신학과 세계 제95호 (2019. 3. 30): 149-179
https://doi.org/10.21130/tw.2019.3.95.149

인공지능과 기독교윤리:
신학과 인공지능 연구와의 대화

유 경 동 *

【주제어】 인공지능, 유형론(독립, 상충, 대화, 통합), 기독교윤리, 신학, 자율

【요약문】 최근 인공지능 연구는 지적 주체인 인간의 다양한 영역을 인공적으로 구성하는 것을 넘어, 인간의 정신을 정보의 형태로 변환하거나 정보를 인간의 생체적이고 물리적인 형태로 재현하는 수준에 이르고 있다. 따라서 인간에 대한 인공지능 분석을 통하여 인간론을 중시하는 신학과의 대화적 모델로서 앞의 네 가지 유형론을 제시하는 본 글은 나름 유의미하다고 본다. 필자가 인공지능과의 간학문적 유형론에서 설명하는 '독립', '상충', '대화', 그리고 '통합유형'에서 '대화유형'은 인공지능 연구에 있어서 신학적이며 철학적인 통찰을 요구하며, 각각 학문적 영역을 통하여 궁극적으로는 인간 정신의 발전을 도모한다는 공동적인 목표를 설정한다. 인공지능과 인간지능의 메커니즘 상의 상관관계를 긍정적으로 파악하며, 양자 공히 기계적 유한성과 실존적 유한성의 특질을 가진다고 본다. 따라서 인공지능의 발전

* 감리교신학대학교 부교수, 기독교윤리학

에 맞추어 인간지능이 나아가야 할 바람직한 모형론 개발을 도모하여 인간의 창의성 개발에 인공지능이 큰 역할을 할 수 있으며, 인간의 덕 윤리의 신장에도 크게 도움이 될 수 있다는 관점이 대화유형이라고 할 수 있다.
이와 같은 이론들은 인공지능에 대한 두려움이나 혹은 기대와 같은 양가적인 감정을 넘어서 적극적으로 인공지능 연구가 인간과 공동체에 미칠 영향을 긍정적인 관점에서 접근하여, 궁극적으로는 인공지능을 통하여 인간성의 개발과 덕 윤리의 신장을 도모하고 나아가 인간의 존재론을 항상 염두에 두는 바람직한 인공지능 연구 윤리를 확립하는데 큰 도움이 될 것으로 기대한다.
인공지능은 '세계 위기 기구'에서도 밝혔듯이 지구적인 위험 요인이다. 지구 멸망의 가능성이 초지능에게 있다고 한다면, 그 일차적인 책임은 과학집단에 있으며, 그리고 동반책임은 그 공동체를 관리 감독하는 국가에게 있을 것이다. 그럼에도 불구하고 그러한 위험을 전가하지 않으면서도 과학집단의 주체가 인간 연구자들임을 감안하여 신학공동체는 과학연구자들과의 지속적인 대화를 하여야 할 것이다. 전인간적 신앙교육과 인간 인지의 개발에 대한 책임이 교회와 신학 공동체에도 있음을 직시하고, 과학 공동체와 공공영역에서 협력을 기대한다.

I. 서론

인공(초)지능[1]에 관한 연구는 이제 공학 분야를 넘어서서 신학에서도

1) 필자는 '인공지능'과 '초지능'을 진보하는 과학기술의 관점에서 본 글에서는 큰 구분 없이 사용함을 밝힌다.

매우 중요한 주제로 부각하고 있다. 전통적인 신학의 주요 주제인 하나님의 형상 개념과 인공지능에 부여하는 인간의 이미지 개념에 대한 존재론적 연구를 필두로, 인공지능의 자율, 의지, 도덕, 책임, 그리고 인공지능 기계 윤리(machine ethics)와 법의 사안까지 관련이 된다.

필자는 최근 인공지능에 관한 다양한 연구들을 독립(independence), 충돌(conflict, 갈등), 대화(dialogue), 그리고 통합(integration)의 네 가지 유형으로 나누어서 신학과 인공지능의 통섭가능성에 대하여 살펴보고자 한다. 필자는 위 네 가지 유형론에서 특히 '대화 유형'에 관심을 가지면서 영어권 학자들의 이론들을 살펴보며, 신학과 인공지능 관련 연구와의 간학문적 소통의 교두보를 마련함으로써 후속학문이 발전하기를 기대한다.[2)]

국내 인공지능에 대한 연구는 각 학문의 영역별로 다양하게 이루어지고 있으며, 유형론에 관하여서는 인문학의 종교학/신학 분야를 포함하여 60여 논문을 찾아볼 수 있다.[3)] 그러나 인공지능에 관한 거시적 관점에서 유형화한 연구는 아직 없기에 이 논문이 후속 연구를 위한 예비적 담론의 역할을 할 수 있으리라고 본다. 다만 이 글은 인공지능과 신학과의 관계에서 위의 유형론에 관심을 가지며, 영어권의 연구물들을 통하여 거시적으로 접근하기 때문에 각 학자들의 연구를 보다 심도 있게 정리하지 못한 한계

2) 필자는 유형론으로 독립(independence), 충돌(conflict), 대화(dialogue), 그리고 통합(integration)으로 구분하며, 이언 바버(Ian Barbour)가 신학과 과학의 통섭을 위하여 사용한 방법론을 따름을 밝힌다. 이언 바버의 유형론에 비판이 없는 것은 아니다. 신학의 주요 주제인 신론이 하나님의 초월을 다루는 것이기에 과학적 사고관으로 접근할 수 없다는 원론적인 논란이 있으나, 필자는 이 논문에서 가급적 가치중립적인 맥락에서 신학과 과학 사이의 소통가능성의 유형론에 제한하고자 하며, 영어권 인공지능 연구가들을 유형론의 각 범주에 배치한 것도 필자의 제한적 관점인 것을 밝힌다. 참고) 이언 바버(Ian Barbour), 이철우 옮김, 『과학이 종교를 만날 때』 (서울: 김영사, 2002).

3) 필자가 지식콘텐츠 검색사이트 *DBpia*를 통하여 '인공지능'과 '유형'을 주제어로 검색하여 본 결과 57편의 논문이 있으며, 신학분야에서는 1편 정도이다. 그리고 유형론도 주로 정책결정과 서비스 유형, 로봇의 분류체계에 관한 내용이다. *DBpia*(누리미디어, 서울), http://proxy.mtu.ac.kr:8080/http/www.dbpia.co.kr/ [2019.04.13. 최종접속].

가 있으며, 추후 보다 미시적인 주제들을 가지고 분석하기를 기대한다.

II. 독립유형

독립유형은 인공지능 관련 연구를 독립적인 분과로 인정하고, 신학이나 전통적인 이론을 중시하면서 상호 인정하는 학문적 태도를 취하는 것을 의미한다고 할 수 있다. 독립유형에서 인공지능에 관한 논지는 주로 과학에 대한 학문적 전문성을 인정하고, 인공지능과 같은 기계장치와 생물학적 항상성을 유지하려는 인간 생체 간의 분명한 차이를 강조하는 것을 볼 수 있다.

스튜어트 암스트롱(Stuart Armstrong)과 카즈 소탈라(Kaj Sotala)에 따르면, 기존의 AI의 미래에 대한 전망은 그것의 위협이나 이점 등에 대하여 편향된 자료와 전제에 의존하고 있음을 비판하며, 그 이유로 전문가들의 관점에 근거한 비전문가적 예측, 또는 비전문가적 예측에 근거한 전문가적 판단의 오류에 있다고 지적한다.[4] 따라서 정확한 예측을 위해서는 각 예측에 적합한 유형을 파악하고, 그 유형 내에서 예측을 해야만 하기 때문에, 인공 지능 자체에 대한 기본적 과학적 발전 가능성에 대해서는 전문 과학자들의 판단이 주가 되어야 하고, 그것이 미치는 사회적 영향이나 인간 지능에 대한 각각의 판단 유형과 기준에 따라 이루어져야 한다고 주장한다.[5]

이리 위더만(Jiří Wiedermann)은 현재 인공지능에 연구에 대한 기본적

4) Stuart Armstrong and Kaj Sotala, “How We’re Predicting AI - or Failing to,” in *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, ed. Jan Romportl, Eva Zackova, and Jozef Kelemen (Switzerland: Springer International, 2015), 11.

5) 위의 책, 11-12.

