리 인간의 직접적 통제와 관리가 없어도 준-독립적으로 행위한다. 인공지능이나 봇(bot) 같이 데

16) 촉진자는 사용자에게 특정 기술의 사용을 유도 또는 촉발시키는 요인을 말하며, 사이성은 사용자와 촉진자 사이를 매개하는 요인이다. 플로리디에 따르면, 디지털 테크놀로지가 등장하기 전 기술의 지배적인 의미는 1차 기술과 2차 기술적인 것이었다. 인간이라는 사용자와 자연의 촉진자 사이를 매개하기 위한 기술로 1차 기술이 등장했으며, 그 다음 단계에서 등장한 2차 기술은 인간 사용자와 자연의 촉진자를 매개하는 기술이 아니라 기존의 기술과 인간 사용자를 다시 매개하는 기술이다.

이터를 자동적으로 수집하여 분석하고 이를 토대로 스스로 의사결정을 하는 자동화된 기계들의 등장은 3차 기술의 가능성을 계속해서 실험한다.

3차 기술 논의에서 특히 주목할 부분은 기술이 갖는 자율성 (autonomy)의 가능성에 있다. 자율성을 획득한 기계라는 개념은 일찍이 노르베르트 위너(Wiener, 1948)의 사이버네틱스, 존 폰 노이만(von Neumann, 1948)이 말한 자가 증식하는 오토마타(automata) 기술, 그리고 앨런 튜링(Turing, 1936)이 상상한 추상 기계에 대한 아이디어로 이어지면서 고안된 개념으로 여기서 기계의 자동화라는 의미는 단순히 특정한 행위를 자동으로 수행한다는 기계 작동의 무한 반복성에 국한되지 않고 자율적으로 행위하는 기계의 존재론적 가능성으로 확장되고 있다(임중수, 2018). 정보통신기술의 발달에 힘입어 정보를 자동적으로 처리할 수 있게 되면서 기계가 인간 고유의 지능 기능까지 모사하여 스스로 의사결정하는 자율적인 기술 시스템이 등장하기 때문이다.

물론 현 단계에서 인공지능이나 로봇이 진정으로 자율적인 행위자인지 단정적인 논의를 하기는 어렵다. 자유의지가 없는 인공지능과 로봇의 행위가 자율적인 것처럼 보이는 외견상의 기능일 뿐이라는 비판도 타당하다. 그러나 3차 기술들이 자율성을 갖고 행위하는 존재로서의 가능성이 점쳐지는 국면에서 앞으로 이러한 기술들과 인간이 공존하는 조건에서 발생하는 문제들을 이해해 가기 위해서 어떤 새로운 시각이 요구되는가 측면에서 자율적인 기계의 출현 맥락에 주목해볼 필요 또한 분명하다. 자율성을 갖춘 기계에 관한 전망에서 자율성의 의미 또한 물론 단순하지 않다. 이 문제는 인공지능을 어떻게 이해할 것인가를 둘러싼 철학적 논의에서도 핵심 쟁점이다. 자율성은 칸트 철학에서처럼 자유라는 조건의 측면에서 파생되는 것으로 파악되기도 하고(김수배, 2008), 자유 의지의 유무 또는 마음의 문제로 환원돼서 논의되기도 한다. 하지만 인공지능과 로봇의 자율성과 인간의 자율성은 같은 발생학적(embryological) 조건을 공유하지 않는다는 점에서 인간과 같은 자율성만을 자율성으로 고집

할 필요는 없다. 다만 인공지능 같은 3차 기술의 자율성을 다루기 위해서는 자율성이라는 개념이 갖는 의미의 수식이나 한정은 어떤 식으로든 필요하다.¹⁷⁾ 가령 지금 보급되고 있는 인공지능 스피커는 외부 상황을 인지하고 인지된 상황에 적절한 즉 타당한 행위 유형을 판단하여 작동되는 자율성을 수행하고 있다고 볼 수 있다. 즉 마음이나 심리적 지향이 없는 기술이라 하더라도 자율적 존재로서의 의미를 부여할 수 있다는 것이다.

따라서 플로리다가 말하는 자율성은 인간의 자율성에 대한 이해와는 조금 다르다. 플로리다는 그의 동료 샌더스와 함께 쓴 <인공적 행위자의 도덕성에 관하여(On the Morality of Artificial Agents)>에서 인간의 자율성과는 다른 개념의 자율성의 의미에 대한 규정을 시도한다 (Floridi & Sanders, 2004). 이에 대해 그들은 도덕적 행위자를 도덕적이라고 평가할 수 있는 행위의 조건을 상호작용적(제1조건), 자율적(제2조건), 적응적 시스템(제3조건)으로 각각 정의하는데 여기서 자율성에 대한 정의는 외부로부터의 자극, 외부의 간섭이 없더라도 시스템 스스로 자신의 내적 상태를 바꿀 수 있거나 내적 상태가 바뀌는 동적이며 마치 유기적인 존재가 수행하는 기능으로 파악된다. 플로리다의 자율성 이해의 초점은 인간의 자율성은 마음, 의식 같은 인간 고유의 요소에 대한 상정이 필요하지만 지금 출현하는 지능적 기계의 경우 그런 전제가 없는 자율성의 가능성을 보여주고 있다. 플로리다가 마음없는 윤리학이나 마음없는 세계를 주목하는 것도 이 때문이다. 윤리 문제에서도 이는 중요한 분기점을 제공하는데, 그것은 마음이 없는 기술적 존재라도 특정 행위

17) 이러한 관점은 인간의 자율성이라는 개념에도 적용되어야 한다. 대부분의 인간은 복잡한 삶의 조건과 시대적 조건과 환경 속에서 순수히 자율적인 존재로 살아가지 않거나 못하기 때문이다. 인간만이 자율적 존재나라는 물음은 큰 의미를 갖지 못하며 이 맥락에서 자율성을 갖는 미디어 테크놀로지 논쟁은 로봇/인공지능이 진정 자율적이거나 아니라는 질문보다는 어떤 자율성이냐에 관한 논의로 초점을 옮길 필요가 있다. 이에 대해서는 다른 기회에 상세히 다루고자 한다. 로봇과 인간의 자율성을 대비시켜 논하는 철학적 고찰로는 포모사(Formosa, 2021)의 논의를 참조.

를 수행함으로써 윤리적 혹은 비윤리적 결과를 산출할 수 있기 때문이다.

더 나아가, 플로리다가 말하는 자율적 시스템에서의 자율성은 꼭 인간과 같은 수준이나 성질의 지능의 실제 여부를 이론적으로 전제하지 않고도 자율적 행위자로 기능하는 기계의 출현이 갖는 의미를 논의하는 데 적합한 이해를 제공한다는 점을 주목해 볼 수 있다. 인공지능이 진정으로 자율적이나 아니라는 논쟁에 집착하는 것은 윤리학적 논쟁에서도 큰 실익이 있는 것은 아니다. 인간이라고 해서 꼭 자율적이지 않으며 인공지능이 단순한 계산만 수행한다 해도 윤리적 문제를 일으키는 세상이 되었다는 점을 파악하는 것이 더욱 중요하다. 인간의 직접적인 통제를 필요로 하지 않으면서 자율적으로 행위하는 기계들 그리고 그 기계들이 연결된 네트워크들은 이미 우리 현실에 자리잡고 있으며 그로 인해서 새로운 윤리적 이슈와 쟁점들이 출현하고 있다.

