

A study on granting legal personality to Artificial Intelligence and responsibility for illegal acts

Yongho Kim*

*Professor, Dept. of Police Science, Kyungil University, Gyeongsangbuk-do, Korea

[Abstract]

This study deals with the legal issues of granting legal personality to artificial intelligence (AI) and discussing responsibility for illegal acts. As AI technology rapidly develops and the scope of AI's autonomous decision-making and actions expands, legal issues are becoming increasingly complex. In particular, given the lack of legal infrastructure to clearly determine responsibility for AI's illegal acts, we explore the possibility of granting AI legal personality to solve the problem.

This paper first explores the concept of legal personality and traditional legal theories, and analyzes the legal and practical issues of granting AI legal personality. To this end, we investigate the possibility of AI as a legal entity for illegal acts committed autonomously through various domestic and international cases and academic discussions. In addition, we compare and review the expected legal advantages and disadvantages of granting AI legal personality, and suggest the possibility of establishing a clear entity responsible for illegal acts by acknowledging AI's autonomous decision-making ability.

▶ **Key words:** artificial intelligence, legal personality, big data, autonomy, inexplicability, unforeseeability, responsibility for illegal acts

[요 약]

본 연구는 인공지능(AI)에게 법인격을 부여하고 위법 행위에 대한 책임을 논의하는 법적 문제를 다룬다. 인공지능 기술이 급속히 발전하면서 인공지능의 자율적인 의사결정과 행동 범위가 확대됨에 따라, 법적 문제가 점차 복잡해지고 있다. 특히, 인공지능의 위법 행위에 대한 책임 소재를 명확히 하기 위한 법적 인프라가 부족한 상황에서 인공지능에게 법인격을 부여하여 문제를 해결할 가능성을 모색한다.

본 논문은 먼저 법인격의 개념과 전통적 법리론을 탐구하고, 인공지능에 법인격을 부여할 경우의 법적 및 실질적 문제를 분석한다. 이를 위해 국내외 다양한 사례와 학문적 논의를 통해 인공지능이 자율적으로 행한 위법 행위에 대해 법적 책임 주체로서의 가능성을 조사한다. 또한, 인공지능에게 법인격을 부여할 때 예상되는 법적 장·단점을 비교 검토하고, 인공지능의 자율적 의사결정 능력을 인정함으로써 위법 행위에 대한 명확한 책임 주체를 설정할 가능성을 제시한다.

▶ **주제어:** 인공지능, 법인격, 빅 데이터, 자율성, 설명 불가능성, 예측 불가능성, 불법 행위 책임

- First Author: Yongho Kim, Corresponding Author: Yongho Kim
- *Yongho Kim (detecpol@kiu.ac.kr), Dept. of Police Science, Kyungil University
- Received: 2024. 12. 06, Revised: 2024. 12. 19, Accepted: 2024. 12. 23.

I. Introduction

인공지능(Artificial Intelligence)은 자율성과 학습 능력 등의 특징을 통해 비약적으로 기술이 발전하였으며, 산업 뿐 아니라 문화, 사회 전반에 큰 영향을 미치고 있다. 이세돌 9단은 딥마인드 기술 회사(DeepMind Technologies Ltd)의 바둑 프로그램 알파고(AlphaGo)의 도전을 받아 대결하였고, 'Microsoft'와 'Delft 공과대학(네덜란드)', 렘브란트 미술관은 2년여간 넥스트 렘브란트(The Next Rembrandt)라는 인공지능 프로젝트를 통해 17세기 화가 렘브란트의 화풍을 그대로 모사한 작품 재현에 성공한 바 있다. 뿐만 아니라, 스마트 폰에서 제공하는 시리(Siri), 인공지능 스피커 '아리', '지니' 등을 통해 서비스를 제공하고 있는 등 이제는 우리 삶에 깊이 관여하고 있다.

그러나 이 과정에서 발생하는 위법 행위, 예를 들어 2020년 12월 AI 스타트업 스캐터랩은 '친구 같은 AI'라는 콘셉트로 20대 여성으로 설정된 인공지능 '이루다'를 출시하였다. 하지만, 출시 이후 윤리적 취약점을 드러내며 논란에 휩싸였는데 논란의 요지는 '소수자를 대상으로 한 혐오 및 차별 발언', '별도 서비스를 통해 무단으로 수집한 대화 데이터 악용', '성적 대화 차단 방안 부재' 등 세 가지였다. 해당 서비스는 이용자가 채팅으로 유도할 경우, 동성애, 인종, 성별, 특정 지역에 대해 차별적인 표현을 서슴지 않았으며, 계좌번호 등의 개인정보를 물어보면 AI 학습 과정에서 필터링 되지 않은 데이터를 그대로 노출하여 실제 계좌번호를 드러냈다고 한다. 스캐터랩은 '연애의 과학'과 '텍스트잇' 등 자사 서비스들을 통해 불특정 다수의 실제 채팅 내용을 수집한 사실이 드러나면서, 개인정보보호 위원회로부터 제재를 받았다[1].