인 질문은 8개로 정리할 수 있다고 보는데, "(i) 지능은 어디로부터 기인하는가?; (ii) 인공적인 인지 체계의 '계산 능력이란 무엇인가?; (iii) 지능에도 '수준차이'가 있는가?; (iv) 지능의 '수준 차이'[를 이해하는 데에 있어서] 인간의 지능은 어떤 역할을 하는가?; (v) 지능에 대한 일반적 기제가 존재하는가?; (vi) 육체를 벗어나는 '완전히 발달한' 지능이 존재할 수 있는가?; 그리고 앞에 제기한 질문들 못지않게 중요한 질문으로, (vii) 지각을 저장하고 연결할 클라우드(저장소)가 존재할 수 있는가?; (viii) 우리는 어떻게 새로운 지식을 창출할 수 있는가?"[6]의 질문이 있다. 위더만은 이러한 질문에 대한 대답은 항상 각 질문이 주어지는 유형과 전제하는 틀 안에서만 정확하게 제시될 수 있다고 지적하면서, 인공지능 연구에 관한 독립적인 범주의 중요성을 강조한다.[7] 위더만의 이와 같은 질문들은 인공지능에 대한 정확한 전문적 이해를 전제하며, 간학문적 접근방법을 예시하는 좋은 모형이 될 수 있다고 본다.

인지과학자 폴 슈와이저(Paul Schweizer)는 인공지능이 과연 인간의 지능과 같은 수준으로 발전될 수 있는지에 대한 연구는 인공지능의 구성체로서 기계적 장치와 인간의 육체적 몸 간의 인터페이스 안에서만 정립 가능하다고 주장하며, 인공지능이 인간 육체와 결합되는 하이브리드적 특징 안에서만 인정할 수 있다고 본다. 인간의 인지 능력이 항상 몸이라는 전체적인 유기적인 물리적 근거를 가지기 때문에, 인공지능 자체는 인간 의식과는 다른 형태를 가지게 된다는 점에서 독립적이라고 할 수 있다.[8]

---

6) Jiří Wiedermann, "Answering Curious Questions about Artificial Intelligence," in *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, ed. Jan Romportl, Eva Zackova, and Jozef Kelemen (Switzerland: Springer International, 2015), 187.

7) 위의 책, 187-188.

8) Paul Schweizer, "Artificial Brains and Hybrid Minds," in *Philosophy and Theory of Artificial Intelligence 2017*, ed. Vincent C. Müller (Switzerland: Springer Nature, 2018), 81, 90.

중국의 컴퓨터 과학자인 유지안 리(Yujian Li)에 따르면, 인지적 능력 중 언어 능력에 집중하여, 인공지능이 언어 이해의 측면에서, 다양한 언어의 상대적 의미를 적합하게 이해하고, 그에 따른 해석이 가능하기 위해서는 인간의 두뇌와 같은 지능이 아니라, 고차원의 기계적 지능을 보유해야 한다고 지적하면서, 인간의 지능과 인공지능의 능력을 구분한다.[9]

라이언 톤켄스(Ryan Tonkens)에 따르면, "인공 도덕 행위자(artificial moral agent, AMA)"이론에 근거하여 "완전 자율적 인공 도덕 주체의 성공적인 개발이 이미 임박"[10]했다고 지적하면서, 인공지능을 포함하는 새로운 기계 윤리(Machine ethics)의 정립이 필요하다고 강조한다. 문제는 기계 윤리를 정립하는 데에 있어서의 윤리는 그것을 개발하는 개발자에 대한 윤리(인간의 윤리적 규범)와 관련되면서도 독립적인 기계만을 위한 윤리적 규범이 정립되어야 한다는 것인데, 톤켄스는 칸트식의 의무론을 기계 윤리에 그대로 적용하면, 오히려 반칸트적 윤리가 도출될 수 있다면서, 독립된 윤리적 원리의 중요성을 강조한다.[11]

존 홀랜드(John H. Holland)는 유기체가 진화과정을 통해 어떻게 복잡성을 획득하며 자연에 적응하는지에 대하여 논의하면서, 적응 기제에 있

9) Yujian Li, "Can Machines Think in Radio Language," in *Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018 Beijing, China, November 2-5, 2018 Proceedings*, ed. Zhongzhi Shi, Cyriel Pennartz, and Tiejun Huang (Switzerland: Springer, 2018) (eBook), 230. https://doi.org/10.1007/978-3-030-01313-4. [2019.04.13. 최종접속]. 유지안 리는 전 세계적으로 언어가 5,000-7,000개 정도가 있으며, 그 중 90퍼센트는 10만 명 미만이 사용하는 언어라고 설명하면서, 미래에 이런 수많은 언어들이 인공지능을 통하여 상관성을 가지기 위하여서는 인간의 두뇌 능력을 넘어서는 그러한 초지능이 필요하다고 강조한다. 같은 책, 231. 유지안 리는 미래에 로봇우주비행사가 우주에서 지구를 바라보며 "지구는 우리는 집이다."라고, 또는 "지구가 정말 아름답다."라고 모로스 부호로 송신이 가능하기 위하여서는 인공지능이 텍스트로 수신하는 고전적 체계와는 완전히 다른 고차원의 지능이 필요하다고 강조한다. 같은 책, 233. 'Fig.1'을 통한 설명이다.

10) Ryan Tonkens, "A Challenge for Machine Ethics," *Minds and Machines* 19(3) (2009), 421.

11) 위의 책, 421.

어서 인공지능은 다른 구조의 다른 연산자를 통해 이루어진다고 설명한다.[12] 인공지능은 다양한 방식의 학습 기제를 바탕으로 여러 데이터를 비교하여 가장 효율적인 결과를 도출하는 방식으로 외부 환경, 또는 정보 간의 복잡성에 대응하는데, 이러한 방식은 자연 유기체와는 다르다고 홀랜드는 강조한다.[13]

지금까지 살펴보았듯이, 독립유형에서는 인공지능과 연관된 전문적 이론들을 먼저 이해하는 것이 요구되며, 성급한 추론보다는 과학적 사고 위에서 인공지능, 초지능, 기계윤리, 정보의 연산자 등과 연관된 원리들에 대한 자연과학적 접근방법이 강조되는 것을 알 수 있다.

III. 충돌유형

충돌유형은 인공지능에 대하여 부정적인 관점이 지배적이라고 할 수 있다. 과학에 대한 가치중립적인 태도는 인간세계의 가치관을 붕괴시킬 수 있으며, 극단적인 경우, 초지능과 같은 혁명적인 기술개발은 과학의 수단과 그 목적을 전도시켜 인류의 멸망을 초래할 수 있다고 경고한다. 특히 인공지능이 반-지능의 역할을 할 수 있다는 가능성도 강조하면서, 인간 본연의 도덕과 윤리의식의 강화를 요구하는 것을 볼 수 있다.

마이클 셔머(Michael Shermer)에 따르면, 인공지능의 발전이 인간에게 실존적인 위협이 될 것이라는 불안은 초지능으로서 인공지능이 인간을 통제할 가능성에 기인한다고 지적한다.[14] 물론 셔머는 인공지능 발달 자체

12) John H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (Cambridge, MA: the MIT Press, 1992), 4.
13) 위의 책, 5.

가 인간의 통제를 벗어날 정도로 빠르지 않으며, 이러한 점진적 발전에 근거해서 보면, 인간이 통제할 수 있는 제한을 개발할 여지도 있다고 지적한다.[15]

닉 보스트롬은 다음과 같이 초지능의 위험성에 대하여 경고한다.

