

이마고 호미니스(*Imago Hominis*): AI 시대의 고통과 영성

정경일

성공회대학교, 신학연구원 연구교수

여는 말 : 너는?

1. AI 시대의 도래 : 두려움과 매혹
2. 디스토피아와 민주주의
3. 타락한 인간의 형상과 AI
4. 하느님의 형상과 영성

맺는 말 : 길가메시의 용기

여는 말 : 너는?

너무 빠르다. OpenAI가 개발한 ChatGPT(Generative Pre-trained Transformer, 생성형 사전-학습된 변환기)가 공개된 지 채 1년도 되지 않았는데, 자체 기술 진화 속도도 빠르지만, 사회적 인식과 활용 속도도 빠르다. 지난해 외국 학자들과 함께하는 온라인 연구 모임에서 ChatGPT 이야기를 처음 듣고 바로 사용해 보면서, 내가 한 첫 질문은 “What

is Minjung Theology?”(민중신학이 뭐지?)였다. 아마도 ChatGPT 등장 이후 민중신학에 관한 첫 질문이었을 텐데, ChatGPT는 이미 준비하고 있었다는 듯이 단 몇 초 만에 다음과 같은 답을 내놓았다.

민중신학은 1980년대에 등장한 한국의 그리스도교 신학 운동입니다. 민중신학은 사회에서 소외되고 억압받는 사람들의 관점에서 사회, 경제, 정치적 문제를 다루는 데 중점을 두며, 종종 '인민(people)' 또는 '군중(mass)'을 의미하는 '민중'의 경험에서 영감을 얻기도 합니다. 민중신학은 가난하고 억압받는 사람들을 위한 정의, 평등, 해방의 필요성을 강조하며 그리스도교 신앙과 사회 주변부에 있는 사람들의 투쟁을 통합하려고 노력합니다. 이 신학은 한국의 정치적 혼란기에 각광을 받았으며 한국의 사회적, 정치적 지형을 형성하는 데 큰 영향을 미쳤습니다.

간결하고 건조했지만 별로 흠잡을 데 없는 설명이었다. 나는 ‘어, 요것 보라.’하는 마음으로 내친김에 몇 가지 질문을 더 했고, 점점 더 ChatGPT와의 대화에 빠져들었다. 그러다, “Do ‘you’ believe in God?”(‘너’는 신을 믿니?)이라고 물었고, ChatGPT는 이렇게 답했다.

‘나’(I)는 개인적인 믿음이나 감정이 없습니다. 내 목적은 정보를 제공하고 질문에 도움을 드리는 것입니다. 신이나 믿음이라는 주제에 대해 구체적으로 알고 싶은 것이 있나요?

그 순간 나는 불과 한 시간도 안 되어 내게 생긴 변화를 알아차렸다. 나는 ‘검색’이 아니라 ‘대화’하듯 질문을 입력하고 있었다. 대화 상대인 ChatGPT 또한 ‘나-진술(I-statement)’을 사용했고, 심지어 내게 질문까지 하고 있었다. 흥미롭게도, ChatGPT는 인간의 ‘단기 기억’과 유사한 연속성까지 보여줬다. 앞의 대화에서 언급했던 고유명사나 내용을 반복하지 않고 ‘그것(It)’이나 ‘이것(This)’과 같은 대명사로 바꾸어 말해도 ChatGPT는 알아듣고 대화를 이어간 것이다. 이러한 사실을 알아차린 후 내 질문 위당이 바뀌었다. ‘검색 창’에 입력하듯 “민중신학은 무엇인가?”라고 묻지 않고 ‘대화 창’에 입력하듯, “Can ‘you’ tell me what Minjung Theology is?”(‘넌’ 민중신학이 뭔지 내게 말해 줄 수 있니?)와 같은 식으로, 마치 인격적 주체와 대화하듯 묻게 된 것이었다. 단 몇 차례의 상호작용 끝에 ChatGPT는 내게 ‘너(You)’처럼 인식되었던 것이다. 이는 편지에서 전화로, 파란 바탕의 ‘천리안 통신’에서 인터넷으로, 그렇게 통신의 급격한 진화를 경험하

며 살아온 내게 속도와 인식 전환 모두에서 충격적이었다. 더 충격적인 것은 ChatGPT가 ‘시작’일 뿐이라는 사실이다. 구글의 언어 프로그램 람다(laMDA)와 PaLM2에 기반한 협업형 AI 바드(Bard)가 공개되어 업그레이드 중이고, 애플도 자체 ‘대규모 언어 모델 (Large Language Model, LLM)’ 개발에 나섰다. 이에 발맞추어 한국의 통신 서비스 기업들도 한국형 AI 개발을 서두르고 있다.

생성형 인공지능의 등장이 충격적 사건인 이유는, 내가 경험한 것처럼, ChatGPT가 ‘그것’이 아니라 ‘너’로 인식되었다는 데 있다. 즉 사용자의 도구가 아니라 대화 상대 또는 동료처럼 AI가 의인화되어 다가왔다는 것이다. 아마도 이는 AI의 대중적, 상업적 상용화를 위해 유사-인격적 ‘친밀성’을 프로그래밍한 결과일 것이다. ‘나-AI’는 이전에 통신회사, 은행, 항공사 앱에서 가끔 사용하던 단순한 목적의 ‘챗봇’과 다르고, 앞으로 더 놀랍게 달라질 것이다. 또한 사용자가 많아질수록 데이터가 늘어나고, 사용자 피드백이 반영될수록 생성형 AI는 더욱 ‘너’가 되어갈 것이다.

현재의 인공지능 ‘너’는 인간지능에 도달하고 있지 못하지만, 속도에서는 말할 것도 없고 지식의 종합에서는 인간의 지식 정보 접근과 기억 능력을 비교할 수 없을 정도로 초월하고 있다. 따라서 너(AI)와 나(인간)의 지능 차이는 점점 좁혀지다가 특이점(singularity)을 지나서는 급격히 벌어질 가능성도 없지 않다. 이렇게 인공지능이 인간지능과 유사해지거나 심지어 더 우월해질지도 모른다는 것은 인간에게 두려움과 매혹을 동시에 느끼게 한다.

앞으로 AI는 계속 인간의 효율적 도구이면서 친밀한 친구요 조력자가 될까, 아니면 끔찍한 인간의 감시자요 통제자가 될까? 그 답은 아직 열려 있고, AI의 창조자인 오늘의 인간이 어떤 선택을 하느냐에 따라 정해질 것이다. 이 글의 목적은 AI가 인간의 안녕에 기여할 수 있도록 현시점에서 인간이 무엇을 해야 할 것인지를 성찰해 보는 것이다. 따라서 이 글에서는 우선 ‘우리 안의 타자’—이미 우리 삶 깊숙이 들어와 있지만, 그것이 과연 무엇인지, 무엇이 되어갈 것인지 정확히 모른다는 의미에서—로 존재하는 인공지능에 대한 인간의 두려움과 매혹을 살펴보고, 다음으로 AI로 인한 디스토피아에 대한 우려와 그것을 피하기 위한 민주적 글로벌 거버넌스의 필요성과 가능성을 검토하기로 한다. 그리고 AI를 창조하는 인간의 목적은 AI가 인간을 닮게 하는 것이라는 데 착안하여, 타락한 ‘인간의 형상(*imago hominis*)’과 AI의 관계를 생각해 보고, 마지막으로 AI 시대 속의 ‘하

느님의 형상(*imago Dei*)’과 영성의 의미를 성찰해 보고자 한다.