2) 윤리학의 존재론적 전환: 피동자 중심 존재 중심의 윤리학

플로리다의 정보윤리학은 인간은 물론 인간 이외의 존재들을 정보라는 근본 범주의 관점에서 바라보는 존재론적 시각을 바탕으로, 존재들 일반에 적용 가능한 일반적인 윤리 이론을 제공하려는 시도라 할 수 있다. 은유하자면 이런 정보적 존재론적 시도의 기원은 노르베르트 위너로부터 찾아볼 수 있는데, 위너는 정보는 존재의 근본 원리이며 물질이라든가 에너지 같은 것들의 상호작용은 정보의 상태가 변화하는 것으로 환원될 수 있다는 관점을 제시함으로써, 인간이나 다른 존재들 모두 정보적으로 이해할 수 있는 존재 일반적으로 바라볼 수 있는 시야를 제공했다. 플로리다 또한 이런 위너의 접근을 계승함으로써 윤리학을 인간의 윤리학이 아닌 정보적 존재 모두의 윤리학으로 재편, 즉 윤리학적 세계를 확장하려는 것이다.

굳이 도덕철학이나 윤리학의 역사 전체를 살펴보지 않더라도 기존의 윤리학은 어디까지나 인간 중심의 윤리학이라는 데에는 큰 이견이 있을

수 없다. 그것이 고전적 자유주의든, 자유주의든, 공화주의든, 공동체주의든, 공리주의든, 심지어 윤리적으로 비난받기 일췌인 신자유주의든 그 무엇이든 간에 어디까지나 인간 중심적 윤리라는 공통성을 갖는다. 서구적 전통에서 있는 인간 중심적 윤리학들은 도덕이나 윤리에 대한 입장과 관점은 다를지언정 누가 어떤 도덕적 해악이나 비윤리적 행위를 수행했는지를 판단함으로써 그 행위를 행한 특정한 주체 즉, 개별 인간 혹은 조직이나 집단, 국가에 책임을 묻는다는 점(Miller, 2001)에서 공통적이다. 행위 주체로서의 인간 중심의 윤리학 아래에서 인간은 동물이라든가, 자동차, 바위, 혹은 컴퓨터나 여타 미디어 장치 등에 도덕적인 책임을 따지지 않는다. 그런 존재들은 스스로 사고하고 행위하는 윤리적 존재가 아니라는 이유다.

그러나 이런 논리들은 정보적 존재들을 도덕과 윤리의 영역 속에서 포섭하는 데 한계를 가진다. 인간이 로봇을 인격적으로 대우해야 하는가라는 질문을 로봇이 인격을 갖는지 여부라는 기준으로 접근하면 로봇에 대한 학대나 비윤리적 활용은 아무런 문제가 되지 않는다. 하지만 만약 로봇이 인간 삶을 영위하는 데 필수적 장치자 관계의 대상으로 기능하고 있다면 역설적으로 그런 로봇을 더 보살피고 보호해야 하는 윤리적 의무가 인간과 기계 사이에 창발한다. 돌봄이나, 육아, 노인 보호 등 사회적 로봇의 쓰임이 대표적이다. 하지만 현 단계의 로봇은 독자적인 인격체로서 사고하는 자신의 판단에 근거하여 행위하는 인격체는 아니기 때문에 인간이 로봇의 행위에 도덕적 책임을 물어야 할 이유는 기각된다. 오직 인간만이 자발적 적극성을 가지고 행위를 하기 때문에 그 행위에 대한 책임을 귀속받는 존재다. 더구나 그런 책임은 엄격한 인과성을 바탕으로 추론될 수 있어야 하기 때문에 누가 어떤 문제를 유발했는지에 관해 인과성의 추적이나 해명이 모호한 이슈에 대해서는 도덕적 윤리적 책임을 묻기도 어렵다(이러한 관점의 윤리철학적 논의로는 Green, 2002, pp. 79-80; Miller, 2001, p. 88; Young, 2009 등을 참조). 이런 난해함

은 정보사회에서 발생하는 거시적인 문제들에 대해서는 더욱 커진다. 예를 들어 플랫폼 자본주의(Srnicek, 2016) 같은 디지털 시대의 거시-윤리적 현상에 대한 윤리적 비판들이 많지만 그런 자본주의의 병폐나 해악을 특정 누군가의 책임으로 몰어야 할지는 난감한 문제다. 그런 도덕적 윤리적 이슈들에는 인간은 물론 인간들의 연합 그리고 다양한 기술적 요인들이 복합적으로 얽혀 있기 때문이다. 오늘날 많은 도덕적 이슈들은 특정 행위자를 원인으로 하기보다 복잡한 행위자들의 하이브리드적인 연결망 차원에서 발생하고 있음을 주목해야 한다.

인간 중심적 윤리학은 애초부터 보편적이지 않다는 점 또한 윤리학이 적용되는 영역을 정보적 차원으로 재정의하려는 존재론적 시도에 특별한 의미를 부여한다. 비판적인 포스트 휴머니즘 이론가 브라이도티(Braidotti, 2019)가 지적했던 것처럼 서구 자유주의 전통의 인간 중심적 사고는 겉으로는 휴머니즘이라는 보편적 가치를 근간으로 삼지만 실제 특정 시기 특정 조건에서 인식된 인간 주체의 일시적 가치에 대한 옹호, 즉 근대 서구가 만들어낸 휴머니즘이라는 특수한 이데올로기로서 자리매김해 왔을 뿐이다. 인간의 가치, 도덕적 삶, 윤리적 기준 같은 도덕적 요소들 또한 인종, 권력, 성과 젠더, 종교, 지역 등에 따른 인간성을 반영하는 차등화된 범주들 속에서 상이한 권력, 이데올로기, 가치들 간의 끊임없는 갈등과 투쟁들을 통해서 설정되고 재설정되기를 반복해온 우연적 구성물이기도 하다. 인공지능이 인간의 가치를 위협할 것이란 디스토피아적 기술 담론들이 옹호하는 인간성이란 보편적 가치는 존재하지 않는다는 것이다(Ferry, 2015, pp. 118-178). 서구 자유주의 사상들, 예컨대 하버마스(W. Habermas)나 마이클 샌델(M. Sendel)처럼 기술에 의한 인간 강화의 시도를 반대하기 위해 인간의 도덕적 주체성, 인간의 도덕적 본성을 보편적으로 가정하려는 시도는 사실상 어떤 경험적 철학적 근거도 없는 공허한 말이다. 인간 윤리라는 것은 한참 전부터 동물 윤리, 생태 윤리 같은 인간 중심성과 충돌하는 윤리적 요청들로부터 도전받

아 왔다. 이런 도전들은 인간 윤리의 한계를 질문하면서 인간 윤리를 넘어서는 보다 확장된 윤리학적 그릇을 요청하고 있으며 이러한 요청은 플로리다 뿐만 아니라 철학, 인류학, 사회학 등 다양한 학제들에서 인간 중심적인 휴머니즘 개념을 초월하려는 이론적 전환들에서 발견되는 추세다 (예를 들어, Bennett, 2010; DeLanda, 2006; Harman, 2018을 보라).

