형사정책연구 Korean Criminological Review
제27권 제4호(통권 제108호, 2016 · 겨울)

형법상 인공지능의 책임귀속

임 석 순*

국 | 문 | 요 | 약

인공지능의 급속한 발전과 눈앞으로 다가온 상용화 가능성으로 인해 형법학을 포함한 규범학도 이제는 어느 정도 미래를 대비하지 않을 수 없게 되었다. 이러한 측면에서 다루어 볼만한 형법적 쟁점으로는 인공지능의 형법적 주체성 인정여부, 인공지능 활용으로 인한 불법결과의 귀속, 위법성조각사유의 사전적 프로그래밍 등이 있다. 다만 적어도 아직까지는 인공지능의 형법적 주체성을 인정하기에는 무리가 있어 보인다. 형사책임의 귀속도 결국에는 불법결과 발생의 근본적 원인이 무엇인지에 대한 인과관계 입증의 문제가 될 것이다.

인공지능에 대해 어떠한 태도를 취하든 그로부터 야기될 수 있는 모든 위험을 완벽하게 통제할 수는 없을 것이다. 그리고 인공지능이 점차 똑똑해질수록 사람의 통제에서 벗어나기 쉬우며 따라서 종래의 이론이나 해석으로는 해결할 수 없는, 지금의 단계에서는 상상하지 못했던 문제상황이 발생할 가능성도 점차 높아질 것이다. 근본적으로는 인간 자체에 대한 (법)철학적 재론까지도 필요하게 될 것이다.

❖ 주제어 : 인공지능, 인공지능의 형법적 주체성, 인공지능의 행위, 인공지능의 형사책임, 정당화상황의 사전적 프로그래밍

* 한국형사정책연구원 부연구위원, 법학박사

Ⅰ. 여는 글

2016년 3월 초 세계정상급 바둑기사 이세돌 9단과 인공지능 알파고(AlphaGo)의 대국은 큰 센세이션을 불러일으켰다. 알파고는 구글 딥마인드(Google DeepMind)[1]에서 개발한 인공지능 바둑 프로그램이다. 첫 대국이 있기 전까지만 해도 적어도 바둑에서는 아직 기계가 인간을 뛰어넘을 수 없다는 견해가 지배적이었지만, 대국이 진행될수록 이세돌 9단의 "1승"이 간절해진 상황으로 전개되었다. 그리고 어느덧 "인공지능에 대한 인간의 숭고한 도전과 승리"라고까지 여겨지기도 했다.[2] 이는 인공지능기술 수준이 적어도 특정 영역에서는 인간의 사고력과 판단력을 뛰어넘을 수도 있음을 단적으로 보여준 사례가 되었다.

알파고 뿐만 아니라 이미 우리의 일상생활은 인공지능과 매우 밀접하게 관련되어 있다. 로봇청소기는 이미 상당히 흔한 가전제품이며, 냉장·냉동기능 외에 다양한 커뮤니케이션 기능을 탑재한 스마트 냉장고가 등장했고,[3] 운전자가 필요 없는 무인자동차의 상용화도 기술적·제도적으로 박차를 가하고 있다.[4] 수십 년 전에 방영되었던 외화 전격Z작전에 등장한 인공지능 자동차 키트나 터미네이터에 등장한 T-800이나 스카이넷 등을 이제는 단순한 공상과학적 상상력의 산물로 치부할 수는 없을 것 같다.

이처럼 적어도 한 가지 방면에서는 사람을 뛰어넘는 인공지능의 판단과 사고를 사람의 그것과 동일한 것으로 법적 가치평가를 내릴 수 있을까? 특정 전문영역에서

---

1) 2010년 영국에서 "딥마인드 테크놀로지(DeepMind Technologies)"라는 이름으로 설립된 인공지능 프로그램 개발회사로서, 2014년 구글이 인수하여 현재의 이름으로 바뀌었다 (https://en.wikipedia.org/wiki/Google_DeepMind, 2016.8.23. 최종검색).

2) 총 다섯 번의 대국 가운데 이세돌 9단은 제4국만을 승리했으며, 단 1승만으로 기적을 일구어 냈다며 국내외에서 상당한 찬사가 쏟아져 나왔다. 연합뉴스, "이세돌의 인간승리…'바둑 괴물' 알파고도 '백기', 2016.3.13.(http://www.yonhapnews.co.kr/bulletin/2016/03/13/0200000000AKR20160313053352007.HTML, 2016.8.23. 최종검색); 허핑턴포스트 코리아, "이세돌의 승리에 대한 해외의 반응", 2016.3.14.(http://www.huffingtonpost.kr/2016/03/14/story_n_9456580.html, 2016.8.23. 최종검색) 등 참조.

3) 머니투데이, "'쇼핑에 음악까지' 삼성 스마트 냉장고 출시...가격은?", 2016.3.30.(http://news.mt.co.kr/mtview.php?no=2016033014233304583, 2016.8.22. 최종검색)

4) 아시아경제, "'두 손 두 발' 다 든 운전자.. 무인차 시대가 왔다", 2015.3.31. (http://media.daum.net/news/view/print?newsId=20150331175016972, 2016.8.22. 최종검색)

활용되는 인공지능은 사람의 사고와 판단을 뛰어 넘어 가장 합리적인 결론을 도출할 수 있을 것이다. 그리고 곧바로 우리의 일상생활에 직접적으로 영향을 미치게 될 것이다. 경우에 따라서는 우리가 낯선 길을 운전할 때 내비게이션의 지시에 "순종"하는 것처럼, 인공지능의 판단에 따라 사람이 행동해야 하는 사례가 발생할 수 있다. 궁극적으로는 특정 분야에서 사람의 사고와 판단을 완전히 대체하는 경우도 발생할 수 있다. 이러한 경우 인공지능의 사고와 판단은 사람의 그것에 버금가는 사회적 파급력을 갖게 될 것이다.