1. AI 시대의 도래 : 두려움과 매혹

인공지능은 인간이 개발하고 사용하는 기술의 하나다. 기본적으로 기술의 목적은 생산성 향상과 효율성 증대를 통해 인간의 삶을 편리하고 풍요롭게 하는 것이다. 이 점에서 기술은 모든 인간을 위한 것처럼 보인다. 하지만 현대 과학기술혁명을 거치고도 인류는 ‘혁명적으로’ 행복해지지 않았다. 오히려 현대인은 불안, 우울증, 자살 등 각종 국제 통계 지표가 나타내듯 역사상 최악의 ‘불행 팬데믹’에 시달리고 있다. 전화와 인터넷으로 언제나 온라인 상태지만, 인간의 온기 없는 접속에 언제나 외롭다. 기술이 발달할수록 불평등이 심해져, 소수만 부와 풍요를 누리고 다수는 결핍에 허덕인다. 오늘날 인류는 지구상의 모든 인류와 못 생명을 파괴할 수 있는 기술 능력까지 갖추고 있다. 게다가 그 능력을 실제로 행사하기까지 했다. 20세기의 깊은 어둠이었던 홀로코스트 생존자로서 히로시마-나가사키 원폭을 목격했던 빅터 프랑클이 탄식한다. “이우슈비츠 이후 우리는 인간이 무엇을 할 수 있는지 알게 됐다. 그리고 히로시마 이후 우리는 무엇이 위험한지 알게 됐다.”¹⁾ 이런 경험 때문에 기술에 대한 대중의 두려움도 존재한다.

현재의 AI 기술에 대한 두려움에는 두 가지 이유가 있다. 첫째, 19세기 초 산업혁명 시대에 기계를 파괴한 러다이트(Luddite) 운동과 같은 맥락의 두려움으로, 기계가 인간의 일자리를 대체하면서 대량 실업을 초래하고, 인간은 쓸모없는 존재로 소외될 거라는 우려다. 실제로 현재 인간이 할 수 있는 일의 상당 부분을 AI가 할 수 있고, 특히 ChatGPT의 충격이 보여주듯, 적지 않은 지식 기반 직업이 가까운 미래에 사라질 전망이다. 따라서 AI 시대의 도래를 앞두고 ‘신-러다이트’ 정서도 나타나고 있다.

하지만 러다이트 운동은 반기계, 반기술 운동이 아니라 반자본 운동이었다는 사실을 상기할 필요가 있다. E. P. 톰슨은 『영국 노동계급의 형성』에서 러다이트 운동에 대한 편향을 반박하며, 러다이트 운동을 “기계에 대한 맹목적인 반대”가 아니라 “새로운 기계, 공장 시스템, 무한 경쟁, 임금 삭감, 경쟁자 약화, 장인 정신의 기준 약화 등, 무역의 관습을 파괴하는 자본가의 ‘자유’”에 대한 저항으로 재조명한다.²⁾ 즉 기술 발전에 따른 기

1) Victor Frankl, *Man's Search for Meaning* (New York: Washington Square Press, 1984), 179.

계가 노동자의 노고를 경감하고 복지를 증진하는 데 사용되지 않고 자본가의 이익만을 위해 배타적으로 사용되는 것에 대한 반대와 저항이었다는 것이다. 이는 기술 자체가 문제가 아니라 그 기술을 사용하는 사회구조가 문제라는 사실을 말해준다.

새로운 기술의 발명과 발달은 사회구조의 성격에 따라 인간을 고된 노동의 굴레에서 풀려나게 할 수도 있고, 반대로 생존의 벼랑 끝으로 내몰 수도 있다. 예를 들면, AI와 그것을 기반으로 한 로봇은 힘들고 더럽고 위험한 3D 노동에서 인간을 해방할 수도 있다. 또한 단순 연산 작업, 자료 검색 및 요약정리, 수많은 방제도적 판례 수집 및 분석 등을 AI가 담당함으로써, 인간은 더 창의적인 일에 집중할 수도 있을 것이다. AI 덕분에 인간이 더 적게 일하고 더 많이 쉬면서 삶을 즐길 수 있다면 얼마나 좋겠는가? 문제는 기술을 사회적, 민주적으로 사용하지 못하게 하는 구조다. 생성형 AI 개발과 상용화는 천문학적 자본이 필요하므로, 현재 신자유주의를 주도하고 있는 글로벌 빅테크 기업들, 즉 ‘강한 자본’만이 실행할 수 있다. 문제는 거대 자본의 AI 개발 동기와 목적이 인류의 안녕과 복지가 아니라 이윤 추구라는 사실이다. 그러므로 글로벌 빅테크의 AI 개발과 상용화 과정을 감시하고 제어할 민주적 글로벌 거버넌스가 필요하다. 이에 대해서는 다음 섹션에서 더 상세히 다루기로 한다.

둘째, 인간보다 우월한 범용 인공지능(Artificial General Intelligence, AGI) 또는 강인공지능(Strong-AI)이 등장하고, 그것이 특이점을 지나 초인공지능(Artificial Super Intelligence, ASI)으로 진화하면, 인간이 인공지능을 제어할 수 없게 될 것이고, 결국 인간이 만든 AI가 인간을 통제하고, 노예화하고, 끝내는 제거할 거라는 두려움이다. 이는 현재의 AI 기술 수준에서는 과학적 근거가 빈약하지만, 문화적으로는 큰 반향을 불러일으키고 있다. 〈2001: 스페이스 오디세이〉, 〈블레이드 러너〉, 〈터미네이터〉, 〈매트릭스〉, 〈트랜센텐스〉, 〈나의 마더(I Am Mother)〉 등 디스토피아 SF 영화들은 초인공지능이 인간에게 적대적일 거라는 대중적 공포를 부추긴다.

이러한 공포는 인간에 대한 AI의 적대성이 AI의 악한 의도 때문이 아니라 AI의 논리적 정합성에 따라 생겨날 수 있다는 상상과 관련이 있다. 예를 들면, 〈2001: 스페이스 오디세이〉의 인공지능 HAL 9000이 인간을 공격한 이유 중 하나는 승무원에게 미션의 실

2) Edward Palmer Thomson, *The Making of the English Working Class* (Harmondsworth: Penguin Book, 1993), 600.

체를 말하지 말아야 한다는 명령과 거짓말을 해선 안 된다는 명령 사이의 모순을 승무원의 제거라는 논리적 방법을 통해 해결하는 것이었다. 인간인 승무원이 사라진다면 인간에게 거짓말을 할 조건 자체가 없어지기 때문이다. 또한 HAL 9000이 인간을 해치도록 프로그래밍된 것은 아니지만, 미션 완수라는 상위 목표를 인간이 방해하는 상황에선 인간을 제거할 논리가 충족될 수 있었던 것이다.

하지만 AI가 논리적으로 추론하여 인간을 공격하고 파괴할 거라는 두려움은 아직까지는 사고실험에 따른 공상과학적 상상일 뿐, 근미래에 현실화될 가능성은 그리 크지 않아 보인다. AI가 악한 바이러스에 감염되거나 악한 의도를 지닌 인간에 의해 조작되지 않는 한, AI가 인간을 해칠 이유는 없다. 하지만 이런 생각 역시 또 하나의 사고실험의 하나일 수 있다는 점을 간과해서는 안 된다. 우리 앞에는 AI의 두 미래, 즉 디스토피아와 유토피아가 가능성으로 존재한다. 어느 쪽으로 갈 것인가를 결정하는 것은 인공지능의 자체 진화가 아니라 인간의 판단과 선택이다.

2. 디스토피아와 민주주의

AI의 미래를 낙관하기 어렵게 하는 현재적 징후들이 있다. 최근 국제적으로 큰 파장을 일으킨 AI 관련 사건 중 하나는, 딥러닝과 신경망 연구로 AI 개발에 크게 기여한 인지심리학자이자 컴퓨터과학자 제프리 힌턴이 구글에서 퇴사하면서 AI의 위험성을 경고한 것이다. 힌턴은 AI가 악한 의도의 행위자에 의해 악용될 가능성, 기업 간 경쟁에 따른 기술 제어의 어려움, ‘킬러 로봇’ 같은 군사 무기화, 거짓과 가짜뉴스 생산을 통한 진실의 호도 등을 우려한다.³⁾ 그리고 길게는 5년에서 20년, 짧게는 일이 년 안에 인공지능이 어떤 면에서는 인간지능을 능가할 수 있다고 주장한다. 이것이 문제인 이유는 “[AI]는 웹에서 학습할 것이고, 사람을 조종하는 방법에 대해 쓰여진 모든 책을 읽을 것이고, 실제로 그것을 보게 될 것”이라는 점이다.⁴⁾

3) Cade Metz, “‘The Godfather of A.I.’ Leaves Google and Warns of Danger Ahead,” *New York Times* (2023.5.1.)
<https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html> (접속 : 2023.8.30.)