이런 문제의식으로 바라볼 수 있는 징후들은 이미 현실에서 벌어지고 있다. 인공지능이나 빅데이터, 로봇 같은 기술들을 새로운 위험사회의 기술적 요인으로 지목하는 시각들이 많지만 실제 그 기술들은 인간 삶에 다양한 편익과 효용을 제공하고 있어서 거꾸로 그러한 기술들이 윤리적 대우를 받지 못할 때 나타나는 도덕적 해악에 대한 주목이 대표적이다. 의료 기기나 보안 장치 혹은 신체에 접합된 스마트 기기들이 해킹을 당하거나 비도덕적 의도로 인하여 영향을 받는 경우들을 생각해볼 수 있다. 인공지능이나 로봇이 학대받는 경우들도 인간의 눈으로만 바라보아서는 그런 기술들과 인간의 윤리적 공존에 적절한 결과를 야기하지 못할 수 있다. 플로리다가 윤리학 세계의 행위자의 범주를 생명 중심적 관점에서 존재 중심적 관점으로 대체하려는 이유에는 이러한 배경이 놓여 있다. 그렇다면 존재 중심의 윤리학이란 구체적으로 도덕과 윤리의 문제를 어떻게 다루자는 것인가? 이에 대해 플로리다는 스스로 자신의 행위에 책임을 질 수 있는 행위자 개념을 폐기하는 데서부터 출발한다. 기존의 윤리학에서 스스로 도덕적인 책임을 질 수 있는 행위자라는 범주에 동물, 인공지능, 로봇, 컴퓨터는 포함되기 어려웠다. 이는 도덕적 행위자를 책임 개념의 차원에서 파악하는 방식으로, 도덕적 책임을 인간 중심적 도덕으로 환원한다. 그러나 도덕적 책임을 도덕적 행위자의 책임으로 환원하려는 윤리학적 접근은 비인간 존재들의 행위들이 출현하고 있는 지금의 상황에서 충분한 해답이 될 수 없다.

여기에는 여러 가지 양상의 문제들이 존재한다. 앞서 언급한 것처럼

인간을 포함하여 이성이나 마음을 갖지 않으면서도 행위하는 존재들 사이의 윤리적 관계를 모색하는 데 있어서 오직 인간만을 도덕적 책임을 질 수 있는 존재로 한정하는 것은 앞으로 전개될 수 있는 비인간 행위자들의 자율성 증대 경향에서 발생하는 윤리적 이슈와 쟁점들에 대처하는 데 적절한 포괄적 시야를 열어주지 못한다. 여기서 쟁점은 인공지능이 진정한 행위자인가 아닌가, 진정한 행위자라면 인간과 같은 마음이나 지능을 갖고 있는가 아닌가 같은 발생학적 대답 여부에 있지 않다. 지금 일상에 배치되는 로봇 기술들은 진정으로 자율적인 행위자라고 보긴 어렵다 해도 그 기계들이 인간에 의해 소비되거나 인간과 관계를 맺어가는 과정에서 도덕적, 윤리적 문제들은 이미 발생하고 있다.

뿐만 아니라 도덕적 책임을 행위의 문제로만 접근할 경우 행위되지 않은 일들로부터 파생되는 도덕적인 이슈와 쟁점을 다루기 어렵다(이에 대해서는 Scheffler, 2001; Singer, 1975를 보라). 예를 들어 사물인터넷 등 지능화된 디지털 테크놀로지들이 정교하게 결합된 스마트 시티에서 벌어지는 어떤 문제들은 그 책임의 주체가 정확히 누구인지 즉 책임의 원인이 어디로 얼마만큼 할당(allocation)되는지가 불분명하다(도덕적 할당 이론에 대해서는 Miller, 2001, 2007을 보라). 이런 상황에서 도덕적 해악은 강한 인과성 논리를 통해서 특정한 원인 제공자를 찾아 그 책임을 온전히 떠넘기기 어려운 성질을 갖는다. 인간과 기계, 기계와 기계 사이의 경계가 모호해지고, 신경과학이나 나노테크놀로지 등 과학기술의 발달로 유기체와 기계 장치 사이의 접점이 점점 더 어떤 이음매나 마찰없이 연결되고 그에 따라 인간, 비인간 존재들은 인터넷 상에서 촘촘히 연결되어가고 있다. 데이비도우(Davidow, 2012)가 과잉연결사회에 담론에서 지적했듯이 오늘날 발생하는 도덕적 해악이나 병폐들은 대부분 다양한 행위자들과 기술적 장치들이 복잡하게 상호작용하는 과정에서 발생하는 것으로, 그것들 간의 고도로 복잡한 상호작용이 전개되면서 거시적으로 특정한 도덕적 윤리적 문제가 발생한다. 반면 전체만 놓고 보면

비도덕적이고 비윤리적인 책임을 따져야 하는 일들이 발생하지만, 각각의 과정들 속에서 대부분의 행위자들은 도덕적으로 중립적인 행위자의 성격을 가질 수 있다(Floridi, 2013b). 미시적 수준에서는 도덕적 이슈에 대해 전적인 책임을 져야 할 이유가 없어 보이기 때문에 그들에게 어떤 책임을 어떻게 물어야 할지 또한 불분명해진다.

플로리디는 이러한 도덕적 조건의 변화에 주목하여, 행위자 중심에서 피동자 그리고 존재 중심(patient-oriented, onto-centric)의 윤리학과 분산된 도덕성(distributed morality) 개념을 중심으로 정보권 시대의 도덕적 책임에 관한 윤리학의 재구성을 시도한다(Floridi, 2013a, p. 102). 피동자란 정보권을 구성하는 일원들로서 정보적 시스템으로 이해가 가능한 모든 존재들에 해당된다.¹⁸⁾ 즉 도덕적 의지나 마음이 없는 정보권 안에서의 존중받아야 할 최소한의 자격을 갖춘 도덕적 존재들이다. 이러한 피동자에는 인간 이외에도 동식물부터 인공지능, 로봇, 사물인터넷, 그 밖의 미디어들 모두가 포함된다. 도덕적 존재들의 범위를 확장한다는 것은 도덕적 고려의 대상이 되어야 할 자격 요건을 최소한으로 낮춘다는 의미다.