인공지능 기술은 비약적으로 발전하고 있으며 인공지능의 법인격 인정 여부 역시 활발하게 논의되고 있다. 본 논문에서는 인공지능의 위법 행위를 직접적으로 묻기 위한 '법인격' 인정 여부에 대해서 논하고자 한다.

II. Definition and characteristics of Artificial Intelligence

1. Definition of Artificial Intelligence

인공지능(Artificial Intelligence)이라는 용어는 1955년 8월 미국의 인지과학자 존 맥카시(John McCarthy)가 그 다음해 여름에 개최될 "다트머스 회의"(Dartmouth conference)를 위해 작성한 자금모금 제안서에서 처음 사

용하였다[2]. 그는 인공지능을 "기계가 사람의 학습이나 지성의 과정을 재현할 수 있는 것"이라고 정의하였다[3]. 국립국어원의 표준국어대사전에서 인공지능의 뜻을 찾아 보면, "인간의 지능이 가지는 학습, 추리, 적응, 논증 따위의 기능을 갖춘 컴퓨터 시스템. 전문가 시스템, 자연 언어의 이해, 음성 번역, 로봇 공학, 인공 시각, 문제 해결, 학습과 지식 획득, 인지 과학 따위에 응용한다"라고 되어 있으며, 동의어로 AI라고 검색된다[4].

우리나라에서는 아직 인공지능에 대한 법률상 개념 정의가 확립되어 있지 않다. 하지만, 유럽연합(EU)은 2024년 8월 1일 인공지능 법(EU AI Act)을 발표하였으며 해당 법 제3조에서 'AI 시스템'은 "다양한 수준의 자율성으로 작동하도록 설계되고 배포 후 적응력을 나타낼 수 있는 기계 기반 시스템을 의미하며, 명시적 또는 암묵적 목표의 경우 수신된 입력을 통해 물리적 또는 가상 환경에 영향을 미칠 수 있는 예측, 콘텐츠, 추천 또는 결정과 같은 결과를 생성하는 방법을 추론한다."라고 정의하고 있다[5].

2. Characteristics of Artificial Intelligence

인공지능은 프로그래밍 된 알고리즘에 의해 스스로 학습하고 결정을 내릴 수 있는 기술이다. 이러한 기술의 기반이 되는 인공지능의 특성은 자율성, 설명 불가능성, 예측 불가능성으로 설명할 수 있다[6].

2.1 Autonomy

자율성(autonomy)이란 '스스로 원칙에 따라 어떤 일을 하거나 자신을 통제할 수 있는 특성'을 의미한다. 특히 인공지능의 자율성은 '선택'과 '결정'의 능력을 의미하며, 이는 인간의 개입 없이 스스로 결정을 내리고 이를 실행할 수 있는 능력을 말한다[7]. 인공지능은 기본적으로 알고리즘 또는 컴퓨터 프로그래밍을 통한 연산 결과가 인간의 조작이나 통제 행위를 일부분 대체한다는 점에서 전통의 도구와 큰 차별점을 가진다[8]. 도로에서 운행되는 일부 자율주행 자동차는 인간의 조작이 없더라도 정해진 목적지까지 운행하여 도착할 수 있는데 이는 인공지능이라는 존재가 자율성을 발현한 것이라 말할 수 있다[6].

2.2 Inexplicability

설명 불가능성(inexplicability)이란 인공지능 알고리즘이 찾아낸 규칙이 어떠한 방식으로 도출된 것인가를 설명할 수 없다는 것이다. 전통적인 알고리즘 규칙은 논리적 규칙에 따라 시간 순서대로 처리하여 결과값을 도출하는 '순차(sequential)' 알고리즘이다. 그러나 인공지능 알고리

좁은 문제 해결 규칙을 사전에 설정하지 않고, 입력값과 결과값이 주어지면 결과값을 도출할 수 있는 규칙을 사후에 찾아내는 방식이다. 이러한 인공지능 알고리즘은 정확한 규칙을 찾아낼 때까지 빅 데이터를 활용하여 현재의 잠정적 규칙에 따른 계산값과 정답을 비교하고, 그 차이를 반영하여 규칙을 개선하는 과정을 반복한다.

인공지능 알고리즘은 전통적 알고리즘과 달리 사람이 아닌 기계가 사후적으로 규칙을 찾아낸다. 이때 '왜 이런 규칙을 찾아냈는가?', '왜 다른 규칙은 찾아내지 않은 것인가?', '기계가 찾은 규칙은 언제나 정확한가?' 등에 대해서 이해하기 어려우며 더욱이 인공지능 알고리즘이 찾아낸 규칙은 학습에 사용된 빅 데이터에 따라 달라질 수 있다. 따라서, 사람은 인공지능 알고리즘으로 얻은 규칙을 문제에 적용하면 정답에 가까운 결과값을 얻는다는 사실은 알지만 왜 인공지능 알고리즘이 그러한 규칙을 찾아내었는지는 설명하기 어렵다[9].