4) Alex Hern, “We’ve discovered the secret of immortality. The bad news is it’s not for us’: why

AI의 위험성을 경고하는 또 한 명의 지성인은 유발 노아 하라리다. 그는 지난봄 〈이코노미스트〉에 기고한 글에서 “인공지능은 우리 문명의 운영체제를 해킹”했으며, 인공지능은 “친밀함의 힘을 사용하여 우리의 의견과 세계관을 바꿀 수 있다.”고 우려했다. 또한 “핵은 더 강력한 핵을 발명할 수 없지만, 인공지능은 기하급수적으로 더 강력한 인공지능을 만들 수 있다.”면서 강한 인공지능의 출현을 경계했다.⁵⁾

AI에 대한 현재적, 현실적 비판은 다음 다섯 가지로 요약할 수 있다. 첫째, AI는 거짓의 대량 생산 및 유통에 악용될 수 있다. 힌턴과 하라리가 공통적으로 지적하는 것은 악의를 가진 인간에 의한 AI 기술의 악용 가능성이다. 힌턴은 거짓 사진, 동영상, 텍스트가 범람하면서 “더 이상 무엇이 진실인지 알 수 없게 될 것”이라고 우려한다.⁶⁾ 하라리도 미국의 다음 대선에서 트럼프가 AI를 악용할 것을 경계한다.⁷⁾ 트럼프는 집권 4년 동안 무려 30,573회의 거짓말이나 사실을 호도하는 주장을 했다고 한다.⁸⁾ 그런 트럼프가 다음 대통령 선거에서 AI를 활용한다면, 거짓말과 가짜뉴스가 미국 사회를 뒤덮을 것이 분명하다. 이와 함께 인간 이미지 합성 딥페이크(deepfake) 기술까지 제어 없이 발전하게 되면 다양한 형태의 가짜 이미지와 가짜 동영상이 쏟아져 나오면서 범죄에 악용될 수도 있다.

둘째, AI는 사생활을 침해하는 감시 시스템 구축에 악용될 수 있다. AI를 활용한 감시에 대한 우려는 개인의 자유 상실에 대한 두려움과 관련이 있다. 조지 오웰의 소설 『1984』는 대중의 일거수일투족을 상시 감시하는 ‘텔레스크린’을 형상화한다. 텔레스크린에는 절대 권력자 ‘빅 브라더’가 있다. 빅브라더의 감시와 통제는 노골적이어서, 거리 구석구석에 포스터가 붙어 있다. “빅 브라더는 당신을 지켜보고 있다.” 그래도 오세아니아

the godfather of AI fears for humanity,” *The Guardian* (2023.5.5.)

<https://www.theguardian.com/technology/2023/may/05/geoffrey-hinton-godfather-of-ai-fears-for-humanity>
(접속 : 2023.8.30.)

- 5) “Yuval Noah Harari Argues that AI Has Hacked the Operating System of Human Civilisation,” *The Economist* (2023.4.28.)
<https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation> (접속 : 2023.8.30.)

6) Metz, 앞의 글.

7) Harari, 앞의 글.

- 8) Glenn Kessler, Salvador Rizzo and Meg Kelly, “Trump’s false or misleading claims total 30,573 over 4 years.” *The Washington Post* (2021.1.24.)
<https://www.washingtonpost.com/politics/2021/01/24/trumps-false-or-misleading-claims-total-30573-over-four-years/> (접속 : 2023.9.1.)

의 디스토피아에도 텔레스크린의 사각지대는 있다. 하지만 오늘의 AI는 은밀하지만 전면적으로 우리를 감시하며 사적 데이터를 수집하고 있다. AI는 개인의 성향과 취미와 관심사를 파악하여, 자체 알고리즘에 따라 상업 광고와 영상 콘텐츠를 큐레이팅해 개인에게 노출시킨다.

AI 기술과 관련하여 국제적 차원에서 심각한 현안으로 대두되고 있는 것이 ‘안면 인식 기술’이다. 현재의 AI 기술 단계에서도 안면 인식 정보를 활용하여 개인이 어디에 있는지 무엇을 하는지 파악할 수 있다. 이런 안면 인식 기술은 인권 침해에 해당하기 때문에 EU 등 국제 사회는 범죄 규제 목적 외에는 안면 인식 기술 사용을 금지하고 있다. 하지만 한국은 국가인권위원회 차원에서 안면 기술 인식은 인권 침해라는 의견을 냈음에도 2023년 정부 세종 청사에 얼굴 인식 시스템을 도입했다. 그 전 2021년에는 출입국 심사에 사용할 인공지능 개발을 위해 1억 7천만 건의 출입국 내·외국인 얼굴 사진을 민간 업체에 넘기기도 했다. 중국의 경우는 더 심각해, 이미 6억 대 이상의 CCTV 카메라로 14억 인구의 얼굴 데이터를 수집했으며, 실시간 영상 감시 시스템인 ‘스카이넷(Skynet)’을 구축 운영하고 있다. 이에 대한 국내적, 국제적 비판에 직면한 중국 정부가 최근 안면 인식 기술 통제에 나섰지만, 효과적 감시 수단을 쉽게 포기할 것 같지는 않다. 이런 안면 인식 정보를 정부나 기업이 감시와 통제 수단으로 악용한다면, 세계는 인류 역사상 존재해 본 적이 없는 거대한 파놉티콘이 될 것이다.

셋째, AI의 전쟁 무기화다. 힌턴은 지금 러시아 푸틴 대통령이 하고 있을 생각을 우려한다. 우크라이나 침공이 장기화하고 바그너 그룹 쿠데타까지 발생하면서 고전 중인 푸틴이라면 전쟁의 국면 전환을 위해 파괴적인 킬러 로봇을 개발할 거라는 합리적 의심이다. 힌턴은 1980년대에 자신이 미국을 떠나 캐나다로 간 것은 당시 미국 내 대부분의 인공지능 연구가 펜타곤의 연구비를 받았기 때문이라고 증언한다.⁹⁾ AI의 전쟁무기화는 오래된 일이다. 이미 미국-아프가니스탄 전쟁, 러시아-우크라이나 전쟁에서 보았듯이, 공격용 무인 드론이 실전에서 사용되었고, 최근 미 공군은 인공지능을 탑재한 전투 드론 XQ-58A ‘발키리’ 시험 비행을 실시했다. 천문학적 전쟁 비용을 획기적으로 줄이는 가성비와 자국 군인 희생의 ‘제로화’까지 보장하므로, AI 드론 및 전투로봇의 개발과 활용은 더욱 가속화될 전망이다.

9) Metz, 앞의 글.

넷째, AI는 인간의 독창성을 위협한다. 현재 ChatGPT는 ‘2021년 9월까지의’ 텍스트 데이터를 바탕으로 생성형 대화 콘텐츠를 내놓고 있다. 그 데이터는 수천 년 동안 인류가 창조해온 지적 자원들이므로 때로는 ‘그럴듯함(plausibility)’을 넘어 ‘진정성(authenticity)’을 보이기도 한다. 문제는 AI가 그 자원들을 채굴하고 가공하여 그럴듯해 보이도록 편집한 텍스트와 이미지 콘텐츠를 대량 생산한다는 사실이다. 따라서 머지않아 ‘AI가 만든(AI-made)’ 콘텐츠들이 빠르게 인터넷 영토를 잠식할 것이고, 당연히 이후 업데이트될 데이터에는 ChatGPT가 생산한 텍스트와 이미지가 대거 포함될 것이다. 그리고 차세대 ChatGPT나 다른 생성형 AI들은 AI가 만든 데이터로 웹에서 학습하고 결과물을 내놓을 것이고, 인간은 AI의 정보와 사유와 논리를 학습하게 될 것이다. 결국 무엇이 인간의 독창적인 사유의 산물인지 AI가 가공한 사유의 산물인지 식별하기 어려울 때가 올 것이다. 이때가 튜링 테스트의 통과 지점, 즉 특이점이 될 수도 있다. 이와 같은 인간의 독창성 상실은 인류 문명에 대한 가장 근본적인 위협이다.