이처럼 인공지능의 본격적·보편적 활용이 목전에 있고, 그 사회적 파급력이 상당할 것으로 예측됨에도 불구하고 인공지능 활용과 관련한 쟁점을 심도 있게 다루고 있는 (형)법이론은 아직 없는 것으로 보인다. 인공지능의 형사책임 및 형사사법적 활용방안에 관한 논의[5)]가 이제 막 시작되었으니, 형사법적 측면에서의 논의도 아직 걸음마 단계라 할 수 있다. 따라서 본 논문에서는 인공지능과 관련해 발생할 수 있는 형법적 쟁점들을 도출하고, 앞으로 인공지능과 관련한 활발한 담론의 장이 열리기를 소망하면서 이러한 쟁점들을 ― 비록 정치하고 깊은 수준에는 못 미치더라도 개략적으로나마 특히 책임귀속 문제와 관련해 ― 논의해 보도록 하겠다.

Ⅱ. 인공지능과 로봇, 그리고 인공지능의 수준

1. 인공지능인가 로봇인가?

인공지능의 실체는 무엇인가? 터미네이터와 같은 "똑똑한 로봇"을 말하는 것인가, 사람의 영혼과 같이 실체가 없는 그 무언가를 말하는 것인가? 인공지능은 말 그대로 사람이 인위적으로 만들어낸 지능이다. 홀로 계산하고 판단하고 결정할 수 있는 능력으로 터미네이터와 같은 로봇도 인공지능을 탑재하고 있고, 최근 화두가 되고 있는 무인자동차에도 인공지능이 탑재되어 있다. 요컨대 로봇이든 무인자동차이든 물리적 실체는 인공지능을 담는 "그릇"에 불과하며, 특히 여러 형법적 쟁점과

5) 윤지영 외, 법과학을 적용한 형사사법의 선진화 방안(VI), 2015.

관련해 중요한 것은 이러한 그릇보다는 그릇 안에 담겨있는 내용물, 즉 "인공지능"이다. 따라서 영화 터미네이터에서 "핵전쟁을 일으켜 인류멸망 프로젝트를 가동시킨 장본인"인 스카이넷(Skynet)도 그 자체로는 별다른 거동기능은 없으나 독자적인 계산, 판단, 결정능력이 있으므로 여기에서 논하는 인공지능의 범주에 포함된다. 요컨대 "로봇", "지능로봇" 또는 "지능형 로봇"을 주제로 하는 문헌들[6]도 결국 그 논의의 방점은 "인공지능"에 찍고 있다고 보아야 할 것이다.

인공지능은 ① 외부환경변화를 인식할 수 있는 능력, ② 독자적 상황판단능력, ③ 자율적 행위결정 능력을 기본요소로 하며, 경우에 따라서는 ④ 사람과 상호작용하는 능력까지도 포함하기도 한다.[7] 우선 외부의 환경변화를 인식한다는 것은 주변의 지형지물이나 온도, 기압 등 인공지능의 역할이나 기능에 따라 감지해야할 사항들을 별도의 명령 없이도 감지하는 것을 의미한다. 그리고 이러한 인식을 토대로 주변 상황을 판단 및 평가하고, 궁극적으로는 스스로 그 상황에 필요한 행위로 나아갈 수 있어야 한다. 상호작용능력은 특히 심리치료용으로 개발된 인공지능에서 필요한 기능이다. 예컨대 자폐증을 앓고 있는 어린이가 모방학습을 할 수 있도록 돕는 인공지능 로봇은 칭찬과 격려를 통해 상호작용을 함으로써 어린이가 사회성과 자립성을 기를 수 있도록 해준다.[8]

2. 인공지능의 수준 : 강한 인공지능과 약한 인공지능

먼 미래나 영화적 상상은 제쳐두고서라도 이미 일상생활에서 접할 수 있는 인공지능은 매우 다양하다. 작게는 로봇청소기에서부터 스마트 냉장고, 무인자동차에 이

6) 고인석, "로봇윤리의 기본 원칙: 로봇 존재론으로부터", 범한철학 제75집, 2014;, 박현섭, "지능형 로봇 산업발전을 위한 전부정책", TTA 저널 제158호, 2015; 신현우, "진화론적 지능형 서비스 로봇에 대한 실천윤리학적 고찰", 윤리연구 제79호, 2010; 윤지영 외, 앞의 보고서; 관계부처 합동, 제2차 지능형 로봇 기본계획(안), 2014; Meyer, Der Einsatz von Robotern zur Gefahrenabwehr, 2014; Hilgendorf, Können Roboter schuldhaft handeln?, 2012; Beck, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 2009 외 다수.

7) 윤지영 외, 앞의 보고서, 21쪽.

8) 윤상석·박성기·최종석, "자폐 범주성 장애 아동의 사회적 기능향상을 용이하게 하는 상호작용로봇", 한국HCI학회 학술대회 2014, 2014.2, 532-534쪽.

르기까지 이제는 인공지능을 빼놓고 현대인의 생활을 논할 수 없을 것이다.