> 그러나 만약 그리고 그렇게 [초지능의 인공 지능] 주체가 스스로 다른 상황을 발견하면, 그 [인공 초지능] 주체가...[어떠한] 계산을 통해 [어떤] 결과를 도출할지에 대한 가능성은, [인공 초지능] 주체가 인간과 협력을 계속하기보다는, 오히려 더 인간종을 파멸시킬 것이라고 한다면, 그것은 매우 위험한 결과로 이어지게 될 것이다. 이러한 예측은 미래의 인공 [지능] 주체가 ... 자신의 초지능을 통해 극단적 수준의 힘과 영향력을 갖춘다면...[더 그러할 것이다.][16]

보스트롬은 또한 초지능의 위험성에 대하여 다음과 같이 말한다. "우리가 처음으로 초지능체를 만들었다고 하면, 우리는 실수를 하나 하게 되는 것인데, 그것은 그 초지능체에게 인류를 멸종시킬 수도 있는 목표를 부여하는 것일 수 있기 때문이다."[17] 보스트롬은 인간에 의하여 만들어진 인공

14) Michael Shermer, "Apocalypse AI," *Scientific American* 316(3) (Mar, 2017), 77. 마이클 셔머는 2014년 "Space X" CEO인 엘론 머스크(Elon Musk)가 보스트롬의 초지능에 대한 글을 읽고 "초지능은 핵무기보다도 더 위험하다."고 트윗한 글, 호킹(Stephen Hawking)이 BBC 방송에서 "인공지능의 개발은 인류의 멸망을 이끌 수 있다."고 언급한 내용, 그리고 마이크로소프트 공동 창업주 빌 게이츠(Bill Gates)가 자신은 "초지능에 대하여 우려하는 측에 낀다."고 한 말을 소개하고 있다. 같은 책, 77.

15) Michael Shermer, "Apocalypse AI," 77.

16) Nick Bostrom, "The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents," *Minds & Machines* 22(2) (May, 2012), 84. 닉 보스트롬은 이와 같은 입장에 이르기까지 다음 학자들과의 논평과 대화에 감사의 뜻을 표시하는데, 그 학자들의 이름은 다음과 같다. Stuart Armstrong, Grant Bartley, Owain Evans, Lisa Makros, Luke Muehlhauser, Toby Ord, Brian Rabkin, Rebecca Roache, Anders Sandberg, 그리고 다른 익명의 토론자이다. 같은 책, 각주 22)를 참조하시오. 84.

17) Nick Bostrom, "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards," *Journal of Evolution and Technology* 9(2) (March, 2002), https://www.jetpress.org/volume9/risks.html. [2019.04.13. 최종접속].

초지능체에 막대한 힘을 부여하게 되면, 인간 스스로 그 수단과 목적을 전도시키는 것이라고 경고하는 것이다.

'세계 위기 기구(Global Challenges Foundation)'의 전 세계적으로 잠재적 위험 요소에 대한 2018년 보고서에 따르면, 인공지능 또한 전 지구적으로 영향을 미칠 위기의 위험 요소 중 하나인데, 특히 인공지능이 초지능으로 구성되는 경우, 그것이 인간에 대한 가치중립적 태도, 또는 우호적 태도의 범위를 넘어설 가능성도 배제할 수 없으며, 이러한 목적이나 결과를 초래할 인공지능 연구라면, 절대 정당화될 수 없다고 주장한다.[18]

정보공학자인 유홍 장(Yuhong Zhang)과 우머 나우만(Umer Nauman)의 공동 연구에서 이들은 인공지능과 인간 지능이 얼마나 다른지에 대해서 논의한다. 장과 나우만은 빅데이터를 구성하는 데에 있어서 인공지능의 알고리즘이 정보를 분류하고, 통합, 정리하는 데에 중요한 역할을 하지만, 문제는 인공지능 자체에 어떠한 한계나 제한이 주어지지 않으면, 인공지능의 '반 지능(Anti-intelligence)'적 특징으로 인해 큰 문제에 처할 위험성도 존재한다고 지적한다.[19] 그 예로, 인공지능이 '반-지능'의 역할을 할 수 있다고 우려하는데, 인공지능에 너무 의지하여 다른 경우를 고려하지 않고 한쪽 이야기만 듣는, 마치 '바보 왕'과 같이 되어버리거나, 초속도의 인공지능 정보에 인간이 사고체계를 통하여 생각할 여유를 갖지 못하게

18) Global Challenges Foundation, "Global Catastrophic Risks 2018," 31-33. https://globalchallenges.org/our-work/annual-report/annual-report-2018 이 '세계 위기 기구'에서 보고하는 인공지능 외 지구적 위험 요소들로는 핵, 생태계 파괴, 유행병, 유성충돌, 태양지구공학, 그리고 그 밖의 잠재적 요소들이 있다고 설명한다. 같은 논문, 14-45. [2019.04.13. 최종접속].

19) Yuhong Zhang and Umer Nauman, "Artificial Unintelligence: Anti-Intelligence of Intelligent Algorithms," in *Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018 Beijing, China, November 2-5, 2018 Proceedings*, ed. Zhongzhi Shi, Cyriel Pennartz, and Tiejun Huang (Switzerland: Springer, 2018) (eBook), 333. https://doi.org/10.1007/978-3-030-01313-4. [2019.04.13. 최종접속].

되거나, 마샬 멕루한(Marshall McLuhan)의 견해처럼 부족화, 탈부족화, 그리고 재부족화가 미디어 시대에도 가속화 되겠지만, 결국 인공지능 시대가 가져올 역기능에 대한 우려를 지울 수 없다고 피력한다.[20]

던컨 퍼브스(Duncan Purves), 라이언 젠킨스(Ryan Jenkins), 브래들리 스트로서(Bradley J. Strawser) 등에 따르면, 인공지능을 포함하는 모든 자율구동의 기계들은 인간의 도덕적 판단과는 전혀 다른 형태의 도덕적 논의를 요구한다고 지적한다.[21] 퍼브스 등은 기계의 경우, 인간과 같은 도덕적 인식 능력과 사고능력을 가질 수 없기 때문에, "원칙적으로 인간의 도덕적 판단을 복제할 수 없다"[22]고 강조한다.

지금까지 필자는 마이클 셔머의 인공지능 통제의 한계, 닉 보스트롬의 초지능의 위험성, '세계 위기 기구'가 염려하는 인공지능 통제의 문제, 유홍 장 등의 반-지능의 일방성, 던컨 퍼브스 등의 인공지능의 도덕성 한계 등에 대하여 살펴보았다. 이와 같은 지적들은 인공지능 연구가 일방적인 과학주의에 편향되지 않고, 전통적인 인간관에서 중시하는 도덕, 윤리, 가치 등에 대하여 좀 더 주목하기를 주장하는 것을 볼 수 있다. 이제 다음에서 인공지능과의 대화유형에 대하여 살펴보자.

IV. 대화유형

대화유형은 인공지능 연구에 있어서 신학적이며 철학적인 통찰을 요구

20) 위의 책, 336-339.

21) Duncan Purves, Ryan Jenkins, and Bradley Strawser, "Autonomous Machines, Moral Judgement, and Acting for the Right Reasons," *Ethical Theory and Moral Practice* 18(4) (2015), 851-852. 던컨 퍼브스는 그의 글에서 주로 "자동 무기 체계(autonomous weapon systems)"에 대한 논란의 여지를 지적한다.

22) 위의 책, 852.

하며, 각각 학문적 영역을 통하여 궁극적으로는 인간 정신의 발전을 도모한다는 공동적인 목표를 설정한다. 인공지능과 인간지능의 메커니즘 상의 상관관계를 긍정적으로 파악하며, 양자 공히 기계적 유한성과 실존적 유한성의 특질을 가진다고 본다. 따라서 인공지능의 발전에 맞추어 인간지능이 나아가야 할 바람직한 모형론 개발을 도모하여 인간의 창의성 개발에 인공지능이 큰 역할을 할 수 있으며, 인간의 덕 윤리의 신장에도 크게 도움이 될 수 있다는 관점이 대화유형이라고 할 수 있다. 다음에서 이러한 대화를 모색하는 이론가들의 관점을 살펴보자.

이언 바버(Ian G. Barbour)는 인공지능 연구에 대한 철학적, 신학적 논의는 인간의 본질이 무엇인가에 대한 논의에 집중해야 하며, 그러한 관점에서 통전적 인간관을 통해, 인공지능 연구에 대하여 신학적, 철학적인 성찰이 필요하다고 강조한다.[23] 바버는 "[인공지능이란] 로보틱스 분야의 형태로서 각 시스템을 둘러싼 환경과의 상호작용을 통해 학습하는 내연 시스템(embodied systems)을 이용하는 [연구분야]"[24]라고 정의하면서, "로봇에게서 감정이나 사회화, 의식 같은 것들에 대한 가능성들이 여전히 문제로 남아있다"[25]고 지적한다.