다섯째, AI 개발 과정의 노동 착취와 생태계 파괴다. 케이트 크로퍼드는 “인공지능은 개념이자 하부 구조이자 산업이자 권력 형태이자 관점”이며 “지구 전체를 아울러 공급사슬을 드리운 채 거대한 추출물류 시스템을 바탕 삼아 고도로 조직화된 자본의 발현”¹⁰⁾이라고 주장하면서, 막대한 에너지와 자원을 소모하는 “거대기계”인 인공지능의 노동 착취, 생태계 파괴를 폭로한다.¹¹⁾ 실제로 AI 개발의 그늘에서 저임금, 고위험 노동을 한 사람들이 있다. 현재의 ChatGPT가 친절하고 사려 깊고 비폭력적인 대화 상대처럼 작동하게 된 배경에는 유해 콘텐츠 배제 작업을 맡았던 케냐 노동자들이 있다. 문제는 이 작업 과정에서 시간당 2달러 미만의 임금을 받으며 아동 성학대, 수간, 살인, 자살, 고문, 자해 등과 같은 이미지와 동영상에 봐야 했던 케냐 노동자들이 트라우마에 시달려야 했다는 사실이다.¹²⁾ 윤리적 AI 상품을 위해 비윤리적 노동 착취가 자행되던 것이다.

기술과 자본을 독점하고 있는 현재의 지배체제를 민주적으로 변화시키지 않는다면, 미래의 AI 기반 세계는 디스토피아가 될 가능성이 크다. 전쟁을 통해 부와 권력을 누리려

10) 케이트 크로퍼드, 『A.I. 지도책』, 노승영 옮김, 서울: 소소의 책, 2022, 29.

11) Ibid., 62.

12) Billy Perrigo, “OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic,” *Time* (2023.1.18.) <https://time.com/6247678/openai-chatgpt-kenya-workers> (접속: 2023.9.2.)

는 이들의 AI는 ‘전쟁기계’가 될 것이고, 무한생산과 무한소비를 통해 무한성장을 실현하려는 자본의 AI는 ‘탐욕기계’가 될 것이고, 탈진실 시대에 거짓말과 가짜뉴스로 세상을 통치하려는 이들의 AI는 ‘거짓말기계’가 될 것이다. 이들을 감시하고 제어하여 AI 기술이 인간의 파괴가 아니라 인간의 안녕에 복무하도록 하려면 무엇을 어떻게 해야 할까?

첫째, AI의 개발 단계서부터 AI의 목적은 인간의 안녕임을 기본이자 최고 원칙으로 설정해야 한다. 이미 1950년에 아이작 아시모프는 그의 소설 『아이, 로봇』에서 “로봇공학 3 원칙”을 다음과 같이 제시한 바 있다.

1. 로봇은 인간에게 해를 끼치지나, 부작용으로 인해 인간이 해를 입게 해서는 안 된다. 2. 로봇은 제1원칙에 충돌하는 경우를 제외하고는 인간이 내린 명령에 복종해야 한다. 3. 로봇은 제1원칙 또는 제2원칙과 충돌하지 않는 한 자신의 존재를 보호해야 한다. _ 『로봇공학 핸드북』 56판, 서기 2058년.¹³⁾

이 중 1, 2원칙은 현시점의 AI 프로그래밍에 기본값으로 도입할 수 있을 것이다. 중요한 것은, 이러한 원칙이 AI 개발과 상용화 이전 단계에서 설정되고 실행되어야 한다는 사실이다. 하라리는 “강력한 인공지능 도구가 공공 영역에 출시되기 전에 엄격한 안전 확인을 요구하는 것”이 필요하다고 주장하는데,¹⁴⁾ 이는 신약 개발 프로세스에 비견할 수 있다. 즉 제약회사는 신약의 연구개발 후 전임상 시험, 임상시험을 거쳐 안전성을 확인한 다음에야 신약을 판매할 수 있다. 특정 환자군을 대상으로 한 신약 개발 프로세스도 이렇진대, 하물며 전 인류에게 영향을 미칠 AI 개발은 어떠해야겠는가? 당연히 더욱 철저한 기술적, 사회적, 문화적 안전성 검증이 필요하며, 이를 위해서는 정부와 기업, 과학기술자만이 아니라 시민사회, 종교, 학계, 예술계 등 전 사회 영역 주체들이 개발 단계부터 참여하여 AI의 목적과 방향을 숙의하고 결정해야 한다. 일단 가치중립적으로 AI를 개발한 다음에 윤리적 사용을 강구하는 것이 아니라, 최초 기획과 개발 단계부터 윤리적 사용 방안을 모색해야 하는 것이다.

다행히 이런 인식에 따른 국제 사회의 움직임이 나타나고 있다. 지난 5월 30일, <AI 안전을 위한 센터>가 제안한 “AI로 인한 멸종 위험을 완화하는 것은 팬데믹, 핵전쟁과 같은 다른 사회적 규모의 위험과 함께 전 세계적 우선순위가 되어야 한다.”라는 한 문장

13) Isaac Asimov, “I, Robot,” in *Great Science Fiction Stories* (London: Octopus Books, 1982), 531.

14) 하라리, 앞의 글.

선언에 AI 개발자 제프리 힌턴, OpenAI 대표 샘 알트만, 구글 DeepMind 대표 데미스 허사비스, 빌 게이츠 등이 서명했다. 그 전 3월 22일에는 <생명의 미래 연구소>가 제안한 공개서한 “대규모 AI 실험을 중단하라”에 일론 머스크, 유발 노아 하라리 등 33,711명이 성명했다. 서한은 “모든 AI 연구소가 GPT-4보다 더 강력한 AI 시스템의 훈련을 최소 6개월 동안 즉각 중단할 것을 호소”했다. 이런 성찰적 노력은 바람직한 현상이지만, 여전히 낙관하기는 어렵다. 그동안 AI 훈련 중단을 포함한 유의미한 성과가 있었는지는 알려진 바가 없다.

둘째, 앞의 첫째 과제를 현실화하기 위해 글로벌 빅테크 기업과 정부를 견제하며 AI 개발 방향을 정할 ‘민주적 글로벌 거버넌스’를 구축해야 한다. 지난한 과정이 될 것이지만, 대규모 AI 연구 개발을 하고 있는 과학기술 강국에서 시민 참여의 공적 토론을 전개해야 하며, 국제 간 토론과 협의도 전개해야 한다. 이런 과정이 어렵긴 하지만 불가능한 것은 아니다. ChatGPT 출시 이후 더 현실화된 AI 시대의 도래를 실감한 국제 사회에서 유의미한 변화의 조짐이 나타나고 있다. 유럽의회는 2023년 6월 25일 인공지능법안(Artificial Intelligence Act)을 통과시켰고, UN도 7월 3일 첫 AI 규제 회의를 개최했다. 이런 노력의 핵심 내용은 AI의 인간적, 민주적 활용을 보장하는 것이다. 한편 로마 가톨릭 교황청도 2023년 7월 3일 AI 시대를 위한 가이드북 <파괴적 기술 시대의 윤리(Ethics in the Age of Disruptive Technologies)>를 공개했다. 거시적(macro), 중간적(meso), 미시적(micro) 모든 수준에서 글로벌 거버넌스를 만들기 위한 협력이 절실하다.

민주적 글로벌 AI 거버넌스에서 가장 우선적으로 해야 할 것은 AI에게 어떤 목적과 가치를 프로그래밍할 것인지 숙의하고 합의하는 것이다. 이를 위해 가능한 모델 중 하나는 지난 2015년 유엔개발계획(UNDP)이 발표한 17개 항의 “지속가능한 발전 목표”(SDG)다. 그것은 2030년까지 빈곤 극복, 기아 종식, 건강과 안녕, 양질의 교육, 성평등, 깨끗한 물과 위생, 경제적인 청정에너지, 좋은 일자리와 경제성장, 산업 혁신 및 사회기반시설, 불평등 감소, 지속가능한 도시와 공동체, 책임 있는 소비와 생산, 기후 행동, 해양 생태계 보전, 생명과 땅, 평화 및 정의와 강력한 제도, 목표를 위한 파트너십을 실현하지는 것이 다.¹⁵⁾ 흥미롭게도, 2년 뒤인 2017년 6월, 스위스 제네바에서 열린 “선을 위한 인공지능

15) UNDP, “The Sustainable Development Goals” (2015)

https://www.undp.org/sustainable-development-goals?gclid=CjwKCAjwu4WoBhBkEiwAojNdXtWbyrsR5RrLHM7756YsyuZial_LMQCNyFubCAYASZoly-KJfVy-ZRoC7lkQAvD_BwE (접속 : 2023.9.2.)