이와 같이 다양한 스펙트럼을 가진 인공지능은 크게 강한 인공지능(strong AI)과 약한 인공지능(weak AI)으로 구분할 수 있다.[9] 영화 터미네이터의 스카이넷이나 영화 AI에 등장하는 주인공 로봇은 자아를 인식한 단계에 이른 인공지능을 강한 인공지능이라고 한다. 다만 강한 인공지능은 기술적 차원에서도 꽤나 먼 미래의 이야기인 것으로 보인다. 즉 아직까지는 강한 인공지능을 법적 논의의 대상으로 삼는 것은 시기적으로 다소 이르다고 할 수 있으며, 오늘날 당면한 또는 가까운 미래에 당면하게 될 문제는 약한 인공지능 또는 "중간단계의 인공지능"[10]으로 인해 발생할 현상이나 결과들이라고 할 수 있다. 따라서 주체성을 담지하기 시작한 인공지능(로봇)의 형사책임 등의 문제는 몇 십 년 후의 쟁점으로 남겨두고, 지금으로서는 아직 그러한 수준에는 이르지 않은 인공지능의 활용으로 인해 촉발될 수 있는 여러 형법적 쟁점만을 다룰 필요가 있다.

한편 현재 상용화되어 있는 로봇청소기나 조향, 균형잡기, 장애물 회피, 자동회귀 등을 위한 기초적 기능만 탑재한 — 오늘날 오락이나 취미용으로도 흔히 접할 수 있는 — 드론은 "매우 약한 인공지능"으로 분류할 수 있다. 설계자가 애초에 이식한 기능 외에는 어떠한 학습능력도 없으며, 몇몇 알고리즘을 제외하고는 독자적 판단능력이나 자율적 행위결정능력이 있다고 볼 수도 없기 때문이다. 이와 같은 매우 약한 인공지능이 보다 높은 수준의 인공지능에서와 같은 새로운 형법적 쟁점을 촉발시키는 일은 상당히 드물 것으로 보인다.

9) 정보통신기술진흥센터, "미국의 인공지능(AI) 기술 R&D 추진 동향", 해외 ICT R&D 정책동향 2015년 3호, 2015, 2쪽; 윤지영 외, 앞의 보고서, 38쪽.

10) 매우 높은 수준의 인지능력과 판단능력, 실행능력과 학습능력을 갖추고 있으나 자아를 인식하지는 못한 단계의 인공지능이라 할 수 있다. 빠른 시간 안에 인간으로서는 시도할 수도 없는 무수한 경우의 수를 시뮬레이션함으로써 대국에서 승리하는 길을 찾아낸 알파고는 적어도 바둑에서만큼은 중간단계의 인공지능으로 분류할 수 있을 것이다.

Ⅲ. 인공지능의 형법적 주체성 : 행위와 책임, 형벌능력

1. 인공지능의 행위

오늘날 형법도그마틱에 따르면 범죄가 되는 것은 사람의 "행위"이다. 그리고 무엇이 행위인가에 대해서는 역사적으로 인과적 행위론에서부터 목적적 행위론, 사회적 행위론, 인격적 행위론 등의 논의가 있어왔다. 인공지능의 형법적 주체성을 문제삼기 위해서는 인공지능의 작동이나 기능이 형법적으로 의미 있는 "행위"이어야 한다.[11)]

인과적 행위론에 따르면 형법적으로 의미 있는 행위는 오로지 "의사에 의한 신체활동(willkürliche Körperbewegung)"이라고 한다.[12)] 이에 따르면 학습을 통해 다변하는 상황에 대처할 수 있는 인공지능의 행동은 적어도 나름대로의 "의지적" 판단에 따른 것이기 때문에 인공지능은 형법적으로 의미 있는 행위주체가 될 수도 있다.[13)] 목적적 행위론에 따르면 목적지향성(Zweckgerichtetheit)이 인간행위의 중심요소가 되고, 행위자의 목적지향적 의지표출이 형법적으로 의미 있는 행위로 인정된다.[14)] 학습능력이 있는 인공지능은 각각의 상황에 따라 자신의 임무를 수행하기 위한 최선의 단계를 독자적으로 정한다고 볼 수 있기 때문에 목적적 행위론에 따를 때에도 인공지능의 행위주체성을 인정할 여지가 존재한다. 다만 그 자신의 판단을 프로그램에 의해 사전에 주어진 선택(choice)에 따라 한 것인지 독자적 평가에 따른 판단(judgement)에 따라 한 것인지를 우선적으로 평가해야 한다는 지적도 있다.[15)] 독자적 판단에 의해 행위를 결정하였다 할지라도 인공지능이 스스로 목적을 의도적으로(willentlich) 설정하였다고 볼 수는 없다는 것이다. 사회적 행위론에 따르면 적어도 인공지능 스스로 자신이 무슨 행동을 하고 있는지 인지하고 자신의 행동이 "사회적으로 중요한" 것임을 알아야 한다.[16)] 끝으로 인격적 행위론은 인간의 인격

---

11) 윤지영 외, 앞의 보고서, 160쪽.

12) 김일수, 한국형법 I [총론 상] 개정판, 1996, 253쪽; 배종대, 형법총론 제12판, 2016, 37/1; 이재상, 형법총론 제7판, 2011, 6/7 참조.

13) Gleß/Weigend, *Intelligent Agenten und das Strafrecht*, ZStW 126(2014), 571.