바버가 제시하는 통전적 인간관은 인간은 "다층적인 심신 연합체로서, 생물학적 유기체인 동시에 응답적인 주체"[26]라고 정의한다. 인간의 다층적 심신 연합 구조를 상정하면서, 바버는 "유물론과 심신이원론"[27]의 맹점을 극복할 수 있다고 주장한다.[28] 바버는 과학과 종교, 또는 과학과 철학

23) Ian G. Barbour, "Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections," *Zygon* 34(3) (Sep. 1999), 361.
24) 위의 책, 361.
25) 위의 책, 361.
26) 위의 책, 362.
27) 위의 책, 362.
28) 위의 책, 362.

사이의 상호작용 및 관련성을 지적하는데, 바버는 철학과 과학이 과학으로 하여금 과학 연구의 맥락의 확장을 가능하게 한다면, 과학은 철학과 신학에 영향을 주어, 과학 자체를 해석하는 방식에 영향을 주는 한편, 철학과 신학이 변혁할 맥락을 또한 제공한다. 따라서 과학과 신학/철학은 각각을 위한 맥락으로서 상호 작용한다고 바버는 주장한다.[29]

대화모델로서 인공지능에 대한 철학적, 신학적 성찰에 대하여, 바버는 인공지능 연구의 목적이 "지적 컴퓨터를 창조하는 것과 인간의 뇌 기능이 어떻게 작용하는지에 대하여 이해하는 것"[30]이라고 설명한다. 바버는 인공지능이 인간의 본성에 얼마나 가까워 질 수 있는지에 대한 예시로서 상징적 인공지능(symbolic AI)을 드는데, 이 개념은 인간의 정신적, 인지적 능력을 정보를 통해 정리함으로써, 물리적 법칙에 환원되지 않는 정보를 상정하여 인공지능의 가능성을 강조하는 입장이다.[31]

이언 바버가 관심을 가지는 '상징적 인공지능'은 앨렌 뉴웰(Allen Newell)과 허버트 사이먼(Herbert A. Simon)의 개념으로, 뉴먼과 사이먼은 모든 지적 행위의 근간은 '정보'라고 가정하면서, "모든 정보는 [특정한] 결과를 도출하는 방식에 있어서 컴퓨터를 통해 처리되며, 우리는 한 체계의 정보를 측정하는 데에 있어서 그 체계가 [어떤 목적을 도출하기 위한 정보가 처리되는] 작업 환경에 의해 발생하는 다양한 변화와 어려움, 복잡성 등에 직면하여 원하는 목적을 달성할 수 있는 능력을 파악해야 한다."고 주장하며, 이러한 입장에서 지적 능력은 정보라는 비물리적인 대상을 바탕으로 형성되고 구성된다고 주장한다.[32]

---

29) 위의 책, 362.

30) 위의 책, 375.

31) 위의 책, 375. 이언 바버는 앨렌 뉴웰(Allen Newell)과 허버트 사이먼(Herbert A. Simon)을 본문에서 언급하며, 출처로 "([1976] 1990)"으로 표기하였다.

32) 필자는 이언 바버가 언급한 앨렌 뉴웰(Allen Newell)과 허버트 사이먼(Herbert A. Simon)의 책을 확인하여 재인용한다. Allen Newell, and Herbert A. Simon, "Computer Science

노린 허츠펠드(Noreen Herzfeld)는 신학적으로 인공지능을 완전히 신학에 적용하기는 어렵다고 보는데, 이러한 관점은 지능에 대하여 어떠한 전제가 인공지능 연구에 들어있는지, 기존의 인공지능 연구 모델에 대한 분석을 통해 제시된다. 기독교적 인간론에 있어서 중요한 영혼의 개념이 인공지능 컴퓨터에 적용될 수 있는가라는 질문을 제시하면서, 허츠펠드는 인공지능의 주체성을 다룬다.[33]

허츠펠드는 일단 지능을 문제 해결 능력, 내연 지능, 관계적 지능으로 구분하여 설명하는데, 먼저 문제 해결 능력을 중시하는 지능에 대한 연구는 대중적으로 가장 잘 알려진 인공지능에 대한 관점으로서, 인간 지능은 "기본적인 일련의 사실들을 기반으로, 그러한 일련의 규칙에 따라 더 복잡한 개념으로 결합됨으로써 재현되는 것"[34]이라고 본다. 이러한 관점에서 구성된 지능 개념에 따르면, 인간의 지능과 인공지능은 메커니즘 상 유사하다고 볼 수 있지만, 그러나 인간의 지능은 의식과 무의식을 통하여 인간의 지식을 활용하여 자신의 지위와 사회적 신분에 변화를 가져올 수 있지만, 인공지능은 단지 입력된 문제해결에만 국한된 것이 아닌가에 대한 비판이 뒤따른다.[35]

---

as Empirical Inquiry: Symbols and Search," in *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden (New York and Oxford: Oxford University Press, 1990), 107.

33) Noreen Herzfeld, "Human and Artificial Intelligence: A Theological Response," in *Human Identity at the Intersection of Science, Technology and Religion*, ed. Nancy Murphy, and Christopher C. Knight (Burlington, VA: Ashgate, 2010), 117.

34) 위의 책, 118. 허츠펠드는 문제 해결로서 지능 개념과 인공지능 연구에 있어서 철학적 배경으로 비트겐스타인(Wittgenstein)과 화이트헤드(Whitehead)를 언급하는데, 별도의 재인용 각주는 없다.

35) 위의 책, 118-119. 허츠펠드는 한 예로서 1997년 인공지능 '딥 불루(Deep Blue)'가 당시 체스 세계 챔피언 게리 카스파로프(Gary Kasparov)를 이겼지만, 그 후 10년이 지난 다음 인공지능은 여전히 체스게임을 벗어나고 있지 못하지만, 카르파로프는 정치인이 되었으며, 심지어 러시아의 대통령 후보가 되었다는 점을 지적하면서, 인공지능의 한계를 꼬집는다. 같은 책, 118. 허츠펠드는 또한 허버트 드레이퍼스(Hubert Dreyfus)가 상징적 인공지능은 '퇴세 하는 연구 프로젝트'라고 비판한 것을 언급한다. Hubert Dreyfus, *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (New

허츠펠드가 인공지능의 한계를 지적하는 내연 지능 개념은 지능 그 자체는 "물질세계뿐만 아니라, 인간 공동체와의 상호작용"[36]을 전제하는 것이므로, 기본적으로 지능은 항상 육체적 특성과 함께 결부되어 정의되고 기능하며, 그러한 의미에서 내연적이라고 할 수 있다. 그렇기 때문에 인공지능을 개발하는 데에 있어서도 인간 지능과 관련되려면, 인공지능에게도 어떤 육체의 형태가 제공되어야만 진정한 의미의 지능으로서 정의된다고 보는 입장이다.[37]

허츠펠드에 따르면, 문제해결이나 내연적 지능 개념은 인공지능이 얼마나 자연적인 인간 지능으로 구성될 수 있다는 가능성에 초점을 두고, 또 그 가능성에 대한 근거로 이용된다면, 관계적 지능은 인공지능의 개념만으로는 설명할 수 없는 인간의 자연적 지능에 대한 독특성을 제공한다고 본다. 관계적 지능의 핵심은 "개별 지능 개념은 무의미하며, 지능은 단지 관계적인 만남을 통해서만 의미를 가진다."고 보는 것이다.[38] 따라서 인공지능이 관계적 지능을 획득하기 위해서는 인공지능을 보유한 인공적 주체가 관계적 특성을 가질 수 있는지에 대한 논의를 요구하는데, 허츠펠드는 "인간 실존의 본질적 특징은 항상 두 개의 불가분의 요소들을 포함하는데, 그것은 자기 초월적 정신과 유한한 피조적 실존"이라고 보며, 인간의 지능에 대한 이해도 이러한 두 가지 특징을 함께 논의할 때에만 신학적으로 유의미한 인간론이 형성된다고 본다.[39]

W. F. 로우리스(W. F. Lawless)와 도널드 소프게(Donald A. Sofge)는 인공지능 연구의 위협성에 대한 경계에도 불구하고, 사실상 현재까지 인

York: Simon & Schuster, 1988), 29에서 재인용.

36) 위의 책, 119.

37) 위의 책, 119.

38) 위의 책, 125. 이와 같은 이론의 배경으로 허츠펠드는 Turing, Damasio, Winograd, and Flores를 언급하는데, 별도의 출처 인용은 없다.

39) 위의 책, 128.