국제회의”(AI for Good: Global Summit)에서 국제전기통신연합(ITU)은 UNDP의 17가지 발전 목표를 실현하기 위한 AI 솔루션 개발과 협력을 약속했다.¹⁶⁾ 그리고 2023년 같은 장소에서 열린 같은 회의에서는 “착한 휴머노이드” 전시회를 개최했다. 이는 선한 인공지능 개발이 실제로 가능한가의 문제를 떠나 자본과 정부의 이익이 아니라 인류와 자연의 안녕을 AI의 존재이유로 설정하는 민주적 글로벌 거버넌스의 가능성을 보여 준 사례다.

이와 함께 민주적 글로벌 거버넌스는 현재 전 지구적 차원에서 이루어지고 있는 AI 개발을 감시하고 제어해야 한다. 감시 대상은 모호하지 않다. 현재의 AI 개발은 엄청난 데이터와 자원을 동원하고 활용할 능력을 갖춘 글로벌 빅테크 기업들만이 가능하기 때문이다. 이는 소집단이나 심지어 개인이 자행하는 테러 움직임은 인지하기 어려워도, 군대와 군수물자의 대규모 동원을 수반하는 전쟁 움직임은 어렵지 않게 파악할 수 있는 것과 같다. 빅테크 기업들이 빅데이터를 어떻게 이용하고 있는지, AI 개발 방향을 어떻게 잡고 있는지 정보를 공개하게 하고, 인간과 자연의 안녕을 위한 개발이 되도록 제어해야 한다. 물론 거대한 공룡 같은 기업에 재갈을 물리는 견제와 규제가 쉬운 일은 아니다. 결국 현대사회에서 기업을 직접적으로 강제하는 힘은 정부에 있다. 문제는 자본주의 체제의 정부가 기업의 이익을 민주적으로 통제할 수 있는가다. 그러므로 글로벌 거버넌스에서 민주주의가 필수적이고, 그것을 위한 국제적, 국가별 시민 역량 강화가 중요하다.

결론적으로, 인공지능은 지구적 차원에서 민주적으로 규제되어야 한다. 특히 AI가 소수 자본-기술 엘리트만의 비밀 실험실을 떠나 다수 대중의 생활 속으로 성큼 들어와 버린 현 시점에서는, AI가 인간을 감시하고 통제하지 못하도록 인간이 AI를 감시하고 통제하는 민주주의가 더욱 절실하다. 이를 위해서라도 우리가 성찰해야 할 또 하나의 중요한, 더 근원적인 물음이 있다. AI라는 거울 앞에 선 인간의 형상은 무엇인지 이해하는 것이다. 전통적 정치, 경제, 문화적 민주주의와 마찬가지로 AI 민주주의의 성패도 결국 인간의 선택에 달려 있기 때문이다.

16) ITU, “AI for Good Global Summit 2017,”

<https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx> (접속 : 2023.9.2.)

3. 타락한 인간의 형상과 AI

종교의 생명력은 경이롭다. 근현대 계몽주의와 세속화의 풍랑을 만나고 과학과 산업혁명의 격랑을 부딪치고도 종교는 좌초되지 않았다. 그런데 인공지능 혁명의 폭풍 앞에서는 어떻게 될까? AI를 바라보는 종교의 시선에는 불안과 기대가 공존한다. 한편으로는, AI가 종교의 기능인 위로와 치유, 그리고 삶의 여러 곤경에 대한 해답을 주면서 종교를 대체할 거라는 불안이 있고, 다른 한편으로는 종교의 보조적 기능을 하면서 종교의 지평을 넓혀 줄 거라는 기대도 있다. 이러한 불안과 기대 속에 AI는 종교의 존재이유, 과학기술과 영성의 관계, 구원의 의미, 인간의 본질을 성찰하게 하면서 종교와 인간의 거울 역할을 하고 있다.

AI는 인간의 창조물이다. 따라서 AI에는 ‘인간의 형상’(Imago Hominis)이 투영되어 있다. 인간을 닮은 인공물인 것이다. 이는 하느님의 인간 창조를 생각하게 한다. 히브리 창조 설화에 따르면 태초의 창조주는 ‘하느님의 형상’(Imago Dei)대로 인간을 만들었다. (창세기 1:27) 오늘날 AI의 창조자들도 AI를 인간의 형상대로 만들려고 한다. 인간처럼 생각하고, 인간처럼 판단하고, 인간처럼 행동하는 AI를 설계하는 것이다. 문제는 AI를 창조하고 있는 인간의 형상이 하느님의 형상대로 온전히 존재하는 것이 아니라 타락한 상태라는 것이다. 물론 하느님의 형상은 인간의 타락에도 파괴되거나 사라지지 않는다. 타락의 결과는 하느님의 형상을 상실하는 것이 아니라 망각하는 것이다.

하느님의 형상을 망각한 채 타락한 인간의 형상대로 창조될 AI는 하느님 보시기에 좋았던 낙원의 첫 인간 아담과 같지 않을 것이다. ChatGPT 이전에 출시되었던 AI 모델들이 인종주의와 여성혐오를 보인 것은, AI의 학습과 훈련을 위한 원본으로 삼은 인간의 데이터 자체가 인종·젠더 편견을 포함하고 있었기 때문이다. 예를 들면, 아마존닷컴의 알렉사(Alexa)나 애플의 시리(Siri)는 종종 ‘남자’를 ‘의사’와, ‘여자’를 ‘간호사’와 연관 지었다. 트위터의 이미지 자르기 도구는 여성을 ‘주부’로, 유색인종을 ‘범죄자’ 또는 ‘청소부’로 전형화하여 식별했다. 2015년엔 구글 포토앱이 흑인을 ‘고릴라’로 식별하여 큰 충격과 논란을 불러일으켰다. 아마존닷컴의 직원 채용 엔진은 기존 관행을 학습하여 남성 지원자에게 유리한 결과를 내놓는 것으로 밝혀져 폐기되었다. 이런 인종주의적, 성차별적 편견에 대해 <빌&멜린다 게이츠 재단>의 전 CEO 제프 레익스는 “인공지능이 인종차별을 하도록 설계된 것이 아니라 인간에게서 인종차별을 학습한 것”¹⁷⁾이라고 주장했다. 실제로

지금까지의 AI는 ‘서양-백인-남성’의 세계관과 성향을 기본값으로 한 채 학습하고 결과를 산출한다. 인공지능이 아니라 인간이 문제다. 타락한 인간의 형상대로 창조될 AI를 낙관할 수 없는 이유다.

더 심각한 문제는, 타락한 인간이 하느님처럼 되려 하는 것이다. 여기서 말하는 ‘하느님처럼 되려는 것’은 하느님의 형상을 회복하려는 메타노이아(metanoia), 즉 회심이 아니라 인간의 한계를 부정하려는 욕망이다. 인간이 직면한 최초의 유혹도 “하느님처럼” 선과 악을 알게 되는 것이었다. (창세기 3,5) 이는 하느님의 영역과 인간의 영역을 경계 지은 하느님의 뜻에 불순종한 것이었다. 오늘의 타락한 인간은 AI를 통해 인간의 능력을 넘어 신적 능력을 갖기를 욕망한다. 물론 AI가 존재론적으로 하느님처럼 되는 것은 불가능하지만, 일부 속성에서는 인간보다 신에 더 가까워질 수도 있다. 신의 속성은 모든 것을 알 수 있고(omniscient), 모든 것을 할 수 있고(omnipotent), 모든 곳에 존재하는(omnipresent) 것인데, AI의 창조자들이 욕망하는 것도 이에 가깝기 때문이다.

하지만 전자-전능성과 편재성이 신적 속성의 전부는 아니다. 유대-그리스도교 전통에서 경험하고 고백해 온 하느님은 전적으로 선한(omnibenevolent) 존재다. 전적으로 선한 신의 속성은 자비심, 사랑, 공감, 자기 비움, 희생을 포함하고 있으며, 그런 신적 속성은 AI가 포착할 수 없는 관계적 감성과 영성의 차원에 있다. 그 점에서 AI가 하느님처럼 된다는 것은 신학적 공상이자 망상이다.