14) 김일수, 앞의 책, 254쪽; 배종대, 앞의 책, 37/15; 이재상, 앞의 책, 6/9 참조.

15) Gleß/Weigend, 앞의 글, 572.

이 외부로 나타난 것이 행위라고 한다.[17] 그런데 사회적 행위론과 인격적 행위론은 각각 사회적 중요성과 인격이라는 개방적 개념을 전제로 한다.[18] 따라서 이들 개념에 대한 확정이 이루어진 다음에야 비로소 인공지능의 형법적 행위능력 여부를 판단할 수 있다. 다만 약한/중간 수준의 인공지능이 자신의 기능이나 작동이 갖는 사회적 의미를 통찰한다고 보기는 어렵다. 또한 도덕적 자기반성능력이나 양심, 형벌감수성 등을 담지하지 못한 인공지능에게 인격이 존재한다고 인정하기도 어려울 것으로 보인다.[19]

2. 인공지능의 형사책임

행위책임을 기초로 하고 있는 오늘날 형법도그마틱은 행위능력을 전제하므로, 애초에 인공지능의 행위능력을 인정할 수 없다면 역시 인공지능의 형사책임을 논하는 것도 무의미하다 할 것이다. 그러나 행위론이 확고불변의 원리도 아니며, 오늘날 태아에게 부분적으로 권리능력을 인정하고 있는 것, 역사적으로 권리능력을 갖는 인격적 주체의 범위가 지속적으로 확장되어 왔던 것에 비추어 볼 때,[20] 형사책임의 개념 및 범주 또한 충분히 변화할 여지가 있다고 보아야 한다.[21] 이러한 맥락에서 언어사용관습과 언어의 유동성을 고려하면 “행위”나 “책임” 등의 법률적 기본개념도 원칙적으로 인공지능에게 적용할 수 있다는 주장도 제기되고 있다.[22]

최근 인간의 뇌와 인공지능에 대한 연구가 활발해지면서 그에 관한 새로운 지식과 함께 책임개념의 변화가능성을 어느 정도 타진해 볼 수 있을 것 같다. 즉 인간의 필수적 속성과 능력에 대한 어떠한 형상이 형법상 유죄판단 가능성의 근거가 되는지를 명확하게 확정하는 것 자체가 애초부터 불가능하기 때문에 권리, 유책성

16) 김일수, 앞의 책, 256쪽; 배종대, 앞의 책, 37/21 이하; 이재상, 앞의 책, 6/17 이하 참조.
17) 김일수, 앞의 책, 264쪽.
18) 배종대, 앞의 책, 37/21 및 37/31.
19) Meyer, 앞의 글, S. 220, 226; Gleß/Weigend, 앞의 글, 569, 579.
20) *Gruber, Rechtsubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehr*, 2012, S. 138.
21) 김영환 “위험사회에서의 책임구조: 자연 재해에 대한 법적 담론”, 홍익법학 제14권 제1호, 2013, 825쪽; 책임개념 및 책임모델의 유동성에 대해 보다 자세한 것은 826-833쪽 참조.
22) Hilgendorf, 앞의 글, S. 122, 125.

(Verantwortlichkeit), 책임(Schuld) 등과 같은 개념을 근본에서부터 되묻는 것도 가능할 것이다.[23] 예컨대 어쩌면 인공지능에게 합리적 판단능력이 있다는 이유로 책임능력을 인정하는 논리를 구성할 수도 있을 것이다.[24] 권리주체는 더 이상 자신 스스로 정한 행위와 의사표명으로 특정할 수 있는 사람인 개인에 국한될 수 없으며, 특히 정보화 시대에 사람은 ID와 같은 인위적 속성으로도 정체성을 규명할 수 있을 것이라는 견해도 조심스럽게 제기되고 있다.[25] 즉 정보기술 시스템이 한편으로는 인격발현의 새로운 영역을, 다른 한편으로는 새로운 사회적 행위를 형성하고 있기 때문에 일정한 조건 하에서 비신체적이고 살아있지 않으며 정보기술적인 인공물에 인격을 부여하는 것을 생각해 볼 수 있다는 것이다.[26] 인공지능에 삽입된 복잡한 연산공식이 기능적으로 인간의 도덕적 자기평가 및 자기조종과 일치한다면 인공지능에 형법적 책임비난을 가하는 것도 가능할 것이라는 견해도 있다.[27]

그러나 아직까지는 인공지능의 형법적 책임을 지는 것은 그것이 "도덕적 인공지능"이 되지 않는 한 적어도 오늘날의 형법관 내지 법철학에 의하면 요원하다고 볼 수밖에 없다.[28] 형법규범에서도 ― 법인처벌에 대한 예외적 양벌규정을 제외하고 ― "…한 자"를 구성요건실현의 주체로 규정하고 있는 것은 형벌규범의 수범자이자 국가형벌권의 대상을 오로지 종으로서 인간으로 한정한 것이기 때문에 인공지능을 형벌규범의 수범자로 받아들일 수 없다.[29] "책임비난의 내적 근거는 인간이 자유롭고, 자신의 책임에 따른 것이며, 도덕적인 자기결정을 의도하였으며, 따라서 스스로 법에 따르고 불법에 저항하도록 결정하고, 법적 당위규범에 따라 자신의 행동을 조정하고 법적으로 금지된 것을 회피할 능력이 있다는 점에 있다"[30]고 한 독일연방법원의 판례는 "인간", "행위", "도덕" 등에 대한 도덕적·철학적 패러다임이 바뀌지 않

23) Beck, 앞의 글, S. 229.
24) Beck, 앞의 글, S. 230.
25) Gruber, 앞의 글, S. 148.
26) Gruber, 앞의 글, S. 157.
27) Gleß/Weigend, 앞의 글, 576.
28) Gleß/Weigend, 앞의 글, 589.
29) *Ziemann, Zur Diskussion über ein Strafrecht für Maschinen*, 2013, S. 184-185; Joerden, *Strafrechtliche Perspektiven der Robotik*, 2013, S. 203.
30) BGHSt 2, 194, 200 = BGH NJW 1952, 593, 594.