공지능 수준은 인간 대 인간의 상호작용을 제대로 이해하고, 모형화 할 수 있는 수준에 이르지는 못했다고 지적한다.[40] 로우리스와 소프게에 따르면, 인간의 지능과 사회성에 근접한 인공지능 수준이 확립되기 위해서는 상호독립성 모델이 정립되고 적용되어야 하는데, 인간의 사회성이나 개인보다는 개인 간의 상호작용을 통해 자율성 개념이 성립되는 것처럼, 인공지능 로봇 또한 최대치의 생산성을 획득하기 위해서는 개체 자율성보다는 개체 간의 상호독립성을 보장하는 체계가 확립되어야 한다고 본다.[41] 따라서, 로우리스와 소프게에 따르면, 근본적으로 인공지능 연구 자체는 인간의 행동과의 연관성을 통해 구성되는 만큼 인간의 자율성과 의사결정 과정, 사회적 관계 및 소통 등이 중요하다고 본다.[42]

스티븐 러셀(Stephen Russell), 아이라 모스코비츠(Ira S. Moskowitz), 아드리엔 래글린(Adrienne Raglin)에 따르면, 인공지능 연구 및 정보과학의 발전을 통해, 인간이 정보와 상호 소통해야 할 필요성이 증가할 것이라고 지적하며, 인공지능의 필요성은 인간이 정보와의 상호 소통에 있어서 제기되는 다양한 문제점들을 교정해야 할 문제들이 늘어나기 때문인데, 인공지능은 인간이 정보와 소통하는 데에 있어서 컴퓨터의 정보처리 능력을 통해, 처리 과정에 보조를 받는 경우로 볼 수 있다고 러셀 등은 설명한다.[43]

---

40) W. F. Lawless, and Donald A. Sofge, “Evaluations: Autonomy and Artificial Intelligence: A Threat or Savior?” in *Autonomy and Artificial Intelligence: A Threat of Savior?*, ed. W. F. Lawless, Ranjeev Mittu, Donald Sofge, and Stephen Russell (Switzerland: Springer International, 2017), 295.

41) 위의 책, 296.

42) 위의 책, 305.

43) Stephen Russell, Ira S. Moskowitz, and Adrienne Raglin, “Human Information Interaction, Artificial Intelligence, and Errors,” in *Autonomy and Artificial Intelligence: A Threat of Savior?*, ed. W. F. Lawless, Ranjeev Mittu, Donald Sofge, and Stephen Russell (Switzerland: Springer International, 2017), 71.

러셀 등은 "정보 이용에 있어서 나타나는 수많은 복잡성은 이미 인간이 간단하게 하는 컴퓨터를 사용한 계산과 인간의 분석에 대한 필요를 이미 넘어섰다"[44]고 지적하면서, 이러한 복잡성의 증가는 정보 처리과정에서의 실수를 포착할 수 없는 인간의 필요를 넘어서, 스스로 자율적으로 계산을 통해 문제를 발견하고 처리하는 인공지능의 정보 처리 능력에 대한 연구를 촉진했고, 인공지능은 그러한 측면에서 인간의 정보 처리 환경에서 도움을 주면서, 자율적으로 정보와 상호 소통하는 주체로 작용한다고 설명하고 있다.[45] 이러한 목적에 부합한 인공지능이 구축되고 개발되기 위해서는 인간이 정보를 어떻게 처리하는지에 대한 근본적 이해가 필요하며, 그렇기 때문에 인간에 대한 다각적 이해가 인공지능 연구에 있어서 없어서는 안 된다고 러셀 등은 주장한다.[46]

펭 타오(Feng Tao) 등에 따르면, 인공지능에 연관된 다양한 과학 기술의 발전은 인간 지능과 인공지능 사이의 상호 보완의 가능성을 제시한다고 본다. 인간과 컴퓨터 사이의 상호작용은 인간의 미적 기준과 창조성에 대해 인공지능의 분석적 능력이 결합함으로써 새로운 형태의 예술적 확장이 가능한데, 이를 위해 펭 타오 등은 창의성에 대한 재정의를 요청한다. 이들은 창의성은 참신함과 독창성의 통합적 개념으로 보며, 이러한 관점에서 보면, 인공지능의 예술적 창의성도 그 자체로 독특성을 가진다고 주장한다.[47]

---

44) 위의 책, 72.
45) 위의 책, 72-73.
46) 위의 책, 87.
47) Feng Tao, Xioahui Zou, and Danni Ren, "The Art of Human Intelligence and the Technology of Artificial Intelligence: Artificial Intelligence Visual Art Research," in *Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018 Beijing, China, November 2-5, 2018 Proceedings I*, ed. Zhongzhi Shi, Cyriel Pennartz, and Tiejun Huang (Switzerland: Springer, 2018) (eBook), 146. https://doi.org/10.1007/978-3-030-01313-4. [2019.04.13. 최종접속].

매트 카터(Matt Carter)는 인간의 지능이 인공적으로 실제 개발될 것인가의 문제는 과학 기술을 넘어, 철학적 논의의 주제라고 주장한다.[48] 매트 카터는 인간의 기본적 인지적 능력에 있어서 인간만의 고유한 정신적 능력 중에 합리적 추론 능력과 언어 사용을 통해, 이러한 능력이 어떻게 컴퓨터적 연산처리 기술과 연관될 수 있는지에 대한 논의가 필요하다고 주장한다.[49] 물론 카터의 대화적 모델은 철저히 인간의 인지적 능력에 대한 올바른 이해가 이루어질 때, 비로소 인공지능 개발의 가능성이 열린다는 입장이다.[50]

도덕 및 종교 철학자인 징롱 펭(Zilong Feng)에 따르면, 덕윤리의 관점에서 인공지능은 점차 인간과 같은 도덕 주체로 변하게 되는데, 이러한 도덕주체로서의 인공지능의 "책임성은 자율성과 민감성의 증가와 더불어"[51] 논의의 주제로 되고 있다고 설명한다. 물론 이러한 윤리적 주체로서 인공지능의 이해가 인간과 인공지능을 동일한 선상에서 논의한다는 뜻은 아니고, 인공지능을 위한 윤리적 규범이 어떻게 규정되는지에 대한 독립적 연구를 포함하는데, 펭은 특히 사회 통합의 측면에서, 인공지능은 추상적 윤리 개념이 아니라, 인공지능의 자기 학습 능력에 기반한 윤리 체계를 바탕으로, 그것이 어떻게 인간의 윤리와 상호작용하는지에 대한 논의가 필요하다고 주장한다.[52]

---

48) Matt Carter, *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence* (Edinburgh: Edinburgh University Press, 2007), 1.

49) 위의 책, 2.

50) 위의 책, 3.

51) Zilong Feng, "Does AI Share Same Ethic with Human Being? From the Perspective of Virtue Ethics," in *Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018 Beijing, China, November 2-5, 2018 Proceedings*, ed. Zhongzhi Shi, Cyriel Pennartz, and Tiejun Huang (Switzerland: Springer, 2018) (eBook), 465. https://doi.org/10.1007/978-3-030-01313-4. [2019.04.13. 최종접속].

52) 위의 책, 465.

알프레도 콜로시모(Alfredo Colosimo)는 인공지능 개발은 단순히 정보 처리를 강화하는 기능적 측면뿐만 아니라, 시스템 생물학 등을 위한 도구적 기능을 보유한다고 주장하면서, 시스템 생물학적 측면에서 인구의 이동이나 환경적 요인이 미치는 경향에 대한 체계적 인과관계를 연구하기 위해서는 인공지능의 시뮬레이션 데이터가 필요할 뿐만 아니라, 시뮬레이션 가능한 상황을 실제로 재현함으로써 다양한 행위자들의 복잡한 역동성을 연구할 수 있다고 본다.[53]

생물공학자인 재키 채펠(Jackie Chappell)과 컴퓨터 공학자 닉 하위스(Nick Hawes)의 공동 연구에 따르면, 기존의 인지 연구의 경우, 동물 인지 기능을 통한 간접적 방식으로 관찰된 데이터를 바탕으로, 인간의 구조적, 생물학적 유사성에 근거한 이론을 제시했는데, 이러한 방식은 궁극적으로 부정확성을 가짐과 동시에, 실제 인간의 인지적 작용에 포함되는 다양한 가능한 요인들을 다각적으로 살펴볼 수 없다는 문제가 있다고 지적하면서,[54] 보다 정확하고 가능한 모델링을 위해서는 인공지능의 계획 기술을 바탕으로, 시뮬레이션이 필요하다고 주장한다.[55]

지금까지 필자는 이언 바버의 인공지능과 신학 간의 대화모델, 노린 허츠펠드의 인공지능의 피조적 실존 개념 적용가능성, 로우리스 등의 인공지능과 인간의 상호 독립성과 협력 체계 구축, 스티븐 러셀 등의 상호 정보를 통한 소통 체계 구축, 펭 타오 등의 상호 보완성을 통한 창의성 개발, 징

53) Alfredo Colosimo, “Multi-agent Simulations for Population Behavior: A Promising Tool for Systems Biology,” in *Systems Biology*, ed. Mariano Bizzarri (NY: Humana Press, 2018) (eBook), 308-309. https://doi.org/10.1007/978-1-4939-7456-6 [2019.04.13. 최종 접속].