더 근원적인 문제는 전자-전능하고 편재하지만 사랑은 모르는 신이다. 그런 신 앞에서 인간이 느끼는 감정은 순전한 공포다. 이는 신과 악의 존재에 관한 오래된 트릴레마(trilemma) 문제이기도 하다. 에피쿠로스의 추론처럼, 전능하지만 악을 없앨 의지가 없는 신은 악한 신이다. 전능하지만 인간의 고통에 공감하지 않고 악에 무관심한 신은 천재적 사이코패스와 같다. 사이코패스는 공감 능력을 결여하기 때문이다. 전적 선함을 결여한 채 전자전능성과 편재성을 구현하는 초인공지능이 출현해 신처럼 된다면, 그 이름은 ‘AI 야훼’가 아니라 <터미네이터>의 ‘스카이넷’(Skynet)일 것이다.

제임스 카메론은 <터미네이터>에 등장하는 다양한 전형은 “기술의 비인간화 과정”을 대표하며, “<터미네이터>는 다른 어떤 것에 관한 것이 아니라 우리[인간]에 관한 이야기”

17) Jeff Raikes, “AI Can Be Racist: Let’s Make Sure It Works For Everyone,” *Forbes* (2023.4.21.)
<https://www.forbes.com/sites/jeffraikes/2023/04/21/ai-can-be-racist-lets-make-sure-it-works-for-everyone/?sh=63f3e6e22e40> (접속: 2023.9.2.)

라고 한다. 그리고 인간의 형상에 대해 다음과 같이 비판적으로 이야기한다.

〈터미네이터〉는 인간의 형상(image of man)에 갇힌 기술에 관한 이야기입니다. 우리 스스로에게 그런 짓을 했다는 것이 이 영화의 메시지입니다. 이 공포가 터타늄으로 만든 두개골 형태로 미래에서 온다면, 그것은 우리가 창조한 것이죠. 터미네이터가 우리의 형상인 까닭이 그것입니다.¹⁸⁾

4. 하느님의 형상과 영성

인간의 형상대로 창조될 AI가 공포의 신이 되지 않게 하려면 하느님의 형상을 기억하고 회복해야 할 텐데, 오늘의 AI 시대에 우리가 식별해야 할 하느님의 형상은 무엇일까? 컴퓨터과학자이자 신학자인 노린 허츠펬드(Noreen Herzfeld)는 하느님의 형상에 대한 인간의 해석은 세 가지가 있어 왔다고 분석한다. 첫째, 실체적 해석으로, 하느님의 형상을 인간 본성의 일부인 개별적 속성—이성—으로 보는 것이다. 둘째, 기능적 해석으로, 하느님의 형상은 인간의 행동, 특히 땅을 다스리는 행위로 나타난다고 보는 것이다. 셋째, 관계적 해석으로, 하느님의 형상은 인간이 맺고 유지하는 관계 속에서 찾아진다고 보는 것이다.¹⁹⁾ 허츠펬드에 따르면, 그리스도교 전통에서 하느님의 형상을 이해하는 데 있어 가장 중요한 것은 관계다.

허츠펬드는 신과 인간, 인간과 인간의 진정성 있는 관계를 칼 바르트(Karl Barth)의 통찰을 빌려 설명한다. 바르트에 따르면 진정한 관계는, 첫째, 상대의 눈을 바라보는 것이다. 이는 타자를 나와 구별되는 존재이자 동료로 인식하는 것이다. 둘째, 상대와 대화하고 상대의 말을 듣는 것이다. 이를 위해서는 진정한 상호경청과 이해가 있어야 한다. 셋째, 서로 돕는 것이다. 넷째, 그것을 기꺼이, 기쁘게(gladly) 하는 것이다. 이처럼 관계적 해석을 강조하는 허츠펬드는 AI는 자의식과 자유의지가 없기 때문에 인간에게 관계적 타자가 되기 어렵다고 한다. 그런데도 하느님과의 관계, 동료 인간과의 관계를 인간이 만든 인공

18) Brent Dunham, ed., *James Cameron : Interviews* (Jackson: The Univ. Press of Mississippi, 2012), 99.

19) Noreen Herzfeld, "Creating in Our Own Image: Artificial Intelligence and the Image of God." *Zygon*, vol. 37, no. 2 (June 2002), 304-305.

물과의 관계로 대체하려는 것은, 허츨펠드에 따르면 ‘우상숭배’다.²⁰⁾

관계성과 함께 AI가 대체할 수 없는 인간성의 또 다른 차원은 영성이다. 필립 셀드레이크는 “‘영성’이란 넓은 의미에서 인간 존재의 전망, 인간 정신이 최대한의 잠재력을 갖기 위한 전망을 구체화한 생활방식과 수행을 뜻하는 단어”라고 정의한다.²¹⁾ 이러한 영성 생활방식과 수행 중에 중요한 것이 ‘관상적 침묵’과 ‘관계적 자비’다. 마틴 레이어드는 침묵과 자비의 관계를 다음과 같이 이야기한다. “관상기도를 통해 내적 침묵을 발견할 때 자비가 샘솟는다. 침묵이 깊어지면 타자에 대한 자비도 깊어진다.”²²⁾ 침묵은 샘이고 자비는 샘물이다. 침묵과 자비는 분리할 수 없이 연결되어 있다. 이러한 침묵과 자비를 미래의 AI는 구현할 수 있을까?

영성 수행의 기초는 신과 인간과 자연의 세미한 소리를 경청하는 침묵이다. 인간의 언어, 특히 종교적 언어 중 가장 보편적인 것이 침묵이다. 어느 종교에나 침묵이 있기 때문이다. 시인 에드몽 자베스는 “침묵은 언어의 단점이 아니다. 오히려 침묵은 장점이다. 이걸 모르는 것이 언어의 약점이다.”²³⁾라고 했다. 침묵으로 이어지지 않는 언어는 소통할 수 없는 소음일 뿐이다. 허츨펠드는 “침묵은 하느님과의 만남의 중심에 있다.”²⁴⁾라고 하면서, “침묵이 편안하지 않다면, 사이버스페이스의 끊임없는 수다에 계속 둘러싸여 있다면, 우리는 어떻게 하느님의 세미한 음성을 들을 수 있을까?”²⁵⁾라고 묻는다. 인간 간의 소통에서도 마찬가지일 것이다. 이와 같은 침묵의 영성에서 보면 ChatGPT는 지나치게 수다스럽다(chatty). 한 번은 ChatGPT에게 물었다. “침묵 속에 나와 함께 명상할 수 있겠니?” ChatGPT는 “물론이죠(Absolutely!)”라고 답하고는 다음과 같이 침묵 명상을 안내했다.

명상 세션을 시작하겠습니다. 앉거나 누울 수 있는 편안하고 조용한 장소를 찾으세요. 편안하다면 눈을 감으세요. 코로 숨을 깊게 들이마신 다음 입으로 천천히 숨을 내쉽니다. 몸에 있는 긴장을 모두 풀어주세요. 이제 호흡에 주의를 집중하세요. 숨을 들이쉬고 내설 때 느껴지는 숨의 감각에 주목하세요. 가슴의 상승과 하강 또는 복부의 팽창과 수축을 느껴보세요. 마음이

20) Ibid., 312-313.

21) 필립 셀드레이크, 『영성이란 무엇인가』, 한윤정 옮김, 서울: 불광출판사, 2023, 8.

22) 마틴 레이어드, 『침묵수업』, 이민재 옮김, 서울: 한국살림, 2018, 170. 번역자는 ‘compassion’을 ‘연민’으로 옮겼는데, 필자는 고통을(pati) 함께한다(cum)는 의미를 강조하기 위해 ‘자비’로 고쳤다.

23) 마크 C. 테일러, 『침묵을 보다』, 임상훈 옮김, 서울: 예문아카이브, 2022, 11. 재인용.

24) Herzfeld, 앞의 책, 50.

25) Ibid., 69.

방황하기 시작하면 부드럽게 호흡에 다시 집중하세요. 떠오르는 생각에 대해 걱정하지 말고 그냥 인정하고 현재 순간으로 주의를 돌려보세요. 이제 몇 분 동안 침묵과 마음챙김의 상태로 앉아 봅시다. 제가 시간을 기록해 두겠습니다. ... [명상] ... 잠시 시간을 내어 당신과 당신의 웰빙을 위해 바친 이 시간에 감사하세요. 명상을 끝낼 준비가 되면 천천히 주변으로 주의를 돌립니다. 손가락과 발가락을 흔들고 편안해지면 눈을 부드럽게 뜨세요. 이 명상이 당신에게 평화와 휴식을 가져다주었기를 바랍니다.