는 한 여전히 유효하다 할 것이다.[31]

3. 인공지능의 형벌능력

인공지능에게 형법상 고유한 행위능력을 인정할 수 없다 하더라도, 법인에 대한 양벌규정과 같이 부분적으로나마 형벌능력을 인정하는 방안도 고려해 볼 수 있을 것이다. 독일형법학계에서 이른바 로봇담론을 주도하고 있는 에릭 힐겐도르프(Eric Hilgendorf) 교수[32]는 필요에 의해 "법인" 개념이 만들어진 것과 마찬가지로 독립적 책임주체로서 이른바 "e-인격(e-Person)" 개념을 도입할 수 있을 것이라고 보고 있다.[33] 이러한 착안은 법인에 대한 형사처벌이 존재하지 않는 독일에서는 아직 시기상조라 할 수 있으나, 우리 형법체계에서는 보다 적극적으로 고려해 볼 수도 있을 것이다. 특히 법인의 행위능력을 인정하는 민사법체계에서는 보다 더 유용하게 차용할 수 있을 것으로 보인다.

그러나 법인과 인공지능은 행위의 발현양태가 다르기 때문에 법인과 같은 구조로 인공지능에게 인격을 부여한다거나 최소한 인공지능의 형벌능력을 인정할 수도 있다는 논리는 성립할 수 없다. 법인의 행위는 그 구성원인 인간의 의사결정을 토대로 한다. 그러나 인공지능은 애초에 이식되었던 프로그램에 더해 스스로 학습을 통해 사고하고 행동한다. 즉 인공지능의 의사결정 및 행동은 — 기본적인 알고리즘을 제외하고는 — 생산되어 작동하기 시작한 이후부터는 개발자 내지 생산자의 의사로부터 단절된다고 볼 수 있다. 뿐만 아니라 현행법상 법인에 대한 처벌은 대부분 법인

31) 인공지능에 형사책임을 귀속시키고자 하는 시도는 현재의 책임개념과는 전혀 다른 책임개념을 개발하지 않는 한 어떠한 설득력도 얻지 못할 것이다(Joerden, 앞의 글, S. 205). 다만 장차 자아를 인식한 강한 인공지능이 출현하게 된다면, 이른바 "인공지능윤리" 내지 "로봇윤리"의 문제가 대두될 것이며, 무엇보다도 인공지능에게 인간도덕에 따른 규칙체계를 이식할 수 있는지 또는 해야 하는지, 이식한다면 어느 수준으로 할 수 있는지 또는 해야 하는지가 중요한 테제가 될 것이다(Gleß/Weigend, 앞의 글, 577.).

32) 2010년 독일 뷔르츠부르크 대학교에 로봇법 연구소(Forschungsstelle RobotRecht)를 설립하여 인공지능 로봇의 사용으로 인해 발생할 수 있는 법적 문제들을 선도적으로 연구하고 있다(윤지영 외, 앞의 보고서, 154-155쪽 참조).

33) Hilgendorf, *Recht, Maschinen und die Idee des Posthumanen*, Telepolis, 25.05.2014.; Fitzi, Roboter als "*legale Personen" mit begrenzter Haftung*, 2013, S. 391-394.

이 소유한 재산에 대한 제재인데, 현행 재산법상 인공지능은 법인과 달리 재산을 소유할 수 없다.[34] 사형이나 구금과 같은 형사처벌은 인공지능의 해체나 파괴, 프로그램 재설치 등을 고려할 수 있는데, 이를 굳이 "형사처벌"로 둘 필요가 있는지는 법인에 대한 형사처벌에 대한 것과 마찬가지로 회의적이다.[35]

애초부터 형벌이 인간에게만 가해진 것은 아니라는 점에서 인공지능의 형벌능력을 고려하는 견해도 있다. 실제로 1474년 스위스 바젤에서는 이른바 동물소송이 진행되어, 닭에게 사형이 선고된 바 있다.[36] 이러한 사례에 비추어 중세의 이른바 "동물형법"을 오늘날 "인공지능형법의 모범"으로 — 물론 당시의 법규 형태와 내용을 그대로 계수하지는 못하겠지만 — 삼으려는 것이다.[37] 그러나 계몽주의 철학과 예방적 형벌론의 등장으로 동물에 대한 형법이 종말을 맞은 지는 이미 오래이며, 앞서 언급한 바와 같이 도덕적·철학적 패러다임이 변화하지 않는 한 단순한 기계이든 인공지능이든 동물과 마찬가지로 인간과 동등하거나 유사한 지위를 갖지는 못할 것이다.[38]

Ⅳ. 인공지능이 야기한 결과에 대한 책임귀속

1. 과실책임과 예측불가능한 위험의 사회적 용인가능성

인공지능에게 형법적 주체성을 인정할 수 없다 하더라도, 인공지능 활용으로 인한 불법결과는 발생할 수 있다. 이러한 경우 그 불법적 결과를 예견하고 회피해야 할 주체는 누구인가, 그리고 인공지능 활용의 반대급부로서 어느 정도의 불법적 결과는 용인할 수 있는 것인가?