피동자들에 대한 플로리디 정보 윤리학의 주목은 새로운 윤리적인 의의를 제공한다. 우선 생명 중심의 윤리에서 정보 중심의 윤리로의 전환은 디지털 생태계라든가, 인공지능/로봇 같은 기술적 존재들 같은 정보권 세계와 그 세계 내에서 우리가 마주하는 모든 것들을 윤리적으로 대할 수 있는 계기를 제공한다. 생명 중심의 윤리는 생명이 있는 것들은 소중하다고 간주되지만 생명이 없는 로봇 같은 존재는 함부로 대할 수도 있다는 비윤리적 사고를 은연중 함축한다. 도덕적 사고나 윤리적 고려를 비-생명

18) 피동자 중심의 윤리학은 사실 동물 윤리나 환경 윤리 분야에서 이미 제기된 것이었다(대표적으로 Singer, 1975). 인공지능, 로봇 윤리와 관련하여 도덕적 행위자와 도덕적 피동자에 관한 철학적 논의를 잘 정리한 문헌으로는 데이비드 궁켈(Gunkel, 2011)의 저서 <기계 질문(The Machine Question)> 1장과 2장을 참고.

적 존재들로부터 적용하게 되면 행위자들은 자신의 도덕적 판단과 그에 입각한 행위를 수행할 때 그로부터 영향을 받는 존재들을 적극적으로 고려해야 한다는 윤리적 요청을 받는다.

다음으로 분산된 도덕성에 관한 플로리디의 윤리학적 구상은 정보권의 특성에 따라 요청되는 윤리학적 모델이다. 플로리디는 정보권 하에서 도덕적 해악은 행위자들 사이의 복잡한 상호작용적 관계에 의해 윤리적 문제로 쟁점화 되고 있다는 점에 주목한다(Floridi, 2013a, 2014). 이 점이 도덕과 윤리에 미치는 영향은 어떤 윤리적 이슈가 개별 행위자에 의해 일어난다기보다는 행위자들의 연합 즉 네트워크에 의해서 발생한다는 점에 있다. 이러한 상황은 도덕적 문제에 대한 책임의 특정한 원인 제공자를 확정하기가 어려우며, 그에 따라 인간, 비인간 존재들이 상호작용하는 과정들 속에서 각각의 행위자들 혹은 행위자처럼 기능한 기계들은 대부분 도덕적으로 중립적인 지위를 가지는 것처럼 보이는 결과를 야기하며, 기술과 인간의 상호작용이 점점 더 복잡한 양상을 그려감에 따라 도덕적 윤리적 문제에 대한 대처는 까다로워진다.

플로리디는 이러한 난관을 돌파하기 위하여 분산된 도덕성(distributed morality)과 분산된 도덕적 행위(distributed moral action) 개념을 도입한다(Floridi, 2013a, p. 137, pp. 261-276). 분산된 도덕성 개념을 통해 플로리디는 대부분의 행위자들이 도덕적으로 중립적인 것처럼 보이지만 실제로는 도덕적 해악에 관여되는 상황의 윤리적 책임을 규명할 수 있는 대안을 제시한다. 분산된 도덕성의 상황에서 원인 제공자를 특정하기가 어려운 상황은 반대로 말하면 전체 행위자들의 작은 참여와 노력을 통해서 도덕적 상황을 개선할 수 있는 역설적 가능성을 갖는다. 가령 영향력 측면에서 무시해도 좋을 개별 행위자들의 상호작용이 가능하다면, 디지털 네트워크 상에서 발생하는 도덕적·윤리적 이슈들에 대한 대처는 효율적으로 가능하다(Floridi, 2013b).

분산된 도덕성 모델은 도덕적으로 중립적인 것처럼 보이지만 실상은

도덕적 해악에 연관되는 행위자들 사이의 상호작용에 주목함으로써, 어떤 도덕적 악이 초래될 때 그것에 책임을 할당하는 문제를 어떻게 이론적으로 재구성할 것인지에 주목한다(Floridi, 2013a, 2013b). 이를 위해 플로리디의 분산된 도덕성 모델은 강한 인과성의 규명에 의존하기보다는 관련성(relevance)에 가까운 약한 인과성(weak causality)에 주목한다. 내가 의도가 있었는지 여부보다는 어떤 연관성이 있었는지에 따라 인과적 해명 책임이 요구될 수 있다. 이러한 접근의 이점은 도덕적 관련성이 있는 주체들의 참여와 연대를 유도할 수 있다는 장점을 제공한다. 이런 접근에서 중요한 점은 의지나 의도가 없더라도 책임에 관해 고려해야 할 필요가 있다는 것이다. 플로리디가 “마음이 없는 윤리학(mindless ethics)” 혹은 “과실이 없는 책임’ 이론의 필요성을 강조하는(Floridi, 2013a, pp. 148-152) 이유다. 그가 윤리학적 개념들을 이렇게 마음, 의식, 의도 없는 행위/행위자의 측면에서 재정의하려는 이유는 다시금 자율성의 문제가 관건이 되기 때문이다. 인간 이외의 알고리즘이나 인공지능, 로봇 같은 인공적인 행위자들이 강한 인과성이 없더라도 도덕적 판단에 개입하는 과정에서 발생하는 윤리적 책임을 어떻게 귀속시킬 것인가를 둘러싼 이론적 발판을 모색하는 것이다.

그렇다면 이 지점에서 왜 우리가 피동자들을 윤리적으로 대우해야 하고, 보다 광범위한 존재들 그리고 그 존재들 사이에 도덕적 관계의 윤리학적 당위와 규범을 설정해야 하는가라는 질문이 제기될 수 있다. 왜 기계들에게도 인간 사이에만 성립하는 윤리적 관계라는 문제를 적용해야 하는가? 이에 대해 플로리디는 고통 같은 질적인 의미를 갖는 윤리학적 개념을 엔트로피(entropy)라는 정보적인 개념으로 대체하여 정보 환경에서 왜 윤리가 보편적으로 적용되어야 하는지에 대한 나름의 논리를 제공한다. 이는 그의 저서 <정보윤리> 4장을 중심으로 제시되어 있다(Floridi, 2013a, pp. 65-84). 그것은 고통 같이 윤리학적으로 우리가 피해야 하는 개념을 엔트로피라는 정보적인 표현 형식으로 대체하는 논

리로, 생명의 개념에서 보다 확장된 존재 일반에 있어서 고통은 정보적인 무언가로 재개념화하려는 시도다.¹⁹⁾ 모든 정보/존재는 자신의 본질적인 가치, 즉 의미를 지니고 있고 존재 자체가 번성할 권리를 갖는다. 만약 어떤 정보나 정보적 존재, 예를 들어 로봇이 인간의 학대나 오용으로 인하여 자신의 존재가 궁핍해지는 상황에 처한다면 엔트로피는 증가한다. 플로리디는 이를 정보윤리 원칙으로 설명한다(Floridi, 2013a, pp. 70-73). 그 원칙은 첫째, 엔트로피는 정보 환경에서 방지되어야 하고, 둘째, 엔트로피는 정보 환경에서 제거되어야 하고, 셋째, 전체 정보 환경은 물론 개별 정보적 존재들의 번영, 즉 그들의 잘 삶(well-being)이 고취되어야 한다는 것이다.

3) 인간의 윤리학적 재개념화와 인간중심주의의 재구성

인포스피어적인 조건에서 인간의 존재 이유나 인간의 의미와 가치, 더 구체적으로 기술과 인간의 관계 사이에서 인간의 윤리적 역할 상은 무엇일까? 플로리디는 이에 대해 호모 포이에티쿠스(homo poieticus)라는 메타포를 통해 정보권에서 인간의 지위와 역할의 윤리적 의미와 가능성을 제시하고 있다(Floridi, 2013a, pp. 161-179).