2.3 Unforeseeability

앞서 설명한 '자율성'과 '설명 불가능성'의 특성은 결국 인공지능의 동작과 결과에 대한 '예측 불가능성(unforeseeability)'으로 이어지게 된다. 인간은 인지적 한계가 있어 인지할 수 있는 범위 내에서 해결책을 모색하지만, 인공지능은 인지적 한계나 계산능력의 한계가 존재하지 않기 때문에 가용할 수 있는 정보를 분석한 후 최적의 결과를 도출하기 때문에 인간이 생각할 수 없는 방법으로 일을 처리할 수 있다.

이와 같이 인공지능은 인간이라면 저지를 수 없는 일을 저지르거나 사고의 발생 가능성을 예측하고 방지하거나 위험을 회피하기 어려운 '예측 곤란'의 상황을 일으킬 수 있다[10].

III. Concept definition and classification of legal personality

1. Definition of the concept of legal personality

우리 민법 제3조를 보면, '권리능력'의 존속기간에 대해서, 제34조에서는 법인의 '권리능력'에 대하여 규정하고 있지만, 권리능력이 무엇인지에 대해서는 아무런 언급이 없다. 이에 대한 일반화된 학설을 보면 권리능력이란 '권리의 주체가 될 수 있는 지위 또는 자격'을 가리키며, '인격' 또는 '법인격'이라고 한다.

오늘날 권리능력은 의무능력을 포함하기에 '권리의무능력'이라는 용어가 더 정확하다고 할 수 있습니다. 그러나 단순히 '권리능력'이라고 표현해도 의무능력을 포함하는 의미로 사용됩니다[11]. 지금의 법체계는 자연인인 사람에게만 제한 없이 권리능력을 부여하고 권리 주체성을 인정하고 법인에게만 제한적인 범위 내에서 인정하고 있다.

2. The legal personality of a natural person

인격이나 법인격은 권리능력과 의무능력, 즉 권리주체나 의무주체가 될 수 있는 지위 또는 자격을 말한다. 우리 민법 제3조와 제34조에서 자연인과 법인의 권리능력을 규정하고 있는데 제3조는 사람은 살아있는 동안 권리와 의무의 주체가 된다고 명시하고 있다. 이 규정은 사람이 출생부터 사망까지 권리능력이 당연히 인정되며, 모든 자연인이 원칙적으로 제한 없이 평등하게 권리능력을 가진다는 것을 설명하고 있다.

한편, 법인은 법률이나 설립 목적에 따라 제한된 권리능력이 부여된다. 이처럼, 법인의 법인격은 시대와 사회, 법 정책에 따라 달라지고 그 당시의 실정법으로 결정한다[12].

3. The legal personality of a juridical person

사단 또는 재단의 재산권 기타 권리의 인정을 위한 법인격 역시 인정되고 있다. 사단이나 재단에 대한 법인 제도는 사회적 필요와 법 정책적 입법 결단에 의해 인정되며, 이는 인공지능의 법인격 부여와 관련하여 가장 유사한 개념적 형태로 볼 수 있다. 오늘날 대다수 법체계에서 법인의 법인격을 부여하고 있으며 우리 민법 역시 다음과 같은 이유에서 법인의 법인격을 인정하고 있다.

첫째, 사단이나 재단은 그 구성원이나 재산 출연자와는 독립된 법적 주체로서 활동하기 위해 존재합니다.

둘째, 사단이나 재단의 재산을 구성원 또는 출연자의 재산과 분리하여, 사단이나 재단이 외부 제삼자에 대한 책임을 지되 구성원이나 출연자의 고유 재산에 대해서는 강제집행을 하지 못하도록 보호하고, 해당 사단 또는 재단 자체의 재산에 대해서만 책임을 지도록 해야 할 필요가 있습니다.

셋째, 자연인은 영구히 생존할 수 없지만, 법인은 인류가 존재하는 한 영구히 지속될 수 있으며, 자금, 정보, 활동 범위 등에서 자연인을 크게 능가할 수 있어 자연인의 한계를 극복할 수 있습니다[13].

4. The legal personality of an animal

현행 우리 법은 권리주체로서 사람과 권리객체로서 물건으로 나누고 있는데 동물은 생명이 있는 자연적 존재임에도 현행법상 물건으로 취급하고 있다. 물론, 독일 민법 제90조의 a에서는 “동물은 물건이 아니다. 동물은 별도의 법률을 통하여 보호된다. 다른 규정이 없는 한 동물에 대해서는 물건에 관한 규정을 준용한다”라고 명시하며, 특수한 법적 지위를 인정한다[14].

최근 반려동물에 관한 관심과 함께 사회적 관념 역시 변화되고 있어 위와 같은 이분법적 틀에서 벗어나고자 지속적인 논의가 진행되고 있는데 사회적 필요성과 합의에 따라 자연인이 아닌 제3의 존재, 즉, 사람과 동일하지는 않지만 일정 부분에 대해서는 법적 지위를 인정할 필요성이 있는 경우라 할 것이다. 이는 인공지능에게도 법인격을 부여할 수 있다는 근거로 삼을만한 것이다.

