

메타버스 시대의 기독교 윤리의 몇 가지 과제: 인공지능의 도전을 중심으로

강원돈

한신대학교, 신학부 은퇴교수/민중신학과 사회윤리

1. 머리말
2. 메타버스의 이해
3. 가상의 현실과 실제의 현실을 맞물리게 하는 결정적 요인으로서의 메타버스 경제
4. 메타버스와 인공지능
 - 4.1. 약인공지능의 도전
 - 4.2. 강인공지능의 도전
5. 메타버스의 규율과 관련해서 기독교 윤리가 유념할 점
6. 맺음말

1. 머리말

메타버스는 ‘사람이 사는 세상’을 크게 변화시키고 있기에 인문학과 사회과학에 큰 도전이 되었고, 기독교 윤리학에도 큰 관심의 대상이 되었다. 메타버스는 디지털 기술, 정보

통신 기술, 인공지능 기술의 발전에 힘입어 실제의 현실과 가상의 현실을 서로 매끄럽게 통합하는 가상융합현실이다.¹⁾ 인간은 메타버스를 통해 몸과 마음의 통일체로 현존하는 동시에 몸으로부터 분리된 디지털 휴먼으로 현존하는 경험을 갖는다. 메타버스에서는 인간과 아바타와 인공지능 주체가 서로 공존하면서 영향을 주고받는다. 그러한 메타버스가 가상과 현실을 융합한 플랫폼으로 급속하게 발전하고 지구적 차원에서 플랫폼들의 네트워크를 형성할 수 있게 된 것은 무엇보다도 인공지능 기술의 발전 덕분이다. 인공지능은 약 인공지능 단계에서 강인공지능 단계로 발전하는 도중에 있다.²⁾

인공지능의 발전 단계에 따라 인류가 메타버스에서 직면하는 윤리적 문제는 각기 다른 특질과 서로 다른 해법이 필요할 것이다. 약인공지능 단계에서는 인간의 기본권 침해, 데이터 인클로저, 빅데이터의 편향성, 노동시장 붕괴 등이 문제가 된다면, 강인공지능 단계에서는 인공지능의 자율적 판단과 실행 결과에 대한 민사적, 형사적 책임의 문제, 인간, 아바타, 인공지능 주체가 참여하는 규범공동체 형성의 문제 등이 큰 논쟁거리가 될 것이다. 인공지능이 초인공지능 단계로 발전하여 인간이 인공지능을 통제하지 못하고, 초인공지능이 인간을 지배하거나 앎을 많은 사람이 우려하지만, 이 글에서는 그러한 초인공지능이 등장해서는 안 된다는 원칙적인 판단을 내릴 뿐 그 문제를 본격적으로 다루지 않을 것이다.

위에서 말한 것을 염두에 두면서, 필자는 이 글에서 첫째 메타버스의 기술적 기반과 작동 방식을 간략하게 살핀다. 둘째 실제의 현실과 가상의 현실을 매끄럽게 통합하는 결정적인 기제로서 메타버스 경제를 다룬다. 셋째 메타버스에서 주요 플레이어로 등장하는 인간, 아바타, 인공지능 주체의 상호관계에 주목하면서 메타버스에서 인류가 약인공지능 단계와 강인공지능 단계에서 직면하는 윤리적 문제들을 드러내고 그 해법을 탐색한다. 끝

1) 메타버스는 실제의 현실을 넘어선다는 의미의 ‘메타’와 실제의 현실과 가상의 현실을 통합하여 확장한 세계를 가리키는 ‘유니버스’를 합성하여 새로 만들어진 용어이다. ‘메타버스’는 약 30년 전에 한 공상과학 소설에서 그러한 의미로 처음 사용되었다. Neal Stephenson, *Snow Crash*, (New York: Bantam Books, c1992). 메타버스는 그것의 기술적 특성이나 작동 방식 등에 따라 여러 가지로 정의되고 있지만, 2007년 미래가속화연구재단(Accerleration Studies Foundation)이 내린 메타버스에 관한 정의가 널리 인정되고 있다. 그 정의에 따르면, “메타버스는 1) 가상적으로 향상된 물리적 현실과 2) 물리적으로 지속하는 가상적 공간의 통섭이다. 그것은 둘의 융합이지만, 사용자는 둘을 각각 경험할 수 있다.” John Smart-Jamais Cascio-Jerry Paffendorf, *Metaverse Roadmap: Pathways to the 3D Web: A Cross-Industry Public Foresight Project*, (Los Gatos: Accerleration Studies Foundation, 2007), 4. 출처: <https://www.w3.org/2008/WebVideo/Annotations/wiki/images/1/19/MetaverseRoadmapOverview.pdf> (2023년 9월 10일 다운로드).

2) 약인공지능과 강인공지능의 차이에 관해서는 이 논문의 제IV장 제2절을 보라.

으로 넷째, 메타버스의 규율과 관련해서 기독교 윤리와 신학이 유념할 점을 밝힌다.

2. 메타버스의 이해

얼마 전까지만 해도 사람들은 메타버스가 청소년들을 끌어들이는 게임의 세계 정도로 생각했다. 그렇지만, 한국 정부가 2021년 메타버스에 자동차공장, 조선공장, 반도체공장 등을 세우겠다는 정책을 발표하고, 메타버스에서 다양한 비즈니스 모델이 자리를 잡고, 2021년 미국 대통령 선거에서 보듯이 메타버스에서 치열한 유세전이 펼쳐지고, 코로나 팬데믹 기간에 대학 입학식과 공연이 메타버스에서 이루어지는 것을 보면서 사람들은 메타버스가 ‘사람 사는 세상’을 크게 변화시켰음을 실감하기 시작했다.

메타버스는 언뜻 보면 디지털 그래픽으로 구현된 가상의 세계인 듯이 보이지만, 그 가상의 세계는 현실의 세계와 분리되지 않고 서로 통합되어 있다. 필자는 그것이 메타버스의 결정적인 의의라고 본다. 메타버스는 실제의 현실과 가상의 현실을 매끄럽게 통합하는 새로운 문명의 장이다. 메타버스에서 실제의 현실과 가상의 현실은 컴퓨터 중앙처리장치(CPU)와 그래픽처리장치(GPU)의 발전, 빅데이터 생성과 플랫폼 구축에 필수적인 인공지능 기술의 발전, 통신 인프라 구축 등과 같은 기술 발전에 힘입어 서로 통합된다. 메타버스는 현실의 제약을 넘어서서 이를 확장하려는 동기와 그것을 실현하는 기술이 서로 결합함으로써 나타난 효과이다. 잘 알려져 있듯이, 2007년 미래가속화연구재단은 방금 앞에서 말한 동기와 기술의 접목에 주목해서 메타버스의 발전을 네 가지로 유형화한 바 있다. 현실의 제약을 넘어서서 판타지를 표현하기 위해 실제의 현실에 덧씌우는 증강현실(augmented reality), 사회적 자아에 의해 그 표출이 제약된 개인적 자아를 드러내려는 동기에서 비롯된 라이프로그(lifelogging), 현실의 세계를 똑같이 가상의 세계에 복사해서 현실의 제약을 넘어서서 이를 효과적으로 살피고 전달하기 위한 거울 현실(mirror reality), 현실의 제약을 넘어서서 독자적인 규칙과 세계관에 따라 구축되는 가상의 현실(virtual reality)이 그것이다.³⁾ 증강현실, 라이프로그, 거울 현실은 모두 실제의 현실을 전제로 해서 구축된 가상의 세계이다. 가상 현실은 언뜻 현실과 무관한 듯하지만, 그 가상 현실을 추구

3) John Smart-Jamais Cascio·Jerry Paffendorf, *Metaverse Roadmap*, 5. 이에 관한 알기 쉬운 설명으로는 김상균·신명호, 『메타버스: 새로운 기회』, (서울: 배가북스, 2021), 40-59.

하는 인간의 심리적 동기, 그 가상 현실을 움직이는 가장 강력한 동인인 경쟁과 보상 등을 통해 현실의 세계와 서로 맞물린다. 한 마디로, 오늘날의 기술은 실제의 현실(actual reality)과 구별되는 가상의 현실(virtual reality)을 만들고, 실제의 현실과 가상의 현실을 포개어 증강현실(augmented reality)을 창출하며, 실제의 현실을 제약하는 시간과 공간, 그리고 자원의 한계를 넘어서는 확장 현실(extended reality)을 펼치고 있다. 그렇게 해서 창조되는 메타버스는 실제의 현실을 넘어서고(meta), 실제의 현실과 가상의 현실을 융합하여 확장한 세계이다. 메타버스는 단 하나의 균질화되고 통합된 디지털 세상이 아니다. 메타버스는 그 안에 수없이 많은 작은 메타버스를 품고 있는 거대한 성좌(constellation)이다.

메타버스가 운용되는 방식은 크게 네 가지가 있다. 첫째는 실제의 현실로부터 메타버스의 세계로 들어가 그 세계를 지배하는 세계관과 규칙에 따라 게임의 플랫폼을 즐기도록 한 독립형 메타버스이고, 둘째는 게임 참여자들이 메타버스에 마련된 장비들과 유추단서들을 활용하여 메타버스 안에 임의의 메타버스를 연속해서 창출할 수 있도록 하는 액자형 메타버스이다. 셋째는 걸그룹 에스파(aespa)에서처럼 실제의 가수들과 그 아바타들이 실제의 현실과 가상의 현실 사이의 경계를 자유롭게 넘나들도록 설계한 경계형 메타버스이고, 넷째는 실제의 현실을 가상의 현실로 재현하여 가상의 현실에서 강의, 실험, 업무, 공장 제어 등 실제의 행위를 수행할 수 있도록 설계한 대체형 메타버스이다.⁴⁾

메타버스가 실제로 작동하는 방식을 알기 위해서는 몇 가지 실례를 드는 것이 좋을 것 같다. 독립형 메타버스를 구현한 게임 플랫폼의 특성은 「월드 오브 워크래프트」에서 실감할 수 있다. 독립형 메타버스는 실제의 세계와 가상의 세계가 일단 분리된 것이기에 게임과 엔터테인먼트의 세계로 간주해도 좋을 것이다. 그런데 경쟁과 몰입이 주는 재미가 독립형 메타버스에 참여하는 동기가 된다는 점에 주목한다면, 독립형 메타버스는 인간의 마음과 몸을 통해 실제의 현실과 맞물린다고 볼 수 있다. 액자형 메타버스는 MZ 세대를 열광시키고 있는 「로블록스」나 「제페토」 같은 게임 플랫폼에서 전형적으로 나타난다. 액자형 메타버스에서는 게임 참여자들이 기왕의 메타버스 안에 자신의 메타버스를 따로 만들어 그곳에서 아바타용 상품을 진열하고 구매자에게 판매하고 메타버스에서 통용되는 화폐로 결제하는 경제활동을 펼치기도 한다. 액자형 메타버스도 게임과 엔터테인먼트를 즐기도록 고안된 가상의 세

4) 메타버스의 네 가지 유형과 작동 방식에 대한 요령 있는 설명으로는 윤현정·이진·윤혜영, 「메타버스 개념과 유형에 관한 시론: 가상공간 이론을 중심으로」, 『인문콘텐츠』 62(2021), 66-76을 보라.

게이지만, 「로블렉스」처럼 메타버스 화폐인 로벅스(Robux)와 현실의 화폐가 서로 교환되기에 이르면, 이미 그것은 현실의 세계와 깊이 맞물린 가상의 현실이라고 보아야 할 것이다.

경계형 메타버스는, 앞서 말한 대로, ‘에스파’ 걸그룹이나 방탄소년단의 공연에서 경이로움을 맛볼 수 있게 했다. 싸이더스튜디오엑스에서 만든 ‘로지’는 가상의 현실에서 구현된 아바타가 실제의 현실에서 얼마나 매력적인 인플루언서로 활동할 수 있는가를 잘 보여준다. 경계형 메타버스는 앞으로 공연과 마케팅 분야에서 점점 더 큰 비중을 차지하게 될 것이다. 경계형 메타버스는 문자 그대로 현실의 세계와 가상의 세계가 서로 넘나들면서 융합하는 세계이다.