---

34) 인공지능에게 인격을 부여한다 하더라도 다시 그에게 재산권을 인정할 것인지는 또 다른 법률적 문제를 야기한다.

35) Ziemann, 앞의 글, S. 189; Gleß/Weigend, 앞의 글, 578.

36) Gleß/Weigend, 앞의 글, 566.

37) Matthias, *Automaten als Träger von Rechten*, 2008, S. 167.

38) Gleß/Weigend, 앞의 글, 567.

가. 예측가능성과 회피가능성

자신의 의심스러운 행동이 타인의 보호이익을 침해할 수 있다는 점을 알 수 있는 자가 그럼에도 그러한 행위를 행한 경우, 즉 구성요건결과의 발생을 예측하고 그에 따라 결과발생을 회피할 수 있었는데 그렇게 하지 않은 경우에는 주의의무를 위반한 것으로 볼 수 있다.[39] 독일연방최고법원 판례[40]는 특히 위험이 잠재된 제품의 생산 및 판매와 관련하여 민사법에서 형사법으로 원용된 매우 강력한 주의의무원칙을 발전시켰다. 그에 따르면 주의깊은 생산자는 제조물의 안전수준이 관련 학문과 기술의 기준에 부합하고 판매 전에 충분한 테스트를 거친 이후에야 비로소 시장에 내 놓아야 한다. 그리고 제품을 생산하여 판매하기 시작한 이후에도 생산자는 소비자의 반응(불만신고)을 지속적으로 관찰하고 배상이나 수리 등 필요한 피드백을 주어야 한다. 설계 및 생산단계에서는 미처 예측하지 못했던 피해나 위험이 발견된 경우 생산자는 소비자에게 그에 대한 경고를 해야 하고 필요한 경우 제품을 리콜할 의무가 있다.[41]

그런데 인공지능은 주변환경으로부터 습득한 정보를 독자적으로 평가하고 그 평가에 따라 사람의 개입 없이 반응하여 자신에게 부여된 임무를 최적의 방법으로 수행하려고 할 것이다. 사용자가 복잡한 인공지능의 반응을 지속적으로 관찰한다 할지라도, 인공지능이 어떠한 방식으로 데이터를 인지하고, 어떻게 해석하며, 어떻게 반응할지를 하나하나 예측할 수는 없다. 사용자뿐만 아니라 생산자나 관리자도 인공지능이 활용되는 상황 안에서 발생할 수 있는 모든 상황을 예측할 수 없으며 모든 예측할 수 없는 위험을 회피하도록 사전에 세밀하게 프로그래밍하는 것도 불가능하다. 또한 인공지능 시스템은 독자적으로 학습을 행하기 때문에 충분히 학습을 행한 인공지능의 판단과 행동을 예측하기는 더욱 어려워진다.[42]

"불법결과의 발생"과 관련해서는 예컨대 인공지능의 사용자에 대해 두 가지의 상반된 논의가 가능할 것이다. 첫째는 인공지능이 행한 불법적이거나 유해한 행동은

39) 배종대, 앞의 책, 153/1; Duttge, MK-StGB, 2. Auflage, 2011, § 15 Rn. 107 ff.

40) BGHSt 37, 106 = BGH NJW 1990, 2560 (이른바 가죽스프레이 판결).

41) Kuhlen, *Strafrechtliche Produkthaftung*, 2012, Rn. 29 ff. 특히 Rn. 39-40.

42) 인공지능이 오류·모순 투성이인 인간의 행동을 학습한다면, 처음에는 (어쩌면) 무결점이었을 인공지능도 '오류'와 '모순'을 학습하게 될 수도 있을 것이다.

자신의 독자적인 정보처리를 근거로 한 것으로 사용자가 예측하거나 회피할 수 없는 것이기 때문에, 자율적으로 행동하는 인공지능의 사용자에게는 결코 그 책임을 귀속시킬 수 없다는 주장이다. 둘째로는 인공지능의 사용자는 인공지능의 소유자이기 때문에 인공지능의 활동으로 인해 발생한 "모든 것"에 책임을 져야 한다, 즉 예측가능성 및 회피가능성과 무관하게 인공지능의 행동으로 인해 발생한 모든 피해에 (무과실)책임을 져야 한다는 주장도 있을 수 있다. 이 두 가지 주장은 생산자나 관리자, 또는 설계자에게도 그대로 적용될 수 있을 것이다. 한편 사용자가 완벽하게 지배할 수 없는 인공지능을 세상에 내놓은 자는 인공지능의 행동에 대한 예측불가능성을 근거로 하여 자신이 미처 예측하지 못했던, 타인에 대한 피해를 유발할 수 있는 오작용(Fehlreaktion)을 부인해서는 안 된다는 주장도 제기되고 있다.[43]

나. 예측할 수 없는 위험의 사회적 용인

오늘날 사회가 아직까지는 생활세계의 기계화·자동화에 무비판적이지는 않으나, 다른 한편으로 인공지능 연구에 박차를 가하고 있는 것은 이미 사회가 그와 같은 최첨단 기술의 투입에 상당한 관심을 두고 있음을 방증하는 것이라 할 수 있다.[44] 또한 일체의 예외를 인정하지 않고 모든 보호이익 침해를 형사처벌하는 것은 인공지능 개발의 동기의식을 저하시켜 인간의 삶에 획기적인 도움을 줄 무한한 잠재적 가능성을 놓치게 만들 수도 있다.[45] 따라서 위험감수의 수준과 주의의무 경감 정도를 판단·결정하기 위해서는 이와 같은 사회적 우려와 기대 양자를 모두 고려할 필요가 있을 것이다.

