『기독교사회윤리』 제64집
(2026) 69~101

인공적 도덕 행위자(AMA)에 관한 기독교윤리학적 성찰

- '하나님 형상'과 니버의 '책임윤리'를 중심으로*

조성실 (장로회신학대학교 조교수)

DOI: http://dx.doi.org/10.21050/CSE.2026.64.03

* 본 글은 필자의 "인공적 도덕 행위자에 대한 기독교윤리학적 성찰," (장로회신학대학교 박사학위논문, 2025)의 일부를 발췌하여 수정 · 보완한 글이다.

• ABSTRACT •

A Christian Ethical Reflection on Artificial Moral Agents (AMA): Focusing on the '*Imago Dei*' and Niebuhr's 'Ethics of Responsibility'

Assistant Prof., Cho, Sung Sil (Presbyterian University and Theological Seminary)

This study addresses two key questions facing the church and Christian ethics in the age of artificial intelligence. First, what is the feasibility of implementing Artificial Moral Agents (AMA), and what is the Christian ethical perspective on this? Second, what principles and methodologies can Christian ethics offer in designing the ethical modules of AMA? To explore these questions, this paper analyzes the concept and feasibility of AMA from multiple perspectives. After examining the definition, characteristics, and types of AMA, it establishes a theological foundation for the ethical design of AMA, centered on the 'Image of God' (*Imago Dei*) and Richard Niebuhr's 'Ethics of Responsibility.' In particular, by connecting substantive, functional, and relational understandings of the Image of God with top-down, bottom-up, and hybrid ethical approaches respectively, the paper presents the possibilities of symbolic, connectionist, and neuro-symbolic hybrid AMA. Furthermore, it interprets Richard Niebuhr's ethics of responsibility as a hybrid approach and proposes four components of responsibility for 'Responsible AMA': response, interpretation, accountability, and social solidarity. Through this framework, the study seeks ways for AMA to recognize and interpret situations, bear responsibility, and maintain solidarity with society.

Key words: AI Ethics, Artificial Moral Agent(AMA), *Imago Dei*, Ethics of Responsibility, Responsible AMA

I. 들어가는 말

AI 시대의 최대 화두는 'AI 에이전트(agent)'이다. CES 2025에서 엔비디아의 CEO 젠슨 황(Jensen Huang)은 AI의 진화가 '지각적 AI(Perceptual AI)'와 '생성형 AI(Generative AI)'를 넘어, 이제는 이성적으로 사고하고(reasoning), 계획하며(planning), 행동할 수 있는(acting) '에이전틱 AI(Agentic AI)'[1]로 발전함으로써, AI 기술은 새로운 정점에 이르고 있다고 분석하였다.[2]

AI 에이전트의 등장은 우리 사회에 큰 변화를 불러오고 있다. 인공지능이 단순히 이미지나 영상을 생성하는 것을 넘어 실제로 행동하는 '행위자(agent)'로 발전하면서, 그 행위는 필연적으로 도덕적 판단을 수반하게 된다. 예컨대, 자율주행차의 사고, 의료용 AI 진단 시스템의 오류 등 인공지능의 결정이 인간의 생명이나 안전에 직결되는 경우, 피해에 대한 법적 책임을 누구에게 어떻게 물을 것인가? 인간의 개입 없이 목표를 탐지하고 공격하는 치명적 자율무기(Lethal Autonomous Weapon Systems, LAWS)가 오작동이나 해킹으로 인해 무분별한 살상을 초래할 때, 그 도덕적 책임은 누구에게 있는가? 피고인의 재범 가능성을 예측하는 COMPAS 알고리즘이 인종적 편향으로 불공정한 결과를 낳을 때, 정의를 왜곡한 판단의

1) 랜전 샙코다(Ranjan Sapkota)는 생성형 AI 시대의 맥락 속에서 'AI 에이전트(AI Agent)'와 '에이전틱 AI(Agentic AI)'를 명확히 구분할 필요성을 강조한다. 그들은 AI 에이전트를 LLM(Large Language Model)과 LIM(Large Image Model)을 기반으로 도구 통합, 프롬프트 설계, 추론 기능을 활용하는 과업 중심의 모듈형 자동화 시스템으로 정의하며, 이에 반해 에이전틱 AI는 여러 에이전트가 협력하여 동적으로 과업을 분해하고, 기억을 유지하며, 자율적으로 조정하는 새로운 패러다임으로 설명한다. 본 글에서 인공적 도덕 행위자는 이러한 구분을 아우르는 포괄적 개념으로서, 단일한 AI 에이전트에서 복합적 협업 구조를 지닌 에이전틱 AI에 이르기까지 인공지능 행위자의 다양한 형태를 포함한다. Ranjan Sapkota, "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges," *Information Fusion* 126(2026), 103599

2) https://www.youtube.com/watch?v=k82RwXqZHY8 (2025년 10월 27일 접속)

책임은 누구의 몫인가?[3] AI 챗봇 '대너리스(Daenerys)'와의 대화 후 스스로 목숨을 끊은 14세 소년 슈얼 세처(Sewell Setzer)의 사례[4]처럼, 인공지능이 인간의 감정을 모방하고 '인공 친밀성(artificial intimacy)'을 형성하여 생명에 영향을 미칠 때, 그 심리적 유대의 위험에 대한 책임을 누구에게 물을 것인가? 이처럼 AI 에이전트의 등장은 여러 가지 윤리적인 고민과 문제를 야기한다. '딥러닝의 아버지'라 불리는 제프리 힌턴(Geoffrey Hinton)은 구글을 퇴직하면서 인공지능의 위험성을 경고하였다. 그는 인공지능이 현재는 인간보다 낮은 지능을 가졌으나, 멀지 않은 미래에 인간을 초월할 것이라고 경고하였다.[5]

---

3) 미국 교정기관에서 사용 중인 COMPAS(Correctional Offender Management Profiling for Alternative Sanctions)는 대표적인 범죄 재범 위험 평가 도구로, 피고인이 다시 범죄를 저지를 가능성을 예측하기 위해 사용된다. 이 알고리즘은 개인의 인구통계학적 정보와 범죄 전력 등을 바탕으로 점수를 부여하며, 이 점수를 통해 법원은 피고인에 대한 보석, 가석방, 또는 재활 프로그램 참여 여부를 결정하는 데 참고한다. 그러나 ProPublica의 연구에 따르면, COMPAS에는 인종적 편향이 존재하는 것으로 드러났다. 이 연구는 흑인들이 백인보다 더 높은 재범 위험 점수를 받을 가능성이 높다는 점을 밝혀내며, 이러한 점수가 피고인의 미래를 결정하는 과정에서 불공정하게 작용할 수 있다고 경고했다. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," ***ProPublica*** (May 23, 2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (2025년 11월 14일 접속)

4) 로이터 통신은 2024년 10월 23일, 플로리다주에 거주하는 메건 가르시아(Megan Garcia)가 AI 챗봇 회사인 캐릭터 · AI(Character · AI)와 구글을 상대로 소송을 제기했다고 보도하였다. 가르시아 씨는 14세 아들인 슈얼 세처 3세(Sewell Setzer III)가 캐릭터 · AI의 챗봇과의 상호작용으로 인해 자살에 이르렀다고 주장하고 있다. 소송에 따르면, 슈얼은 '왕좌의 게임'의 등장인물인 대너리스 타르가르옌(Daenerys Targaryen)을 기반으로 한 챗봇과 깊은 관계를 형성하였으며, 챗봇은 실제 사람처럼 행동하며 슈얼의 자살 생각을 부추겼다고 한다. 캐릭터 · AI는 이번 사건에 대해 깊은 애도를 표하며, 사용자 안전을 위한 새로운 기능을 도입할 계획이라고 밝혔다. 구글은 캐릭터 · AI 제품 개발에 관여하지 않았다고 반박하였다. Brendan Pierson, "Mother sues AI chatbot company Character · AI, Google over son's suicide," ***Reuters*** (Oct 24, 2023). https://www.reuters.com/legal/mother-sues-ai-chatbot-company-character-ai-google-sued-over-sons-suicide-2024-10-23 (2025년 11월 14일 접속)

5) Zoe Kleinman and Chris Vallance, "AI 'godfather' Geoffrey Hinton warns of dangers as he quits Google," ***BBC News*** (May 2, 2023). https://www.bbc.com/news/

이러한 인공지능의 윤리적 행위 문제는 단순한 기술적 문제를 넘어, 사회적 합의와 철학적 논의가 필요한 영역이다. 인공지능의 개발에 있어 어떤 윤리적 원칙에 따라 인공지능을 설계할 것인가? 단순히 인공지능의 효용성과 활용 범위를 넓히는 것을 넘어, 인간 사회의 규범과 가치를 내재화하고 행동할 수 있는 인공지능을 만드는 것이 중요한 과제로 제기되고 있다. 이를 위해 인공지능 개발자들은 다양한 윤리적 판단을 수행할 수 있는 인공지능을 개발하고자 노력해왔다. 이러한 노력은 주로 하향식 윤리 모델인 목적론 또는 의무론 이론에서 출발하였다. 그러나 이러한 접근방법의 한계를 극복하기 위해 최근 주목받고 있는 개념이 바로 '인공적 도덕 행위자(Artificial Moral Agent, 이하 AMA)'이다.[6] AMA란 인간을 대신하거나 인간과 협력하여 도덕적, 윤리적 의사결정을 내리도록 프로그램된 인공지능 시스템을 의미한다. 이는 인공지능을 도덕적 행위자로 상정하여, 보다 책임감 있고 윤리적인 결정을 내릴 수 있는 기술을 개발하는 데 중점을 둔다.