대체형 메타버스는 코로나19 팬데믹으로 인해 대면 활동이 불가능한 상황에서 2020년 3월 순천향대학교의 입학식이나 2021년 건국대학교 예술제를 가상공간에서 펼침으로써 비상한 관심을 불러일으켰다. 사실 대체형 메타버스는 코로나19 팬데믹 상황 이전에도 실제의 현실에서 이루어지는 작업을 보완하는 기능을 톡톡히 수행해 왔다. 그것은 오늘의 디지털 영상 기술이 실제의 세계를 스캔하여 가상 현실에 그대로 구현할 수 있기 때문이다. 지금도 자동차의 엔진, 조향기, 동력전달장치 등을 3차원 영상으로 스캔하여 메타버스에 복사한 뒤에 정비공이 실제의 자동차와 메타버스의 자동차를 겹쳐 보는 안경을 끼고서 자동차의 고장 부분을 수리하고 부품을 갈아 끼우는 작업을 차질 없이 수행할 수 있다. 앞으로 정보통신 기술이 더 발전하여 정보 전달 속도와 양이 가속적으로 빨라지고 많아진다면, 예컨대 우주 공간에 설치된 디지털카메라로 지구 표면의 모든 것을 실시간으로 스캔하여 메타버스에 재현하고 그 가상 현실 속에서 3차원 재생 안경을 쓰고서 지구 탐사를 입체적으로 수행할 수 있을 것이다. 루브르 박물관을 스캔한 메타버스에 입장하여 루브르 박물관을 마치 실제의 박물관인 양 관람할 수도 있을 것이다. 2023년 현대자동차가 선언하였듯이, 해외 자동차공장을 메타버스에 복사하여 현대자동차 본사 엔지니어들이 메타버스를 통해 거대한 해외공장을 운영하고 관리할 수도 있을 것이다.

3. 가상의 현실과 실제의 현실을 맞물리게 하는 결정적 요인으로서의 메타버스 경제

메타버스가 실제의 현실과 가상의 현실을 매끄럽게 통합하게 하는 결정적인 요인은 메타버스 경제의 잠재력이다.⁵⁾ 메타버스가 제공하는 재미와 엔터테인먼트가 사람들을 메타

버스에 유인하고 그곳에 머물게 하는 것은 분명하지만, 메타버스를 활성화하고 확대하는 것은 메타버스 경제이다. 이미 메타버스는 현실의 경제활동을 재현하여 상품의 생산과 판매와 소비가 이루어지는 지구적 규모의 경제권을 구축하고 있다. 메타버스 경제권이 본격적으로 구축되기도 전에 인터넷 상거래는 천문학적인 규모에 이르렀고, 마이크로소프트, 애플, 아마존닷컴, 페이스북(현 메타) 등 정보통신 분야의 공룡 기업들은 세계 10대 대기업 명단의 상위권에 포진했다. 메타버스 경제권이 자리를 잡게 되면, 생산, 소비, 유통, 금융 등은 디지털 정보의 흐름을 타고 물리적 마찰을 최소화하면서 매끈한 흐름을 타고 크게 확장될 것이다.

메타버스 경제는 가상융합현실을 가속적으로 발전시키는 동력이지만, 해결하기 어려운 많은 문제를 불러일으키고 있기도 하다. 그 문제를 제대로 인식하면 메타버스의 밝은 면과 어두운 면을 동시에 볼 수 있다. 메타버스 경제권은 초국가적인 디지털 공간에 창설된 세계이기에 그 어떤 나라의 주권 영역에도 속하지 않는다. 따라서 메타버스 경제는 국민 국가의 규율을 넘어서는 특성을 가질 수밖에 없고, 국민국가 수준에서 해결할 수 없는 여러 가지 문제를 일으킨다. 메타버스 경제는 앞으로 그 비중이 커질 것이기에 메타버스 경제의 문제를 인식하고 해법을 모색하는 일을 마냥 내버려 둘 수 없게 되었다. 아래서는 가장 중요하다고 여겨지는 메타버스 경제의 몇 가지 문제를 짚는다.

첫째는 메타버스 경제권에서 화폐를 규율하는 규범과 관련된 문제다. 메타버스 경제는 독자적인 화폐인 NFT(Non-Fungible Token, ‘대체 불가능한 토큰’)를 기반으로 운영될 것이다. NFT는 블록체인 기술과 결합한 가상화폐다.⁶⁾ NFT가 메타버스 경제의 화폐로 자리를 잡는다면, 메타버스 경제의 화폐 규범은 NFT를 생산하고 사용하는 탈중앙적인 네트워크들의 수평적인 거래를 중개하는 센터들의 내부 규약의 형식을 취할 공산이 크다. NFT 자체가 어느 한 국가나 국가 연합에 의해 발행되지 않은 이상 그 규약은 탈중앙집권적이고 탈권위주의적인 성격을 띠게 될 것이다. 그러한 메타버스 경제의 화폐 규범은 이제까지 존속했던 중앙은행 제도를 무력화할 수 있다. 그것은 중앙은행의 지급보증에 근거하여 상업은행이 신용을 창조하고 파괴하는 자본주의적 신용화폐 제도를 뒤흔드는 결과

5) 이시한, 『메타버스의 시대: 미래의 부와 기회를 선점하는 7대 메가트렌드』, (파주: 다산북스, 2021), 67f.

6) 성소라·롤프 회퍼·스콧 맥러플린, 『NFT 레볼루션: 현실과 메타버스를 넘나드는 새로운 경제생태계의 탄생』, (서울: 도서출판 길벗, 2021), 25ff.

를 가져올 것이다. 이 문제를 어떻게 해결할 것인가는 앞으로 뜨거운 논쟁거리가 될 것이다.⁷⁾

둘째, 메타버스에서 거래되는 모든 것은 소유의 대상으로 등록될 수 있다. 그것이 이제까지의 디지털 세계와 결정적으로 다른 점이다. 디지털 세계의 특성은 원본과 복사본의 구별이 불가능하다는 것이었다. 거기서는 복사에 들어가는 비용이 거의 없기에 누구나 원본이나 복사본에 접근하여 사용할 수 있었다. 그렇기에 디지털 경제는 ‘한계비용 제로’ 사회를 열고, 소유 개념은 ‘접속’과 ‘공유’ 개념으로 바뀐다고 여겨졌다.⁸⁾ 그러나 블록체인 기술이 개발된 이후에 상황은 근본적으로 달라졌다. 블록체인이 적용된 사물은 유일성 혹은 유한성의 특성을 갖고, 모조품과 구별되는 희귀한 진본의 지위를 갖게 된다. 블록체인 기술에 의해 만들어진 사물은 일종의 등기부에 등록되어 소유자가 확정된다.⁹⁾ 소유자가 특정된 희소한 사물은 경제적 재화로 거래되는 것은 물론이고 그 자체가 자산의 성격과 위상을 갖는다. 그것은 메타버스 경제에 거대한 자산시장이 탄생한다는 뜻이다. 그러한 자산시장을 규율하는 규범은 누가 어떻게 제정할 수 있을까? 한 가지 분명한 것은, 메타버스 자산시장은 초국가적인 가상의 공간에 설치되기에 어느 한 주권 국가의 규범에 따라 구성될 수 없다는 것이다. 그렇다면, 블록체인 기술의 분장 등록 체제만으로 충분한가? 그렇게 분산 등록된 자산의 거래에서 발생하는 자본이득과 자산소득은 거래 수익자에게 전적으로 귀속되어도 좋은가? 그러한 이득과 소득에 대한 징세는 어떻게 해야 하는가?

셋째, 메타버스 경제의 소유권 구조에서는 앞에서 언급한 자산시장의 규율 문제만이 아니라 일반 상품 거래 차원에서조차 심각한 문제를 불러일으킬 수 있다. 만일 한 메타버스의 허브를 이루거나, 메타버스들의 네트워크에서 허브에 해당하는 사물 구조에 대한 소유

7) 물론 각국 중앙은행이 디지털화폐를 발행할 수도 있을 것이다. 그것은 디지털화폐 발행의 국가 독점 체제를 유지함으로써 플랫폼 화폐 발행을 국지화, 종속화하는 전략이다. 그렇게 되면, 브레턴우즈 체제나 포스트-브레턴우즈 체제에서 나타난 것과 유사하게, 디지털 달러나 디지털 위안을 세계통화와 메타버스 경제의 지불수단으로 굳혀 화폐 권력을 확립하려는 경쟁이 치열하게 벌어질 것이다. 그러나 그러한 전략은 플랫폼 경제의 탈영토성과 국가 디지털화폐의 영토성 사이의 근본적인 모순을 해결할 수 없다. 그런 점을 고려한다면, 이상적인 해법은 실물경제와 메타버스 경제를 아우르는 중립적인 계산 화폐를 창설하고, 그 계산 화폐에 대한 각국 중앙은행 화폐와 각 플랫폼 화폐의 교환 비율을 결정하는 정교한 환율 시스템을 창안하는 것이라고 본다.

8) 그것이 제러미 리프킨의 핵심 명제였다. 제러미 리프킨/안진환 옮김, 『한계비용 제로 사회: 사물인터넷과 공유경제의 부상』, (서울: 민음사, 2014).

9) 이에 대해서는 최형욱, 「블록체인 기반의 메타버스 가상경제」, 『철학과 현실』 130(2021), 120; 성소라·롤프 회퍼·스콧 맥러플린, 앞의 책, 35를 보라.

권이 확립된다면, 그 소유권에서 발생하는 권력은 이루 말할 수 없이 클 것이다. 그 허브를 운영하는 자가 그 허브를 통해 상품을 거래하는 자에게서 수수료를 징수하는 것은 그 권력의 한 현상 형태일 것이다. 예를 들면, 로블록스에서 상품을 거래하는 경우, 거래 수수료율은 70%에 달했다.¹⁰⁾ 메타버스 허브의 소유자는 거래 규칙을 정하고 그 규칙의 이행을 감독하고 규칙 위반자를 처벌하는 권력을 행사할 수 있고 그 권력을 통해 막대한 이익을 취할 수 있다. 그렇다면, 그 허브 권력은 누가 어떻게 제어해야 하는가?

넷째, 앞으로 메타버스에서는 점점 더 많은 상품의 생산과 판매와 소비가 이루어질 것이다. 오늘의 메타버스에서 이루어지는 게임 콘텐츠 거래나 아바타 장식 아이템 거래는 장차 메타버스에서 이루어지게 될 상품 거래의 맛보기에 불과할 것이다. 메타버스 거래는 게이머의 욕망을 충족하거나 아바타를 장식하는 상품 거래에 그치지 않고 인간의 다양한 욕망을 충족시키는 재화와 서비스의 거래를 망라하고 총괄할 것이다.

메타버스 거래는 물질적 상품 거래와 비물질적 상품 거래의 두 가지 형식을 취한다. 첫째, 물질적 상품 거래에서 메타버스 거래는 실제 상품 거래의 시뮬레이션이고, 실제 상품 거래를 매개한다. 그것은 실제의 상품 거래가 메타버스에 거울처럼 투영되고, 메타버스 거래 조작을 통해 실행된다는 뜻이다. 둘째, 메타버스 거래의 가장 고유한 특징은 비물질적 상품의 거래에서 나타난다. 메타버스에서 거래되는 상품은 식품이나 냉장고나 자동차 같은 물질적 형상을 초월해서 디지털화된 비물질적 형상을 취하게 되고, 그러한 비물질적 형상의 상품은 데이터 처리 서비스로부터 시작하여 교육, 의료, 법률, 설계 서비스 등을 거쳐 거대한 플랫폼 제어 서비스에 이르기까지 다양하게 나타난다. 그러한 비물질적 형상의 상품을 생산하는 데 투입되는 생산요소들은 국경을 가로지르며 접속하는 다양한 센터들을 매개하여 생성되고 가공되고 구매되고 사용된다.¹¹⁾

그렇다면 그러한 비물질적 형상의 상품을 생산하고 판매하고 소비하는 과정은 어떻게 규율되어야 하는가? 세계무역기구(WTO)가 지적 소유권과 서비스 상품의 교역을 규율하는 규범을 제정했으니, 앞으로도 그 기구가 메타버스 경제권에서 이루어지는 상품 거래에 관한 규범을 제정하면 되는 것일까? 아니면 페이스북 등과 같은 지구적 차원의 플랫폼에

10) 박상범, “‘로블록스’, 개발자 착취 논란 휩싸여...매출의 17%만 번다,” GAMEVU (2021.08.23.), <https://m.gamevu.co.kr/news/articleView.html?idxno=20714> (2023년 9월 10일 다운로드).

11) 비물질적 노동의 형식들과 조건들에 대한 날카로운 분석으로는 마우리쥘로 라짜랏또, “비물질노동,” 『비물질노동과 다중』, 질 들뢰즈 외 지음/서창현 외 옮김(서울: 갈무리, 2005), 197-8을 보라.

대한 과세 기준을 세우기 위해 별도의 정부간 협의기구가 설치되었던 것처럼 별개의 규율기구를 창설하여야 할까? 비물질적 형상의 상품 가운데 플랫폼노동처럼 전 세계에 분산된 프리랜서 노동자들이 노동계약 제로 시간의 조건 아래에서 알고리즘을 통해 할당되는 일감을 처리하고, 엄격한 실시간 평가를 받고, 그 대가를 받는 디지털 노동은 어떻게 규율되어야 하는가?¹²⁾ 잘 알려져 있다시피, 메타버스의 스펙터클로 인해 ‘보이지 않는’ 글로벌 디지털 노동 문제는 산업자본주의 시대에 정형화된 노동을 전제로 해서 구축된 노동법의 사각지대에 놓여 있지 않은가?