플로리디에 따르면, 호모 포이에티쿠스 개념은 자연 자원의 착취자(exploiter)로서의 호모 파베르(home faber), 그리고 경제적 생산, 유통, 소비를 위해 살아가는 존재로서의 호모 에코노미쿠스 그리고 윤리적 돌봄과 책임으로부터 자유로운 유희적 인간으로서의 호모 루텐스(home

19) 엔트로피 개념은 에너지를 정보를 보유한 시스템 상의 무질서도(disorder) 또는 무작위성(randomness)이 얼마나 있는지를 뜻하는 개념(Floridi, 2013a, p. 66)이다. 플로리디는 이를 열역학적 개념과 연관이 있지만 물리학에서 말하는 엔트로피와는 다른 개념으로 정의하기 위해서 형이상학적 엔트로피(meta-physical entropy)라고 표현하고 있다. 형이상학적 엔트로피 개념은 정보의 양적 측면과 질적 측면 즉 의미론적 측면을 모두 다루기 위한 개념이다. 정보의 손실이나 파괴가 일어나면 이런 형이상학적 엔트로피가 발생하는 것으로 이해할 수 있다.

ludens)와는 다르다. 호모 파베르적 인간에게는 그동안 자연이나 생명 그리고 기계를 비윤리적 목적으로 활용하는 인간의 모습이 반영되어 있으며 호모 에코노미쿠스 또한 부와 자본의 증식, 유통, 소비에 매몰되는 경제적 인간의 윤리적 한계 그리고 호모 루덴스에는 타자에 대한 배려나 자신의 놀이/유희에 따른 도덕적 해악에 대한 책임성이 결여되어 있다. 이는 인간 중심적인 인간에 대한 메타포가 갖는 윤리적 한계라 할 수 있다. 이에 반해 정보권의 구성주의적 특성, 즉 인간이 기계를 만들지만 동시에 그 기계들이 인간에게 다시 영향을 미치는 구성주의를 강조하는 플로리다에게, 인간은 인간이 유일한 윤리적 실천과 판단의 담지자가 아니라도 정보권을 보호하고 번영시켜야 할 도덕적 의무와 윤리적 책임을 갖는 중요한 존재로서 여전히 그 의미가 없지 않다(Floridi, 2013a, pp. 161-179). 이러한 에코-포이에틱한 인간은 정보권 안의 모든 존재들을 창조적으로 관리해야 하는 윤리적 책임을 자각하는 인간을 의미하며 정보권의 환경을 도덕적으로 구성하는 과정에 적극적으로 동참하는 인간이라는 의미를 갖는다.

플로리다는 이러한 인간상을 그리스 로마 신화의 데미우르코스(demiurge), 즉 ‘만드는 자’라는 비유를 통해서 부연한다. 데미우르코스는 물질 세계를 창조하는 역할을 맡은 신으로 과거 희랍 신화에서는 신 중에서도 최고의 신으로 간주되었다. 일반적으로 기독교의 창조론에서와 같이 창조주가 존재하고 있는 질료들(물질들)을 가지고 자연과 인간을 만들었지만 그 정신만은 데미우르코스에게 의해 만들어졌다. 이러한 데미우르코스 비유를 통해 플로리다는 정보권 시대에 인간의 역할이 축소되거나 배제되는 기술소외론적 기술 혐오론적 관점들과는 정반대로 인간의 역할에 대한 더 큰 의무와 책임을 부여하고 있다.²⁰⁾ 정보권에서 인간은

20) 기술소외론은 인간과 기술을 분리하고 인간의 삶에서 기술을 배제해야 한다는 전통적, 소극적 휴머니즘 논의로 귀결된다는 점에서, 인간과 기술의 공존이라는 맥락에서는 그 이론적 효용과 쓸모가 제한적이다.

환경에 대한 더 큰 영향력을 행사할 수 있으며 그에 따라 피동자 혹은 존재들에 대한 더 많은 의무와 책임을 갖게 된다. 탈-인간 중심적 입장과 상기의 견해는 일견 형용모순이 아니냐고 생각하기 쉽지만, 예를 들어 인공지능이 진정한 자율적 행위자로 진화해 간다고 해서 인간의 역할이 그만큼 반비례해서 축소되는 것은 아니다. 인간은 인공지능이라는 아이를 낳아 키우는 만큼, 그 존재를 돌볼 도덕적 의무와 윤리적 역할은 여전히 중요하다. 즉 인간이라는 존재의 윤리학적 의미는 더욱 풍성해지고 그 존재 의의는 계속 지속될 수 있다.

이런 측면에서 기술이 행위자로 진화해 간다 해도 인간이 가진 힘의 존재론적 의미는 축소되지 않는다고 보는 것이 플로리디의 입장이다. 탈-인간 중심주의를 받아들인다 해도 그것이 휴머니즘의 의미를 축소시키는 것은 아니라는 것이다. 더구나 호모 포이에티쿠스적 인간은 서구 휴머니즘에 근간하는 윤리학적 접근이 갖는 실천적 한계의 극복을 지향한다는 점에서 역설적이기는 하지만 달라진 기술적 조건 속에서 휴머니즘의 지속 가능성을 모색하는 윤리적 기획이다(Durante, 2017). 이런 의미에서 전통적인 휴머니즘이나 고전적 인문학을 강조하면서 기술의 비윤리성을 무작정 비판하는 인본주의 전략이 오히려 휴머니즘의 미래 가능성을 제약하는 측면도 생각해볼 수 있다.²¹⁾ 요컨대 이런 견해가 시사하는 바는 우리 인간의 의미를 소비자로 축소시키지 않음으로써 정보권 환경을 창조, 관리, 감독할 수 있는 윤리적 책임을 지닌 주체로 재인식할 수 있

21) 이런 측면에서 인간의 가치나 권리를 옹호하는 입장에서 현재 기술기업이나 알고리즘, 인공지능, 디지털 플랫폼(산업)을 비판하는 - 광범위한 의미에서 - 미디어 기술 비평이 오히려 기술과 공존하는 인간적 삶의 조건이 어떠해야 하는가에 관한 실천적 전망을 열어주지 못하는 측면이 크다고 본다. 그것은 인간을 내세우면서, 기술을 윤리적으로 약용하는 사회 구조나 질서 그리고 정치경제학적 힘들의 문제를 은폐하는 데 비의도적으로 기능하기 때문이다. 이런 입장을 취하고 있는 기술 비평가로는 예브게니 모로조프를 들 수 있는데, 그는 디지털 시대에 기술 비평 문화가 사실상 “구글을 인간화하기”를 영무새처럼 반복하면서, 당대 기술 비평이란 게 “그저 구글에 관해 글을 쓰는 일” 이상의 의미를 보여주지 못하고 있다고 비판한다(이에 대해서는 Morozov, 2015를 참고).

는 계기를 열어준다.