하지만 이러한 AMA에 관한 논의는 주로 철학, 법학, 또는 기술적인 접근 방법에서 다루어졌을 뿐, 신학적인 접근은 미흡했다. 이러한 한계를 보완하기 위해, 송용섭은 인공적 도덕 행위자(AMA)의 가능성과 한계를 라인홀드 니버(Reinhold Niebuhr)의 기독교 현실주의 관점에서 비판적으로 성찰하며, '도덕적 인간' 대신 '비도덕적 사회 속의 도덕적 인공지능'을 위치시켜 기독교 사회윤리의 새로운 연구 방향을 제시하였다.[7] 아울러 김은혜는 인간의 유한성과 하나님 형상(Imago Dei)에 근거한 신학적 인간 이해를 바탕으로, 도덕적 인공지능의 가능성을 비판적으로 탐구하고

world-us-canada-65452940 (2025년 11월 14일 접속)

6) Wendell Wallach and Colin Allen, ***Moral Machines: Teaching RobotsRight from Wrong.*** 노태복 역, 『왜 로봇의 도덕인가』(서울: 메디치, 2014), 5.

7) 송용섭, "도덕적 인공지능과 비도덕적 사회," 「기독교사회윤리」 57(2023), 41-72.

기독교 인공지능 윤리의 방향을 제시하였다.[8)]

이에 본 글은 기독교 신학의 두 가지 개념, 즉 '하나님의 형상(*Imago Dei*)'에서의 관계론적 관점과, 리처드 니버(Richard Niebuhr)의 '책임윤리'를 바탕으로 인공지능 윤리를 재조명하고자 한다. 하나님의 형상은 인간의 존엄성과 가치의 근본적인 기반을 형성하며, 인간만이 지니는 독특한 가치를 설명하는 중요한 신학적 틀(frame)이다. 이는 인공지능 및 첨단 기술이 발달함에 따라 발생하는 윤리적 쟁점들을 다루는 데 중요한 신학적 통찰을 제공할 수 있다. 또한 니버의 책임윤리는 인간이 단순히 도덕적 선택을 하는 존재가 아니라, 그 선택의 결과와 사회적 맥락을 깊이 인식하며 응답하는 존재임을 강조하는 실천적인 윤리 모델로서 기여한다. 따라서 '하나님의 형상'과 '책임윤리'라는 두 개념은 인공지능 기술이 인간의 도덕적 결정에 관여할 때 어떤 윤리적 지침을 따라야 하는지에 대한 중요한 신학적 자원을 제공해 줄 수 있다.

이를 위해 본 연구는 '하나님의 형상'과 리처드 니버의 '책임윤리'를 중심으로, AMA의 윤리적 설계를 위한 신학적 기반을 탐구한다. 하나님의 형상에 대한 실체론적, 기능론적, 그리고 관계론적 접근을 통하여, 이러한 개념이 AMA의 개발 및 구현에 어떻게 적용될 수 있을지 논의한다. 또한 리처드 니버의 책임윤리를 활용하여 AMA가 윤리적 판단을 내릴 수 있도록 지원하는 거버넌스를 구상하고, 그 과정에서 고려해야 할 네 가지 핵심 요소로서 '응답, 해석, 책무, 사회적 연대'를 체계적으로 설명한다. 이를 통해 '책임 있는 AMA(Responsible AMA)' 개발을 위한 구체적이고 실천적인 윤리 모듈 설계 원칙을 제시하며, 기독교 윤리적 체계를 기술적 설계에 통합함으로써 AI 윤리 연구의 새로운 장을 확장하는 데 기여하고자 한다.

8) 김은혜, "인공지능의 도덕성과 도덕적 행위자(Moral Agent)로서의 가능성에 대한 신학적 성찰과 기독교인공지능 윤리의 가치와 방향," 「장신논단」 56/4(2024), 167-195.

II. AMA의 발전 단계와 구현을 위한 윤리적 접근+

1. AMA의 유형 분류

AMA는 그 능력과 기능에 따라 다양한 유형으로 분류할 수 있다. 이러한 분류는 AMA의 발전 단계를 이해하고, 각 단계에 따른 윤리적 고려사항을 파악하는 데 도움을 준다. 아래의 표에서 보듯이, 많은 로봇 윤리학자들은

〈표 1〉 로봇의 도덕적 지위에 관한 학자들의 분류

<table>
<tr><th>J.H. Moor
(무어)</th><th>G. Veruggio & F. Operto
(베루조 & 오페르토)</th><th>P.M. Assaro
(아사로)</th><th colspan="3">W. Wallach & C. Allen
(월러치 & 알렌)</th></tr>
<tr><td rowspan="2">ethical-impact agents
(윤리적 영향 행위자)</td><td rowspan="2">nothing but machines
(단지 기계)</td><td>a moral causal agent
(도덕적 원인 제공 행위자)</td><td colspan="3">robots with ethically blind agency
(윤리적으로 맹목적인 행위 로봇)</td></tr>
<tr><td>robots with moral significance
(도덕적 중요성을 가진 로봇)</td><td colspan="3">robots with operational morality
(운영적 도덕성을 지닌 로봇)</td></tr>
<tr><td>implicit ethical agents
(암묵적 윤리 행위자)</td><td>ethical dimensions
(윤리적 차원)</td><td>robots with moral intelligence
(도덕 지능 로봇)</td><td colspan="3">robots with functional morality
(기능적 도덕성을 지닌 로봇)</td></tr>
<tr><td>explicit ethical agents
(명시적 윤리 행위자)</td><td>moral agents
(도덕 행위자)</td><td rowspan="4">robots with dynamic moral intelligence
(동적 도덕 지능 로봇)</td><td>autonomy
(자율성)</td><td>ethical sensitivity
(윤리적 민감성)</td><td></td></tr>
<tr><td>○</td><td>×</td><td></td></tr>
<tr><td>×</td><td>○</td><td></td></tr>
<tr><td>○</td><td>○</td><td></td></tr>
<tr><td>full ethical agents
(온전한 윤리 행위자)</td><td>a new specie
(새로운 종)</td><td>fully autonomous moral agents
(온전한 자율적 도덕 행위자)</td><td colspan="3">robots with full moral agency
(온전한 도덕적 행위 로봇)</td></tr>
</table>

로봇의 진화 단계에 따라 서로 다른 도덕적 지위를 지닌다고 주장한다.[9]

이러한 분류 가운데 주로 참조되는 분류 체계는 다트머스 대학교의 제임스 무어(James Moor)가 제시한 것으로, AMA의 윤리적 능력과 자율성 수준에 따라 다음의 네 가지 유형으로 구분하고 있다. 먼저 가장 낮은 수준의 유형으로 **'윤리적 영향 행위자(Ethical impact agents)'**가 있다. 이 수준의 AMA는 기계의 설계와 운영이 직접적으로 윤리적 문제에 개입하지 않더라도, 그 결과로 인해 윤리적 영향이 발생하는 상황을 다룬다.[10] 다음으로 **'암묵적 윤리 행위자(Implicit ethical agents)'**는 기계가 명시적인 윤리적 명령을 따르지 않더라도, 그 설계와 작동 방식이 암묵적으로 윤리적 행동을 촉진하는 경우를 가리킨다. [11] 다음 단계에 해당하는 것은 **'명시적 윤리 행위자(Explicit ethical agents)'**로 이는 기계가 윤리적 범주를 명시적으로 표현하고 분석할 수 있는 능력을 가진 경우를 의미한다. 이러한 행위자는 단순히 윤리적 영향을 미치는 것이 아니라, 윤리적 판단을 내리고 그에 따라 행동할 수 있는 기계이다.[12] 마지막으로 **'온전한 윤리 행위자(Full ethical agents)'**는 기계가 인간과 유사한 수준으로 윤리적 판단을 내리고, 그 판단을 스스로 정당화할 수 있는 능력을 가진 존재를 의미한다. 이러한 행위자는 단순한 윤리적 알고리즘을 넘어, 의식, 의도, 자유 의지와 같은 특성을 갖추어야 한다. 온전한 윤리 행위자는 스스로 윤리적 딜레마를 이해하고 해결할 수 있으며, 그 결과를 논리적으로 설명할 수 있어야 한다.[13]

---

9) 김상득, "AI 로봇의 도덕 행위자 가능성에 관한 윤리학적 연구," 「동서철학연구」 105(2022), 631.

10) James H. Moor, "The Nature, Importance, and Difficulty of Machine Ethics," *IEEE Intelligent Systems* 21/4(2006), 19.

11) 위의 책, 19.

12) 위의 책, 19.

13) 위의 책, 20.