5. The legal personality of Artificial Intelligence

법인에 대한 법인격을 부여한 것과 같이 입법적 방식을 통한다면 인공지능에 대한 법인격 역시 가능하다. 하지만, 이때 인공지능의 법인격 부여에 관한 근거가 무엇인지 명확하게 설명되어야 한다. 그 근거에 대해서 '법인의제설'과 '법인실제설'이 검토될 수 있다[15]. 법인의제설에 따르면, 인공지능을 법적으로 인정하는 법을 제정함으로써 인공지능에게 법인격을 부여할 수 있다. 또한, 법인실제설에 의하더라도 AI는 단순한 도구가 아닌 자율성을 가진 존재로서, 사회적 실재성을 갖추고 있으며, 유기체로 간주할 여지도 있다[16]. 그래서 어느 입장에서도 법인격을 부여 받을 수 있는 충분한 근거가 인정된다[17].

IV. Granting legal personality to Artificial Intelligence

1. A view that affirms the legal personality of Artificial Intelligence

권리의 주체성이 인간에서 법인으로 확대된 것과 같이 사회의 요구와 필요에 따라 인공지능 역시 법인격을 부여할 수 있다는 견해이며 관련되는 근거로는 법인의제설 및 법인실제설과 같이 법인 이론을 통하여 입법적으로 인공지능에 법인격을 부여할 수 있다는 주장이다.

이미 법인처럼 권리능력을 부여하였듯이, 인공지능에게 법인격을 부여하는 것이 법적으로 문제 되지 않는다. 따라서, 인공지능의 보편적 사용과 그 사회적 효용성이 인

정될 때 인공지능이 계약의 이행이나 불이행에 대한 책임을 부담할 수 있는 경우라면, 인공지능 자체에 대하여 독립적인 법인격을 부여하는 것도 가능하다고 볼 수 있다[18]. 또한, 인공지능의 자율성이 강조되면서 책임 재산을 소유할 필요성이 정책적으로 인정될 경우, 제한적으로라도 인공지능에게 책임 목적의 재산을 소유할 수 있는 권리 주체성을 부여하는 방향으로 나아가야 한다[19]. 따라서, 인공지능의 권리능력, 의사능력, 행위능력은 사회적 필요성과 요구에 따라야 한다[13].

2. A view that denies the legal personality of Artificial Intelligence

만약 인공지능에 법인격을 인정할 때 발생하는 법적 문제점이 크고 사회적 혼란이 야기될 수 있다. 또한, 자율성이라는 특성 역시 운영자가 이미 설계한 알고리즘에 따른 작동에 불과하기 때문에 이를 근거로 법인격을 부여하기 어렵다는 주장과 견해가 있다[20]. 또한, 인공지능의 창조자인 인간과 피조물이 같은 격으로 보이는 것은 윤리적으로 수용하기 어렵고 인공지능의 권리능력 시기와 종기가 불명확하다는 이유 등에 따라 인공지능에게 법인격을 부여할 필요 없다고도 한다[21].

3. Overseas cases on discussion of Artificial Intelligence legal personhood

3.1 EU

2017년 초 유럽의회는 입법 권고 결의안으로 EU 집행위원회에 “장기적으로 로봇에 대한 특수한 법적 지위를 생성하여 적어도 정교한 자율 로봇이 발생시키는 손해에 대해 전보할 책임이 있는 전자인의 지위를 부여하는 방안을 고려해 볼 것”을 제안하였다[22]. 하지만, 결의안은 공상과학(Science-Fiction) 같은 이야기라고 비판한 바 있다[23].

EU 집행위원회 산하 '인공지능에 관한 고등전문가 그룹(High-Level Expert Group on Artificial Intelligence, HLEG)'은 인공지능 시스템의 개발 및 사용이 인간의 복리에 이바지해야 한다는 윤리 원칙을 설정하였으며, 인공지능은 단지 인간을 위한 도구일 뿐이라는 점을 명확히 하였고, '전자인의 지위' 또는 '전자 인격' 창출에 대한 결의안에는 따르지 않았다[24]. 고등전문가 그룹은 인공지능 시스템이 자율적인 의사결정을 하지만 권리주체가 아니라 권리객체라고 했으며, 윤리적·법적 행위의 대상은 인간뿐이라고 주장하며 인공지능 시스템은 소프트웨어이자 권리객체 또는 도구에 불과하다 하였다[25].