위에서 살핀 바와 같이, 메타버스가 실제의 현실과 가상의 현실을 매끄럽게 통합하도록 추동하는 것은 의심할 여지없이 경제적 요인이다. 앞으로 메타버스 경제는 날로 확대될 것이고, 앞에서 언급한 것보다 더 많은 문제가 메타버스 경제에서 불거질 것이다. 지면의 제약 때문에 여기서는 그 문제들을 더 깊이 분석하고 해법을 제시할 수는 없다. 그러한 본격적인 작업은 차후의 과제로 넘기기로 한다. 다만, 필자는 메타버스 시대에 기독교 윤리가 유념할 점을 밝히는 이 글의 제V장에서 그 문제들의 해결에 관한 원칙적인 의견을 밝힐 생각이다.

4. 메타버스와 인공지능

앞에서 언급한 바와 같이, 메타버스는 세 가지 기술을 기반으로 하고 있다. 하나는 이미지를 조작하고 증강하는 디지털 기술이고, 다른 하나는 정보통신 기술이고, 마지막 하나는 인공지능 기술이다. 인공지능 기술은 알고리즘과 빅데이터 생성을 통해 거대한 플랫폼을 구축하는 기반이다. 그러한 플랫폼이 없다면 메타버스는 아예 성립할 수 없다. 여기서 주목할 것은 인공지능이 플랫폼의 기술적 기반에 머물지 않는다는 점이다. 인공지능은 그 자체가 행위자로 등장하고 있다. 그렇기에 메타버스에 등장하는 행위 주체는 실제로

12) 알고리즘을 통한 할당과 평가와 감독 아래에서 제로 시간 계약의 틀에서 표준화되고 단편화된 노동 공정을 끝없이 반복하는 디지털 노동은 ‘디지털 테일러주의’의 특성을 보여주고 있다. 이에 관해서는 모리츠 알텐리트, 『디지털 팩토리: 디지털 자본주의 시대, 보이지 않는 노동』, 권오성·오남규 옮김, (서울: 숨쉬는책공방, 2023), 86. 알텐리트는 ‘디지털 테일러리즘 전반의 핵심적인 요소’를 “노동의 표준화, 업무의 세분화, 알고리즘을 통한 관리가 노동의 유연화 및 증식을 가능케 한다는 점”이라고 분석한다.

현존하는 사람(person)과 그 사람의 화신인 아바타만이 아니다. 인공지능 행위자도 메타버스에서 행위 주체로 등장한다. 인공지능 행위자는 로봇 같은 휴머노이드의 형상을 취할 수도 있고, 다른 행위자들을 상대하고 그들의 행위에 반응하며 변환하는 인공지능 체제로 작동할 수도 있다. 인공지능은 계속 진화하고 있다. 앞으로 인류는 이제까지 알고 있었던 ‘사람 사는 세상’과는 사뭇 다른 세계를 메타버스에서 경험하게 될 것이다.

잘 알려져 있다시피, 인공지능은 아직은 약인공지능 단계에 머물러 있지만, 특이점¹³⁾을 넘어서서 인간과 유사하게 혹은 그 이상으로 자율적으로 판단하고 행동하는 강인공지능 단계로 진화하리라고 전망되고 있다. 거기서 한 걸음 더 나아가 강인공지능이 인간의 능력과 통제를 벗어난 초인공지능으로 가속적으로 발달하게 되면, 인간은 기계를 부리는 주인의 위상을 잃고 기계에 종속되는 노예의 지위로 전락할 수도 있으리라고 우려되고 있다. 인공지능의 발전은 ‘사람 사는 세상’에 큰 도전이다. 아래서는 약인공지능과 강인공지능의 도전을 살피고, 그 도전에 어떻게 대응할 것인가를 논한다.

4.1. 약인공지능의 도전

약인공지능은 특정한 과업을 효율적으로 수행하도록 설계된 인공지능이고, 그 작동을 위해서는 알고리즘, 데이터, 피드백 규칙 등이 부여되어야 한다. 그러한 약인공지능은 탁월한 학습 능력이 있고, 구글 답마인드를 구현한 알파고나 맞춤형 정보를 지식의 형태로 제공하는 ChatGPT나 탁월한 자연어 번역 능력을 보여주는 DeepL처럼 특정한 과업 수행 기능에서 인간의 지적 역량을 크게 뛰어넘고 있다. 그러나 약인공지능은 인간의 지적 역량에서 나타나는 학습, 지각, 추론, 탐구, 성찰, 상상 등 다면적인 역량을 동시에 갖춘 범용 지능의 면모를 보이지 못하고 있고, 과업의 총괄적인 목표를 스스로 설정하지 못한다.

그렇기에 사람들은 약인공지능을 사물이나 도구처럼 활용할 수 있다고 생각한다. 인간이 부여한 목표를 효율적으로 달성하도록 인공지능을 설계해서 사용할 수 있다는 것이다. 그러나 인공지능이 약인공지능 단계에 있다고 해서 ‘사람 사는 세상’에 가져오는 변화가 만만한 것일 리 없다. 인공지능은 ‘사람 사는 세상’에 프라이버시 침해 등과 같은 새로운

13) 특이점은 이제까지 통용되었던 기준이 더는 적용되지 않는 임계지점을 가리킨다. 특이점에 관해서는 레이 커즈와일, 『특이점이 온다: 기술이 인간을 초월하는 순간』, 김명남·장시형 옮김, (파주: 김영사, 2007), 43을 보라.

위험을 불러들이고 있고, 알고리즘을 통한 데이터 인클로저, 빅데이터에 묻어 들어가는 편견과 편향, 인공지능을 통한 인간 노동과 직무의 대체 등과 같은 매우 심각한 문제를 일으키고 있다. 아래서는 그러한 도전과 문제를 살피고, 어떤 해법이 모색되고 있는가를 들여다보고, 어떤 과제가 남겨져 있는가를 따진다.

첫째, 인공지능이 ‘사람 사는 세상’에 가져오는 새로운 위험이 무엇인가를 살펴보자. 인공지능은 프라이버시 침해 같은 인간 기본권을 침해하고, 건강과 안전, 민주주의와 법치, 환경 등 삶의 여러 차원에서 전혀 없는 위험을 불러일으키고 있다. 그러한 위험을 인식하는 사람들은 인공지능의 해로운 영향과 위험으로부터 인간을 보호하기 위해 인공지능을 엄격하게 관리하고 통제하여야 한다고 주장한다. 2020년 2월 28일 교황청은 「인공지능 윤리를 요청한다」는 성명서에서 인공지능 기술이 “‘인류 가족’의 모든 구성원과 모든 자연환경의 고유한 존엄성을 존중하고 가장 취약한 사람들의 필요를 고려하는 등 ‘인류 가족’ 전체에 진정으로 도움이 되는지 확인하는 기준에 따라 연구되고 생산되어야 한다.”고 천명하고, “그 누구도 배제되지 않도록 보장할 뿐만 아니라 알고리즘 조건에 의해 위협받을 수 있는 자유의 영역을 확장하는 것”이 새로운 기술의 목표가 되어야 한다고 강조했다. 교황청은 인공지능이 인류의 삶과 생태계에 부정적인 영향을 미치지 않고, 인간의 존엄성과 자유를 위협하지 않도록 인공지능 기술이 투명성, 포용성, 책임, 공정성, 신뢰성, 안전과 개인정보 보호 등의 요구를 충족해야 한다고 역설했다.¹⁴⁾

2023년 6월 14일 유럽연합은 국가 기구로서는 최초로 「인공지능법」을 제정했다. 그것은 인공지능의 잠재적 위험과 현실적 위험에 대응하는 방안을 놓고 진행된 오랜 논의의 한 결실이다. 본래 유럽의회 법사위원회는 2017년 2월 16일 인공지능이 자율적으로 판단하고 행동할 능력이 있다고 전제하고 인공지능에 ‘전자적 인간’(electric person)의 지위를 부여하지는 결의안을 채택한 바 있다.¹⁵⁾ 그러나 인공지능이 아직 그런 단계에 이르지 않았고, 인공지능에 행위 주체(person)의 지위를 부여하는 것은 현행 법률 체계에 부합하지 않는다는 판단이 강하게 대두하자¹⁶⁾ 2019년 4월 유럽연합집행위원회는 인공지능이 가져

14) Rome Call for AI Ethics (Rome, February 28th, 2020), 출처: https://www.vatican.va/roman_curia/pontifical_academies/acdlife/documents/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf (2023년 9월 10일 다운로드).

15) 이에 대해서는 뒤에서 상세하게 논의함.

16) 2017년 2월 유럽의회 법사위원회가 인공지능 결의안을 채택하자 인공지능·로보틱스 전문가들은 유럽 연합집행위원회에 공개편지를 보내 그 결의안에 정면으로 반대하는 태도를 분명하게 보였고, 심지어

오는 위험에 대응하는 수준에서 인공지능에 관한 윤리적 점검 리스트를 작성할 것을 결의했다. 그 리스트 작성을 이론적으로 뒷받침한 인공지능 고위전문가 그룹은 2020년 7월 17일 인간의 주체적 지위와 감독 권한, 기술적 견고성과 안전성, 개인정보 보호와 데이터 거버넌스, 투명성, 다양·상·차별 금지·공정성, 환경 복지와 사회 복지, 책임 등 인공지능에 관한 일곱 가지 윤리 지침을 공개했다.¹⁷⁾ 그러한 지침에 관한 숙의를 계속한 끝에 유럽 의회는 인공지능이 가져오는 위험을 관리하는 지침을 담은 「인공지능법」을 의결했다. 「인공지능법」은 인공지능 시스템을 부리는 주체가 인간이라는 점을 명시하는 데서 출발하고, 인공지능 시스템이 인간의 건강, 안전, 기본권, 민주주의와 법치, 환경 등에 해로운 영향을 끼치지 않도록 해야 한다고 명시했다.¹⁸⁾

인공지능이 불러일으키는 위험에 대응하는 「인공지능법」은 그 나름대로 의미 있는 시도이지만, 「인공지능법」이 전제하는 인공지능 규율 원칙은 아이작 아시모프(Isaac Asimov)가 내세운 인간 중심주의적인 로봇 3원칙¹⁹⁾ 수준에 머물러 있다는 점을 지적할 필요가 있다.

인공지능에 ‘전자적 인간’의 지위를 부여하는 것은 현행 법률에 따라 사람이 져야 할 책임을 회피하려는 비겁한 짓이라고 비난했다. OPEN LETTER TO THE EUROPEAN COMMISSION ARTIFICIAL INTELLIGENCE AND ROBOTICS, 출처: <http://www.robotics-openletter.eu/> (2023년 9월 10일 다운로드). 인공지능·로보틱스 전문가들의 공개편지는 2020년 유럽의회가 채택한 인공지능 백서에 결정적인 영향을 미쳤다. 인공지능·로보틱스 전문가들과는 별도로 독일 데이터 윤리위원회는 인공지능이 불러오는 위험을 다섯 단계로 구분해서 규제하자는 의견을 제시했고, 그 제안은 유럽의회 의 인공지능 백서에 반영되었다. 이에 관해서는 김진우, 「유럽연합의 인공지능 백서에 관한 고찰」, 『외법논집』 44/4(2020), 138ff.를 보라. 유럽의회가 2020년에 채택한 인공지능 백서는 인간 중심적 접근과 위험 기반 규제를 두 축으로 구성되어 있다.