이러한 측면에서 우리 사회가 위험을 동반한 새로운 기술을 받아들일 준비가 되어 있다면, 그리고 인공지능의 사용으로 인해 예측할 수 없는 피해가 발생할 수 있음을 일반적으로 받아들일 수 있다면, 결과적으로 침해가 발생하였다 할지라도 인공지능의 생산자나 운영자, 설계자를 형사처벌해서는 안 된다는 의견이 제시되고 있다.[46] 사회 또는 공동체가 어떠한 혁신적 기술을 이용함으로 인해 일정한 이익을

43) Gleß/Weigend, 앞의 글, 581.
44) Gleß/Weigend, 앞의 글, 583.
45) 윤지영 외, 앞의 보고서, 162쪽.

보고 있음에도 그로부터 불가피하게 발생할 수 있는 위험을 온전히 생산자나 운영자에게 전가하는 것은 공평하지 않다는 것이다.[47] 이러한 경우에는 형사처벌보다는 보험이나 민사상 위험책임(Gefährdungshaftung)을 고려하는 것이 보다 타당하다고 한다.[48]

이러한 논지에서 인공지능 서비스를 이용하는 것이 사회적으로 매우 유용하다거나 반드시 필요하다는 공감대가 형성된다면, 주의의무를 경감함으로써 인공지능의 적극적 개발 및 활용 가능성을 열어둘 수 있을 것이다. 다만 사회 전체적인 관점에서 인공지능 활용의 유용성과 필요성이 인정된다 할지라도 그 활용가능성을 극대화하기 위한 주의의무 경감 수준을 어떻게 정할 것인지, 또는 활용가능성 극대화와 주의의무 수준 사이의 적절한 접점을 어디로 정할 것인지는 여전히 해결해야 할 중요한 문제로 남는다.[49]

적절한 접점으로는 우선 판매자로 하여금 사용자에게 포괄적 경고의무 및 안내의무를 이행하도록 하면서, 생산자는 자신이 생산한 인공지능의 사후 경과를 지속적으로 관찰하고 피해가 발생한 경우 피해보고를 접수하여 신속하고 적절한 피드백을 이행할 의무를 부과하는 것을 고려해 볼 수 있다.[50] 즉 설계·제작·판매 단계에서는 주의의무를 감경하고, 사후 관리단계에서의 주의의무를 강화하는 것이다. 예컨대 판매되어 사용 중인 무인자동차가 유해한 "행위" 방식을 보인 사례가 드물지 않거나 사용자의 사소한 오·남용에 의해 심각한 해악이 발생한 경우, 무인자동차의 판매를 금지하거나 경우에 따라서는 모든 판매제품에 대한 리콜을 명하도록 해야 한다.[51] 그리고 이때 생산자나 운영자, 설계자가 법적으로 의무화된 조치를 전혀 취하지 않으면, 의무불이행에 대한 형사처벌할 수 있도록 하는 것이다. 요컨대 온전히 지배할 수 없는 인공지능을 설계·제작·판매함으로써 생산자나 운영자, 설계자는 일반대중에 예측할 수 없는 위험을 야기하거나 증대시켰음에도, 제조물을 지속적으로 관리·

46) Kuhlen, 앞의 글, Rn. 27 ff.

47) 김호기, "개발위험의 항변과 형법적 제조물책임", 형사정책연구 제27권 제1호, 2016, 192쪽; Gleß/Weigend, 앞의 글, 583.

48) Gleß/Weigend, 앞의 글, 584.

49) Gleß/Weigend, 앞의 글, 583; 윤지영 외, 앞의 보고서, 162쪽.

50) Kuhlen, 앞의 글, Rn. 34 ff. 참조.

51) Gleß/Weigend, 앞의 글, 585.

감독하고 사후적으로 인지된 위험으로부터 야기될 수 있는 결과를 예방하고, 기왕에 발생한 결과를 제거할 의무를 이행하지 않았음을 이유로 형사처벌하는 것이다.[52)]

다. 형법적 제조물책임

인공지능과 같은 첨단과학기술이 적용된 제조물은 작동방법, 기능뿐만 아니라 그 생산공정도 매우 복잡하고 전문적이어서 생산과정에 오류가 발생할 가능성이 전혀 없다고 장담할 수 없다. 오히려 제조물의 성능과 생산공정의 복잡해지고 전문화될수록 오류가능성도 그에 비례해 증가한다. 뿐만 아니라 제조업자들은 미래의 손해배상을 대비해 보험을 들어두게 마련이고, 무엇보다도 제조자가 잠재적 손해배상금액을 제품가격에 미리 반영해 놓는 편법을 막을 방도도 없다. 결국 첨단 제조물의 위험성을 소비자 내지 사용자가 떠안게 되는 꼴이다. 이처럼 민사상 손해배상법제만으로 새롭고 복잡한 제조물의 위험성을 담보하고 소비자의 안전욕구를 충분히 충족시켜준다고 보기는 어렵다. 이와 같은 왜곡된 배상책임 구조와 인공지능이 사람에게 미칠 영향, 잠재적 위험성, 기능적 복잡성, 취급의 전문성, 소비자의 안전욕구 증대 등을 고려하면 형법적 제조물책임의 도입은 필요해 보인다.[53)]