인간의 의미에 대한 플로리디의 재정의는 그의 정보윤리학이 비판적으로 극복하고자 하는 인간 중심주의 사상과의 관계 측면에서 인간 중심주의의 탈피나 소멸이라기보다는 인간 중심주의의 재구성이라 할 수 있다. 왜냐하면 플로리디의 논의 전반에는 존재론의 확장과 이를 수용하는 윤리학의 체계를 이야기할 뿐, 어디에서도 기술과 인간 사이의 우열 관계나 위계가 어떻게 결말지어질 것이란 시나리오는 제시되어 있지 않기 때문이다. 오히려 인간이 갖는 역할의 재개념화에 대한 그의 주목은 인간과 생명이라는 전통적 범주를 넘어서 비인간 존재 일반을 고려하는 윤리학적 확장 안에서도 여전히 인간의 역할이 갖는 중요성과 가치를 강조하려는 시도일 수도 있다. 그렇다면 여기서 또 다시 그의 정보윤리학은 인간의 의미를 둘러싸고 새로운 논쟁으로 연결될 수 있는 가능성을 남긴다. 그가 말하는 호모 포이에티쿠스라는 역할이 제대로 수행되기 위한 인간의 특성과 조건은 무엇인지가 불분명하게 남아 있기 때문이다. 그렇다면 결국 그러한 인간 상의 가능성은 이념, 자본, 민족, 인종, 젠더, 종교, 그리고 지정학적 요인들에 의해 위계적으로 설정된 조건과 그러한 조건이 야기하는 차별과 배제의 문제, 즉 넓은 의미에서 인간 주체의 보편성을 강조하는 광의의 휴머니즘 이론에 내재된 윤리학적 한계를 고스란히 물려받을 개연성도 부정하기 어렵기 때문이다. 이런 점에서 플로리디의 인간에 대한 재개념화 그리고 이와 연관된 많은 포스트 휴머니즘 담론에서의 새로운 인간 개념에 관한 논의는 인간성의 의미를 새롭게 해석하거나 확장하는 '선언적 방식'을 넘어서 기존의 정치, 사회, 문화이론들과의 면밀한 검토 속에서 보다 비판적으로 검토되어야 할 필요를 남긴다.

이와 관련하여, 필자가 보기에 플로리디의 정보윤리학의 성격은 아마도 플로리디의 학문적 행보와 긴밀하게 연결되어 그 지향을 드러내 보이는 것 같다. 그는 정보철학과 정보윤리학에 대한 이론적 작업뿐만 아니라, 유럽과 글로벌 관련 기구들 안팎에서 빅데이터 윤리나 의료 윤리 같

은 대단히 실무적이고 기능적인 논의 공간에도 적극 참여하는 현실-참여적 이론가에 가깝다. 이러한 그의 실천적 지향은 그의 정보윤리학의 성격 자체에서도 내재되어 있는데, 그것은 그의 인간 중심주의에 대한 비판이 공상과학 영화 속 로봇의 등장 가능성을 짐치거나, 인류를 초월하는 지적 존재를 예견하는 담론들에 대해 별로 관심이 없고, 인간이 무엇을 할 것인가라는 대단히 실천철학적 관심에서 정보윤리학의 대안을 모색하는 것 같다는 것이다. 여기서 발견되는 바는 플로리디의 정보철학은 기술, 정보, 인간에 대해 어떤 본질적 혹은 근본주의적인 가정이나 전제도 도입하지 않는다는 점이나 인포스피어를 우리가 어떻게 가꿔갈 것인가라는 프래그머티즘적 지향을 논의 곳곳에서 드러낸다는 점이다. 기술이 정확히 어떤 존재로까지 나아갈지 그 실체적 전망을 뒤로 하고 대신 기술적 존재들과의 적극적 공존을 전제하고 그러한 전제를 기반으로 하는 정보권을 가꿔가자는 그의 윤리학적 비전은 그러한 관심을 매우 직접적으로 뒷받침한다.

4. 결론

플로리디의 정보윤리학은 속류적인 인공지능 논쟁이나 헐리우드 영화 속 로봇윤리 논쟁같이 새롭게 등장한 도덕적, 윤리적 문제들을 단순히 특정하게 분절되어 있는 영역에 적용되는 응용윤리적인 접근 또는 인공지능이나 로봇을 개발, 보급하는 주체들, 즉 기술 기업들과 정부 등 제도화된 행위자들의 관점에서 새롭게 빚어지는 갈등과 마찰을 기능적으로 해결하려는 실무적 차원에서만 접근하는 것이 아니라, 보다 저층에서 우리의 삶을 구성하는 정보적 세계에 걸맞은 보편 윤리학의 토대를 하나의 대안으로 보여준다. 그의 윤리학은 존재론적인 토대 위에서 윤리적 관계를 어디까지 전망할 수 있는 것인가를 질문하고 있다는 점에서 기존의 인간 중심

적 윤리학과는 다른 결을 보여준다.

물론 그의 이론적 시도에 대한 이러한 평가가 온전히 타당한 것은 아니다. 인공지능 윤리에 대한 많은 기능적 분석들이나 정책적 처방들이 이론과 원리적 측면에 대해 언급하고 있지 않다고 해서 그들의 논의들이 비원칙적이거나 보편적 적용 가능성이 없다고 말할 수는 없다. 대신 플로리디의 정보윤리학의 미덕은 응용윤리학적 접근이나 기술산업적 요구에 따른 실무적 접근과 차별화된다는 점에서 찾기보다는 오히려 인포스피어를 윤리적으로 가꿔야 한다는 당위적 요청을 근간에 설정하고, 그러한 요청을 실현해가기 위해 요청되는 타당하고 또한 실천적 힘을 담보한 윤리학적 논리들을 적극 모색한다는 데서 찾아볼 수 있다. 마음 없는 윤리학, 피동자 중심-존재 중심의 윤리, 분산된 도덕성 모델 개념들은 바로 그런 윤리학적 원리를 설명하고 있다.

물론 그러한 윤리학적 원리들은 추상적으로 제출된 것으로서, 구체적으로 어떻게 적용될지는 아직 불투명하고 모호해 보인다. 더 근본적인 반론은 플로리디가 비인간 존재들에게 자동적으로 도덕적 지위를 부여하고, 인간-비인간 관계/연합이 갖는 의미에 도덕적 당위를 기계적으로 적용하고 있다는 점에서, 사실과 당위의 차원을 잘못 연결하는 오류를 범하고 있다는 비판도 가능하다. 도덕의 의미 그리고 도덕이 적용되는 대상의 속성에 대한 판단 기준과 관련하여 플로리디의 정보 윤리학의 추상화 방법론(Foridi, 2008)은 윤리학적 원리로서 불충분하다는 한계도 안고 있다.

그럼에도 불구하고 플로리디의 정보윤리학의 시사점은 적지 않다. 무엇보다 윤리적 세계의 범위를 어디까지 설정할 것인가와 관련하여 그의 정보권 관점은 온라인과 오프라인의 경계가 해체되고 정보가 영향을 미치는 범위가 급속도로 확장되는 오늘날 중요한 설득력을 갖는다. 특히 디지털 장치와 네트워크에 의존하는 사회적 커뮤니케이션 시스템 안에서 인간이 커뮤니케이션하는 대상은 이미 인간이 아닌 지능화되고 있는 컴

퓨터 장치로 치환된 오늘날의 의사소통적 조건에서 대상 자체를 윤리적인 타자로 정의해야 할 필요는 더욱 커지고 있다.