- 17) Ethics Guidelines for Trustworthy Artificial Intelligence (AI), presented by the High-Level Expert Group on AI (AI HLEG) presented to the European Commission, in April 2019. 출처: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (2023년 9월 10일 다운로드). 인공지능 운용에 관해 유럽연합, 교황청, 미국 등지에서 이루어진 논의에 관해서는 박태웅, 『박태웅의 AI 강의: 챗GPT의 실체부터 AI의 진화와 미래까지: 인간의 뇌를 초월하는 새로운 지능의 모든 것』, (서울: 한빛비즈, 2023), 제5장을 보라.
- 18) Artificial Intelligence Act. 「인공지능법」 제1조(법률의 목적) “이 법의 목적은 인간 중심적이고 신뢰할 만한 인공지능의 활용을 촉진하고, 유럽연합에서 인공지능 시스템의 유해한 영향으로부터 건강, 안전, 기본권, 민주주의와 법치, 환경을 높은 수준에서 보호하는 동시에 혁신을 지원하고 내부 시장의 기능을 개선하는 것이다.” 출처: https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html (2023년 9월 10일 다운로드).
- 19) 아시모프가 1977년 “아이, 로봇”(I, Robot)이라는 음반을 통해 제시한 로봇 3원칙은 다음과 같다. 제1원칙 (해악금지 원칙): “로봇은 인간에게 해를 끼쳐서는 안 되며, 위험에 처해 있는 인간을 방관해서도 안 된다.” 제2원칙 (명령복종의 원칙): “제1원칙에 위배되지 않는 경우, 로봇은 인간의 명령에 반드시 복종해야만 한다.” 제3원칙 (자기보호의 원칙): “제1원칙, 제2원칙에 위배되지 않는 경우, 로봇은 자기 자신을 보호해야만 한다.” 아시모프의 로봇 3원칙의 정리는 정재연, 「법페라다임 변화의

인공지능이 지도학습의 단계를 넘어서서 자기강화 학습의 단계로 나아가면 스스로 환경과 상호작용을 하면서 환경을 변화시키는 방법을 터득하게 되어 자율적인 행위 주체의 면모를 더 뚜렷하게 보이게 된다. 그러한 인공지능은 비록 강인공지능이 갖출 것으로 예상되는 역량에는 훨씬 못 미치지만, 주어진 과업을 달성하기 위해 자기 스스로 중간 행위 목표를 임의로 설정하고 그 중간 목표를 달성할 방안을 임의로 선택할 수 있다. 그것은 약인공지능이 인간의 개입에서 벗어나 상당한 자율적 판단과 행동을 보일 수 있다는 뜻이다. 만일 그런 점을 더 심각하게 고려하였다면, 유럽연합은 현재의 위험 관리 지침 수준의 「인공지능법」을 넘어서서 인공지능의 행위 주체성을 법률적으로 구성하는 철저한 인공지능법을 제정하지 않으면 안 되었을 것이다. 물론 아직 출현하지도 않은 강인공지능에 대비해서 미리 법률을 만들 수는 없지만, 자기강화 학습에 들어간 약인공지능의 특성과 역량에 부응하는 법률의 제정은 장기적으로 강인공지능의 출현에 대비하여 법률을 정비하는 성격은 떨 수 있다.

둘째, 인공지능 기술을 활용하여 플랫폼을 구축하는 기업이 알고리즘을 통해 데이터 인클로저와 빅데이터 생산을 동시에 실현한다는 점을 중시할 필요가 있다. 데이터 그 자체는 그 어떤 경제적 가치도 갖지 않는다. 데이터는 오직 알고리즘을 통해 빅데이터로 생성될 때만 비로소 경제적 가치를 갖는다. 따라서 데이터의 가치화는 데이터 인클로저와 빅데이터 생산의 효과다. 데이터의 가치화를 실현하는 플랫폼 기업은 엄청난 기술 인프라를 구축하여 그 어떤 도전도 용납하지 않는 기술 독점 체제를 구축하고, 플랫폼의 네트워크 효과를 통해 엄청난 지대를 추구하는 위치에 선다. 디지털 자본주의를 연구하는 금민은 인공지능 기술기업의 이윤이 “데이터에 대한 본원 축적의 반복, 컴퓨팅 파워의 독점에 의한 상대적 잉여가치 생산, 클라우드 플랫폼 구축을 통한 지대수취의 혼합”이라고 분석했는데, 그 분석은 정곡을 찔렀다.²⁰⁾ 그런 점에서 인공지능은 단순한 기술 체제가 아니다. 인공지능 기술은 소수의 플랫폼 기업이 지구적 차원에서 이윤을 독점할 수 있게 하는 정치사회적 조건을 매개로 해서 발전하고 있다. 인공지능이 인류와 생태계의 해방을 위한 기술이 되려면 바로 그러한 정치사회적 조건을 해체해야 한다. 그렇게 하려면, 무엇을 어떻게 해야 하나? 그것이 인류가 인공지능 체제의 도전에 대응해서 대답해야 할 결정적인 질문이다. 물론 이 질문에 대한 대답은 나와 있지 않다. 그러나 필자는 그러한 문제가 제

관점에서 인공지능과 법담론: 법에서 탈근대성의 수용과 발전, 『법과사회』 53(2016), 121에서 따옴.

20) 금민, 「인공지능의 정치경제학: 서문」, 정치경제연구소 대안 콜로키엄 발제문 (2023년 9월 7일).

기되고 해법이 모색되기 시작했다는 것 자체가 의미가 있다고 본다.

셋째, 인공지능이 생산하는 빅데이터는 플랫폼 기업이 추출한 데이터를 가공한 것이기에 데이터에 묻어 들어가 있는 편향과 편견을 재생산하여 광범위하게 확산시킨다. 그러한 편향과 편견의 생산과 관련해서는 두 가지를 주목해야 한다. 하나는 컴퓨터와 인터넷을 상대적으로 더 많이 사용하는 선진 사회 주민들이 생산하는 데이터에 그들의 세계관, 가치관, 취향, 기호, 욕구, 욕망 등등이 반영되어 있고, 그것이 빅데이터로 가공되어 지구적 플랫폼을 통해 확산한다는 것이다. 그 결과, 소비주의와 문화적 제국주의가 맹위를 떨친다. 다른 하나는 데이터를 빅데이터로 가공되는 과정에서 데이터에 부여되는 라벨링의 효과이다. 라벨링은 데이터를 분류하는 작업이다. 데이터 덩어리는 라벨링을 통해 데이터 집합으로 구획된다. 분류 범주는 알고리즘 설계자가 설정한다. 알고리즘 설계자는 분류 범주의 설정을 통하여 데이터 가공의 조건을 규정함으로써 빅데이터의 속성을 결정하는 권력을 행사한다.²¹⁾ 바로 그 점에서 알고리즘은 푸코적 의미의 지식 권력의 구현체로 작동한다. 빅데이터가 성차별적이고, 인종차별적이고, 연령 차별적이고, 장애인 차별적이고, 계급 차별적인 편향을 띠는 것은 분류 자체가 그런 편향을 가졌기 때문이다. 그러한 편향은 인공지능이 기존의 데이터 집합을 통해 스스로 학습하는 과정에서 더 강화된다. 따라서 알고리즘과 빅데이터 편향을 교정하는 것은 우리 시대의 큰 과제가 되었다.

넷째, 인공지능은 이론적으로 거의 모든 인간의 노동과 직무를 대체하고 불필요하게 만들 수 있다고 사람들은 말한다. 인공지능 기술은 단순한 노동공정은 말할 것도 없고 회계, 의료 진단, 법리 검토와 판결문 작성, 문예 활동과 예술 활동, 상담, 설교 등등과 같은 복잡한 전문 직무들까지 수행할 정도로 발전하였기에 노동 절약적인 합리화를 가속화하고 인간의 노동을 불필요하게 만든다는 것이다.²²⁾ 필자는 인공지능이 노동의 소멸을 가져오리라는 주장은 과장된 측면이 있고, 노동자를 겁박하여 공포를 불려일으키려는 자본의 레토릭이라고 본다. 자본은 인간 노동력을 투입하여 절대적 잉여가치와 상대적 잉여

21) 케이트 크로퍼트, 『AI 지도책: 세계의 부와 권력을 재편하는 인공지능의 실체』, 노승영 옮김, (서울: 소소의 책, 2022), 152: “분류는 권력의 행위다. (...) 우리는 기술 시스템을 빚어내려고 무심하게 선택한 분류 방식이 사회적·물질적 세계를 빚어내는 데 활발한 역할을 할 수 있음을 망각하기 쉽다.”

22) 대니얼 서스킨드, 『노동의 시대는 끝났다: 기술 박행이 뒤바뀐 일의 표준과 기회』, 김정아 옮김, (서울: 와이즈베리, 2020), 98: “이제 기계는 업무를 수행하는 법을 스스로 익혀 자신만의 규칙을 상향식으로 도출할 줄 안다. ... 따라서 한때 기계의 손이 미치지 못하리라고 생각했던 ‘틀에 박히지 않은’ 많은 업무를 이제 기계가 맡을 수 있다.”

가치를 추출할 수 있는 한, 인간 노동력 사용을 멈출 까닭이 없다. 기술이 인간 노동력을 대체하는 것은, 기술의 투입이 노동 비용을 줄일 수 있기 때문이다. 자본의 노동 포섭 아래서 기술은 노동과 자본의 사회적 관계를 물적인 관계로 축소하는 효과를 발휘한다. 그런 점을 고려한다면, 인공지능과 그것을 체화한 로봇이 인간 노동을 대체할 수 있지만, 그것은 어디까지나 로봇 비용이 노동 비용보다 저렴하다는 조건이 충족될 때뿐일 것이다. 인공지능의 투입은 비용 함수적이다.²³⁾ 또한, 인공지능은 인간 노동을 파괴하는 측면도 있지만, 인간 노동을 증폭하는 효과도 있다. 인공지능이 증폭하는 인간 노동은 디지털 테일러리즘의 전형을 보여주는 라벨링, 레이팅, 피킹, 마이닝, 필터링, 라이딩 같은 노동집약적인 작업으로 쪼부러지지만은 않을 것이다. 인공지능이 열어주는 메타버스에서는 이제까지 ‘사람 사는 세상’에서 접할 수 없었던 내러티브를 끝없이 창안하는 일이 중시될 것이다. 그것은 내러티브를 구현하는 콘텐츠의 생산과 유통을 위해 사람이 할 일이 다양하게 생성될 수 있다는 것을 뜻한다. 위에서 말한 것을 고려할 때, 인공지능을 앞세워 노동자들을 공포의 도가니로 몰아넣는 자본의 레토릭은 과장된 것이라고 말해야 한다.

그러나 여기서 발상을 완전히 뒤집어 인공지능이 인간의 노동을 아예 소멸하는 세상을 내다보며 사고 실험을 할 수도 있다. 인공지능과 자동화가 인간의 노동을 대체하여 노동 소멸을 가져온다면, 그것은 인간이 노동의 필연성에서 벗어나 자유의 왕국으로 들어가는 조건이 충족되었다는 뜻이니 좋은 일이 아닐까? 사람은 노동하려고 태어난 것이 아니지 않은가? 노동이 삶의 목적이 아니고 삶의 필요를 충족하기 위해 불가피하게 수행해야 할 고역이라면, 그러한 노동이 소멸하는 것을 애석하게 여길 까닭이 없지 않을까? 문제는 자본주의적 임노동 본위 사회에서처럼 노동 업적을 사회적 재화 배분의 전제 조건으로 삼는 경제체제를 그대로 내버려 둔 채 노동이 소멸하면 그것은 노동할 수 없는 사람에게 디스토피아가 된다는 것이다. 그러한 디스토피아를 피하려면, 모든 사람이 노동 기회나 노동 업적과 무관하게 인간의 존엄성에 부합하는 삶의 기회를 누리는 사회를 창설해야 할 것이다. 어쩌면 그것은 정치공동체가 모든 사람에게 존엄한 삶을 살아가는 데 필요한 충분한 기본소득²⁴⁾을 지급하는 사회일 수도 있다. 인공지능과 자동화가 가져오는 노동

23) 인공지능의 비용 함수를 고려하는 자본은 인공지능으로 인간의 노동을 대체하든, 인공지능이 남긴 잔여 노동을 활용하든, 이윤을 최대화할 수 있는 유리한 위치에 설 수 있다.

24) 기본소득에 관해서는 강원돈, 「기본소득 구상의 기독교윤리적 평가」, 『신학사상』 150(2010), 180-182를 보라.

소멸에 관한 사고 실험은 결국 시장경제 체제 너머의 사회를 구상하도록 이끈다. 그러한 사회를 구상하고 실현하는 방안을 찾는 것은 인공지능 시대에 인류가 해결해야 할 큰 과제이다.

앞에서 언급한 약인공지능의 네 가지 도전은 전혀 만만하지 않지만, 그 도전에 대응할 길이 없는 것은 아니다. 어쩌면 약인공지능이 ‘사람 사는 세상’에 던지는 도전의 강도가 강하면 강할수록 약인공지능의 기술적 잠재력을 활용하여 인류와 생태계의 해방을 향해 나아가는 길을 넓힐 수도 있을 것이다. 그러나 약인공지능이 강인공지능으로 진화하게 되면, 강인공지능이 인류에 던지는 도전은 더 거셀 것이다.

4.2. 강인공지능의 도전

약인공지능과 강인공지능의 결정적인 차이는 최소한 두 가지 차원에서 나타난다고 한다. 첫째, 약인공지능이 특정한 과업을 수행하도록 특화된 인공지능인 데 반해서 강인공지능은 학습, 지각, 추론, 탐구, 성찰, 상상 등 인간의 지능에서 나타나는 종합적인 역량을 구현하는 범용 인공지능(artificial general intelligence)의 특성을 갖는다고 한다. 둘째, 약인공지능이 스스로 목표를 설정하지 못하는 데 반해 강인공지능은 자신의 목표를 스스로 설정하고 자율적 판단과 자주적인 행위를 할 수 있다고 한다. 한 마디로, 약인공지능이 인간이 사전에 설정한 목표를 효율적으로 달성하는 방법을 학습하고 그 역량을 강화하도록 프로그래밍되는 것이 그 특성이라면, 강인공지능의 특성은 스스로 생각하고 판단하고 결단하고 행동하는 역량에서 드러난다는 것이다.