2021년 4월 유럽 집행위원회는 유럽의회에 「인공지능에 관한 통일규범(인공지능법)의 제정 및 일부 연합제정법들의

개정을 위한 법안(Proposal for a Regulation laying down harmonized rules on artificial intelligence(Artificial Intelligence Act) and amending certain Legislative Acts)을 발의하였으며, 해당 법안의 주요 내용은 EU 내 인공지능의 영업 목적의 활용 및 EU 밖 인공지능의 결과물이 EU 내에서 활용될 때 본 법을 적용하도록 하고 있다. 따라서 EU 회원국은 각국의 법률의 범위 내에서 민사책임 문제를 논의해야 하는 문제가 있다[26].

3.2 United States of America

EU의 관련 법안을 제정하자 미국은 인공지능 관련 규제의 필요성에 따라 2019년 “인공지능에서 미국의 리더십 유지에 관한 행정명령 13859호(Executive Order 13859, Maintaining American Leadership in Artificial Intelligence)”, 2020년 기존의 AI 정책을 확대하면서 국가 AI 이니셔티브법(National AI Initiative Act of 2020), 2022년에 알고리즘의 편향된 데이터나 차별의 강화로 인해 발생하는 사회적 문제를 방지하기 위한 “알고리즘 책임 법안(A Proposal for Algorithmic Accountability Act of 2022)”이 상·하원에서 발의하였다[27].

3.3 Japan

일본은 인공지능 기술 발전을 촉진하고 생성된 데이터의 활용에 초점을 맞추어 관련 법률 등을 정비하고 있다. 「개인정보의 보호에 관한 법률(個人情報の保護に関する法律)」에서 개인정보의 보호를 위한 내용을 규정하고 있고, 「정보처리의 촉진에 관한 법률(情報処理の促進に関する法律)」에서 신기술의 개발과 활용 기반 구축을 위한 내용을 규정하고, 「관민 데이터 활용 추진기본법(官民データ活用推進基本法)」은 국가, 지방공공단체, 독립행정법인 또는 기타의 사업자가 관리, 이용, 제공하는 ‘관민 데이터’의 유통·활용의 추진 등을 목적으로 2016년에 제정하였다. 또한, 디지털 기술의 활용을 촉진하기 위해 행정 절차와 관련된 민간 절차를 연계하도록 하는 「정보통신기술을 활용한 행정의 추진 등에 관한 법률(情報通信技術を活用した行政の推進等に関する法律)」을 개정하여, 정보통신기술 촉진을 위한 환경을 정비하였다[26].

V. AI's responsibility for illegal acts

1. Legal status of Artificial Intelligence in Korea

현행 민법 제3조는 “사람은 생존한 동안 권리와 의무의 주체가 된다.”라고 규정하고, 권리와 의무의 주체에 대해

서도 ‘자연인’과 ‘법인’만 인정한다. 인공지능이 마치 스스로 생각하고 판단하는 자율적인 존재인 것으로 보이지만, 사실은 소프트웨어와 하드웨어로 구성된 물건일 뿐이다. 인공지능이 비약적으로 발전하여 사람과 유사한 판단과 행동을 한다, 하여도 지금의 법체계에서는 인간과 동일한 권리와 의무의 주체로 보기는 어렵다. EU, 미국, 일본 등 해외 동향 역시 인공지능에 법인격을 부여하지 않고 있으며 우리나라 역시 법인격을 부여하지 않고 있다.

인공지능은 산업과 문화, 우리 일상에서 상용화되고 고도화되고 있지만 어떠한 법적 의무를 부과할지 사실상 명확하지 않다. 따라서, 인공지능이 야기할 수 있는 법익 침해에 대해 현행법상 어떠한 책임을 물을 수 있을지에 대해 검토할 필요가 있다.

2. Liability for damages caused by illegal acts of Artificial Intelligence

인공지능은 사전에 데이터 학습을 통해 설계된 프로그래밍 알고리즘에 따라 작동되고, 비약적인 발전으로 산업 현장뿐 아니라 문화, 예술 등 분야를 가리지 않고 활용되고 있으나, 만약 오작동이나 위법 행위로 인해 사고가 발생한다면 그 책임 주체가 누구인지 명확하지 않다. 만약 인공지능의 위법 행위로 사고가 발생한다면 책임 주체를 불법행위자, 제조업자, 프로그램 개발자 등으로 검토해 볼 수 있으며[28], 인공지능의 위법 행위 논란 관련하여 책임의 법리를 살펴보면 제조물책임법, 민법 제750조(불법행위의 내용), 제756조(사용자의 배상책임), 제758조(공작물 등의 점유자, 소유자의 책임)[29], 자동차손해배상보장법, 자율주행자동차법 및 개인정보보호법 등이 있으나 특별법상 불법행위책임을 그대로 적용할 수 있지 등에 대해서는 관련 입법을 통해 개선되어야 할 것으로 보인다. 그러나, 인공지능의 위법 행위에 대한 책임을 위해 법인격 부여에 대한 부분을 제외하더라도 손해 배상책임을 귀속시킬 책임 주체 지정에 대한 부분은 입법을 통해 보완되어야 할 것이다.