54) Jackie Chappell, and Nick Hawes, “Biological and Artificial Cognition: What Can We Learn about Mechanisms by Modelling Physical Cognition Problems Using Artificial Intelligence Planning Techniques?,” *Philosophical Transactions: Biological Sciences*, 367(1603), (Oct.2012), 2724.

55) 위의 책, 2724.

롱 펭의 인공지능 자기학습을 통한 윤리적 체계 확충 가능성, 콜로시모 등의 시뮬레이션 학습체계를 통한 인공지능의 환경요인 분석능력 기능 개발 등에 대하여 살펴보았다. 이와 같은 이론들은 인공지능에 대한 두려움이나 기대와 같은 양가적인 감정을 넘어서 적극적으로 인공지능 연구가 인간과 공동체에 미칠 영향을 긍정적인 관점에서 접근하여, 궁극적으로는 인공지능을 통하여 인간성의 개발과 덕 윤리의 신장을 도모하고, 나아가 인간의 존재론을 항상 염두에 두는 바람직한 인공지능 연구 윤리를 확립하는 데에 큰 도움이 될 것으로 기대한다. 이제 다음에서 통합유형에 대하여 살펴보자.

V. 통합유형

대화유형이 인공지능연구에 대한 간학문적 소통의 체계를 긍정적으로 모색하는 방법론이라면, 통합유형은 한 걸음 더 나아가 인공지능 연구는 인간의 신경체계의 작동과 비슷한 의식수준에 도달하는 획기적인 과학적 혁명에 도달할 것이라는 청사진을 제시한다. 그렇게 되면 인간의 생활 전반에 걸쳐 인공지능 위에 건설되는 과학적 유토피아도 가능하다.

폴 샤프샥(Paul Shapshak)은 인공지능 연구의 패러다임이 인간 정신의 생물학적 특징이 컴퓨터나 다른 인공적 기능을 선험적으로 초월한다는 기존의 상식을 넘어서, 컴퓨터 기술이 인간 신경 기능 연구에도 크게 공헌하고 있음을 주장하면서, 인공지능은 단순히 하나의 인공적 구조물이 아니라, "다양한 인간이 만들어낸 방법들, 체계, 언어에 컴퓨터 공학 기술에 덧붙여진"[56] 분야로서, 뇌과학 연구와 인공지능 연구는 함께 통합적으로 이해해야 한다고 주장한다. 이를 위해 샤프샥은 "거울 뉴런 회랑"이라는 개

념을 소개하는데, 이 개념은 인공지능을 가능하게 하는 학습 알고리즘의 다층적 구조를 통해, 입력-처리-출력이라는 단순한 기능을 하는데, 이러한 단순성은 마치 주어진 정보를 처리하고, 이에 대한 판단과 실행이라는 합리적 추론, 또는 사고 (지능적 사고)와 유사하다.[57]

기존의 알고리즘은 컴퓨터 공학의 한 분야로서 인공지능은 항상 인간 신경계를 모방하는 인공적 구조물이라는 관점이 강했다면, 거울 뉴런 회랑을 통해 샤프샥은 뇌과학과 인공지능 연구 사이의 상호적 공헌 가능성에 더 집중한다고 주장하며, 뇌과학이 거울 뉴런을 연구함으로써 얻게 되는 인간 지능에 대한 특징은 인공 지능을 통해 드러난 복합적 알고리즘을 통한 시뮬레이션을 가능하게 하고, 이 시뮬레이션이 인간 신경 기능과 얼마나 일치하는지를 통해 간접적으로 인간의 신경 기능도 이해할 수 있는 가능성이 생긴다는 것이다.[58]

> 거울 뉴런은 뇌의 다양한 부위에서 발생하며, [뉴런 간의] 넓은 범위의 상호 접속을 통해, 특정한 감각-운동 활동들에 영향을 미치며, 그러한 활동을 통제하는 동시에 복사하고 재현한다. 거울 뉴런이 있는 여러 뇌 영역들은 [일반적으로] 배후 전운동 피질, 일차 운동 피질, 복측 전운동 피질의 구분 분절, 그리고 특히 두정엽의 외측하부 영역과 복측 두정엽 내 영역을 포함한다. 거울 뉴런은 다른 개체가 [자신과] 비교할 만한 행동, 또는 동일한 행동을 하는 것을 목격할 때뿐만 아니라, 다른 개체가 특정한 운동 행위를 할 때 활성화되거나 그러한 감각활동에 포함된다.[59]

샤프샥은 인공지능 연구 및 뇌과학 연구가 서로 통합함으로써, "거울 뉴

56) Paul Shapshak, "Artificial Intelligence and Brain," *Bioinformation* 14(1) (Jan, 2018), 38.
57) 위의 책, 39.
58) 위의 책, 39.
59) 위의 책, 39.

런 회랑"이라는 형태의 패러다임이 구축된다고 보는데, 거울 뉴런의 작용에 대한 과학적 데이터가 쌓일수록 그 기능이 인간 인지와 의식에 얼마나 큰 역할을 하는지를 통해, 이론적인 인지 능력에 대한 연구에 경험적이며 관측 가능하고 유의미한 자료를 획득하는 한편, 특정한 거울 뉴런이 단독적으로 작용하는 것이 아니라, 다양한 거울 뉴런 군의 연결 및 상호작용이 마치 거울 회랑처럼, 뇌 활동 전체를 비추며, 통합적이며 포괄적으로 작용한다고 주장한다.[60]

캐서린 테시어(Catherine Tessier)에 따르면, 로봇의 자율성은 항상 로봇과 인간 조정자, 그리고 로봇-인간조정자 간 상호작용의 세 가지를 통해서 구축된다고 보며, 그러한 관점에서 인공지능의 자율성은 위험한 것으로 간주하는 것은 인간중심적 관점에 지나지 않는다고 지적한다.[61] 테시어는 인공지능 등 로봇의 자율성에 대한 재정의가 필요하다고 보며, 자율성은 "로봇의 의사 결정 기능과 인간 사이의 공유되는 권한의 틀 안에서 구성되는 상대적 개념"[62]이라고 정의한다. 테시어는 로봇은 "컴퓨터에 의하여 조정되며, 물리적 공간에서 움직이는 기계"라고 정의하며, '자율적'이란 표현을 쓰기 위하여서는 "세상과 자신, 그리고 상황을 이해하면서" "자신의 이해에 근거하여 목적을 수행하기 위한 다양한 행위의 [선택적] 유형을 독립적으로 조합하고 선택하는 능력을 갖추어야만"하는 것을 전제한다고 강조한다.[63]

---

60) 위의 책, 39-40. 물론 샤프샥은 거울뉴런 체계가 그대로 인공지능 연구에 동일하게 적용될 수 있는지에 대한 연구가 계속되어야 한다는 점을 지적한다. 그러나 뇌기능을 특정한 뉴런에 개별적 작용에 한정하지 않고, 거울 뉴런 간의 확장적 작용에 대한 논의를 제공한다는 점에서, 거울 뉴런 회랑 개념이 통합적 모델의 중요한 기틀이 된다고 볼 수 있다. 같은 논문, 40.

61) Catherine Tessier, "Robot Autonomy: Some Technical Issues," *Autonomy and Artificial Intelligence: A Threat of Savior?*, ed. W. F. Lawless, Ranjeev Mittu, Donald Sofge, and Stephen Russell (Switzerland: Springer International, 2017), 179.

62) 위의 책, 179.