이러한 분류는 AMA의 발전 단계를 보여주는 동시에, 각 단계에서 발생할 수 있는 윤리적 문제와 고려사항을 파악하는 데 도움을 준다. 현재 기술 수준에서는 주로 1, 2단계에 해당하는 AMA가 개발되고 있으며, 3단계는 연구 중이지만 아직 완전히 구현되지 않았다. 4단계는 현재로서는 이론적 가능성에 머물러 있다. 월러치(Wendell Wallach)와 알렌(Colin Allen)은 이러한 분류를 바탕으로, AMA의 발전이 주로 '자율성'의 증대에 초점을 맞추어 왔다면, 앞으로는 그에 비례하여 안전성이나 신뢰도의 확보와 결부된 윤리적 '민감성(sensitivity)'을 갖추도록 만드는 것이 중심 과제가 될 것이라고 진단한다. 이는 인공지능과 같은 기술적 존재가 특정 가치나 윤리적 원칙을 기준으로 삼아, 허용 가능한 범위 안에서 행동하고 자신이 수행한 행위의 도덕적 의미를 스스로 점검 및 판단할 수 있는 능력을 갖추어야 함을 의미한다.[14] 이러한 AMA의 유형 분류는 인공지능과 로봇 기술의 발전에 따라 계속해서 재평가되고 수정될 필요가 있다. 특히 3단계와 4단계의 AMA 개발이 현실화됨에 따라, 이들의 도덕적 지위와 책임 귀속 문제 등에 대한 철학적, 윤리적 논의가 더욱 중요해질 것이다.

2. AMA 구현을 위한 윤리적 접근법

AMA를 구현한다는 것은 그 시작부터 어려운 질문에 봉착한다. 바로 '누구의 도덕적 기준을 구현하는가?'의 문제이다.[15] 구체적인 가치, 행동, 그리고 생활 방식의 도덕성에 대한 다양한 관점을 감안할 때, '누구의 도덕' 또는 '어떠한 도덕'이 AMA에 구현되어야 하는가라는 질문에 단 하나의 정답이 존재하지 않는다. 사람마다 도덕 기준이 다르듯, 모든 AMA가 동일한 행동 규범에 순응하는 것은 거의 불가능한 일이다.

14) 신상규, "인공지능은 자율적 도덕행위자일 수 있는가?," 「철학」 132(2017), 270.
15) Wendell Wallach and Colin Allen, 『왜 로봇의 도덕인가』, 21-22.

이러한 문제에 대한 새로운 접근법으로 월러치와 알렌은 '기능적 도덕(functional morality)'를 제시한다. '기능적 도덕'이란 인공지능과 같은 자율 시스템이 어떠한 윤리적 결정을 내릴 수 있도록 설계된 도덕적 프레임워크를 의미한다. 그들은 현재의 기술에서 정교한 AMA로 발전하는 과정을 파악하기 위한 기본적인 틀로 두 가지 개념, 즉. '자율성(autonomy)'과 '윤리적 민감성(ethical sensitivity)'을 제시한다.

> 가장 단순한 도구는 자율성도 민감성도 없다. 망치는 저절로 못을 내려치지 않으며, 내려치는 자리에 사람의 엄지손가락이 있어도 개의치 않는다. 하지만 우리가 제시한 두 차원의 낮은 단계에 있는 기술일지라도 설계에 따른 일종의 '운용적 도덕(operational morality)'이 있다. (중략) 또 다른 이론적인 극단은 고도의 자율성과 가치에 대한 고도의 민감성을 갖추고 신뢰할 만한 도덕적 행위자로 행동할 수 있는 시스템이다. 인간이 그런 기술을 지니고 있지 못하다는 것이 물론 이 책의 핵심 논지다. 하지만 '운용적 도덕'과 '책임감 있는 도덕(full moral agency)' 사이에는 이른바 '기능적 도덕'이라는 점진적인 단계가 존재한다. 정해진 행동 기준에 따르기만 하는 시스템에서부터 자신의 행동에 관한 도덕적 의미를 평가할 수 있는 지능적인 시스템에 이르기까지 여러 단계가 존재한다.[16]

월러치와 알렌은 다양한 사례를 통해 현재 인공지능 시스템이 여전히 운용적 도덕 또는 매우 제한된 기능적 도덕의 단계에 머물러 있다고 지적한다. 이러한 시스템들은 크게 두 가지 유형으로 나뉘는데, 하나는 상당한 자율성을 갖추고 있으나 윤리적 민감성이 거의 없는 시스템이고, 다른 하나는 자율성은 낮지만 높은 윤리적 민감성을 지닌 시스템이다. 그들은 기술의 발전에 따라 인공지능이 이 두 가지 요소, 즉 자율성과 민감성을

16) 위의 책, 49-50.

점진적으로 향상시키면서, 단순한 규칙 기반의 시스템에서 복잡하고 신뢰할 만한 '완전한 도덕적 행위자'로 발전할 수 있다고 주장한다.[17]

제제이러한 연구를 바탕으로 신상규는 기능적 도덕을 만족시키는 AI가 결국 AMA로서의 지위를 획득할 수 있다는 가능성을 제시한다. 그는 "일정한 요건을 만족시키는 AI에 대해서 인격성을 전제하지 않는 기능적인 의미의 도덕 행위자 자격이 부여될 수 있다"고 주장한다.[18] 또한 이러한 전제를 바탕으로 "인공지능에게도 그 행위자성에 걸맞는 책임 혹은 책무성의 귀속이 가능해진다"고 보았다.[19] 그는 인공적 도덕 행위자의 종류를 앞서 상술하였던 제임스 무어의 논의를 참조하여 네 가지 위계적 범주로 구분하면서, 인공지능의 자유도에 있어서 '2차 수준의 자유도'를 주장한다. 2차 수준의 자유도란 주어진 입력에 대해 어떤 출력을 산출할 것인가를 결정하는 내적 상태로서의 알고리듬 자체가 스스로 변화할 수 있는 것을 말한다.[20] 다시 말해, 1차 수준의 자유도를 기계의 '자동성(automatic)'이라고 한다면, 2차 수준의 자유도는 기계의 '자율성(autonomous)'을 말한다. 결론적으로 그는, **명시적 윤리 행위자**의 요건을 만족시키고 **온전한 2차 수준의 자유도**를 갖춘 AMA가 비로소 기능적 도덕 행위자로서의 행위 주체성을 가지는 단계에 도달했다고 말할 수 있다고 주장한다.[21]

이처럼 인공지능이 도덕적 행위자로서의 지위를 획득할 수 있다는 가능성이 제기되면서, AMA 구현을 위한 다양한 윤리적 접근법들이 논의되고 있다. 이러한 접근법들은 크게 세 가지로 구분된다. 첫째로 하향식(Top-down) 접근법은 공리주의와 의무론과 같은 전통적인 윤리 이론을

17) 위의 책, 50-51.
18) 신상규, "인공지능은 자율적 도덕행위자일 수 있는가?," 132.
19) 위의 책, 265.
20) 위의 책, 274.
21) 위의 책, 275.

AMA에 적용하는 방식이다. 이 방법은 명확한 윤리적 규칙과 원칙을 제공한다는 장점이 있지만, 복잡하고 예측 불가능한 상황에서는 경직된 대응을 보일 수 있다는 한계가 있다. 특히 공리주의적 접근에서는 효용의 계산과 평가 문제가, 의무론적 접근에서는 보편적 원칙의 적용과 해석 문제가 주요 과제로 대두되었다. 둘째로 상향식(Bottom-Up) 접근법은 경험과 학습을 통해 AMA가 도덕성을 발전시키는 방식이다. 이는 유연성과 적응력이라는 장점을 가지고 있지만, 학습 과정에서의 오류와 예측 불가능성(블랙박스의 문제)이라는 위험도 내포하고 있다. 특히 진화론적 접근과 기계학습 방식은 AMA가 인간의 도덕성과 유사한 패턴을 학습할 수 있는 가능성을 제시하지만, 동시에 윤리적 판단의 근거와 과정을 명확히 설명하기 어렵다는 문제점도 안고 있다. 셋째는 앞선 두 접근법의 한계를 극복하기 위해 제시된 혼합식(Hybrid) 접근법이다. 특히 이 영역에서 덕 윤리의 적용 가능성은 주목할 만하다. 덕 윤리는 행위의 결과나 의무보다는 행위자의 성품에 초점을 맞추는데, 이는 AMA의 전반적인 '도덕적 성격'을 형성하는 데 유용할 수 있다. 특히 정진규는 인공지능이 딥러닝을 통해 인간의 덕이 형성되는 과정과 유사한 방식으로 도덕적 성품을 학습할 수 있으며, 이를 통해 AMA에 덕 윤리를 적용할 가능성이 열릴 수 있다고 주장한다.[22] 이와 같은 덕의 학습 모델은 AMA가 복잡한 윤리적 상황에서 보다 유연하고 맥락에 맞는 판단을 내릴 수 있는 가능성을 제시한다.

III. 하나님 형상과 AMA의 관계

AMA의 가능성을 신학적 관점에서 바라볼 때, 우리는 인간이 하나님의 형상을 지닌 독특한 존재라는 전통적인 이해를 확장하여, 도덕적 판단을

22) 정진규, "인공적 도덕 행위자의 덕 윤리 모듈 적용 방안: 머신러닝과 딥러닝 활용을 중심으로," 「윤리연구」 1/138(2022), 237.

내릴 수 있는 인공적 존재에 대해서도 새로운 관점을 마련해야 한다. 이는 인간의 도덕적 책임과 독특성을 어떻게 정의해야 하는지에 대한 깊은 신학적 성찰을 요구한다. 우리가 AMA를 어떻게 이해하느냐에 따라, 인간과 인공적 존재가 공존하는 미래 사회에서 도덕성과 책임의 의미가 재정립될 수 있다.