결국 인공지능을 설계·제작·운용하거나 이용하는 자가 발생한 위법 행위에 대하여 책임져야 한다는 것이 현실적인 방안이다. 하지만, 이러한 방안도 궁여지책일 뿐이다. 인공지능은 ‘약한 인공지능(weak AI)’, ‘강한 인공지능(strong AI) [30], ‘초지능(superintelligent AI)’ 등으로 나누어지는데, 강한 인공지능 이상의 단계에서는 인공지능이 사전에 설정된 대로 작동하지 않고 변수에 대비하기 위한 새로운 데이터 학습을 통해 알고리즘을 수정하여 작동하게 되고 이때 발생할 수 있는 위법 행위 등에 대해서는 설계·제작·운용·이용하는 사람 역시 결과를 예측할 수

없으므로 이에 대한 책임을 물을 수 있는지에 대해서도 대비해야 한다.

3. Criminal liability for illegal acts committed by Artificial Intelligence

형사 책임을 논할 때 반드시 행위능력을 전제로 해야 하므로 지금의 법체계와 인공지능에 대한 인식으로 볼 때 형사 책임에 대해서 논하는 것은 의미가 없다고 본다. 하지만, 인공지능의 비약적 발전과 태아에게 부분적으로 권리능력을 인정하거나 권리능력을 갖는 인격적 주체의 범위가 확장하는 점[31] 등에 비추어 볼 때 인공지능에 대한 형사 책임 여부에 대해서도 논의할 수 있다. 하지만, 형사 책임을 위해서는 우선 인공지능에 법인격을 부여해야 하는 것이 우선이다. 형법 규범에서도 “~ 한 자”라고 행위 주체를 사람으로 한정하고 있으므로 인공지능을 형법 규범의 당사자로 인정할 수 있다. “책임 비난의 내적 근거는 인간이 자유롭고, 자신의 책임에 따른 것이며, 도덕적인 자기결정을 의도하였으며, 따라서 스스로 법에 따르고 불법에 저항하도록 결정하고, 법적 당위 규범에 따라 자기 행동을 조정하고 법적으로 금지된 것을 회피할 능력이 있다는 점에 있다”라고 한 독일연방법원의 판례와 같이 ‘인간’, ‘행위’, ‘도덕’ 등에 대한 도덕적·철학적 패러다임이 바뀌지 않는 한 여전히 유효하다 할 것이다[32].

4. Responsibility for intentional or negligence of Artificial Intelligence

인공지능은 학습된 데이터를 바탕으로 설계된 알고리즘에 따라 결과가 도출된다. 따라서, 의도적으로 오류가 발생하게 설계·제작하여 위법 행위가 발생하고 이로 인한 피해가 발생했을 때 이를 인공지능의 고의·과실로 볼 수 있느냐에 대한 논의가 필요하다. 정상적으로 설계·제작을 하였음에도 인공지능의 동작 과정에서 발생한 오류로 인해 타인에게 손해를 입혔다면 이는 인공지능의 오작동인가? 아니면, 설계에 참여한 사람의 과실로 볼 것인가에 대해서는 명확하게 구분 짓지 못한다.

과실이란, ‘마땅히 기울여야 할 주의를 다하지 않아, 어떤 결과의 발생을 미리 내다보지 못한 일’을 의미한다.

결국 과실을 판단하는 핵심은 ‘손해를 피하기 위한 조치를 다 하지 않았다’라는 것이다. 그러나, 인공지능의 수준이 발전함에 따라 앞서 언급한 바와 같이 예견가능성이 점점 감소하게 되어 과실을 인정할 가능성도 그에 비례하여 줄어들게 된다. 인공지능의 작동 결과에 대한 예견가능성은 줄어들 수 있지만 아직 어떠한 주의의무를 인공지능을

활용하는 사람에게 부과할 것인지에 대해서는 지속해서 논의해야 할 것을 보인다[33].

VI. Conclusions

스마트 폰에서 사용자에게 다양한 음성 서비스를 제공하는 인공지능 시리(Siri) 또는 환자를 진료하고 치료와 관련된 의견을 제시하는 IBM의 인공지능 왓슨 포 온콜로지(watson for oncology) 등 인공지능은 우리 생활에서 다양한 형태로 편의를 제공하고 있으며 때로는 우리 생활에 아주 밀접하게 연결되어 있다.

하지만, 우리에게 편의를 제공하는 인공지능의 동작으로 인한 발생하는 위법 행위에 대하여 그 책임 소재를 어떻게 물어야 할지 꾸준히 논의가 진행되고 있지만 명확한 해답을 내지 못하고 있다.

법인의 권리주체를 인정하는 것과 같이 인공지능에게 법인격을 부여하고 권리주체로 인정하자는 의견이 존재하지만, 현실적인 문제를 해결하지 않고는 이를 수용할 수 없다.