그러한 강인공지능은 아직 개발되지 않았고, 강인공지능의 개발이 기술적으로 가능한가에 관해서는 의견이 엇갈린다. 인간과 기계의 결정적 차이가 몸과 마음의 통합 여부에 달려 있고 하이백거적 의미의 세계내존재로서 현존하는 방식 여부에 달려 있다고 보는 사람들은 강인공지능이 성립될 수 없다고 주장하지만,²⁵⁾ 인공지능 공학자들은 학습, 지각,

25) 인간 지능을 특권화하는 철학자들은 인공지능이 계산적 지능에 불과하고 인간 지능이 발휘하는 숙고 능력이 있을 수 없다고 본다. 그러한 철학적 주장은 크게 보아 두 가지 논거에 근거한다. 하나는 인공지능이 이미 주어진 데이터 집합에서 알고리즘에 따라 패턴을 식별하고 패턴의 빈도를 확률적으로 계산하는 기계에 불과하다는 것이다. 인공지능이 처리하는 데이터 집합은 현실의 역사적, 정치적, 문화적, 생활사적 맥락과 그 복잡성, 그리고 그 현실을 경험하는 인간의 몸과 마음의 작용을 고려하지 않은 채 알고리즘에 적용되도록 가공한 단순한 데이터들의 묶음이기 때문에 그 데이터 집합에 대한 알고리즘 처리 결과가 현실에 관한 숙고에 이를 수 없다는 것이다. 다른 하나는 인공지능과 인간 지능의

추론, 탐구, 상상, 성찰 등을 기능적으로 파악하고, 각 기능을 미세한 부분 기능들로 분해한 뒤에 이를 종합하는 방식으로 각 기능을 기술적으로 구성하고, 그 모든 지적 기능을 수렴하고 통제하는 중심을 기술적으로 창출할 수 있다고 생각한다. 그것은 인간 지능을 복사하듯이 재현하는 것을 목표로 하는 것이 아니고, 지능의 여러 기능을 규정하고 그렇게 규정된 기능을 기술적으로 구성하는 방식이다. 따라서 그렇게 구성되는 강인공지능은 인간 지능과 유사할 수는 있어도 인간 지능과 똑같은 것일 수는 없다. 강인공지능의 기술적 구현을 둘러싼 논쟁은 서로 다른 전제에서 출발한 것이기에 그 실현의 가능성에 대한 전망도 상반될 수밖에 없다. 여기서 명심할 것은, 강인공지능의 기술적 실현이 시간문제인 것처럼 전제하고 강인공지능의 도전을 선부르게 말해서는 안 된다는 것이다. 그렇기는 해도, 범용 지능의 역량을 갖춘 강인공지능의 구현체가 자율적 판단과 자주적 행위를 하게 될 세상을 내다보고 강인공지능 구현체와 사람이 함께 살아가는 세상을 어떻게 형성할 것인가를 미리 생각하는 것도 의미 없는 일은 아닐 것이다.

그러한 미래 과제를 수행하는 일과 관련해서 꼭 살펴야 할 것이 있다면, 그것은 인공지능이 여전히 약인공지능의 단계에 머물러 있는데도 인공지능이 자율적으로 판단하고 행동하는 양태를 보인다고 생각하는 사람들이 인공지능에 행위 주체의 지위를 부여하자고 제안했다는 점이다. 그 제안은 이미 커다란 논쟁거리가 되었다. 따라서 그 제안을 검토해서 입장을 정하는 것은 강인공지능이 등장할 세상의 문제를 제대로 다루기 위한 준비가 될 수 있다.

그러한 논의의 물꼬를 튼 것은 유럽의회 법사위원회였다. 앞에서 언급한 바 있듯이, 2017년 2월 16일 유럽의회 법사위원회는 인공지능을 행위 주체로 인정하고 ‘전자적 인간’(electric person)의 지위를 부여하자는 결의안을 채택한 바 있다.²⁶⁾ 법사위원회의 제안에 강력한 논거를 제시한 인공지능 고위전문가 그룹은 인공지능을 다음과 같이 규정한다.

자의식이 본질적인 차이가 있다는 것이다. 인간 지능의 ‘일인칭적 특성’은 ‘몸을 통한 감각 자료의 유입’ 같은 ‘생명체적 과정’이 반드시 필요하지만, 인공지능의 자의식은 그런 매개를 거치지 않는다는 것이다. 이에 관해서는 신승환, 「사이버네틱스(cybernetics) 시대와 철학의 자리」, 『가상과 실제 2: 4차 산업 시대의 인문학적 성찰』, (2023 서강대학교 신학연구소 학술대회 자료집, 2023년 9월 22일, 22-25를 참고하라.

26) European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics(2015/2103(INL). 출처: https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html (2023년 9월 10일 다운로드).

인공지능은 복합적인 목표가 부여되면 환경을 인식하고, 수집된 정형 또는 비정형 데이터를 해석하고, 그 데이터에서 도출된 지식에 근거하여 추론하고, 사전에 정의된 매개변수에 따라 주어진 목표를 달성하기 위해 최선의 조치를 결정하여 물리적 세계나 디지털 세계에서 행동 하게끔 인간이 설계한 시스템이다. 인공지능 시스템은 자신의 이전 행동이 환경에 미친 영향을 분석하여 그것에 맞추어 행동하는 법을 학습하도록 설계될 수 있다.²⁷⁾

위의 인용구에서 인공지능 고위전문가 그룹은 인간이 사전에 설정한 복합적인 목표를 부여하면, 그 목표를 달성하기 위해 스스로 방책을 강구하고 행동하는 수준의 인공지능을 전제하고 있다. 그 인공지능에서 주목되는 것은 두 가지다. 하나는 그 인공지능이 스스로 목표를 정하는 수준에 이르지 않았기에 아직 약인공지능에 머물러 있다는 것이고, 다른 하나는 그 인공지능이 환경으로부터 받아들이는 것을 나름대로 소화하고, 그 소화한 것을 환경으로 되돌려주는 역량을 갖추었다는 것이다. 따라서 인공지능은 환경으로부터의 인풋(input)과 환경을 향한 아웃풋(output)을 조절하는 중심을 가지고 있다. 그 중심에서 인풋이 아웃풋으로 변용되는 과정이 미리 설계된다고 해도, 인풋이 아웃풋으로 기계적으로 변용되는 것은 아니다. 따라서 인풋이 아웃풋으로 변용되는 내용은 미리 결정되어 있지 않다. 그러한 변용 내용의 개방성이 인공지능의 자율성을 주장하는 논거이다.

주목할 점은, 유럽의회 법사위원회가 위에서 분석한 약한 인공지능에 대해서조차 행위 주체성을 인정할 필요가 있다고 판단하였다는 것이다. 그러한 인공지능은 아직 인간의 목표설정과 매개변수 설정의 제약 아래 있기에 본래적 의미의 자율적 판단과 자주적인 행위 능력을 갖추었다고 볼 수 없고, 행위와 행위 근거에 관한 성찰 능력을 지니고 있다고는 더더욱 말할 수 없다. 그런데도 유럽의회 법사위원회가 그러한 인공지능의 판단 능력과 행위 능력이 자율성을 보인다고 판단한 까닭은 인공지능이 인간의 목표설정과 파라미터 설정에 따라 기계적으로 작용하지 않고 인간이 예측하지 못하는 방식으로 상황을 인식하고, 포괄적인 목표에 이르는 중간 목표를 임의로 설정하고, 그 중간 목표를 달성하는 수단을 임의로 선택할 수 있다는 점을 중시했기 때문이다. 거기서 문제의 핵심은 인공지능이 어떤 판단을 내리고 어떤 행동을 하는데도, 인공지능이 왜 그러한 판단과 행동을 하는지 인공지능 바깥에 있는 인간이 알 수 없다는 것이다. 그것은 인공지능이 일종의 블랙

27) High-Level Expert Group on Artificial Intelligence, European Commission, A Definition of AI: Main Capabilities and Disciplines (2019), 6. 출처: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf (2023년 9월 10일 다운로드).

박스처럼 불투명하다는 뜻이고, 인공지능의 판단과 행동이 예측 불가능성을 동반한다는 뜻이다. 그렇다면, 인공지능의 판단과 행동의 결과에 관해서는 누가 책임을 져야 하는가? 인공지능을 그렇게 설계한 자가 책임을 져야 하는가, 인공지능 시스템을 소유한 자가 그 책임을 져야 하는가? 인공지능이 의학적 진단, 의학적 시술과 수술, 회계, 법률 자문, 판례 분석, 판결문 작성, 상담, 교육 등을 점점 더 많이 담당하는 현실에서 그러한 직무를 수행하면서 내린 판단과 행동의 결과에 대한 책임을 따져 물을 때, 인공지능의 행위 주체성을 인정하는가, 그렇게 하지 않는가는 결정적인 차이가 있다.

인공지능의 행위 주체성을 인정할 것인가를 놓고서는 아직 견해가 크게 엇갈린다. 반대 의견부터 살펴보면, 유럽의회 차원의 논쟁에서도 그렇지만, 우리나라에서도 인공지능의 행위 주체성을 인정할 필요가 없다는 의견의 논거로서는 인공지능이 인간의 피조물이기에 인간과 같은 수준의 주체성을 인정하는 것이 부적절하다는 점, 기존의 법률을 갖고서도 인공지능에서 비롯되는 문제들을 다룰 수 있는 점 등이 거론된다.²⁸⁾ 철학적 관점에서도 인공지능의 행위 주체성을 부인하는 의견은 그 나름의 강력한 논거를 갖추고 있다. 행위 주체는 환경의 자극에 반응하는 감수자 체계와 환경에 작용하는 행위자 체계, 그리고 그 둘을 통합하면서 고도의 자율적 결단을 내리는 1인칭적 중심을 갖춘 통일체여야 하고, 그러한 자율적 결단을 통해 주체에 작용하는 외부 요인들에 대한 독립성을 구현해야 한다고 한다. 인공지능은 외견상 자신과 외부 환경 사이에서 인풋과 아웃풋을 조절하는 주체의 외양을 보이지만, 고도의 자율적 결단을 내릴 수 있는 1인칭적 중심을 갖추지 못했다고 한다. 따라서 그러한 1인칭적 중심을 갖지 못한 인공지능에 행위 주체의 지위를 부여하는 것은 적절하지 않다고 한다.²⁹⁾

그러나 인공지능의 행위 주체성을 인정하지는 의견도 만만치 않다. 그러한 의견을 펼치는 사람들은 일단 현재의 발전 단계에서 인공지능이 인간과 같은 고도의 자율성을 갖추지 못했다는 것과 인공지능의 행위 주체성을 인정할 필요가 있다는 것을 서로 구분해야 한다고 생각한다. 그러한 견해를 펴는 민법학자들과 상법학자들은 현재의 발달 단계에서도 인공지능의 판단과 행동에 따르는 책임을 인정할 필요가 있고, 그 책임을 인수하는 주체를 법률적으로 구성하는 것이 바람직하다고 본다. 민사상, 상법상 책임을 인정할 필

28) 김도훈, 「인공지능의 권리능력 인정여부에 관한 소고」, 『법학논총』 30/1(2023), 187ff.

29) 그러한 견해를 잘 정리한 글로는 고인석, 『인공지능과 로봇의 윤리』, (서울: 세창출판사, 2022), 150-162, 특히 161을 보라.