차운페이 친(Chaunfei Chin)에 따르면, 인공지능이 인간의 인지능력과 같은 수준에 도달하여 인간적 수준의 의식을 가질 수 있는가를 따지는 연구(강한 인공지능, 또는 강한 인공적 의식, strong artificial consciousness)에 대하여 기존의 철학적 논의는 주로 인공지능이 인간 수준에 도달할 수 없다는 불가능에 초점을 두었다고 지적한다.[64] 친에 따르면, 이러한 불가능론자들의 논리는 기본적으로 의식은 인공적으로 구성될 수 없는 본질적 근간이 있다고 보고, 의식을 환원할 기본적 단위를 구성할 수 없기 때문에, 인공적 의식을 가진 기계를 만들 수 없다는 전제에 근거한다.[65] 그러나 차운페이 친은 의식에 대한 철학적 관점을 경험적이며 인식론적 틀을 벗어나, 현상적 틀에서 보면, 인공지능에 의한 의식은 다양한 현상적 수준의 복합적 작용으로 설명 가능하며, 그러한 의미에서 인공지능이 인간 수준의 의식에도 이를 수 있는 가능성이 있다고 주장한다.[66] 따라서 이러한 의미의 관점의 전환을 통해, 인공지능에 대한 철학적 인식에 나타나는 세 가지 문제, 즉 "설명적 수준의 문제, 주관성의 문제, 그리고 [인공지능의 의식을

63) 캐서린 테시어는 로봇은 라우몬드(Laumond, 2012)의 해석을 빌린다. 위의 책, 179. 그리고 자율성의 개념은 미국 '국방과학 위원회(Defence Science Board, 2016)'의 정의를 빌리고 있다. 위의 책, 181에서 재인용. Jean-Paul Laumond, *La robotique: une recidive d'Hephaistos*, Lecon inaugurale prononce au, (College de France, 2012); Department of Defense, *Summer Study on Autonomy* (2016), PDF, https://www.hsdl.org/?abstract&did=794641 [2019.03.15. 최종접속] 참고로 보리스 갈리츠키(Boris A. Galitsky)와 애나 파르니스(Anna Parnis)는 자폐증 어린이와 기계의 학습 작용의 유사성을 통해 인공지능과 인간 연구가 통합적으로 이루어져야 한다고 주장한다. Boris A. Galitsky, and Anna Parnis, "How Children with Autism and Machines Learn to Interact," *Autonomy and Artificial Intelligence: A Threat of Savior?*, ed. W. F. Lawless, Ranjeev Mittu, Donald Sofge, and Stephen Russell (Switzerland: Springer International, 2017), 195.

64) Chaunfei Chin, "Artificial Consciousness: From Impossibility to Multiplicity," *Philosophy and Theory of Artificial Intelligence 2017*, ed. Vincent C. Müller, (Switzerland: Springer Nature, 2018), 4. 차운페이 친은 Bishop, Gamez, Mcdermott 등 많은 학자들의 이론을 분석하여 이와 같은 함의에 이르는데 필자는 재인용은 생략한다. 재인용 출처는 다음을 참고하시오. 같은 책, 16-18.

65) 위의 책, 5-6. 차운페이 친은 Block과 Chalmer의 이론에 대하여 반대하는 관점을 취하는데 재인용 출처는 같은 책을 참조하시오, 16-18.

66) 위의 책, 15.

기본적인 의식에 포함하는] 새로운 분류법의 도덕적 중요성"[67]을 해결할 수 있다고 주장한다.

아팀타예바(L. Atymtayeva), 코자크메트(K. Kozhakhmet), 보르트소바(G. Bortsova) 등에 따르면, 인간의 지능에 대한 연구가 인공지능의 연구와 통합적으로 이루어져야 하는 이유를 정보 보안을 확보하기 위함인데, 아팀타예바 등의 연구팀은 인공지능을 이용할 경우, 정보 보안을 위한 작업이 더욱 저비용의 효율적으로 이루어질 것이라고 전망한다.[68] 아팀타예바 등은 이론적으로 이를 위한 지식 기반이 확보되어야 한다고 보는데, 이러한 지식 기반은 인간 인지의 특정 도메인의 구획으로 구분된 특징과 마찬가지로, 특정 정보 처리 과정에 특화된 각 도메인으로서 지식 기반을 의미하며, 이렇게 할 경우, 정보 처리에 대한 과정과 보안을 필요로 하는 정보에 대한 보호가 동시에 효율적으로 가능할 것이라고 주장한다.[69]

켄지 이와다테(Kenji Iwadate), 이쿠오 스즈키(Ikuo Suzuki), 미치코 와타나베(Michiko Watanabe), 마사히토 야마모토(Masahito Yamamoto), 마사시 후루카와(Masashi Furukawa) 등에 따르면, 인공지능과 인간의 자연적 소통 능력 사이에는 분명한 차이가 존재하지만, 인공지능이 기존의 방식대로 주어진 정보에 대한 처리를 수동적으로 처리하는 형태를 넘어, 주체적으로 정보를 처리하고 학습에 있어서 그것을 통제할 능력을 얻기 위해서는 인간의 신경 체제를 모방하는 형태의 정보 처리 체계를 구축해야 한다고 주장한다.[70]

---

67) 위의 책, 15.

68) L. Atymtayeva, K. Kozhakhmet, and G. Bortsova, "Building a Knowledge Base for Expert System in Information Security," *Soft Computing in Artificial Intelligence*, ed. Young Im Cho, Donghan Kim, and Eric T. Matson, (Switzerland: Springer International, 2014), 57.

69) 위의 책, 59.

70) Kenji Iwadate, Ikuo Suzuki, Michiko Watanabe, Masahito Yamamoto, and Masashi Furukawa, "An Artificial Neural Netword Based on the Architecture of the Cerebellum

지금까지 필자는 폴 사프샥이 강조하는 인공지능과 뇌과학의 통섭, 캐서린 태시어의 인공지능 자율 가능성, 차운페이 친의 인공지능의 인간의식 도달 가능성, 아팀타예바 등의 인공지능을 통한 정보 보안의 체계적 확립, 그리고 켄지 이와다테 등의 인공지능의 인간 신경체계 모방 가능성 이론들을 통하여 인공지능과 인간지능과의 '통합유형'에 대하여 살펴보았다. 이제 다음에서 결론으로 위 네 가지 유형에 대한 신학적 관점, 특히 기독교윤리학적 입장에서 몇 가지 제언을 하고자 한다.

VI. 결론

필자가 인공지능 연구분야의 복합적이며 간학문적 특성에 대한 이해가 부족한 상황에서 유형화한 시도는 한계가 있다. 그러나 최근 인공지능 연구는 지적 주체인 인간의 다양한 영역을 인공적으로 구성하는 것을 넘어, 인간의 정신을 정보의 형태로 변환하거나 정보를 인간의 생체적이고 물리적인 형태로 재현하는 수준에 이르고 있다. 따라서 인간에 대한 인공지능 분석을 통하여 인간론을 중시하는 신학과의 대화적 모델로서 앞의 네 가지 유형론을 제시하는 본 글은 나름 유의미하다고 본다. 필자가 이 글을 통하여 기독교윤리학의 관점에서 제시하고자 하는 제언은 다음과 같다.

첫째, 인공지능연구와의 간학문적 관점에서 신학적 인간학의 연구가 매우 중요하다고 본다. 인간의 자유의지와 그 한계를 지적하는 신학적 인간학은 하나님의 창조성에 대한 인간의 피조성을 지적하는 데에 있다고 본다. 인공지능에게 부여하는 자율성의 한계와 책임, 또한 기계 윤리의 확

for Behavior Learning," *Soft Computing in Artificial Intelligence*, ed. Young Im Cho, Donghan Kim, and Eric T. Matson, (Switzerland: Springer International, 2014), 143.

립은 신학적 인간학이 기여할 매우 중요한 영역이라고 판단한다. 인공지능의 알고리즘이 인간의 도덕적 인식을 대신할 수 있다는 인공지능 주체성도 강조되는 현 시점에서, 자율과 의지, 윤리와 도덕, 그리고 책임과 법의 사안들에 대하여 기독교윤리학은 관심을 갖고 지속적인 간학문적 연구가 필요하다고 본다.