인공지능의 위법 행위에 적용할 수 있는 법률을 살펴보면, 민법상의 일반 불법행위책임부터 제조물책임, 공작물 책임, 자동차손해배상보장법상의 자동차 사고 책임까지 다양한 법률이 적용될 수 있습니다. 그러나 누구에게 위법 행위의 책임을 묻는가에 관한 문제는 일반 불법행위책임의 성립 요건인 고의·과실, 위법성, 인과관계, 손해, 책임의 주체 등을 근본적으로 따져야 하며, 이러한 접근 방식으로는 결국 종래의 법 해석에 의존할 수밖에 없습니다. 이럴 때 인공지능의 위법 행위에 대한 논의는 이루어지기 어렵습니다.

인공지능이 발생시킨 손해에 대한 책임을 인정하려면 피해자가 일반 불법행위책임에 따라 고의, 과실, 위법성, 인과관계, 손해 등을 증명해야 합니다. 그러나 인공지능의 설계나 작동 원리 등은 일반인이 증명하기 어렵고, 작동이 대부분 숨겨진 알고리즘에 따르기 때문에 효과적인 구제를 기대하기 어렵습니다.

따라서, 종래의 방법으로 해결되지 않는다면, 인공지능의 위법 행위에 적용할 수 있는 책임 규정을 도입해야 한다.

위법 행위에 대한 손해 배상의 1차 책임은 인공지능을 통해 이익을 얻는 이용자가 되어야 하며, 인공지능의 위험성을 유형별로 명시한 특별법 제정이 이루어져야 한다.

그리고, 1차 책임과 별도로 배상에 대한 문제 해결을 위해서 자동차 보험과 같은 형태의 보험제도를 운영하여 위험과 책임을 분산시켜야 한다.

형법적 차원에서 인공지능의 행위능력, 형사 책임, 형벌 능력 등을 논의해 보았으나, 확고한 인본주의적 형법관과 형법 규정, 기술 수준 등의 이유로 인공지능의 형법적 주체성을 인정하기에는 다소 무리가 있다.

그러나 반대로, 인공지능과 사람 간의 책임 귀속 문제가 발생할 수 있습니다. 인공지능의 합리적 판단과 명령에 따라 사람이 단순히 기계적으로 행동하여 발생한 위법 행위에 대해서는 인공지능의 형법적 주체성을 인정할 수 없으므로, 그 배후에 있는 사람에게 형사 책임을 물어야 하는 문제가 제기됩니다. 또한, 형사 책임을 누구에게 귀속시킬 것인가라는 실체법적 문제는 결과 발생의 근본적·결정적 원인이 무엇인지를 절차법적으로 입증하는 문제로 이어질 것입니다.

마지막으로, 인공지능에 의한 불법행위에 대한 손해를 명확히 규정하고 적용하기 위한 윤리적인 기준과 법적 규제가 필요하다. 인공지능이 자율적으로 의사결정하고 행동하는 경우, 책임과 규제를 적용하기 어렵기 때문이다. 따라서 인공지능의 불법행위에 대한 책임을 규정하고, 윤리적인 기준과 법적 규제를 마련하는 과정이 필요하다. 이를 통해서 인공지능의 발전과 윤리적, 법적 이슈를 함께 고려한 적절한 환경을 조성할 수 있을 것이다.