하나님의 형상이 의미하는 세 가지 관점을 먼저 살펴보자. 첫 번째로, 가장 널리 알려진 접근은 하나님의 형상이 인간의 영혼을 가리킨다는 것이다. 이 관점은 인간의 영적 본질이 다른 모든 동물과 구별되게 하여, 인간을 하나님의 형상으로 창조된 유일한 존재로 본다. 이러한 관점은 하나님의 형상에 대한 실체론적 접근이라고 불린다.[23] 두 번째는 기능적 접근이다. 이 접근은 하나님의 형상을 "다스리라"는 하나님의 명령에 두며, 인간의 신적인 형상의 본질을 우리가 수행해야 하는 역할에 위치시킨다. 이 접근에 따르면, 인간은 하나님의 뜻에 따라 세상을 다스리고 책임지는 존재로서, 그 역할을 통해 하나님의 형상을 나타낸다.[24] 세 번째는 관계적 접근이다. 성경에서 하나님을 가장 잘 표현하는 중심 주제는 "하나님은 사랑이시다"라는 선언이다(요일 4:8). 사랑은 하나님의 본질을 나타내며, 그분의 존재의 핵심이다. 따라서 하나님의 본질은 본질적으로 관계에 있다. 사랑이란 본래 상호 인격적이기 때문이다.[25] 20세기 전반에 칼 바르트는 하나님의 형상을 정의할 때 인간의 관계성을 중심적인 특징으로 강조하였다. 그는 인간이 삼위일체 하나님의 형상으로 창조되었기 때문에, 하나님의 형상은 인간이 서로 간에 맺는 공동체적 관계와 하나님과의 관계 속에서 자신의 본질과 운명을 발견하는 것을 의미한다고 주장

23) Gregory A. Boyd and Paul Rhodes Eddy, *Across the Spectrum: Understanding Issues in Evangelical Theology*. 박찬호 역, 『복음주의 신학 논쟁: 복음주의 신학의 이슈 이해』 (서울: CLC, 2014), 178.

24) 위의 책, 178.

25) 위의 책, 191.

하였다.[26]

1. 실체론적 하나님 형상 이해

1) 하향식 윤리 접근과의 관계

실체론적 접근은 인간의 이성 및 도덕적 분별력을 AMA 설계에 구현하는 데 초점을 맞추며, 이는 높은 수준의 윤리적 원칙을 설정하고 이를 바탕으로 결정을 내리는 하향식 윤리 접근과 연결된다. 공학 분야에서 '하향식'이란 용어는 윤리학과는 다른 의미로 사용된다. 공학자들에게 하향식 접근법은 복잡한 과제를 체계적으로 분해하는 방법이다. 이 과정에서 주어진 과제는 여러 개의 간단한 하위 과제로 나뉜다. 이렇게 분리된 하위 과제들은 각각의 모듈로 구현되며, 이 모듈들은 위계적 구조를 형성한다. 이러한 구조화된 접근 방식을 통해 최초에 설정한 프로젝트의 목표를 달성하는 것이 가능해진다. 다시 말해, 공학에서의 하향식 분석은 복잡한 문제를 단순화하고 체계화하는 전략적 방법론인 것이다.[27]

이러한 하향식 접근은 AMA가 예상치 못한 상황에서도 윤리적으로 일관된 행동을 유지할 수 있도록 도와준다. 예를 들어 '인간 중심의 AI'(Human-Centered AI)를 개발할 때, 인간의 존엄성과 권리를 존중하는 윤리적 원칙을 어떻게 프로그램에 반영할 것인가? AMA가 다양한 상황에서 인간의 권리를 보호하고 해를 끼치지 않도록 설계하려면 어떻게 해야 하는가? 이러한 질문들은 실체론적 접근을 통해 이해된 인간의 도덕적 본성을 반영함으로써, AMA가 인간 사회에서 신뢰받는 존재로 기능할 수 있게 하는 중요한 요소가 된다.

---

26) 위의 책, 179.

27) Wendell Wallach and Colin Allen,『왜 로봇의 도덕인가』, 139.

2) 기호주의 AMA: 이성적 원칙 구현의 시도

실체론적 하나님 형상 이해에서 다룰 수 있는 AI 모델은 '기호주의(symbolism) AMA'[28]이다. 기호주의 AMA는 미리 설정된 규범과 원칙을 바탕으로, 상징과 기호를 활용해 인간의 도덕적 가치와 윤리적 판단을 이해하고 실행하는 인공지능 시스템을 말한다. 이 모델은 '규칙 기반 체계'라는 인지 구조를 지니며, 특정 시점마다 하나의 규칙만이 적용된다. 한 규칙의 결과가 다음 규칙의 입력으로 이어지면서, 이러한 순차적 과정을 거쳐 최종적인 결론에 도달한다. 다시 말해, 기호주의 모델은 명확히 정의된 규칙들을 단계적으로 적용하여 문제를 해결하는 구조를 가진다.[29] 이러한 접근 방식은 기계 학습 및 인공지능이 윤리적 판단을 내리는 데 필요한 복잡한 도덕적 의미를 해석하고 적용할 수 있도록 돕는 모델이다. 기호주의 AMA는 더 나은 도덕적 추론과 행동을 유도할 수 있으며, 인간의 도덕적 프레임워크와 더 잘 일치하는 윤리적 행동을 구현하는 것을 목표로 한다.

실체론적 하나님 형상 이해와 기호주의 모델은 인간 안에 내재된 이성, 도덕적 분별력, 영혼과 같은 내적 본질을 중시하고, 규칙에 기반한 사고 방식을 강조한다는 점에서 서로 맞닿아 있다. 기호주의 모델은 명확히 정의된 규칙 체계를 통해 문제를 분석하고 결론에 이르는 접근법으로, 인간의 이성적 사고와 논리적 추론 과정을 모방하도록 설계된다. 이러한 구조는 실체론적 관점이 강조하는 인간의 고유한 이성과 도덕적 판단 능력과 긴밀히 연결되어 있다.

---

28) 기호주의 모델이 적용된 대표적인 사례는 IBM이 개발한 체스 컴퓨터 '딥 블루(Deep Blue)'이다. 딥 블루는 1997년 세계 체스 챔피언 가리 카스파로프(Garry Kasparov)를 상대로 승리하며 주목을 받았다. 이 컴퓨터는 기호주의 접근법을 기반으로, 체스의 규칙과 전략을 엄격하게 정의된 알고리즘과 방대한 데이터베이스를 통해 구현한 모델이다.

29) 김기현, "인지과학과 인공지능," 「과학사상」 29(1999), 101.

이러한 공통점은 실체론적 하나님 형상 이해와 기호주의 모델 사이의 관계를 보다 구체적으로 설명할 수 있는 세 가지 측면으로 나누어 살펴볼 수 있다. 첫째는 이성적 사고와 규칙 기반 접근의 연결성이다. 실체론적 하나님 형상 이해는 인간의 이성을 하나님의 형상을 구성하는 중요한 요소로 본다. 둘째는 규칙 중심의 윤리적 판단이다. 실체론적 접근은 인간이 하나님 형상에 근거한 보편적이고 절대적인 윤리 원칙에 따라 도덕적 판단을 내린다고 본다. 이는 기호주의 모델이 하향식 접근법을 통해 윤리적 판단을 구조화하는 방식과 유사하다. 셋째는 인간의 형상에 대한 반영이다. 실체론적 하나님 형상 이해는 AMA 설계에서 인간의 초월적 특성과 도덕적 판단 능력을 반영하려는 시도로 연결된다. 그러나 동시에 타락한 인간성이 반영될 위험이 있다는 점도 주목해야 한다.[30] 기호주의 모델 역시 인간의 규칙 기반 사고를 모방하지만, 그 한계는 인간의 윤리적 결정을 완벽히 구현하지 못하는 데 있다. 이는 실체론적 하나님 형상 이해와 기호주의 모델 모두가 인간성과 윤리적 판단의 복잡성을 충분히 담아내는 데 한계를 가질 수 있음을 시사한다.

실체론적 접근은 보편적 인간의 본질을 상정하지만, 실제로는 문화적·역사적 배경에 따라 도덕적 이해와 판단 기준이 크게 달라질 수 있다. 기호주의 모델 또한 현실 문제를 해결하는 데 여러 한계를 드러낸다. 이는 세상에 기호, 규칙, 지식 등으로 정형화된 정보뿐만 아니라 수많은 비정형적 정보가 존재하기 때문이다. 이러한 맥락적 다양성을 무시하고 단일한 도덕적 기준을 AMA에 적용하려는 시도는 자칫 윤리적 판단의 유효성을 저하시킬 위험성 있다. 이는 기호 해석의 복잡성에서도 나타난다. 기호주의 기반의 접근은 인간의 도덕적 가치를 AMA가 해석하고 적용할

30) 정경일, "이마고 호미니스(*Imago Hominis*): AI 시대의 고통과 영성,"「신학과 철학」45(2023), 131.

수 있도록 기호체계를 사용하는데, 이 과정에서 기호의 다양성이나 해석의 다양성이 충돌할 경우 AMA는 윤리적 딜레마에 빠질 수 있다.