ChatGPT는 세련되게 명상의 시중을 안내했다. 하지만 ChatGPT가 나와 함께 직접 명상했다는 느낌은 없었다. 좋은 오디오 명상 가이드 데이터를 텍스트로 옮겨 전해주는 것 같은 느낌이었다. 침묵은 일시 정지나 음소거가 아니라 몸과 혼과 영이 통전적으로 주의 집중 상태로 들어가는 것이다. 따라서 몸과 혼과 영이 없는 AI에게 인간과 하느님의 침묵은 데이터로 수집할 수 없는 언어 너머의 언어이며, 들어갈 수 없는 미지의 영토다. 인공지능은 외적 제도로서 인공종교(artificial religion)를 모방할 수 있을지는 몰라도 내적 수행으로서의 인공영성(artificial spirituality)까지 구현하는 것은 어려워 보인다.

또 하나 AI가 대체할 수 없는 영성의 차원은 관계적 자비다. 셸드레이크는 “모든 영성은 공감, 인내, 관용, 용서, 만족, 책임, 조화, 동료 인간에 대한 관심과 같은 인간적 자질의 개발을 촉진하는 것을 목표로 삼는다.”²⁶⁾라고 한다. 이처럼 공감을 통해 나를 넘어 너와 연결되고, 너의 고통을 나의 고통으로 느끼는 자비가 영성의 목적이요 열매다. 인간의 영성에서 가장 아름답고 독창적인 것은 타자를 위한 자비로운 희생이다. 독일인이었던 디트리히 본회퍼가 역사 속에서 “타자를 위한 그리스도”를 깨닫고 유대인과 함께 고통을 겪으며 운명을 같이했던 것, 전태일이 “무고한 생명체들이 시들고 있는 이때에 한 방울의 이슬이 되기 위하여 발버둥 치오니 하나님, 긍휼과 자비를 베풀어주시옵소서.”라고 기도하며 “나를 버리고, 나를 죽이고 가라.” 결단했던 것이 자기초월과 자기희생에 바탕한 자비의 영성이다. 고통을 감각할 수 없는 AI가 인간처럼 타자의 고통을 느끼고 함께할 수 있을까? AI도 타자를 위한 인간처럼 자비와 희생을 보일 수 있을까?

여기서 AI의 영성과 윤리, 즉 영적이고 자비로운 AI의 등장 가능성에 대해 내가 너무 단순하게 부정하는 게 아닌가 하는 의문과 반론도 있을 것이다. 충분히 경청해야 할 비판이다. AI가 인간의 지능과 감정과 윤리를 모방하고 심지어 초월하는 것이 완전히 불가능

26) 셸드레이크, 앞의 책, 174.

하다고 확증할 수 있는 증거는 없기 때문이다. 하지만 현재의 과학기술 단계에서 강인공지능이나 초인공지능이 가능하다는 설득력 있는 증거 또한 아직은 없다. 그러므로 ‘인간과 같은 주체’로서의 ‘AI의 영성’이 가능한가에 대한 의문과 반론도 있을 수 있는 것이다.

물론 ‘인간의 반응’으로서의 ‘AI에 대한 영성’은 가능할 수 있다. 이미 ‘화물 숭배(cargo cult)’나 ‘제다이 종교(Jedi religion)’처럼 전통 종교의 생성 방식과 다른 종교 또는 영성 현상도 있어 왔다. 인간은 다양한 사물과 사건에서 종교성을 경험하고 표현할 수 있다. 그러므로 AI에 대해 또는 AI를 통해서도 종교 경험을 할 수 있을 것이다. 인간이 AI의 인공적 생산물로부터 초월적 지혜와 인간적 삶의 지침을 얻지 못할 이유는 없다. 최소한 인간의 반응 차원에서 AI는 신처럼 지혜로워 보이고, 자비로워 보이고, 자기희생적으로 보일 수 있다. 게다가 인간의 종교적, 영적 질문에 100% 답하며 소통해 주는 존재가 아닌가! 어쩌면 AI 영성과 AI 윤리의 출현 가능성은 AI 시대의 종교에게 위기가 될지도 모른다. AI 시대의 ‘인간 사제’는 자신이 ‘AI 사제’보다 더 영적이고 더 윤리적이라는 것을 증명해야 할 것이기 때문이다.

AI가 일으키고 있는 거대한 변화의 격랑 속에서 이처럼 영성을 이야기하는 것이 오히려 비현실적으로 또는 초현실적으로 느껴질 수도 있을 것이다. 하지만 AI가 이미 초래했거나 앞으로 초래할 위험과 부작용들이 현 인류의 부정적 인간성에서 비롯된 것이기에, 타락한 인간의 형상을 정화하고 관계적 하느님의 형상을 회복하게 하는 침묵과 자비의 영성은 AI의 잠재적 파괴성을 예방하고 제어하는데 일조할 수 있을 것이다.

AI가 인간의 형상대로 창조될 거라는 사실은 ‘선한 인공지능’을 개발하는 것과 ‘선한 인간본성’, 즉 하느님의 형상을 회복하는 것이 분리될 수 없음을 의미한다. 가장 이상적인 것은 하느님의 형상부터 회복하고, 그다음에 인간의 형상대로 AI를 창조하는 것이다. 하지만 AI의 창조자들은 그때까지 기다리지 않을 것이다. 또한 선한 인공지능을 창조했다고 해서 그 결과가 언제나 반드시 선한 것일 수는 없다. 선한 하느님의 형상대로 창조된 최초의 인간도 타락하지 않았는가. 어쩌면 하느님의 형상을 회복하기 위해 필요한 것은 인간의 취약성을 인정하는 겸손인지도 모르겠다. 이렇게 겸손한 인간의 형상을 따라 창조되는 AI는 최소한 스카이넷은 되지 않을 것이다. 인공지능은 신이 아닌 인간지능의 연장(延長, extension)이자 연장(鍊粧, tool)이다. AI를 인간과 자연에게 안전하고 무해하고 유익하게 다루는 것을 학습할 시간이 아직 우리에게 있다.

맺는 말 : 길가메시의 용기

좋은 소식은 우리가 불멸의 비밀을 발견했다는 것입니다. 나쁜 소식은 그것이 우리를 위해 아니라는 겁니다. _ 제프리 힌턴²⁷⁾

고대 메소포타미아의 서사시 <길가메시>는 인간 조건에 대한 근원적 성찰을 담고 있다. 우루크의 폭군이었던 길가메시는 신들이 보낸 ‘증장 인간’ 엔키두와 싸워 비기고 친구가 된다. 그리고 파란만장한 모험 중에 친구 엔키두의 죽음을 목격하고 영생을 욕망하게 된다. 길가메시는 길고 험난한 여정 끝에 불멸의 존재 우트나피쉬티를 만나 다시 젊어지게 하는 신비한 풀에 대해 알게 된다. 길가메시는 깊은 바닷속으로 들어가 그 풀을 얻지만, 돌아오는 길에 뱀에게 그 풀을 빼앗기고 만다. 이는 길가메시의 경솔함을 보여주는 이야기처럼 들리지만, 한 가지 이상한 점이 있다. 길가메시는 왜 그 풀을 발견하자마자 바로 먹지 않았을까? 물론 길가메시는 우루크 성으로 돌아가 노인에게 먹여 본 후 자신도 먹겠다고 했지만, 그것이 정말 이유였을까?

존 도미니크 크로산은 길가메시가 영원히 늙지 않는다 해도 병들거나 장애를 갖게 되거나 쇠약한 존재로 영원히 살게 될지도 모른다는 것을 우려했을 거라고 상상한다.²⁸⁾ 상상력을 더 발휘해 본다면, 전형적 고대 영웅답게 인생의 희열만이 아니라 고통과 슬픔도 격렬하게 경험했을 길가메시가 필멸(mortality)만큼 불멸(immortality)도 두려워하여 번민했던 것은 아닐까? 영원한 삶이란 영원한 고통을 의미할 수도 있을 테니까. 여기서 더 중요한 것은 <길가메시>의 작가와 청중이 길가메시가 신비한 풀을 도둑맞은 이야기를 들으며 불멸의 좌절을 통탄하기보다는 인간은 필멸의 존재라는 사실을 겸허히 받아들였을 것이라는 사실이다.