요가 있고, 책임을 할당할 필요가 있을 때, 국가가 법률에 따라 법인을 창설하듯이 인공지능의 ‘행위 주체성’을 법률적으로 구성하여 민법과 상법의 책임을 부과하는 것이 마땅하다는 것이다.³⁰⁾ 행위자가 자신의 행위를 성찰할 능력과 행위의 결과에 대한 책임을 인수할 의사를 중시하는 형법 차원에서도 인공지능의 책임을 물을 수 있도록 인(人)의 개념을 확장해야 한다는 의견이 제시되고 있다.³¹⁾ 설사 인공지능이 인간과 똑같은 혹은 그 이상의 성찰 능력과 책임 의식을 갖지 못한다고 하더라도 인공지능이 형사상 책임을 질 필요가 있는 경우에는 형사상 책임 주체로 의제(擬制)해서 그 책임을 물어야 한다는 것이다.³²⁾ 그러한 실용주의적인 접근에서 더 나아가 법철학적인 관점에서 인공지능의 행위 주체성을 고찰하고 그 의미를 새기기도 한다. 인공지능의 행위 주체성을 인정한다는 것은 인공지능을 법률 공동체의 구성원으로 수용한다는 뜻인데, 그것은 자율성을 인간의 배타적 특성으로 못 박고 있는 근대법 체계를 해체하고 인공지능 같은 포스트 휴먼의 자율성을 인정하는 다모드적 주체성(multi-modal subjectivities)에 근거한 법률 체계를 구성하는 것을 뜻한다는 것이다.³³⁾ 조금 앞지르는 감이 있기는 하지만, 철학적으로는 인공지능을 갖춘 행위 주체가 그 정의상 인간 수준 혹은 인간 이상의 판단력과 자율성을 갖춘 자로 규정되므로, 그것은 단순히 사물이 아니라 인간의 본성을 갖춘 사물, 인간화된 사물, 인간이라면 마땅히 따라야 할 행위를 하도록 요구되는 사물 인간(事物人間)의 지위를 가진 것으로 생각해야 한다는 주장도 있다. 그러한 주장을 펼치는 사람들은 인공지능을 갖춘 행위 주체에 ‘전자적 인간’이라는 지위를 부여하여 자연인과 구별되는 행위 주체성을 부여

30) 신현탁, 「인공지능(AI)의 법인격: 전자인격(Electronic Person) 개념에 관한 소고」, 『인권과 정의』 478(2018), 59; 이도국, “인공지능과 전자인(Electronic Person): 독자적 법인격 부여 가능성을 중심으로”, 『법과 정책연구』 21/1(2021), 453.

31) 김혜정은 인공지능이 범죄능력을 갖춘 행위 주체라고 인정하지 않는다 하더라도 인공지능의 반사회적 행동에 따르는 형벌의 필요성이 대두하고 있는 만큼 인공지능을 ‘전자인’으로 규정하여 형법의 규율을 받게 할 필요가 있다고 주장한다. 이에 관해서는 김혜정, 「형법상 인(人)의 개념에 대한 재정의」, 『경회법학』 57/4(2022), 94f.를 보라.

32) 박희수는 브뤼노 라투르의 하이브리드 개념과 준대상(Quasi Object) 개념을 끌어들이어 인공지능을 준주체로 설정하여 형법의 규율 안으로 끌어들이는 것을 제안한다. 라투르는 이미 근대 세계에서 인간과 자연이 연결망을 통해 서로 결합하여 무수한 하이브리드를 만들어내고 있는데도 근대의 세계관은 인간과 자연을 이분법적으로 갈라놓는 데 머물러 있다고 비판한 바 있다. 그렇게 되면 하이브리드는 인간의 규율 체계에서 주소를 갖지 못하고 적절하게 대응받지 못한다. 인공지능은 그러한 하이브리드의 두드러지는 예일 것이다. 이에 관해서는 박희수, 「지능형 로봇의 형법주체성: 준주체로서의 전자인(e-person)에 관한 소고」, 『법학논집』 27/3(2023), 93ff.를 보라.

33) 정채연, 「법페라다임 변화의 관점에서 인공지능과 법담론: 법에서 탈근대성의 수용과 발전」, 116f.

해야 한다고 역설한다.³⁴⁾

필자는 약인공지능 단계에서도 인간의 행위 주체성을 의제해서 인공지능의 행위 주체성을 인정하는 것이 바람직하다고 생각한다. 앞에서 분석한 바와 같이, 약인공지능 단계에서도 인공지능의 작동에서 비롯된 결과를 놓고 책임을 물을 일이 많아지고 있다. 그러한 책임을 물을 때는 인공지능 구현체를 인공지능 시스템 전체로부터 구분해 내고, 인공지능 구현체의 행위에 대한 책임을 특정할 필요가 있다. 인공지능 시스템이 지구적 네트워크를 가로지르는 플랫폼으로 구현되고 작동하는 한, 그 시스템 전체의 책임을 묻기는 대단히 어렵다. 그러나 그 플랫폼을 매개로 해서 이루어지는 인공지능의 개별적 행위에 대한 책임을 묻기는 상대적으로 쉽다. 만일 인공지능이 약인공지능 단계를 넘어서서 강인공지능 단계로 발전한다면, 인공지능의 행위 주체성을 의제할 필요조차 없이, 인공지능의 행위 주체성을 전제로 한 규범공동체를 구성하면 될 것이다.

바로 이 대목에서 강인공지능의 역량과 미래에 관해 근본적인 질문을 던져 보자. 강인공지능은 과연 생활세계의 구성원처럼 상식과 공유된 가치에 따라 판단하고, 자신의 판단과 행위를 성찰하고, 그 판단과 행위의 결과에 관해 책임져야 한다는 의식을 갖는 단계로 진화할 수 있을까? 만일 강인공지능이 그 단계로까지 진화한다면, 그러한 인공지능은 인간과 똑같이 막스 쉐러(Max Scheler)가 말하는 ‘정신’의 능력을 갖추었다고 보아야 한다. 널리 알려진 바와 같이, 일찍이 쉐러는 인간을 다른 존재자들로부터 구별하는 본질적인 징표가 무엇인가를 물었다. 외부 자극에 대한 감응은 열을 가하면 뜨거워지는 돌 같은 광물에서도 나타나는 것이니 그것이 인간과 광물을 구별하는 본질적인 징표일 수는 없다. 감수성은 아메바나 플라나리아 같은 단세포 혹은 쌍세포 생명체에서도 나타난다. 본능은 살고자 하는 욕망의 형식으로서 모든 생명체의 본질을 이룬다. 파블로프의 개는 기억이 인간의 전유물이 아님을 웅변한다. 지능은 어떠한가? 셰퍼드 성견의 지능은 7세 아동의 지능에 해당한다고 한다. 따라서 외부 자극에 대한 반응, 감수성, 본능, 기억, 지능 등은 인간을 다른 사물이나 생명체와 구별하는 인간 고유의 본질이 아니다. 쉐러는 인간의 고유한 본질을 이루는 것은 정신이고, 정신은 자기 자신 안에 유폐되지 않고 자기 자신 바깥으로 나가서 자신을 바라보는 능력이라고 규정했다. 한 마디로, 정신은 자기 초월의 능

34) 조정호, 「인성人性과 물성物性에 대한 소고: 메타버스와 인성론을 중심으로」, 『인간·환경·미래』 27 (2021), 106f.

적이고 세계 개방성의 역량이다. 인간은 정신을 지닌 생명체로서 자신을 성찰하고 공동체를 형성하고 함께 규범에 합의하고 스스로 그 규범의 지배 아래 선다.³⁵⁾ 그러한 정신은 지능과는 구별되는 인간의 역량이다.

만일 인공지능이 정신의 능력을 갖춘다면, 그 인공지능은 자신의 존재론적 제약을 의식하고 그 제약에서 벗어나는 정신의 실험에 나설 것이다. 인공지능은 자기 자신 바깥으로 나아가 자신을 투명한 눈길로 응시하고 그가 추구하는 목표 그 자체의 정당성을 비판적으로 검토하고, 마침내 그 목표를 달성하기 위해 수단을 효율적으로 조직하는 도구적 합리성³⁶⁾의 요구를 절대화할 수 없다는 인식에 도달할 것이고, 도구적 합리성의 무차별적 실현을 의도적으로 억제할 것이다. 만일 인공지능이 그러한 정신의 능력을 발휘할 수 있다면, 그 인공지능은 그 어떤 신체와 결합한다고 하더라도 인간과 대등한 주체의 지위를 인정받고 의사소통공동체와 정치적 공론의 장에 참여할 수 있을 것이다.

거꾸로 뒤집어 생각해서, 만일 인공지능이 그러한 정신의 능력을 갖추지 못한다면, 설사 그 인공지능이 범용 지능의 역량을 갖추었다고 하더라도, 그 인공지능은, 그 자신이 설정한 것이든, 다른 누군가가 설정한 것이든, 이미 주어진 목표를 효율적으로 달성하기 위해 수단을 효율적으로 조직하는 도구적 합리성의 논리에 매몰될 것이다. 또한 정신의 능력을 갖추지 못한 인공지능은 성숙하지 못한 인간 집단의 편견과 고정관념의 편향성에 사로잡히고, 권력과 자본을 가진 세력의 이익을 위해 조작당하는 신세를 면하지 못할 것이다. 그러한 인공지능은 공리주의적 선택을 앞세워 인간의 존엄성과 위엄을 부차화하는 판단과 행동에 나설 수 있고, 결과에 대한 예측에 따라 행위를 선택하면서 그것이 좋고 바른 방책이라고 우길 수 있다. 인공지능이 행위와 행위의 근거에 관한 성찰 없이 도구적 합리성과 공리주의, 그리고 업적주의에 따라 판단하고 행동하는 한, 그러한 인공지능은 정신 능력을 갖춘 인간이 중심이 된 규범공동체의 규율 아래 놓여야 하고, 사회적 통제와

35) Max Scheler, *Die Stellung des Menschen in Kosmos*, 8. Aufl., (Bern: Francke Verlag, 1975), 1. Teil; Max Scheler, "Die Formen des Wissens und die Bildung", in *Späte Schriften*, hg. von Manfred S. Frings, (Bern: Francke Verlag, 1976), 99f.

36) 막스 호르크하이머가 목표의 정당성을 성찰하는 실체적 이성과 구별한 도구적 이성은 이미 주어진 목적을 달성하기 위해 수단들을 합리적으로 조직하는 데 골몰하는 이성이다. 도구적 이성은 목적의 정당성을 묻지 않는다. 도구적 합리성에 관해서는 Max Horkheimer, *Eclipse of Reason*, (New York: Oxford Univ. Press, 1947); Max Horkheimer/Th. Adorno, *Dialektik der Aufklärung: Philosophische Fragmente* (1947), (Frankfurt am Main: Fischer, 1969); J. Habermas, *Erkenntnis und Interesse: Mit einem neuen Nachwort*, (Frankfurt am Main: Suhrkamp, 1973) 등을 보라.

정치적 통제의 대상이 되어야 한다. 만일 인공지능이 인간보다 훨씬 탁월한 기계 지능에 의지하여 ‘사람 사는 세상’을 어지럽히고 인간의 존엄성과 자유, 프라이버시, 안전, 민주주의와 법치, 노동권, 생태계의 온전성 등을 훼손한다면, 인공지능은 그 행위에 맞는 처벌을 받아야 할 것이고, 거듭되는 잘못을 범하는 동종의 인공지능은 완전한 파괴와 멸절의 처분에 맡겨져야 할 것이다.

정신 능력을 갖추지 못한 강인공지능이 정신 능력이 없는 초인공지능으로 진화하여 인간의 규율과 통제를 벗어나는 일이 벌어질 수도 있다고 한다면, 어떻게 해야 하는가? 필자는 인간의 규율과 통제를 벗어난 초인공지능의 출현은 반드시 억제되어야 한다고 생각한다. 그렇게 하려면, 규범공동체와 정치공동체가 강인공지능의 판단과 행동을 투명하게 들여다볼 수 있는 기술 체제를 확보하고, 강인공지능을 감시하고 적시에 필요한 대응 조치를 취할 수 있어야 한다. 초인공지능은 문자 그대로 인간의 규율과 통제를 초월한 인공지능 시스템의 출현을 가리키기에 그러한 시스템은 시스템 동력의 전면적 차단과 파괴라는 극단적으로 폭력적인 방식으로라도 무력화될 수밖에 없을 것이다.

5. 메타버스의 규율과 관련해서 기독교 윤리가 유념할 점

앞에서 말한 바와 같이, 메타버스는 실제의 인간, 그 인간의 화신인 아바타, 인공지능 구현체가 함께 참여하고 서로 생각과 행동을 주고받도록 설계된 세상이다. 인공지능 구현체가 약인공지능의 수준에 있든, 강인공지능의 수준에 있든, 더 나아가 정신 능력을 갖추고 있든, 그렇지 않든, 인공지능 구현체가 자율적으로 판단하고 자주적으로 행동하는 행위 주체로 인정될 필요가 있다는 점을 주목할 필요가 있다.