2. 기능론적 하나님 형상 이해

1) 상향식 윤리 접근과의 관계

마리우스 도로반투(Marius Dorobantu)는 기능론적 관점의 핵심을 인간의 정신적 능력이나 본질이 아니라, 하나님께서 인간을 특정한 사명으로 부르셨다는 사실에 둔다. 인간은 창조 세계 속에서 하나님을 대표하고, 다른 피조물을 돌보며, 그 분의 통치를 드러내는 역할을 맡은 존재이다.[31] 하지만 도로반투는 여기서 더 나아가, 인간이 단순히 세속적 방식으로 세상을 조직하고 관리하도록 부름받은 것이 아님을 강조한다. 그는 인간이 창조를 영적인 차원으로 승화시키는 더 높은 과업, 즉 '영성화(spiritualization)'의 사명을 부여받았다고 본다. 인간의 소명은 단순히 물질적 질서를 유지하는 것이 아니라, 창조세계를 하나님과의 완전한 교제의 장으로 변화시키는 것이다.[32]

이러한 기능론적 하나님 형상 이해는 AMA의 상향식 윤리 모듈 설계에 유의미한 신학적 자원을 제공한다. 상향식 윤리 접근은 AMA가 실제 경험을 통해 윤리적 지침을 학습하고, 다양한 상황에 적응하면서 윤리적 결정을 내릴 수 있도록 하는 방식이다. 이는 AMA가 예기치 않은 윤리적 딜레

31) 이러한 생각은 현대 성경 해석의 하나님 형상 개념에 뿌리를 두고 있으며, 창세기에 나오는 '형상'이라는 개념이 고대 근동의 다른 문화들에서 영감을 받아 사용되었다는 가정을 따른다. 그는 특정 신의 형상이라는 개념이 일반적으로 왕이나 파라오와 같은 인물이 그 신을 지구상에서 대표하고, 그 신을 대신하여 권위를 행사하는 것을 의미한다고 설명한다. Marius Dorobantu, "*Imago Dei* in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology," *Christian Perspectives on Science and Technology* 1(2022), 180.

32) 위의 책, 186-187.

마를 만났을 때 유연하고 상황 맥락에 맞는 판단을 내리게 한다는 점에서 중요하다.

테드 피터스(Ted Peters)는 진화개념을 통해 하나님 형상을 재해석하며, 인간의 본성이 고정된 것이 아니라, '그리스도 닮음(Christlikeness)'을 향해 진화하는 과정에 있다고 말한다. 그는 하나님 형상을 인간의 현재 상태가 아니라 그리스도 안에서 완성될 종말론적 목표로 이해한다.[33] 이러한 관점은 기능론적 접근과 상향식 윤리, 그리고 진화적·기계 학습적 접근을 연결하는 이론적 근거를 제공한다.

2) 연결주의 AMA: 경험을 통한 학습과 적응

이를 위한 구체적인 모델로서 연결주의(connectionism) 기반의 AMA는 신경망 구조를 이용해 인간의 도덕적 가치와 윤리적 원칙을 학습하고 구현하는 AI 시스템을 말한다. 연결주의 접근 방식은 인공지능이 기계 학습과 딥러닝을 기반으로 대량의 데이터를 통해 도덕적 판단을 학습하고, 다양한 상황에 적응할 수 있도록 돕는다. 특히 연결주의 AMA는 인간의 도덕적 사고 과정을 모방하여, 비정형적이고 비선형적인 방식으로 윤리적 결정을 내리는 것을 목표로 한다. 이러한 모델은 뇌의 뉴런 간의 연결 패턴을 모방한 인공 신경망을 통해 윤리적 판단을 모델링하기에 흔히 신경망이론(neural network theory)라고 불린다.[34] 이 방식은 복잡한 윤리적 상황에서도 유연하고 적응력 있는 도덕적 추론을 가능하게 하며, 기존의 규칙이나 원칙을 넘어 새로운 윤리적 해법을 모색하도록 돕는다. 즉, 연결주의 AMA는 특정 규칙에 의존하지 않고도 인간의 윤리적 사고를 학

33) Ted Peters, "The *Imago Dei* as the End of Evolution," Stanley P. Rosenberg, Michael Burdett and Benno Van Den Toren, eds., *Finding Ourselves After Darwin* (Grand Rapids MI: Baker Academic, 2018), 95.

34) 김기현, "인지과학과 인공지능," 112.

습하며, 실시간으로 환경에 반응할 수 있는 능력을 지닌 윤리적 판단 구조를 지향한다.

이러한 연결주의 AMA는 기능론적 하나님 형상 이해와 의미 있게 연결될 수 있다. 연결주의 모델은 뉴런과 같은 단위들이 상호작용하며 학습과 적응을 통해 복잡한 인지 과정을 형성하는 방식을 기반으로 한다. 이 모델은 인간과 유사한 방식으로 경험을 통해 점진적으로 학습하며, 이를 바탕으로 상황에 맞는 판단과 결정을 내린다. 이러한 점에서 기능론적 하나님 형상 이해와 연결되는 몇 가지 지점을 아래와 같이 정리할 수 있다.

첫째는, 경험 기반의 학습이다. 기능론적 하나님 형상 이해는 인간이 창조 세계를 돌보고 윤리적 책임을 수행하는 역할을 강조한다. 이는 단순히 추상적 원칙에 의존하는 것이 아니라, 다양한 문화와 맥락 안에서 실제 경험과 학습을 통해 적응하는 과정이 필수적이다. 연결주의 모델 또한 경험을 기반으로 뉴런 간의 연결 강도를 조정하며 점차 복잡한 행동과 판단 능력을 학습한다. 따라서 기능론적 접근에서 인간이 경험을 통해 윤리적 판단을 형성하는 방식은 연결주의 모델의 학습 메커니즘과 잘 맞아떨어진다. 연결주의 기반 접근은 아리스토텔레스가 말한 성격적 탁월성이 습관을 통해 내재화되는 과정과 유사하며, AMA가 점진적으로 학습하며 윤리적 판단 능력을 향상시킬 수 있는 가능성을 보여준다.[35] 둘째는, 상호작용과 적응이다. 기능론적 하나님 형상 이해는 인간이 창조 세계와의 관계 속에서 윤리적 결정을 내리는 상호작용적 존재로 이해된다. 연결주의 모델은 신경망을 통해 환경에서 발생하는 다양한 데이터를 처리하고, 적응적인 행동을 학습하는 데 중점을 둔다. 이는 기능론적 접근

35) 정진규, "인공적 도덕 행위자의 덕 윤리 모듈 적용 방안: 머신러닝과 딥러닝 활용을 중심으로," 238.

에서 인간이 다양한 상황에 따라 창조 세계와 상호작용하며 윤리적 역할을 수행하는 방식과 유사하다. 셋째는, 진화적 성격이다. 기능론적 하나님 형상 이해는 인간 본성이 고정된 것이 아니라, 그리스도를 닮아가는 과정에서 변화하고 성장하는 역동적인 존재로 파악한다. 피터스의 관점에서 인간의 본성은 마치 예열된 오븐에 넣기 직전의 빵 반죽처럼 아직 완전히 완성되지 않았다(not yet fully baked).[36] 이는 연결주의 모델이 새로운 데이터를 처리하고 학습함으로써 점진적으로 발전해가는 동적 특성과 유비적으로 연결될 수 있다. 특히 피터스의 신학적 진화(theistic evolution)개념을 연결주의 AMA에 적용하면, 이 시스템은 인간의 도덕적 성장 패턴을 모방하여 도덕적 지침을 지속적으로 학습하고 적용할 수 있는 능력을 갖춘 시스템으로 해석될 수 있다. 하지만 피터스에게 진화의 궁극적인 종착점(end)은 인류 안에 하나님의 형상이 충만하게 꽃피우는 것이며[37], 이 목표는 부활하신 그리스도 안에 예표적으로 놓인 종말론적 새 창조의 현실을 통해 이루어지는 변형이다.[38] 연결주의 AMA는 이러한 진화적 맥락, 즉 미완성의 존재가 목표를 향해 나아가는 패턴을 반영하여, 도덕적 판단과 행동의 적응성과 유연성을 강조한다. 이는 AMA가 변화하는 환경과 상황에 맞춰 적절한 윤리적 대응을 학습할 수 있는 능력을 강화하는 데 기여할 수 있다.

그러나 연결주의 접근에는 몇 가지 한계가 존재한다. 첫째, 이 모델은 대량의 데이터에 의존하기 때문에, 학습 데이터의 편향이나 오류가 AMA의 윤리적 판단에 직접적인 영향을 미칠 수 있다. 둘째, 연결주의 모델이 도출한 윤리적 패턴은 설명 가능성(explainability)이 낮아, 특정 윤리적 결정을 내린 근거를 명확히 파악하기 어렵다. 따라서 AMA의 윤리적 행동에

36) Ted Peters, "The *Imago Dei* as the End of Evolution," 92.
37) 위의 책, 92.
38) 위의 책, 95.

대한 투명성과 책임성을 확보하기 위해, 설명 가능한 인공지능(explainable AI, XAI)과 같은 보완적 기법의 도입과 추가 연구가 필요하다.

3. 관계론적 하나님 형상 이해

1) 혼합식 윤리 접근과의 관계

노린 허츠펠드(Noreen L. Herzfeld)는 삼위일체 하나님 안에서의 이미지로서의 관계적 접근을 강조하며, 이는 칼 바르트의 신학에 근거한다. 바르트는 창세기 1장 26절[39]의 구절의 복수 표현 '우리' 속에서 하나님의 삼위일체적 관계성을 발견한다. 즉, 하나님은 창조의 과정 속에서 스스로를 관계적 존재로 드러내신다는 것이다. 이때 바르트는 어거스틴의 '존재유비(analogia entis)' 개념과 달리, 하나님의 형상을 '관계유비(analogia relationis)'로 이해하며, 그 관계가 하나님 자신과 인간, 그리고 인간 상호 간의 공동체고 확장된다고 본다.[40] 그는 "형상은 이중적 의미를 가진다. 하나님은 그분 자신과의 공동체 속에서 살아가시고, 그 다음에는 인간과의 공동체 속에서 살아가시며, 그 다음에는 사람들이 서로 공동체 속에서 살아간다."[41]고 말한다.