둘째, 인간의 몸을 이식하는 로보틱스의 연구와 기술은 트랜스휴머니즘을 이끌고 있는데, 인간생체에 도움이 되는 이런 획기적인 발전은 전통적인 인간의 몸(육신)에 대한 사고의 변화를 요구할 수밖에 없다고 본다. 이런 맥락에서 육체의 부활과 하나님의 나라에 대한 믿음이 흔들리지 않으면서도 과학주의에 함몰되지 않는 기독교세계관의 구축이 필요하다고 본다. 하나님의 창조와 과학기술의 창발성에 대한 이론적 통섭의 가능성이 없지는 않지만, 그렇다고 기계가 인간의 생체를 대신할 수 있을 것이라는 과학만능주의에 대하여 필자는 부정적인 입장이다. 하나님의 나라에 대한 종말론적 소망은 이 세계를 무의미화 하는 것이 아니라, 그 유한성을 하나님의 형상을 통하여 회복하며, 이것이 또한 복음의 본질이기 때문이다. 따라서 초지능 세계의 도래를 바라보며, 그 초지능이 인간을 지배할 것을 두려워하는 것이 아니라, 그 초지능을 개발하고자 하는 인간의 과학주의가 인간을 파멸로 이끌지 않도록 4차산업의 혁명에 비견하는 인간의 영적 혁명도 전제된다고 본다.

셋째, 신학과 신앙공동체의 역할이 중요하다고 본다. 본문에서도 살펴보았지만, 인공지능은 '세계 위기 기구'에서도 밝혔듯이, 지구적인 위험요인이다. 지구 멸망의 가능성이 초지능에게 있다고 한다면, 그 일차적인 책임은 과학집단에 있으며, 그리고 동반책임은 그 공동체를 관리 감독하는 국가에게 있을 것이다. 그럼에도 불구하고 그러한 위험을 전가하지 않으면서도 과학집단의 주체가 인간 연구자들임을 감안하여 신학공동체는

과학연구자들과의 지속적인 대화가 필요하다고 본다. 전인간적 신앙교육과 인간 인지의 개발에 대한 책임이 교회와 신학 공동체에도 있음을 직시하고, 과학 공동체와 공공영역에서 협력을 도모하여야 할 것이다. 이론적으로는 인공지능과의 유형론에서 '대화유형'이 그 역할을 일부분 감당할 수 있으리라고 본다. 특히 인공지능과의 간학문적 통섭을 통하여 덕 윤리의 발전을 위하여 애쓰는 신학/교회 공동체가 되기를 기대한다.

참고문헌

Armstrong, Stuart, and Kaj, Sotala. “How We’re Predicting AI - or Failing to.” *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*. Edited by Jan Romportl, Eva Zackova, and Jozef Kelemen. Switzerland: Springer International, 2015.

Barbour, Ian. “Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections.” *Zygon* 34(3) (Sep 1999). 361-398.

Bizzarri, Mariano, eds. *Systems Biology*. NY: Humana Press, 2018. (eBook) https://doi.org/10.1007/978-1-4939-7456-6 [2019.04.13. 최종접속].

Boddington, Paula, Millican, Peter, and Wooldridge, Michael. “Minds and Machines Special Issue: Ethics and Artificial Intelligence.” *Minds and Machines* 27(4) (2017). 569-574.

Boden, Margaret A. *AI: Its Nature and Future*. NY and Oxford: Oxford University Press, 2016.

_____. eds. *The Philosophy of Artificial Intelligence*. NY and Oxford: Oxford University Press, 1990.

Bostrom, Nick. “The Super Intelligence Will: Motivation and Instrumental Relationality in Advanced Artificial Agents.” *Minds and Machines* 22(2) (May 2012). 71-75.

Bostrom, Nick, and Ćirković, Milan M., eds. *Global Catastrophic Risks*. NY and Oxford: Oxford University Press, 2008.

Carter, Matt. *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh: Edinburgh University Press, 2007.

Chappell, Jackie, and Hawes, Nick. “Biological and artificial cognition: what can we learn about mechanisms by modelling physical cognition problems using artificial intelligence planning techniques?.” *Philosophical Transactions: Biological Sciences* 367(1603) (Oct 2012). 2723-2732.

Cho, Young Im, Kim, Donghan, and Matson, Eric T., eds. *Soft Computing in*

Artificial Intelligence. Switzerland: Springer International, 2014.

Fisher, Christopher L. *Human Significance in Theology and the Natural Science: An Ecumenical Perspective with Reference to Pannenberg, Rahner, and Zizoulas*. Eugene, Oregon: Pickwick, 2010.

Global Challenges Foundation. "Global Catastrophic Risks 2018." https://globalchallenges.org/our-work/annual-report/annual-report-2018 [2019.04.13. 최종접속].

Holland, John H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA: the MIT Press, 1992.

Lawless, W. F., Mittu, Ranjeev, Sofge, Donald, and Russell, Stephen, eds. *Autonomy and Artificial Intelligence: A Threat of Savior?*. Switzerland: Springer International, 2017.

Mercer, Calvin, and Trophen, Tracy J., eds. *Religion and Transhumanism: The Unknown Future of Human Enhancement*. Santa Barbara, Denver, and Oxford, England: Praeger, 2015.

Müller, Vincent C., eds. *Philosophy and Theory of Artificial Intelligence 2017*. Switzerland: Springer Nature, 2018.

Murphy, Nancy, and Knight, Christopher C. *Human Identity at the Intersection of Science, Technology and Religion*. Burlington, VA: Ashgate, 2010.

Purves, Duncan, Jenkins, Ryan, and Strawser, Bradley. "Autonomous Machines, Moral Judgement, and Acting for the Right Reasons." *Ethical Theory and Moral Practice* 18(4) (2015): 851-872.

Ringle, Martin. "Mysticism as a philosophy of artificial intelligence." *Behavioral and Brain Sciences* 3(3) (Sep 1980): 444-445.

Romportl, Jan, Zackova, Eva, and Kelemen, Jozef, eds. *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*. Switzerland: Springer International, 2015.

Russell, Robert J., Murphy, Nancy, Meyering Theo C., and Arbib, Michael A., eds. *Neuroscience and the Person: Scientific Perspectives on Divine Action*.

Berkeley, CA: Center for Theology and the Natural Sciences, 2002.

Shapshak, Paul. “Artificial Intelligence and Brain.” *Bioinformation* 14(1) (Jan. 2018): 38-41.

Shermer, Michael. “Apocalypse AI.” *Scientific American* 316(3) (Mar 2017): 77-77.

Shi, Zhongzhi, Pennartz, Cyriel, and Huang, Tiejun, eds. *Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018 Beijing, China, November 2-5, 2018 Proceedings*. Switzerland: Springer, 2018. (eBook) https://doi.org/10.1007/978-3-030-01313-4 [2019.04.13. 최종접속].

Simeonow, Plamen L., Smith, Leslie S., and Ehresmann, Andrée C., eds. *Integral Biomathics: Tracing the Road to Reality*. Berlin and Heidelbarg: Springer-Verlag, 2012.

Tonkens, Ryan. “A Challenge for Machine Ethics.” *Minds and Machines* 19(3) (2009). 421-438.

Abstract

Artificial Intelligence and Christian Ethics: A Dialogue between Theology and Artificial Intelligence Research

Kyoung-dong Yoo, Ph. D.
Department of Ethics and Society
Methodist Theological University

Research on the artificial intelligence, going beyond the field of engineering, is emerging to be a very important subject in theology as well. With the ontological research of the idea of God's image, a traditionally important subject in theology, and on the idea of human image given to the artificial intelligence at its forefront, autonomy, will, moral, responsibility of artificial intelligence, machine ethics of artificial intelligence and issues of law are now all relevant.

The author will endeavor to divide the recent researches on the artificial intelligence into four topics of independence, conflict, dialogue, integration and examine the possibility of consilience between theology and artificial intelligence. The theories of the english speaking researchers will be categorized in order to establish a bridgehead for interdisciplinary communication of researches related to theology and artificial intelligence,

with a hope that follow-up studies will develop.

Since this article, with the interest on the aforementioned typologies, takes macroscopic approach towards the relationship between the artificial intelligence and theology, a limit to the depth of presentation for each researcher's research exists, and hopes to analyze with more microscopic topics later on.

【Key Words】

Artificial Intelligence, Typologies (independence, conflict, dialogue, and integration), Theology, Christian Ethics, Autonomy