자본과 권력, 민족과 인종, 젠더와 성 등 위계적이고 불균등한 사회적 조건 안에서 기술이 차이와 다양성을 넘어서 불평등과 차별을 야기하는 일들을 막기 위해서는 인간 중심주의에 호소하는 인공지능 윤리 담론들이 사실상 윤리적 세탁(ethics washing)의 도구로 전락하는 일을 막기 위해서도(Vincent, 2019) 유행하는 휴머니즘적 인공지능 윤리라는 일종의 수사학적 스타일에 대한 경계는 필요하며, 이러한 필요와 관련하여 플로리디의 정보 윤리학에 내재된 인간 중심주의에 대한 비판은 인공지능 윤리가 윤리를 앞세운 다른 무언가로 변질되지 않도록 하는 실질적 장치를 마련해가는 데도 특별히 유익한 자극을 제공한다.

마지막으로 플로리디의 정보윤리학이 윤리학적 대안으로서 갖는 가장 강력한 장점은 기술적 실체의 의미에 대한 본질주의적 사고방식으로부터 자유롭고, 지금의 정보사회에 대한 윤리적 책임을 회피하지 않는 인간 존재의 의미를 재정의를 하려는 노력에 있다고 보여진다. 인공지능 윤리는 순수하게 기술적 차원의 문제도, 정치적 차원의 문제도 아니다. 그렇다면, 인공지능 윤리학의 적절한 자리는 인간과 기술의 관계맺음에 주목하고 그 의미를 윤리학적 차원에서 모색하는 과정에 있어야 한다. 그러한 인공지능 윤리를 위해서 이원론적 사고로부터 일정한 전환은 불가피해 보인다. 이런 점에서 플로리디 정보윤리학을 온전히 지금 시점에서의 기술의 발전 정도나 우리의 이해 정도에 국한시키는 것은 적절치 않아 보인다. 그런 점에서 플로리디 정보 윤리학은 기술의 발전 정도와 그에 대한 우리의 이해가 변화해가는 과정에서 계속 살펴봐야 할 윤리학적 텍스트라 할 것이다.

끝으로, 이 글의 한계에 대해서 언급하고자 한다. 첫째, 이 글은 플로리디의 정보윤리학의 열개와 내용을 소개하는 데 일차적 취지를 두었다. 연구라고 보기에는 미흡한 이런 탐색적 작업을 시도한 까닭은 인공지

능 윤리에 대한 기능적, 제도적, 실무적 차원의 논의들은 차고 넘치는 반면 윤리학적 원칙이나 기준에 대한 질문들이 없어 보인다는 연구자의 소박한 학술적 문제의식이 있다. 따라서 이 글은 플로리다의 정보윤리학을 기존의 윤리학적 이론들과 엄밀히 대조하거나, 그의 정보윤리학의 핵심 개념이나 사상을 집중적으로 파고들어 재해석하지는 못했다. 이는 이미 철학계에서 부분적으로 진행되고 있는 것으로 알지만, 차후에 기회가 된다면 플로리다의 논의를 포함 정보윤리학의 이론적 최신 지형의 특성과 경향을 비판적으로 고찰하는 작업을 시도해 보고자 한다.

둘째, 이 글은 플로리다 정보 윤리학의 필요성과 가치를 인간 중심주의에 대한 비판적 문제의식과 실체론적 관점에서 관계론적 관점에서의 윤리학적 전환이라는 두 가지 측면에서 접근했다. 그러나 전자의 경우 인간 중심주의 사상에 대한 비판적 논의는 철학, 정치학, 문화이론과 정체성, 페미니즘, 사이보그 논의에 이르기까지 매우 방대하면서 다채롭게 펼쳐져 있는 주제다. 이를 한편의 글에서 매우 엄밀히 갈라볼 것은 갈라보고 합칠 것은 합쳐보면서 플로리다의 정보윤리학과의 이론적 관계를 지형화하는 것은 간단치 않은 주제였다. 플로리다 정보윤리학의 성격을 드러내기 위해 과도한 단순화와 축약이 있을 수 있다는 점을 밝힌다.

참고문헌

- 강영안·이상현 (2013). 포스트휴머니즘에 관한 철학적 성찰. <지식의 지평>, 15권, 150-171.
- 김병선 (2007). 정보기술의 존재론: 인공지능은 현존재가 될 수 있는가? <사회과학논총>, 26권 1호, 257-272.
- 김수배 (2008). 칸트의 도덕철학과 역사철학의 긴장 관계 -“자율성” 개념을 중심으로. <칸트연구>, 21호, 1-31.
- 김왕배 (2021). ‘사회적인 것’의 재구성과 ‘비(非)인간’ 존재에 대한 사유. <사회와 이론>, 통권 제40호, 127-146.
- 김용찬 (2020). 도시의 디지털화 : 인공지능 기반 ‘디지털 도시’의 커뮤니케이션 이슈들. <언론정보연구>, 57권 4호, 95-149.
- 김재희 (2021. 7. 13). 인포스피어와 정보철학 [전자매체본]. <Horizon>, URL: <https://horizon.kias.re.kr/18165/>
- 김진석 (2019). <강한 인공지능과 인간>. 파주: 글항아리.
- 김환석 (2018). 새로운 사회학의 모색: 탈인간 중심주의. <경제와 사회>, 통권 제117호, 236-261.
- 박남기 (2020). 인공지능과 윤리적 이슈. <언론정보연구>, 57권 3호, 122-154.
- 박남기·김윤정·장은채 (2021). 사회적 능력이 인공지능 챗봇의 사회적 사용과 자기노출에 미치는 영향 : 사용자의 외로움 정도와 챗봇에 대한 지각된 인간다움의 매개효과. <한국언론학보>, 65권 5호, 367-401.
- 배영·최향섭. (2021). <지능정보사회의 이해>. 파주: 나남출판.
- 손승혜 (2020). 우리는 어떤 로봇을 만들어가고 있는가? 2010-2019 텔레비전 뉴스에 등장한 인공지능 로봇의 전달 방식과 특성에 관한 탐색적 연구. <한국언론학보>, 64권 6호, 202-237.
- 손화철 (2018). 인공지능 시대의 과학기술 거버넌스. <철학사상>, 68권, 267-299.
- 송하연·김지은·김태년 (2020). 인공지능과 헬스 커뮤니케이션. <언론정보연구>

- 구), 57권 3호, 196-238.
- 신상규 (2016). 자율기술과 플로리디의 정보 윤리. <철학논집>, 45권, 269-296.
- 이호영·문아람·양기문·변성혁·문상현·유경한 ... 임종수 (2020). <알고리즘 데이터 이용의 사회문화적 영향>(정책연구20-19-04). 진천: 정보통신정책연구원.
- 이희은 (2021). 기계는 권력의 지도: 인공지능과 자동화된 불평등. <문화과학>, 통권 105권, 127-142.
- 임종수 (2018). 오토마타 미디어: AI 미디어의 커뮤니케이션 양식을 위한 시론. <언론과 사회>, 26권 4호, 33-84.
- 허유선·이연화·심지원 (2020). 왜 윤리인가: 현대 인공지능 윤리 논의의 조망, 그 특징과 한계. <인간·환경·미래>, 통권 24권, 165-209.
- 홍대식 (2020). 인공지능 기술과 유료방송 미디어 규제에 대한 도전. <언론정보연구>, 57권 3호, 81-121.
- Bell, D. (1976). *The coming of post-industrial society: A venture in social forecasting*. New York, NY: Basic Books.
- Bennett, J. (2010). *Vibrant matter: A political ecology of things*. Durham, NC: Duke University Press.
- Braidotti, R. (2019). *Posthuman knowledge*. Cambridge, UK: Polity Press.
- Candea, M. (Ed.) (2010). *The social after Gabriel Tarde: Debates and assessments*. London, England: Routledge.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, 24, 505-528.
- Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics.