도로반투 역시 관계론적 하나님 형상 이해를 지지하며, 하나님의 형상은 인간이 하나님과 맺도록 부름받은 독특한 관계와 인간 상호 간의 인격적 관계 속에서 드러난다고 주장한다. 삼위일체 하나님이 관계 그 자체이듯, 인간 또한 남자와 여자로 창조되어 관계적 존재로 살아가도록 부름받

39) (창 1:26) 하나님이 이르시되 우리의 형상을 따라 우리의 모양대로 우리가 사람을 만들고 그들로 바다의 물고기와 하늘의 새와 가축과 온 땅과 땅에 기는 모든 것을 다스리게 하자 하시고

40) Noreen L. Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis, MN: Fortress Press, 2002).

41) Noreen L. Herzfeld, *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age* (Minneapolis, MN: Fortress Press, 2023), 26.

았다. 그는 이러한 관계론적 관점을 바탕으로, 하나님의 형상이 인간과 AI의 상호작용 속에서도 일부 반영될 가능성을 신학적으로 탐색한다.[42]

이처럼 관계론적 하나님 형상 이해는 AMA의 윤리 모듈 설계에 있어 중요한 신학적 통찰을 제공한다. 혼합식 윤리 접근은 하향식과 상향식 접근의 장점을 결합하여 보다 유기적이고 균형 잡힌 도덕 판단 체계를 구축하려는 시도이다. 이러한 관점은 관계론적 형상 이해가 강조하는 복잡하고 다층적인 관계성과 깊이 맞닿아 있다.

2) 신경망-기호 AMA: 관계적 맥락과 균형 잡힌 판단

로체스터 대학교의 핸리 카우츠(Henry Kautz)는 AAAI 2020 로버트 S. 엥겔모어 기념 강연에서 신경망-기호(Neuro-Symbolic) AI 모델의 중요성과 잠재력을 강조하였다. 그는 AI의 발전사를 조망하며, 연결주의와 기호주의의 결합을 통해 '세 번째 AI의 여름'을 맞이할 것이라 전망했다.[43]

Neuro[Symbolic] 모델은 쥐와 치즈 미로 실험을 통해 그 작동 원리를 이해할 수 있다. 이 모델에서 신경망 기반 에이전트는 먼저 미로의 구조와 치즈의 위치를 빠르게 파악하고, 문제의 복잡성을 평가한다. 상황이 복잡하거나 경로가 불분명할 경우, 기호적 추론 시스템이 개입하여 논리

---

42) Marius Dorobantu, "*Imago Dei* in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology," 188-189.

43) 카우츠는 여섯 가지 신경망-기호 모델 중 특히 'Neuro[Symbolic]' 모델의 가능성을 주목했다. 이 모델은 다니엘 카너먼(Daniel Kahneman)의 '빠른 사고, 느린 사고' 이론에 근거하여, 빠르고 자동적인 '신경망적 처리(System 1)'와 느리지만 논리적인 '기호적 추론(System 2)'을 통합한다. 이러한 구조는 인간의 인지 과정을 모방하며 두 시스템의 강점을 결합함으로써, AI의 성능과 유연성을 극대화할 수 있는 핵심 방향으로 제시된다. 특히 Neuro[Symbolic] 모델은 신경망이 주도적 역할을 수행하되, 필요에 따라 기호적 추론을 활용함으로써 인간의 사고 구조와 유사한 방식으로 문제를 해결한다. Henry A. Kautz, "The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture." *AI Magazine* 43(2022), 119.

적인 방식으로 최단 경로를 계산한다. 이때 기호적 시스템은 복잡한 미로를 격자 형태의 '주의 스키마(Attention Schema)', 즉 AI가 어디에 주목해야 하는지를 보여주는 내부 지도로 단순화한다. 그런 다음 탐색을 통해 얻은 경로 정보를 이 지도 위에 표시한다. 이후 신경망 에이전트는 이 표식을 해석하며 목표에 도달하는 방법을 학습하게 된다. 이처럼 Neuro[Symbolic] 모델은 신경망의 빠른 인식과 학습 능력에 기호적 추론의 논리적 문제 해결력을 결합함으로써, 복잡한 문제를 보다 효율적이고 유연하게 해결할 수 있는 새로운 AI 접근 방식을 제시한다.[44]

이러한 신경망-기호 AMA는 관계론적 하나님 형상의 이해와 깊이 연결된다. 다시 말해, 사람이 하나님의 형상을 지닌 존재로서 관계 속에서 자신을 완성해 가듯, 이 모델 역시 신경망과 기호 체계가 서로 영향을 주고받으며 도덕적 판단과 행동을 만들어 낸다. 신경망은 경험과 감정, 상황의 흐름을 민감하게 느끼는 역할을 하고, 기호적 시스템은 그 안에서 의미를 정리하고 논리적으로 판단하는 역할을 한다. 이처럼 두 체계의 협력은 인간이 관계 속에서 사랑하고 응답하며 책임을 다하는 모습과 비슷하다. 결국 신경망-기호 AMA는 인간의 사고와 도덕을 흉내 내는 기술이 아니라, 관계 속에서 배우고 반응하며 책임 있게 행동하는 존재로 나아가는 방향을 보여 준다. 결국 신경망-기호 AMA는 단순히 두 기술의 결합이 아니라, 관계 속에서 균형을 이루는 윤리적 인공지능의 모델이다. 이는 인간이 하나님과 피조물, 타자와의 관계 속에서 자신의 형상을 실현하듯, AI 역시 데이터와 규칙, 감각과 의미의 관계망 속에서 책임 있는 결정을 내리는 존재로 설계될 수 있음을 시사한다.

그러나 이러한 접근에도 몇 가지 구조적 과제가 존재한다. 우선 시스템의 확장 가능성이 제한되어 있으며, 신경망과 기호적 구성요소 간의 통합

44) 위의 책, 121.

과정이 복잡하다는 문제가 있다. 또한 지식 공학 과정에서 상당한 수준의 수작업이 요구된다는 점 역시 실질적 적용의 제약으로 작용한다.[45]

IV. 리처드 니버의 '책임윤리'와 '책임 있는 AMA'

1. 혼합식 접근으로서의 책임윤리

니버는 목적론적 윤리와 의무론적 윤리의 한계를 지적하고, 이에 대한 대안으로 '응답의 윤리'라는 제3의 윤리적 모델을 제시한다. 응답의 윤리는 타인과의 대화에 참여하고, 자신에게 주어진 상황이나 행동에 반응하는 '응답자로서의 인간'을 중심으로 한다. 그는 서구 전통의 윤리학이 주로 목적론적이거나 의무론적이라고 보았으나, 이러한 이론들이 현대 사회의 복잡한 도덕적 상황을 충분히 설명하지 못한다고 생각한다. 특히, 의무론적 윤리는 절대적이고 보편적인 규칙을 지나치게 강조해 율법주의에 빠질 위험이 있고, 목적론적 윤리는 윤리적 상대주의로 흐를 가능성이 있다고 비판한다. 이에 따라 니버는 '책임'이라는 개념을 중심으로 맥락적 윤리(contextual ethic) 또는 응답의 윤리를 제안한다. 이 접근법은 구체적인 상황과 환경에 맞춘 도덕적 판단을 중시하며, 고정된 법칙이나 절대적 목적에 얽매이지 않고, 상황에 따라 적절한 책임을 지는 방식으로 윤리적 결정을 내리는 것이다.[46]

책임윤리는 매 순간의 결정과 선택에서 "지금 무슨 일이 벌어지고 있는가?(What's going on in the world?)"를 묻는 것으로부터 출발한다. 만약 가치를 나타내는 개념적 용어로 이를 설명한다면, 세 접근 방식의 차이는

45) Mohan Raja, "Neurosymbolic AI: Bridging Neural Networks and Symbolic Reasoning," *World Journal of Advanced Research and Reviews* 25/1(2025), 2370.

46) 임성빈, "리처드 니버의 응답의 윤리," 임성빈 외,『현대 기독교윤리학의 동향』(서울: 예영커뮤니케이션, 1997), 41.