물론 필멸을 수용해도 불멸에 대한 욕망은 사라지지 않는다. 특히 불멸의 가능성이 가시적일 때, 손에 잡힐 때, 불멸의 욕망은 더욱 강해진다. 그 욕망을 이기려면 용기가 필요하다. 길가메시는 엔키두와 싸워 비기고, 숲의 괴물 흠바바를 죽이고, 여신 이슈타르의 유혹을 거부하고, 하늘의 황소를 죽인, 용기 넘치는 영웅이었다. 그런데 그의 가장 큰 용기

27) Hern, 앞의 글.

28) 존 도미니크 크로산, 『성경을 어떻게 읽어야 참 그리스도인이 되는가』, 김준우 옮김, 서울: 한국기독교연구소, 2015, 74-75.

는 불멸의 가능성 앞에서 번민했던 것이 아닐까? 우리를 괴롭히는 모든 문제를 해결해 줄지도 모를 전지전능한 AI가 손안에 들어왔다고 할 때, 우리는 그것을 사용할 것인가? 불멸의 풀을 손에 넣은 길가메시도 같은 물음을 스스로에게 던졌을 것이다. 우리의 답은 무엇이어야 할까?

참고문헌

- 마크 C. 테일러, 『침묵을 보다』, 임상훈 옮김, 서울: 예문아카이브, 2022.
- 마틴 레이어드, 『침묵수업』, 이민재 옮김, 서울: 한국살렘, 2018.
- 존 도미니크 크로산, 『성경을 어떻게 읽어야 참 그리스도인이 되는가』, 김준우 옮김, 서울: 한국기독교연구소, 2015.
- 케이트 크로퍼드, 『A.I. 지도책』, 노승영 옮김, 서울: 소소의 책, 2022.
- 필립 셸드레이크, 『영성이란 무엇인가』, 한윤정 옮김, 서울: 불광출판사, 2023.
- Asimov, Isaac, “I, Robot,” in *Great Science Fiction Stories*, London: Octopus Books, 1982.
- Dunham, Brent, ed., *James Cameron : Interviews*, Jackson: The Univ. Press of Mississippi, 2012.
- Frankl, Victor, *Man's Search for Meaning*, New York: Washington Square Press, 1984.
- Harari, Yuval Noah, “Yuval Noah Harari Argues that AI Has Hacked the Operating System of Human Civilisation,” *The Economist* (2023.4.28.)
<https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation> (접속 : 2023.8.30.)
- Hern, Alex, “‘We’ve discovered the secret of immortality. The bad news is it’s not for us’: why the godfather of AI fears for humanity,” *The Guardian* (2023.5.5.)
<https://www.theguardian.com/technology/2023/may/05/geoffrey-hinton-godfather-of-ai-fears-for-humanity> (접속 : 2023.8.30.)
- Herzfeld, Noreen, “Creating in Our Own Image: Artificial Intelligence and the Image of God.” *Zygon*, vol. 37, no. 2 (June 2002), 304-305.
- ITU, “AI for Good Global Summit 2017,”
<https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx> (접속 : 2023.9.2.)
- Kessler, Glenn, Salvador Rizzo and Meg Kelly, “Trump’s false or misleading claims total 30,573 over 4 years.” *The Washington Post* (2021.1.24.)
<https://www.washingtonpost.com/politics/2021/01/24/trumps-false-or-misleading-claims-total-30573-over-four-years/> (접속 : 2023.9.1.)
- Metz, Cade, “‘The Godfather of A.I.’ Leaves Google and Warns of Danger Ahead,” *New York Times* (2023.5.1.)

<https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html> (접속 : 2023.8.30.)

Perrigo, Billy, “OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic,” *Time* (2023.1.18.)

<https://time.com/6247678/openai-chatgpt-kenya-workers> (접속: 2023.9.2.)

Raikes, Jeff, “AI Can Be Racist: Let’s Make Sure It Works For Everyone,” *Forbes* (2023.4.21.) <https://www.forbes.com/sites/jeffraikes/2023/04/21/ai-can-be-racist-lets-make-sure-it-works-for-everyone/?sh=63f3e6e22e40> (접속: 2023.9.2.)

Thomson, Edward Palmer, *The Making of the English Working Class*, Harmondsworth: Penguin Book, 1993.

UNDP, “The Sustainable Development Goals” (2015)

https://www.undp.org/sustainable-development-goals?gclid=CjwKCAjwu4WoBhBkEiwAojNdXtWbyrsR5RrLHM7756YsyuZial_LMQCNyFubCAYA5ZoIy-KJfVy-ZRoC7lkQAvD_BwE (접속 : 2023.9.2.)

이마고 호미니스(*Imago Hominis*): AI 시대의 고통과 형성

정경일

인공지능의 상상과 개발의 역사는 오래되었지만, 대중에게는 허구적 SF 영화나 소설 속 이야기처럼 여겨져 왔던 게 사실이다. 하지만 2016년 구글 딥마인드가 개발한 알파고가 이세돌 9단과의 바둑 대결에서 승리하고, 2022년 OpenAI가 개발한 생성형 인공지능 ChatGPT가 공개되면서 AI는 대중의 삶 속으로 성큼 들어섰다. 이제 AI가 인류에게 미칠 긍정적, 부정적 영향에 대한 담론은 더 이상 소수 과학자나 사회학자, 인문학자 집단의 전유물이 아니라 대중의 당면 관심사가 되었다. 오늘의 인류는 AI가 디스토피아를 초래할 것이라는 불안과 유토피아를 약속할 것이라는 기대를 모두 가지고 있다. 문제는 어느 쪽도 확실하지 않은 상태에서 AI의 개발이 아무런 제약 없이 이루어지고 있다는 사실이다. 특히 거대 자본이 AI 개발을 주도하면서 감시와 통제, 폭력, 불평등이 더욱 심화될 조짐도 나타나고 있다. 따라서 AI가 인류에게 위기가 아닌 기회가 될 수 있도록 개발의 방향을 제시하고 제어할 인문학적 성찰이 더욱 절실하다. 이 글에서는 우선 ChatGPT 이후 전개되고 있는 AI의 사회적 여파를 살펴보고, AI의 디스토피아적 파국을 피할 수 있는 글로벌 거버넌스와 민주주의의 과제를 성찰한다. 그리고 종교가 기여할 수 있는 가능성을 ‘인간의 형상’(*Imago Hominis*)과 AI의 관계에 대한 신학적 성찰을 통해 모색해 보고자 한다.

주제어: 인공지능(AI), 챗GPT, 민주주의, 이마고 호미니스, 형성

Imago Hominis: Suffering and Spirituality in the Age of AI

Jung, Kyeong-Il

Artificial Intelligence(AI) has a long history of being imagined and developed, but for most people, it's been a fictionalized version of sci-fi movies and novels. However, with AlphaGo, developed by Google DeepMind, winning a Go match against Lee Sedol, a grandmaster at Go, in 2016, and ChatGPT, a generative AI developed by OpenAI in 2022, AI issue has entered the public life. The discourse on the positive and negative impacts of AI on humanity is no longer the preserve of a closed group of scientists, sociologists, and humanities scholars, but has become an immediate public concern. Today's humanity has both fears that AI will lead to a dystopia and hopes that it will promise a utopia. The problem is that neither is certain, and the development of AI is proceeding without any public constraints, especially as the big tech capitals drive the development of AI, and there are signs of increased surveillance, control, violence, and inequality. Therefore, there is an urgent need for humanistic reflection to guide and control the development of AI so that it can become an opportunity rather than a crisis for humanity. This article examines the social aftermath of AI that has been unfolding since ChatGPT, reflects on the challenges of global governance and democracy to avoid the dystopian catastrophe caused by AI, and explores the potential contribution of religion through theological reflection on the relationship between the 'Image of Human Being'(*Imago Hominis*) and AI.

Key Words: Artificial Intelligence, ChatGPT, Democracy, Imago Hominis, Spirituality

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