- Philosophy & Technology*, 27(1), 61-77.
- Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., ... & Whittaker, M. (2019). *AI Now 2019 Report*. New York, NY: AI Now Institute.
- Davidow, W. H. (2012). *Overconnected: The promise and threat of the internet*. Harpercollins.
- DeLanda, M. (2006). *A new philosophy of society: Assemblage theory and social complexity*. London, England: Continuum.
- Durante, M. (2017). *Ethics, law and the politics of information: A guide to the philosophy of Luciano Floridi*. Dordrecht, Netherlands: Springer.
- Ferry, L. (2015). *La révolution transhumaniste*. Paris, France: Plon.
- Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, 18(3), 303-329.
- Floridi, L. (2013a). *The Ethics of Information*. Oxford: Oxford University Press.
- Floridi, L. (2013b). Distributed morality in an information society. *Science and Engineering Ethics*, 19, 727-743.
- Floridi, L. (2014). *The Fourth Revolution: How the infosphere is reshaping human reality*. Oxford: Oxford University Press.
- Floridi, L. (2017). A plea for non-naturalism as constructionism. *Minds and Machines*, 27, 269-285.
- Floridi, L., & Sanders, J. W. (2001). Artificial evil and the foundation of computer ethics. *Ethics and Information Technology*, 3(1), 55-66.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349-379.
- Formosa, P. (2021). Robot autonomy vs. human autonomy: Social

- robots, Artificial Intelligence(AI), and the nature of autonomy. *Minds and Machines*, 31(4), 595-616.
- Green, M. (2002). Institutional responsibility for global problems. *Philosophical Topics*, 3(2), 79-95.
- Gunkel, D. J. (2011). *The machine question: Critical perspectives on AI, robots, and ethics*. Cambridge, Mass: MIT Press.
- Hagendorff, T. (2019). *The ethics of AI ethics: An evaluation of guidelines*. arXiv preprint arXiv: 1903.03425.
- Harman, G. (2018). *Object-oriented ontology: A new theory of everything*. London, UK: Pelican Books.
- Höffe, O. (1993). *Moral als preis der moderne : Ein versuch uber wissenschaft technik und umwelt*. 김시형(역)(2013). <학문윤리학>. 서울: 시와진실.
- Jobin, A., Marcello, I., & Effy, V. (2019). Artificial intelligence: The global landscape of ethics guidelines. arXiv preprint arXiv: 1906.11668.
- Kurzweil, R. (2005). *The singularity is near*. New York, NY: Viking Books.
- Latour, B. (1993). *Nous n'avons jamais été modernes : Essai d'anthropologie symétrique* (C. Porter, Trans.). *We have never been modern*. Cambridge, Mass: Harvard University Press. (Original work published 1991).
- Latour, B. (2005). *Reassembling the social - An introduction to actor-network-theory*. Oxford, England: Oxford University Press.
- Mazlish, B. (1995). *The fourth discontinuity: The co-evolution of humans and machines*. New Haven, CT: Yale University Press.
- Miller, D. (2001). Distributing responsibilities. *The Journal of*

- Political Philosophy*, 9(4), 453-471.
- Miller, D. (2007). *National responsibility and global justice*. Oxford, England: Oxford University Press.
- Morozov, E. (March 2015). Taming tech criticism. *Baffler*, (27), URL: <https://thebaffler.com/salvos/taming-tech-criticism>
- Pepperell, J. (2009). *The posthuman condition: Consciousness beyond the brain*. Bristol, UK: Intellect.
- Samuel, A. L. (1960). Some moral and technical consequences of automation: A refutation. *Science*, 13(3429), 741-742.
- Scheffler, S. (2001). *Boundaries and allegiances: Problems of justice and responsibility in liberal thought*. Oxford, England: Oxford University Press.
- Singer, P. (1975). *Animal liberation: A new ethics for our treatment of animals*. New York, NY: New York Review.
- Srnicek, N. (2016). *Platform capitalism*. Cambridge, England: Polity Press.
- Stiegler, B. (2017). *La société automatique. Tome 1, L'avenir du travail* (D. Ross, Trans.). *Automatic society. Vol. 1: The future of work*. Cambridge, England: Polity Press. (Original work published 2015).
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, 49(236), pp.433-460.
- Vincent, J. (2019, April 3). The problem with AI ethics. The Verge. Retrieved from <https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech>
- Vinge, V. (1993). *Vernor Vinge on the singularity*. Retrieved from <https://mindstalk.net/vinge/vinge-sing.html>

von Neumann, J. (1948). The general and logical theory of automata. In: Jeffress, L.A. (ed.) *Cerebral Mechanisms in Behavior - The Hixon Symposium*, pp. 1 - 31. John Wiley and Sons, New York.

Wiener, N. (1948). *Cybernetics: Or control and communication in the animal and the machine*. Cambridge, MA.: MIT Press.

Young, I. (2009). *Responsibility for justice*. Oxford, England: Oxford University Press.

투 고 일 자: 2021년 12월 30일

심 사 일 자: 2022년 05월 02일

게재확정일자: 2022년 05월 02일

Abstract

Renewing Ethics in the Age of Artificial Intelligence

Focusing on Luciano Floridi's Information Ethics

Yong-Min Ryu

Assistant professor, Dept. of Mass Communication of Inje University

This article aims to examine the possibilities and limitations of Luciano Floridi's information ethics as an alternative framework for universal ethics in the age of artificial intelligence. His ethics shows the value as a universal in terms of that it is a patient-oriented, being-centered ethics, which goes beyond the limitations of human-centered ethics that presuppose a strict hierarchy between humans and non-human beings. His information ethics allows us to critically recognize the shortcomings of western-humanist strategies revealed in popular artificial intelligence ethical narratives, such as artificial intelligence for humans or artificial intelligence for a good society, while at the same time solving the problem of the moral status of technological beings. In this regard, it provides implications for the prospect for media ethics in the era of human-machine communication in that it fills the practical limits of the substantive point of view.

KEYWORDS Artificial Intelligence, Artificial Intelligence Ethics, Information Ethics, Actors, Anthropocentrism, Humanism, Luciano Floridi