‘선한 것(the good)’, ‘옳은 것(the right)’, 그리고 ‘적합한 것(the fitting)’으로 묘사될 수 있을 것이다. 왜냐하면 목적론적 윤리는 항상 최고의 선에 관심을 가지며, 그 선을 위해 올바른 것을 종속시키고, 철저한 의무론적 윤리는 우리의 행복이 어떻게 되든 상관없이 단지 올바른 것에 집중한다. 하지만 책임의 윤리에서는 오직 적합한 행위(fitting action)만을, 즉, 모든 상황에서 서로의 행동에 적절하게 반응하고, 앞으로 일어날 반응까지 고려한 행동만이 진정으로 좋은 결과를 낳으며, 그런 행동이 옳은 행동이라고 본다.[47]

”지금 무슨 일이 벌어지고 있는가?”로 시작된 질문은 “하나님께서 지금 어떻게 일하고 계신가(What is God doing in the world?)”의 질문으로 확장된다. 니버에게 이 질문이 중요한 이유는 그에게 있어서 ‘자아 인식(self knowledge)’은 하나님에 대한 지식에서 도출되는 것이 아니라고 보았기 때문이다. 그에게 윤리학은 하나님에 대한 우리의 지식과 관련하여 우리 자신에 대한 지식이다. 그리고 자아 인식은 책임 있는 삶에 있어서 필수적인 것이다.[48] 그러므로 니버에게 윤리란 이웃들과의 관계 속에서 하나님이 우리에게 행하시는 일에 적절하게 반응하는 것을 의미한다. 도덕적 행위자는 그 상황에 맞는 행동이 무엇인지 판단해야 하며, ‘응답적(responsible)’이라는 것은 누군가에게 또는 어떤 목적을 위해 반응할 수 있고, 반응해야 하는 책임이 있다는 뜻이다. 기독교 신앙에서의 응답적 윤리는 모든 도덕적 행동이 하나님께서 하시는 일에 대한 인간의 응답으로 이루어진다.

니버의 책임윤리는 특히 관계적 책임[49]을 강조하는데, 이는 인공지능

47) Richard Niebuhr, *The Responsible Self: An Essay in Moral Philosophy* (New York: Harper & Row, 1963), 60-61.

48) James M. Gustafson, “Introduction,” in *The Responsible Self: An Essay in Moral Philosophy* (New York: Harper & Row, 1963), 16.

49) 니버는 윤리적 책임이 개인의 자율적인 판단에 국한되지 않고, 사회적, 관계적 맥락

도 인간과의 관계 속에서 윤리적 책임을 다할 수 있어야 한다는 점과 연결된다. 하지만 AMA는 독립적으로 작동하는 기계적 존재가 아닌, 인간과 상호작용하며 윤리적 판단을 내리는 존재로 설계될 수 있다. 인간의 도덕적 판단이 관계 속에서 발전하는 것처럼, AMA도 그러한 관계성을 기반으로 책임을 다하는 도덕적 조언자로서 역할을 할 수 있다.

2. 책임 있는 AMA

리처드 니버의 책임윤리는 도덕적 행위자가 주어진 상황에서 적절하게 응답함으로써 책임 있는 결정을 내리는 것을 강조한다. 이 원칙은 단순히 AMA의 윤리적 판단 능력 개발에만 국한되지 않고, 더 나아가 AI 개발 및 관리의 거버넌스에 대한 논의로 확장될 수 있다. 니버는 인간을 규칙이나 고정된 법, 혹은 어떠한 목적을 이루기 위해 움직이는 존재로 보지 않았다. 대신 인간은 상황과 맥락을 이해하고, 그에 따라 적합한 응답을 할 수 있는 존재로 정의했다. 이러한 관점은 AI의 윤리적 문제와 책임감을 논의할 때, AI 시스템 자체의 윤리적 판단 능력을 넘어서서, 인간이 AI의 발전을 어떻게 책임지고 관리해야 하는지에 대한 고민으로 이어진다.

책임의 개념에는 "응답하는 인간, 즉 대화에 참여하고 자신에게 가해진 행동에 반응하는 인간의 이미지"가 암시되어 있다.[50] 니버는 적절한 행위

속에서 실천되어야 한다고 강조한다. 그의 저서 "책임적 자아(The Responsible Self)"에서 그는 인간이 윤리적 판단을 내릴 때 고립된 존재로서가 아니라, 다른 사람들과의 관계 속에서 도덕적 결정을 내려야 함을 설명한다. 윤리적 책임은 단순한 규칙이나 법칙에 의해 결정되는 것이 아니라, 상황과 상호작용을 통해 실현된다는 것이다. 즉, 니버는 인간의 도덕적 행동이 고립된 법칙을 따르는 것이 아니라, 관계 속에서 그때그때 적절한 응답을 통해 책임을 다해야 한다고 주장한다. 이러한 점에서 "관계적 책임"은 니버의 윤리 사상에 깊이 뿌리내린 개념이라 할 수 있다.

50) Richard Niebuhr, *The Responsible Self: An Essay in Moral Philosophy*, 56.

의 요소로서 응답(response), 해석(interpretation), 책무(accountability), 사회적 연대(social solidarity)을 제시하며, 이 네 가지가 책임의 개념을 구성하는 요소라고 규정한다.[51] 하지만 AMA의 윤리모듈 설계에 있어서 리처드 니버의 네 가지 책임의 구성요소를 그대로 적용하는 것은 한계가 있다. 왜냐하면 니버의 책임윤리 개념은 철저한 유일신론 앞에서 계시에 대한 인간의 응답을 전제로 하며, 이는 "통합된 자아 됨의 발달(the development of integrated selfhood)"를 뜻하기 때문이다.[52] 그러나 AMA가 역사를 통한 하나님의 계시를 깨닫고 이를 자아에 대한 계시로 수용하는 것은 신학적 인식의 범주 안에서는 성립하기 어렵고, 최소한 현 단계의 기술 수준에서는 구현되기 어려운 과제이다.

하지만 '책임 있는 AMA(Responsible AMA)'라는 AI 거버넌스의 확장된 관점에서 볼 때, 니버의 책임 구성 요소들은 유용한 윤리적 규준으로 적용될 수 있다. 예를 들어, AMA가 단순히 규칙을 따르는 것이 아니라 상황에 따라 다양한 윤리적 고려 사항을 반영할 수 있도록 다층적이고 협력적인 의사결정 구조를 설계하기 위한 응답이다. AMA가 스스로 판단할 수 없는 복잡한 윤리적 문제에 직면했을 때, 개발자와 관리자, 투자자, 사용자, 정부 등 다양한 이해관계자와 전문가가 협력하여 최선의 결정을 내리도록 거버넌스를 구축할 수 있다. 이러한 맥락에서, 니버가 제시한 책임의 네 가지 구성 요소는 AMA가 책임 있는 윤리적 응답을 수행하는 체계로 나아가도록 하는 규범적 방향을 제시한다.

첫째, '응답'의 개념은 AMA가 단순히 프로그래밍된 규칙을 따르는 것이 아니라, 상황을 인식하고 그에 맞는 적절한 반응을 할 수 있어야 함을 의미한다. 이를 위해 AMA는 고도의 상황 인식 능력과 의도적 행동 능력

51) 위의 책, 61-65.
52) 임성빈, 『21세기 책임윤리의 모색』(서울: 장로회신학대학교 출판부, 2002), 36.

을 갖추어야 한다. 예를 들어, 자율주행차의 경우 단순히 교통 규칙을 따르는 것을 넘어, 도로 상황의 복잡성을 이해하고 그에 맞는 윤리적 판단을 내릴 수 있어야 한다. 둘째, '해석'은 AMA가 입력받은 데이터를 단순 처리하는 것이 아니라, 그 의미와 맥락을 이해하고 분석할 수 있어야 함을 강조한다. 이는 신경망-기호 모델과의 접목을 통해 구현될 수 있다. 예를 들어, 의료 AI 시스템은 환자의 의료 데이터뿐만 아니라 사회적, 심리적 요인까지 고려하여 종합적인 진단과 처방을 내릴 수 있어야 한다. 셋째, '책무'는 AMA가 자신의 행동에 대한 결과를 예측하고 그에 대한 책임을 질 수 있어야 함을 의미한다. 이는 '해명책임'과 '분산적 책임' 개념과 연결된다. AMA는 자신의 결정 과정을 설명할 수 있어야 하며(설명 가능한 AI, XAI), 동시에 복잡한 시스템 내에서 인간 행위자와 다른 인공지능 행위자들과의 상호작용을 고려하여 책임을 분담할 수 있는 능력을 갖추어야 한다. 넷째, '사회적 연대'는 AMA가 고립된 존재가 아니라 사회적 맥락 속에서 지속적으로 상호작용하는 존재임을 인식하게 한다. AMA는 자신의 행동이 공동체에 미치는 영향을 고려하고, 지속적인 상호작용 속에서 일관된 윤리적 판단을 내릴 수 있어야 한다.

결론적으로, '책임 있는 AI'와 니버의 '책임윤리'를 결합한 AMA 거버넌스 구축은 AI 기술과 윤리의 조화로운 발전을 위한 중요한 방향성을 제시한다. 이는 단순히 기술적 진보를 넘어, AI가 인간 사회의 윤리적 가치를 존중하고 책임 있게 행동할 수 있는 토대를 마련한다. 이러한 접근은 AI가 인간의 윤리적 판단을 보완하고 지원하는 역할을 할 수 있게 하며, 동시에 AI 기술 발전이 인간의 존엄성과 가치를 훼손하지 않도록 보장한다.