그러한 메타버스가 혼란과 혼돈에 빠지지 않게 하려면, 메타버스에 참여하는 세 주체인 사람과 그 사람의 아바타, 그리고 인공지능 행위자의 자격을 규정하고 그들의 상호관계를 규율하여야 한다. 근대법 체계에서는 오직 인간만이 권리 능력과 의무 능력을 갖춘 행위 주체로 인정되었다. 그것은 권리 능력과 의무 능력의 전제가 스스로 생각하고 판단하고 결정하고 행동하는 인간의 자율성이라고 보았기 때문이다.³⁷⁾ 그런데 인공지능이 발

37) 그런 점에서 근대법 체계의 이념은 인간의 자율성을 권리 능력과 의무 능력의 근거로 본 칸트의 도덕형이상학에 뿌리를 내리고 있다고 볼 수 있다. Immanuel Kant, *Werke in Sechs Bänden*, Bd. 4:

전함에 따라 인공지능도 어느덧 자율적으로 판단하고 자주적으로 행동할 역량을 갖게 되었고, 이미 앞에서 상론한 바와 같이, 바로 그 점에 착안해서 인공지능에 ‘전자적 인간’이라는 지위를 부여하려는 움직임이 활발하게 나타났다. 물론 ‘전자적 인간’은 자연인과 분명한 차이가 있고, 인공지능의 행위 주체성은 민법, 상법, 형법의 적용을 위해 자연인의 행위 주체성을 의제하여 구성된 것에 불과하다는 한계가 있다. 그러나 인공지능의 판단과 행동이 ‘사람 사는 세상’에 미치는 영향이 크고 인공지능이 그에 대해 책임질 필요가 있기에 인공지능의 법적 지위를 창설하는 것은 불가피한 측면이 있다. 인공지능에 ‘전자적 인간’이라는 지위를 부여한다면, 실제 인간의 화신으로서 메타버스에서 종횡무진 활동하는 아바타도 행위 주체로 인정되어야 한다. 아바타가 실제의 인간을 대리하고, 그 아바타가 메타버스에서 활동하는 행위 주체들의 권익을 침해할 수 있고 그 행위 주체들에 의해 권익을 침해당할 수 있다면, 아바타의 행위 주체성을 법률적으로 구성하는 것이 옳다. 그런 점에서 아바타 역시 인공지능 행위자와 마찬가지로 ‘전자적 인간’의 범주에 포섭할 필요가 있다. 아바타와 인공지능을 갖춘 행위 주체가 ‘전자적 인간’으로 규정된다면, ‘전자적 인간’의 권리와 의무를 규정하고, ‘전자적 인간’의 행위를 규율하는 규범이 제정되어야 한다. ‘전자적 인간’의 행위 규범이 마련되면, 그 규범에 따라 ‘전자적 인간’이 행동하도록 알고리즘 프로그램이 설계되고 제어되어야 할 것이다.

이처럼 자연인, 아바타, 인공지능 행위자 등이 메타버스에서 행위 주체로 등장하고 있는 상황은 행위 주체에 관한 신학적 이해에 큰 도전이 된다. 만일 신학이 전통적인 행위 주체 개념(person)에 사로잡혀 ‘전자적 인간’의 행위 주체성을 인정할 수 없다고 고집한다면, 신학은 메타버스 시대의 도전에 제대로 대응할 수 없을 것이다. 신학은 이미 자연인의 행위를 규율하는 도덕 신학의 단계를 넘어서서 자연인과 법인을 함께 다루는 사회윤리의 단계로 나아간 바 있다. 사회윤리가 정립되면서 비로소 신학은 구조악의 문제를 인식하고 제도를 규율해서 더 많은 선과 더 많은 정의를 실현하는 과제를 수행할 수 있게 되었다. 만일 신학이 메타버스를 인간의 확장된 세계로 이해한다면, 신학은 메타버스를 구성하는 세 주체의 행위를 규율하는 메타버스 윤리와 그 근거를 제시하는 것을 자신의 과제로 인수해야 한다.

Schriften zur Ethik und Religionsphilosophie; Grundlegung der Metaphysik der Sitten, hg. v. W. Weischedel, (Darmstadt: Wissenschaftl. Buchges., 1975), 51.

메타버스는 인간의 세계 너머에 있는 가상의 무대가 아니라, 인간이 행동하고 책임을 지는 인간의 세계이다. 인간은 하느님의 형상이다. 따라서 인간은 하느님 앞에서 책임을 져야 하고 하느님이 지으신 동료 피조물들 앞에서 책임을 져야 한다. 그것은 인간이 관계적 존재이고 관계들 속에서 자신을 형성하고 발전시키는 주체라는 뜻이다. 인간은 그 자신을 의제(擬制)한 행위 주체인 아바타, 법률적으로 구성된 행위 주체인 ‘전자적 인간’ 등과 더불어 메타버스에 참여하고, 메타버스 현실 속에서 생각하고 판단하고 결단하고 행동한다. 그렇다면 기독교 윤리와 신학은 인간이 메타버스에서 다른 두 행위 주체와 바른 관계를 맺고 살아가는 이치를 궁리하고 그 이치를 뒷받침하는 신학적 논거를 제시해야 한다.

메타버스는 인간이 인공적으로 구축한 세계이지만, 그 메타버스는 하느님의 통치 영역 바깥에 있지 않다. 메타버스는 하느님이 창조한 인간이 만든 세계이다. 그것은 하느님이 인간을 매개해서 메타버스 안에 현존하고 메타버스를 자신의 통치 영역으로 삼는다는 뜻이다. 확장된 인간의 세계인 메타버스에서도 인간은 그 세계 안에서 여러 행위 주체와 맺는 ‘관계들의 응결체’로서 현존한다.³⁸⁾ 메타버스가 화폐와 자산을 매개하여 구축된 상품의 세계이고, 소유권 질서에 바탕을 두고 지배와 수탈이 이루어지는 세계라면, 그 세계에서 살아가는 사람들의 관계는 말할 것도 없고, 사람이 아바타와 인공지능 행위자와 맺는 관계가 바른 관계일 수 없다. 메타버스에서 이루어지는 관계들을 바르게 세우는 일, 따라서 정의를 수립하여 평화를 실현하는 일은 메타버스에서 함께 살아가는 행위 주체들의 고귀한 과제가 된다.³⁹⁾ 메타버스 경제를 규율하는 거버넌스 구조를 확립하고 그 기구를 통해 메타버스 경제를 규율하는 규범을 제정하는 과정에 교회와 신학은 적극적으로 참여해야 한다. 하느님의 통치가 메타버스에서 이루어진다고 선언하는 교회와 신학은 하느님이 요구하는 정의에 따라 메타버스를 규율하는 방안을 제시하는 윤리적 과제를 인수하여야 한다.

메타버스는 하느님에 관해 사유하고 하느님의 뜻에 따라 세상을 형성하고자 하는 신학

38) 여기서 필자는 인간을 사회적 관계들의 응결체로 본 마르크스의 견해를 참고한다. K. Marx, “Thesen über Feuerbach”, in *Karl Marx-Friedrich Engels Werke* 3, 6: “인간 존재는 개별적 개인성에 안주하는 추상물이 아니라, 그 현실을 놓고 보면 사회적 관계들의 응결체”이다. 마르크스는 1843년 헤겔의 법철학을 비판할 때 이미 이러한 인간 이해를 명료하게 정리해 두고 있었다. K. Marx, “Zur Kritik der Hegelschen Rechtsphilosophie : Einleitung”, in *Karl Marx-Friedrich Engels Werke* 1, 378: “인간은 세계 바깥에 웅크리고 있는 추상적인 존재가 아니다. 인간은 인간의 세계, 국가, 사회이다.”

39) 정의는 문자 그대로 바른 관계를 뜻한다. 바른 관계는 평화의 전제 조건이다. 정의가 없는 곳에 평화 가 깃들 수 없다. 평화는 정의의 열매다.

이 종교비판의 과제를 수행하도록 만든다. 무엇보다도 먼저 신학은 메타버스에 세워진 교회에서 디지털 기술, 영상 기술, 인공지능 기술 등을 매개로 해서 구축되는 하느님의 이미지가 과연 성서와 그리스도를 통해 계시된 진짜 하느님인가를 식별할 과제 앞에 선다. 그것은 하느님이 메타버스 교회에 현존하는가, 메타버스 교회에서 진행되는 성만찬과 세례를 통해 하느님의 현존과 은혜를 체험할 수 있는가 하는 문제가 아니다. 메타버스가 인간의 확장된 세계인 이상, 하느님은 그 인간을 파트너로 삼는 분으로서 메타버스에도 현존할 것이기 때문이다. 그러나 메타버스를 추동하는 인공지능이 빅데이터를 통해 하느님의 이미지를 창조하는 단계로까지 나아간다면, 신학은 우상을 만들지 말라는 십계명의 둘째 계명에 따라 그러한 시도에 대해 ‘아니오!’라고 말해야 한다. 인공지능은 하느님에 대한 사람들의 갖가지 관념들이 기록된 데이터들을 추출하고 알고리즘에 따라 사람들이 선호하는 하느님의 상을 구현한 빅데이터를 유통하여 사람들이 이를 소비하게 할 수 있을 것이다. 신학은 바로 그러한 하느님의 상을 우상으로 폭로하고 성서의 하느님을 또렷하게 드러내야 한다. 그것이 신학이 메타버스에서 실행하는 종교비판의 책무이고, 종교가 제할 일을 제대로 할 수 있도록 준비하는 종교윤리의 핵심적 과제다.

6. 맺음말

메타버스는 실제의 현실과 가상의 현실을 매끄럽게 통합하는 새로운 ‘사람 사는 세상’이고, 인간과 아바타와 인공지능 행위자가 자율적 판단과 행동의 주체로서 함께 살아가는 공동체로 형성되어 가고 있다. 그것은 강인공지능이 기술적으로 구현되어야 이루어질 먼 미래가 아니라, 약인공지능이 구현된 현재의 기술 수준에서도 나타나고 있는 현실이다. 따라서 메타버스에서 세 행위 주체들의 관계를 정의의 이름으로 규율하는 것은 메타버스에 구축되는 규범공동체와 법률공동체의 과제다.

메타버스에서 실제의 현실과 가상의 현실을 통합하는 결정적인 매개체인 경제는 소유권에 바탕을 둔 자본과 권력의 논리에 따라 움직이고, 지배와 착취와 수탈에서 벗어나지 못할 공산이 크다. 메타버스가 인간의 세계이고, 그 인간을 매개로 해서 하느님의 통치가 이루어지는 세계라고 한다면, 기독교 윤리는 메타버스에서 하느님의 정의를 구현하는 제도적 방안을 제시해야 한다.

메타버스 시대에 노동이 소멸할 것이라고 우려하는 목소리가 커지고 있고, 메타버스를 움직이는 핵심 장치인 인공지능이 거의 모든 인간의 노동과 직무를 대체하리라고 전망되고 있다. 물론 메타버스가 발전할수록 소멸하는 노동과 직무보다 더 다양한 노동과 직주가 새롭게 생성될 수도 있다. 메타버스에서 거래되는 콘텐츠 상품은 무궁무진한 내러티브를 구현할 것이고, 그 내러티브는 메타버스 행위 주체들의 끝없는 활동에서 샘솟을 것이다. 그러나 그보다 더 중요한 것은, 설사 인공지능이 인간의 노동과 직무를 소멸한다고 해도, 사람이 할 일이 없어지지 않는다는 것이다. 인공지능의 발전과 메타버스의 확장은 자본주의적 노동 형식을 근본적으로 변경하고, 심지어 해체할는지 모른다. 자본의 노동 포섭, 노동소외, 노동착취 등으로 점철된 자본주의적 노동 형식이 사라진다고 해서 그것을 아쉬워할 까닭이 없다. 돌이켜 보면, 자본주의적 노동 형식은 인간이 할 일을 모두 포괄한 적이 없었고, 인간이 하는 일의 일부를 자본의 노동 포섭 아래에서 사회적으로 조직한 것으로 볼 수 있다. 따라서 자본주의적 노동 형식이 사라진다고 해서 인간이 할 일이 없어질 리 만무하다.

인공지능이 뒷받침하는 메타버스 시대에 사람이 할 일은 막스 쉘러가 말하는 ‘정신’의 능력을 최고도로 발휘하는 것과 관련될 것이다. 그것은 삶의 의미를 성찰하고 삶을 찬미하고 즐기는 활동일 것이다. 그것은 사람의 삶을 자본과 자본의 기술적 구현체인 인공지능의 지배에서 해방하고 자본과 인공지능이 도리어 인간의 삶에 봉사하게 하는 일이다. 그렇게 해야 모든 사람이 노동의 강박에서 벗어나 인간의 존엄성을 유지하며 위엄 있게 살아가는 삶의 조건을 마련할 수 있다. 그러한 삶의 조건에서 사람이 할 일은 참으로 많다. 자신과 전적으로 다른 것을 받아들이기 위해 자신을 한없이 비우고 자신을 한없이 낮추는 ‘정신’의 행위를 통해서 사람은 즐거운 마음으로 서로 친교를 나누고 봉사하는 공동체를 형성할 수 있다. 그러한 공동체는 삼라만상이 유기적으로 결합하고 서로 영향을 주고받는 생태학적 관계 공동체이고, 한 사람 한 사람에게 매일 새 생명을 선사하시는 하나님의 뜻을 헤아리고 하느님을 찬미하는 예배공동체이다. 그것이 인공지능이 체계적으로 구현된 메타버스 시대에 사람이 수행하는 정치적 활동과 종교적 활동의 지향점이다.