REFERENCES

- [1] Kang Eun-kyung, Will AI 'Iruda 2.0' escape controversy over hate and personal information?, Biz.Hankook, Available: <https://www.bizhankook.com/bk/article/23538>[Accessed: 2-Dec-2024].
- [2] John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," *AI Magazine*, vol. 27, no. 4, pp. 12-14, Dec. 2006.
- [3] Rosario Girasa, *Artificial Intelligence as a Disruptive Technology*, Palgrave, p. 7, 2020.
- [4] National Institute of Korean Language, "Standard Korean Language Dictionary," Available: <https://stdict.korean.go.kr/search/searchResult.do>[Accessed: 2-Dec-2024].
- [5] EU Artificial Intelligence Act §3(1), 'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.
- [6] Haewon Lee, "Artificial Intelligence and Legal Personality - From the Perspective of Tort Liability," *Beopjo*, vol. 70, no. 4 (no. 748), pp. 208-245, Aug. 2021.
- [7] Jinmyung Jung and Sangyong Lee, *Research on Civil Law Tasks for Preparing for the AI Society*, Ministry of Justice, pp. 17-18, Dec. 2017.
- [8] Ryan Abbott, *The Reasonable Robot*, Cambridge University Press, pp. 34-35, 2020.
- [9] David Gunning, *Explainable Artificial Intelligence (XAI)*, DARPA-BAA-16-53, Aug. 2016.
- [10] Jongmo Yang, "Characteristics and Legal Regulation of AI Risks," *Hongik Law Review*, vol. 17, no. 4, Hongik University Law Research Institute, Dec. 2016, p. 544.
- [11] Hoyoung Song, "Formation and Development of Legal Personality - Is a New Concept of Legal Personality Necessary?," *Journal of Property Law*, vol. 38, no. 2, Korean Society of Property Law, Aug. 2021, p. 26.
- [12] Kyungkyu Lee, *Recognition of Legal Personality for Non-Human Entities and Legal Status of AI*, *Journal of Legal Studies*, vol. 21, no. 1, Inha University Legal Research Institute, 2018, p. 428.
- [13] Sungjin Lee, *Recognition of AI as a Legal Personality, Theory and Practice of Civil Law*, vol. 23, no. 3, Association of Theory and Practice of Civil Law, 2020, p. 76.
- [14] Jinwoo Kim, "A Study on the Introduction and Realization of the Need for an Electronic Person System for AI," *Justice*, no. 171, Korean Institute of Law, 2019, p. 19.
- [15] Geonwoo Kim, "Autonomous AI Robots as Legal Subjects I: Significance and Perspectives," *Sungkyunkwan Law Review*, vol. 30, no. 2, 2018, pp. 226-229.
- [16] Sungwoong Lee, "Discussion on AI Firms' Eligibility for Commercial Status under Commercial Law," *Beopjo*, no. 728, Law Society, 2018, pp. 258-259.
- [17] Hyuntak Shin, "Legal Personality of AI - Considerations on the Concept of Electronic Person," *Human Rights and Justice*, no. 478, Korean Bar Association, 2018, p. 54.
- [18] Boeun Jang, "Legal Assessment of Actions Using AI - A Preliminary Study for Civil Law Discussion," *Journal of Foreign Law*, vol. 43, no. 4, Hankuk University of Foreign Studies Legal Research Institute, 2019, p. 64.
- [19] Joonggi Lee, "Legal Treatment of Robots with AI: Focusing on Legal Perception and Responsibility of Autonomous Vehicle Accidents," *Hongik Law Review*, vol. 17, no. 3, 2016, p. 3.
- [20] Choonghun Lee, "Validity of Contracts Using AI," *Journal of Culture, Media, and Entertainment Law*, vol. 14, no. 1, Chung-Ang University Cultural Media and Entertainment Law Research Institute, 2020, pp. 72-73.
- [21] Dokook Lee, "AI and Electronic Person - Focus on the Possibility of Granting Independent Legal Personality," *Law and Policy Research*, vol. 21, no. 1, Korean Association for Law and Policy Studies, Mar. 2021, p. 448.

- [22] European Parliament, "Civil Law Rules on Robotics," 2015/2103 (INL), Feb. 2017.
- [23] Jinwoo Kim, "Liability of AI Systems - Focusing on the Need for an Electronic Person System," *Journal of Central Law*, vol. 23, no. 4, Central Law Society, Dec. 2021, p. 17.
- [24] European Commission, *Creating Trust in Human-Centered Artificial Intelligence*, COM(2019) 168 final; Expert Group on Liability and New Technologies, *Liability for AI and Other Emerging Technologies*, 21 Nov. 2019, Key Finding 8, p. 37 et seq.; European Commission, *White Paper on Artificial Intelligence - A European Approach to Excellence and Trust*, Brussels, 19 Feb. 2020, COM(2020) 65 final.
- [25] HLEG, *Draft Ethics Guidelines for Trustworthy AI*, 18 Dec. 2018, p. I.
- [26] Seokcheon Jang and Byungsam Lee, "Liability for Damages from AI Tort," *Theory and Practice of Law Research*, vol. 11, no. 3, Korean Theory and Practice of Law Society, Aug. 2023, pp. 287-305.
- [27] Soyoung Park, "Legislative Trends and Implications for AI FATE (Fairness, Accountability, Transparency, and Ethics)," *Issues and Points*, National Assembly Research Service, 2023, no. 2111.
- [28] Yongju Kim, "Liability for Patent Infringement by AI Robots," *Journal of Legal Studies*, vol. 21, no. 2, Inha University Legal Research Institute, Jun. 2018, p. 74.
- [29] Seil Ko, "Legal Principles of Tort Liability for AI," *Journal of Legal Studies*, vol. 29, no. 2, Chungnam National University Legal Research Institute, May 2018, p. 85.
- [30] Sangyong Lee, "Legal Personality of AI," *Journal of Civil Law*, no. 89, Korean Civil Law Society, 2019, p. 3.
- [31] Gruber, *Legal Subjects and Partial Legal Subjects in Electronic Commerce*, 2012, p. 138.
- [32] Seoksoon Lim, "Criminal Responsibility of AI," *Criminal Policy Research*, vol. 27, no. 4, Korean Institute of Criminology, 2016, p. 76.
- [33] Kyungjin Choi, "AI and Tort Liability," *Journal of Information Law*, vol. 25, no. 2, Korean Association of Information Law, Aug. 2021, p. 49.

Authors



Yongho Kim received the B.S. degree in Law from Dong-a University in 2007, M.S. and Ph.D. degrees in Information Security from Pukyong National University in 2011 and 2019. In addition, he worked in

cybercrime investigation and digital forensics for 24 years at the Korea National Police Agency. Dr. Kim joined Kyungil University's Professor of Police Science Department in 2022. He is interested in criminal investigations, digital forensics, and cybercrime investigations.