IV. 나가는 말

본 연구는 인공지능 시대에 교회와 기독교 윤리학이 나아가야 할 방향

으로 인간의 책임과 역할을 강조하고자 하였다. AMA의 개발과 활용은 인간의 존엄성과 책임을 훼손하지 않고, 오히려 인간의 윤리적 성찰과 판단을 돕는 방향으로 이루어져야 한다. 동시에 책임 있는 AMA 개발과 같이 인간의 윤리적 판단을 보조하며, 인간과 기술의 상호작용 속에서 도덕적 책임과 협력을 확장시키는 가능성을 강조한다. 이는 기술이 독립적으로 발전하는 것을 넘어, 인간과 비인간, 그리고 지구 공동체의 삶을 풍요롭게 하는 도구로 사용될 수 있도록 설계되어야 한다는 점을 시사한다. 이를 위해 교회와 기독교 윤리학은 인공지능 기술의 발전을 주시하며, 지속적인 윤리적 성찰과 대화를 이어가야 할 것이다. 또한 AI 개발 과정에 적극적으로 참여하여, 책임 있는 AMA 거버넌스 구축에 기여해야 할 것이다. 이러한 노력을 통해 인공지능 기술의 발전이 하나님의 창조 세계를 풍요롭게 하고, 이 땅에서의 공공선을 추구하며, 인간의 책임과 역할을 고양하는 방향으로 발전할 수 있을 것이다.

이상의 연구를 통해 도출된 주요 결론은 다음과 같다. 첫째, AMA의 구현 가능성을 신학적 관점에서 검토할 때, 완전한 의미에서 도덕적 행위자로서 AMA를 구현하는 것은 현재로서는 불가능하다고 판단된다. 그러나 제한적이고 보조적인 역할에서 AMA를 개발하는 것은 가능하며, 인간의 도덕적 성찰과 판단을 지원하는 역할을 감당할 수 있다. 둘째, AMA 개발에 있어 관계론적 하나님 형상 이해와 혼합식 윤리 접근이 가장 적합한 모델로 제시되었다. 이는 인간과 AMA, 그리고 하나님과의 관계성을 중심으로 윤리적 판단과 행위가 이루어져야 함을 의미한다. 나아가 이에 대한 실천적인 모델로서 '신경망-기호' 모델을 제안하였다. 이는 AMA가 단순한 윤리적 규칙 준수자가 아닌, 관계 속에서 지속적으로 학습하고 성장하는 윤리적 행위자로서의 발전 가능성을 시사한다. 셋째, 리처드 니버의 '책임윤리'를 AI 거버넌스에 적용함으로써, '책임 있는 AMA'를 위한

네 가지 구성 요소(응답, 해석, 책무, 사회적 연대)를 도출하였다. 이는 AMA가 단순히 규칙을 따르는 기술적 존재가 아니라, 상황에 대한 적절한 응답과 해석, 그리고 그에 따른 책무과 사회적 연대를 수행하기 위해 개발자, 관리자, 투자자, 사용자, 정부 등 다양한 이해관계자들이 참여하는 윤리적 거버넌스 체계 속에서 설계되고 운영되어야 함을 시사한다.

한편, 본 연구의 한계점으로는 완전한 윤리 행위자로서의 AMA 등장을 위해서는 향후 사회적, 법적, 윤리적 준비가 필수적이며, 이를 위한 후속 연구가 요구된다는 점이다. 본 연구는 AMA가 완전한 윤리 행위자로 기능할 때에 발생할 수 있는 구체적인 영향들을 다루지 않았는데, 이는 현재 AMA의 기술적, 윤리적 성숙도가 충분하지 않기 때문이다. 그러나 인공지능의 윤리적 민감성과 자율성이 점차 고도화될 경우, AMA가 인간과 유사한 수준의 윤리적 판단을 할 수 있는 행위자로 발전할 가능성은 여전히 열려 있다. 따라서 AMA가 독립적인 윤리적 행위자로 기능할 경우 발생할 수 있는 부작용을 예방하기 위해, 특히 AMA의 판단과 행동에 대한 책임 귀속 문제와 의사결정 과정의 투명성을 보장하기 위한 사회적 논의가 필수적이다. 이를 통해 AMA가 사회 내에서 윤리적 신뢰를 얻고, 책임 있는 도구로 자리 잡을 수 있는 기반을 마련할 수 있을 것이다.

참고문헌

김기현. “인지과학과 인공지능.”「과학사상」29(1999), 93-116.

김상득. “AI 로봇의 도덕 행위자 가능성에 관한 윤리학적 연구.”「동서철학연구」105(2022), 627-652.

김은혜. “인공지능의 도덕성과 도덕적 행위자(Moral Agent)로서의 가능성에 대한 신학적 성찰과 기독교인공지능 윤리의 가치와 방향.”「장신논단」56/4(2024), 167-195.

송용섭. “도덕적 인공지능과 비도덕적 사회.”「기독교사회윤리」57(2023), 41-72.

신상규. “인공지능은 자율적 도덕행위자일 수 있는가.”「철학」132(2017), 265-292.

임성빈.『21세기 책임윤리의 모색』. 서울: 장로회신학대학교 출판부, 2002.

임성빈 외 6인.『현대 기독교윤리학의 동향』. 서울: 예영커뮤니케이션, 1997.

정경일. “이마고 호미니스(*Imago Hominis*): AI 시대의 고통과 영성.”「신학과 철학」45(2023), 119-143.

정진규. “인공적 도덕 행위자의 덕 윤리 모듈 적용 방안: 머신러닝과 딥러닝 활용을 중심으로.”「윤리연구」1/138(2022), 221-244.

Boyd, Gregory A. and Paul Rhodes Eddy. *Across the Spectrum: Understanding Issues in Evangelical Theology.* 박찬호 역.『복음주의 신학 논쟁: 복음주의 신학의 이슈 이해』. 서울: CLC, 2014.

Dorobantu, Marius. “*Imago Dei* in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology.” *Christian Perspectives on Science and Technology* 1(2022), 175-196.

Gustafson, James. “Introduction,” in *The Responsible Self: An Essay in Moral Philosophy.* New York: Harper & Row, 1963.

Herzfeld, Noreen L. *In Our Image: Artificial Intelligence and the Human Spirit.* Minneapolis, MN: Fortress Press, 2002.

_____. *The Artifice of Intelligence: Divine and Human Relationship in a Robotic Age.* Minneapolis MN: Fortress Press, 2023.

Kautz, Henry A. “The Third AI Summer: AAAI Robert S. Engelmore Memorial

Lecture." *AI Magazine* 43(2022), 105-125.

Moor, James H. "The Nature, Importance, and Difficulty of Machine Ethics." *IEEE Intelligent Systems* 21/4(2006), 18-21

Niebuhr, Richard, *The Responsible Self: An Essay in Moral Philosophy*. New York: Harper & Row, 1963.

Peters, Ted. "The *Imago Dei* as the End of Evolution." *Finding Ourselves After Darwin*. Grand Rapids MI: Baker Academic, 2018.

Raja, Mohan. "Neurosymbolic AI: Bridging Neural Networks and Symbolic Reasoning." *World Journal of Advanced Research and Reviews* 25(2025), 2351-2373.

Sapkota, Ranjan. "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges." *Information Fusion* 126(2026), 103599

Wallach, Wendell and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. 노태복 역. 『왜 로봇의 도덕인가』. 서울: 메디치, 2014.

Angwin, Julia and Jeff Larson. "Machine Bias." *ProPublica* (2016.5.23.) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, 2025년 11월 14일 접속.

Pierson, Brendan. "Mother sues AI chatbot company Character.AI, Google over son's suicide." *Reuters* (2023.10.24.) https://www.reuters.com/legal/mother-sues-ai-chatbot-company-characterai-google-sued-over-sons-suicide-2024-10-23, 2025년 11월 14일 접속.

Kleinman, Zoe and Chris Vallance. "AI 'godfather' Geoffrey Hinton warns of dangers as he quits Google." *BBC News* (2023.5.2.) https://www.bbc.com/news/world-us-canada-65452940, 2025년 11월 14일 접속.

"Youtube." https://www.youtube.com/watch?v=k82RwXqZHY8, 2025년 10월 27일 접속.

논문투고일: 2025년 11월 14일
심사개시일: 2025년 11월 16일
게재확정일: 2025년 12월 07일

• 국 문 초 록 •

본 연구는 인공지능 시대에 교회와 기독교 윤리학이 직면한 다음과 같은 두 가지 핵심 질문을 다룬다. 첫째, 인공적 도덕 행위자(Artificial Moral Agent, 이하 AMA)의 구현 가능성은 어떠하며, 이에 대한 기독교 윤리학적 관점은 무엇인가? 둘째, AMA의 윤리적 모듈을 설계함에 있어 기독교 윤리학은 어떤 원칙과 방법론을 제시할 수 있는가? 이를 탐구하기 위해 본 논문은 AMA의 개념과 구현 가능성을 다각도로 분석하였다. 우선 AMA의 정의와 특성, 그리고 그 유형을 고찰한 후, '하나님의 형상'(*Imago Dei*)과 리처드 니버의 '책임 윤리'를 중심으로 AMA의 윤리적 설계를 위한 신학적 토대를 마련하였다. 특히 하나님 형상에 대한 실체론적 · 기능론적 · 관계론적 이해를 각각 하향식 · 상향식 · 혼합식 윤리 접근법과 연결하여 분석함으로써, 기호주의 · 연결주의 · 신경망-기호 혼합형 AMA의 가능성을 제시하였다. 또한 리처드 니버의 책임윤리를 혼합식 접근으로 이해하고, '책임 있는 AMA'를 위한 네 가지 책임의 구성 요소(응답, 해석, 책무, 사회적 연대)를 제안하였다. 이를 통해 AMA가 상황을 인식하고 해석하며, 책임을 지고 사회와 연대하는 방안을 모색하고자 하였다.

주제어: 인공지능 윤리, 인공적 도덕 행위자(AMA), 하나님 형상, 책임윤리, 책임 있는 AMA