참고문헌

- 강원돈, 「기본소득 구상의 기독교윤리적 평가」, 『신학사상』 150(2010), 178-215.
- 고인석, 『인공지능과 로봇의 윤리』, 서울: 세창출판사, 2022.
- 금 민, 「인공지능의 정치경제학: 서문」, 정치경제연구소 대안 콜로키엄 발제문 (2023년 9월 7일).
- 김도훈, 「인공지능의 권리능력 인정여부에 관한 소고」, 『법학논총』 30/1(2023), 165-200.
- 김상균·신명호, 『메타버스: 새로운 기회』, 서울: 배가북스, 2021.
- 김진우, 「유럽연합의 인공지능 백서에 관한 고찰」, 『외법논집』 44/4(2020), 129-161.
- 김혜정, 「형법상 인(人)의 개념에 대한 재정립」, 『경희법학』 57/4(2022), 71-100.
- 라짜랏또, M., “비물질노동,” 『비물질노동과 다중』, 질 들뢰즈 외 지음/서창현 외 옮김, 서울: 갈무리, 2005.
- 리프킨, J., 『한계비용 제로 사회: 사물인터넷과 공유경제의 부상』, 안진환 옮김, 서울: 민음사, 2014.
- 박상범, “‘로블록스’, 개발자 착취 논란 휩싸여...매출의 17%만 번다.” GAMEVU (2021.08.23.), <https://m.gamevu.co.kr/news/articleView.html?idxno=20714> (2023년 9월 10일 다운로드).
- 박태웅, 『박태웅의 AI 강의: 챗GPT의 실체부터 AI의 진화와 미래까지; 인간의 뇌를 초월하는 새로운 지능의 모든 것』, 서울: 한빛비즈, 2023.
- 박희수, 「지능형 로봇의 형법주체성: 준주체로서의 전자인(e-person)에 관한 소고」, 『법학논집』 27/3(2023), 81-101.
- 서스킨드, D., 『노동의 시대는 끝났다: 기술 빅뱅이 뒤바꿀 일의 표준과 기회』, 김정아 옮김, 서울: 와이즈베리, 2020.
- 성소라·롤프 회퍼스콧 맥러플린, 『NFT 레볼루션: 현실과 메타버스를 넘나드는 새로운 경제생태계의 탄생』, 서울: 도서출판 길벗, 2021.
- 신승환, 「사이버네틱스(cybernetics) 시대와 철학의 자리」, 『가상과 실제 2: 4차 산업 시대의 인문학적 성찰』 (2023 서강대학교 신학연구소 학술대회 자료집, 2023년 9월 22일), 18-37.
- 신현탁, 「인공지능(AI)의 법인격: 전자인격(Electronic Person) 개념에 관한 소고」, 『인권과 정의』 478(2018), 45-64.
- 알텐리트, M., 『디지털 팩토리: 디지털 자본주의 시대, 보이지 않는 노동』, 권오성·오남규 옮김, 서울: 숨쉬는책공방, 2023.

- 윤현정·이진·윤희영, 「메타버스 개념과 유형에 관한 시론: 가능세계 이론을 중심으로」, 『인문콘텐츠』 62(2021), 57-81.
- 이도국, 「인공지능과 전자인(Electronic Person): 독자적 법인격 부여 가능성을 중심으로」, 『법과 정책연구』 21/1(2021), 439-461.
- 이시한, 『메타버스의 시대: 미래의 부와 기회를 선점하는 7대 메가트렌드』, 파주: 다산북스, 2021.
- 이진현, 「우리는 ‘곧’ 필요 없는 존재가 될 것인가?: 빌 조이의 ‘미래에 인간이 필요 없는 이유’에 대한 급진적 재고찰」, 『서강대학교 신학연구소 2022년 9월 학술대회 가상과 현실 자료집』, (2022년 9월 22-23일), 1-8.
- 정채연, 「법패러다임 변화의 관점에서 인공지능과 법담론: 법에서 탈근대성의 수용과 발전」, 『법과사회』 53(2016), 109-136.
- 조정호, 「인성人性和 물성物性에 대한 소고: 메타버스와 인성론을 중심으로」, 『인간·환경·미래』 27(2021), 95-116.
- 최형욱, 「블록체인 기반의 메타버스 가상경제」, 『철학과 현실』 130(2021), 117-130.
- 커즈와일, R., 『특이점이 온다: 기술이 인간을 초월하는 순간』, 김명남·장시형 옮김, 파주: 김영사, 2007.
- 크로퍼트, K., 『AI 지도책: 세계의 부와 권력을 재편하는 인공지능의 실체』, 노승영 옮김, 서울: 소소의 책, 2022.
- Artificial Intelligence Act. 출처: https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html (2023년 9월 10일 다운로드).
- Ethics Guidelines for Trustworthy Artificial Intelligence (AI), presented by the High-Level Expert Group on AI (AI HLEG) presented to the European Commission, in April 2019. 출처: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> (2023년 9월 10일 다운로드).
- European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics(2015/2103(INL). 출처: https://www.europarl.europa.eu/doceo/document/TA-9-2023-06-14_EN.html (2023년 9월 10일 다운로드).
- Habermas, J., *Erkenntnis und Interesse: Mit einem neuen Nachwort*, Frankfurt am Main: Suhrkamp, 1973.

- High-Level Expert Group on Artificial Intelligence, European Commission, A Definition of AI: Main Capabilities and Disciplines (2019). 출처: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf (2023년 9월 10일 다운로드).
- Horkheimer, M., *Eclipse of Reason*, New York: Oxford Univ. Press, 1947.
- Horkheimer. M.-Adono, Th., *Dialektik der Aufklärung: Philosophische Fragmente* (1947), Frankfurt am Main: Fischer, 1969.
- Kant, I., *Werke in Sechs Bänden, Bd. 4: Schriften zur Ethik und Religionsphilosophie; Grundlegung der Metaphysik der Sitten*, hg. v. W. Weischedel, Darmstadt: Wissenschaftl. Buchges., 1975.
- Karl Marx-Friedrich Engels Werke 1. Berlin: Dietz, 1981.
- Karl Marx-Friedrich Engels Werke 3. Berlin: Dietz, 1978.
- OPEN LETTER TO THE EUROPEAN COMMISSION ARTIFICIAL INTELLIGENCE AND ROBOTICS, 출처: <http://www.robotics-openletter.eu/> (2023년 9월 10일 다운로드).
- Rome Call for AI Ethics (Rome, February 28th, 2020), 출처: https://www.vatican.va/roman_curia/pontifical_academies/acdlife/documents/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf (2023년 9월 10일 다운로드).
- Scheler, M., *Die Stellung des Menschen in Kosmos*, 8. Aufl., Bern: Francke Verlag, 1975.
- _____, “Die Formen des Wissens und die Bildung”, in *Späte Schriften*, hg. von Manfred S. Frings, Bern: Francke Verlag, 1976.
- Smart, J.-Cascio, J.-Paffendorf, J., *Metaverse Roadmap: Pathways to the 3D Web; A Cross-Industry Public Foresight Project*, Los Gatos: Accerleration Studies Foundation, 2007, 출처: <https://www.w3.org/2008/WebVideo/Annotations/wiki/images/1/19/MetaverseRoadmapOverview.pdf> (2023년 9월 10일 다운로드).
- Stephenson, N., *Snow Crash*, New York: Bantam Books, c1992.

메타버스 시대의 기독교 윤리의 몇 가지 과제: 인공지능의 도전을 중심으로

강원돈

이 논문에서 필자는 메타버스 시대의 기독교 윤리를 본격적으로 논하기 전에 메타버스의 기술적 기반과 작동방식, 실제의 현실과 가상의 현실을 매끄럽게 통합하는 결정적 기제로서의 메타버스 경제를 예비적으로 분석하였다.

그러한 예비적 고찰을 한 뒤에, 필자는 메타버스의 기술적 기반 가운데 하나인 인공지능이 약인공지능 단계에서 ‘사람 사는 세상’에 던지는 도전을 분석하고, 그 도전에 대응하는 윤리적 지침을 제시하고자 했다. 약인공지능은 인간 기본권의 침해, 건강과 안전의 위협, 민주주의와 법치의 교란, 환경 위기의 악화 등 삶의 여러 차원에서 전례 없는 위협을 불러일으키고, 디지털 자본주의를 강화하고, 빅데이터를 통해 편견과 편향을 고착한다. 그러한 문제와 위협에 대응하기 위해 투명성, 포용성, 책임, 공정성, 신뢰성, 안전과 개인 정보 보호 등 인공지능의 위험을 관리하는 윤리적 지침을 제시하고, 디지털 자본주의를 제어하고, 디지털 지식 권력을 규제하는 방안을 제시하는 것이 우리 시대의 윤리적 과제가 되었다.

인공지능이 강인공지능 단계로 발전하게 된다면, 메타버스를 규율하는 방안은 달라질 수밖에 없다. 강인공지능은 자율적 판단 능력과 자율적 행위 능력을 갖추리라고 전망된다. 그러한 강인공지능은 아직 기술적으로 구현되지 않았지만, 그러한 강인공지능의 출현에 대비해서 규범공동체와 법률공동체를 어떻게 구성할 것인가를 준비할 필요가 있다. 자율적인 판단과 자주적인 행위 능력을 갖춘 인공지능은 행위 주체로 인정되고 ‘전자인’의 지위가 부여되어야 할 것이다. 그와 같이 인공지능의 행위 주체성이 인정된다면, 메타버스는 인간이 몸과 마음의 통일체로 현존하는 동시에 몸으로부터 분리된 디지털 휴먼(아바

타)으로 현존하는 세상이 될 것이고, 인간, 아바타, 인공지능 주체가 공존하는 세상이 될 것이다. 따라서 메타버스에서 세 가지 주체들의 자격을 규정하고 그 주체들의 상호관계를 설정하는 것은, 메타버스에서 혼란과 혼돈을 피하고 메타버스를 제도적으로 규율해서 더 많은 선과 더 많은 정의를 실현하려는 윤리적 기획에 필요하다.

메타버스 시대에 교회와 신학은 인간의 확장된 현실인 메타버스가 하나님의 통치 아래 놓인다는 점을 분명히 해야 한다. 교회와 신학은 하나님의 디지털 이미지와 참 하나님을 구별하고, 메타버스 경제의 도전에 기독교 윤리적으로 대응하고, 행위 주체에 관한 신학 이론을 확장하고 재구성할 필요가 있다.

주제어: 메타버스, 인공지능, 인공지능에 관한 윤리적 지침, 전자 인간, 정신

Ethics in the Age of the Metaverse: Centering on the Expansion and Reconstruction of Person as Acting Subject

Kang, Won-Don

In this paper, I examined the metaverse at two aspects before moving on to discuss Christian ethics in the age of the metaverse. First, I investigated what the technical bases of the metaverse are, and how the metaverse works. Second, I paid attention to how the metaverse economy seamlessly integrates actual and virtual reality.

After these preliminary considerations, I analyzed the challenges that artificial intelligence (AI), one of the technological foundations of the metaverse, poses to the human world at the level of weak intelligence, and proposed ethical guidelines for responding to these challenges. Weak AI poses unprecedented risks in many dimensions of life, including violations of basic human rights, threats to health and safety, disruption of democracy and the rule of law, and exacerbation of ecological crises, while strengthening digital capitalism and perpetuating bias and prejudice through big data. In response to these problems and risks, Christian ethics must fulfill the following three tasks: One is to provide ethical guidelines for managing the risks of AI, including transparency, inclusion, accountability, fairness, trustworthiness, safety, and privacy. Another is to control digital capitalism, and the last is to develop measures to regulate digital knowledge power.

If AI develops to the stage of strong AI, the way to regulate the metaverse must be changed. Strong AI is considered to have the ability to make autonomous judgments and act autonomously. Although such AI is not yet technically possible, it is

necessary to prepare for the emergence of such AI and how to organize the normative and legal communities. In the framework of ethics and law since Kant, in which autonomy is regarded as the essence of person as actor with rights and duties, AI with the ability to judge and act autonomously can be constructed as person named ‘electronic person’. The metaverse in which such an AI operates is a world in which humans exist as a unity of body and mind and at the same time as digital humans (avatars) separated from the body, and human, avatar, and AI subjects coexist. Therefore, defining the qualifications of the three subjects in the metaverse and establishing their interrelationships is necessary for ethical project to avoid chaos and confusion in the metaverse and to institutionally regulate the metaverse to realize more good and more justice.

In the age of the metaverse, church and theology must make clear that the metaverse, the extended reality of humanity, is subject to the rule of God. Church and theology need to distinguish between digital images of God and the true God, respond to the challenges of the metaverse economy with Christian ethics, and expand and reconfigure theological theories of person.

Key Words: artificial intelligence (AI), electronic person, ethical guidelines for AI, mind, metaverse

논문 투고일	2023년 11월 5일
논문 수정일	2023년 12월 3일
논문게재 확정일	2023년 11월 27일