그런데 현행「제조물 책임법」은 여전히 민사상 손해배상책임만을 규정하고 있다. 본 법률 제4조 제2항을 형법적 제조물책임 인정근거로 보아야 한다는 견해[54)]가 있으나, 죄형법정주의원칙상 본 법률에는 제조물관리의무 위반에 대한 처벌규정이 없기 때문에 아직은 형법적 제조물책임의 법적 근거가 마련되었다고 볼 수는 없다. 형법적 제조물책임을 인정하기 위해서는 제조업자 등에게 사실상 무과실의 형사책임을 인정해야 할지, 민법상 제조물책임과 같이 주의의무위반에 대한 입증책임을 전환해야할지 등에 대한 논의가 선행되어야 할 것이다.[55)]

---

52) Gleß/Weigend, 앞의 글, 585, 587.

53) 전지연, “형벌론적 관점에서 본 형법적 제조물책임의 필요성”, 형사정책 제13권 제1호, 2001, 105쪽 참조. 인공지능 활용으로 인한 형법적 제조물책임 발생에 대한 몇 가지 예로는 윤지영 외, 앞의 보고서, 162-163쪽.

54) 김호기, 앞의 글, 180-181쪽 : 제조물 책임법 제4조 제2항은 “제조자에 대하여 일정한 작위의무를 부과하는 행위규범으로서의 의미도 갖고 있”기 때문에 제조자에게 작위의무로서 형법적 제조물계속감시의무의 발생근거가 된다고 한다.

55) 형법적 제조물책임의 필요성에 대한 보다 자세한 논의는 전지연, 앞의 글; 하태훈, “缺陷製造物로

2. 사람과 인공지능 사이의 책임귀속

가. 사람 배후의 인공지능

산업사회의 대량생산체계에서 노동자 개인은 그 시스템이 어떠한 알고리즘에 의해 운영되고, 자신에게 어떠한 이유와 경로로 해당 임무가 배정되는지 알기가 쉽지 않다. 또한 자신의 행동이 장차 어떠한 결과를 가져올지 예측하는 것도 사실상 불가능한 경우가 많다. 지금까지는 산업의 자동화가 노동력 대체의 의미만을 갖고 있었다면, 인공지능이 투입되는 미래시대에는 경영의 대체까지도 가능할 것이라고 예측할 수 있다. 인간보다 연산능력이 훨씬 뛰어난 인공지능이 시장의 각종 요소를 면밀히 분석하여 최적의 경영판단을 내리고, 사람이 개별영업을 수행하게 되는 것이다. 이처럼 자동화된 시스템에 따라 인공지능이 분석하고 판단하여 결정한 작업명령에 따라 — 별다른 의심 없이 "기계적으로" — 행동한 사람이 결과적으로 범죄적 행동을 하였거나, 범죄적 결과를 발생시켰을 때에는 누구에게 형사책임을 귀속시킬 것이며, 누구를 어떠한 근거로 처벌할 것인가 하는 문제가 발생하게 될 것이다.[56] 이러한 경우 인공지능의 명령에 따라 행동한 사람은 적어도 인공지능이 합리적이고 적법하게 판단하여 명령하였을 것이라는 신뢰를 갖거나 경우에 따라서는 일체의 가치판단이 배제된 상태에서 행동하는 사례도 있을 것이다. 즉 인공지능의 명령에 따라 행동한 단순 노동자에게 범죄적 결과에 대한 확고한 예측가능성이나 주의의무가 있었다고 보기는 어려울 것이다.

인공지능에게 행위능력과 책임능력, 형벌능력을 인정한다면 범죄적 결과에 대한 근본적으로 예측했어야 하고, 범죄적 결과를 회피하기 위한 방안을 마련했어야 할 인공지능에게 불법결과에 대한 책임을 귀속시키는 것은 논리적으로 별다른 문제가 없어 보인다. 다만 인공지능에게 형법적 주체성을 인정한다 하더라도 형법의 구체적 구현방법, 즉 형벌의 종류에 대한 문제가 남는다. 이와 달리 인공지능에게 형법적 주체성을 부정한다면, 결국 다시 그 인공지능 뒤에 있는 "사람"에게 책임을 귀속시키는 논리를 개발해야 할 것이다.

인한 법익침해와 그 형사책임", 형사법연구 제17호, 2002; 김호기, 앞의 글 등 참조.

56) Joerden, 앞의 글, S. 206.

나. 인공지능 배후의 사람

인공지능의 작동에 의해 범죄적 결과가 발생했다 할지라도 인공지능의 형사책임을 부정할 수밖에 없다면 결국 그 배후에 있는 사람, 즉 인공지능의 사용자나 관리자, 생산자, 또는 설계자에게 책임을 물을 수밖에 없을 것이다.[57] 이러한 경우 누구에게 책임을 귀속시킬 것이며, 어떠한 조건 하에서 인공지능의 범죄적 행동에 대한 책임을 귀속시킬 수 있을 것인지, 귀속시켜야 하는지를 면밀히 따져봐야 할 것이다.

우선 생산자나 관리자가 생산과정에서 또는 관리(A/S)과정에서 인공지능에 의도적으로 악성 프로그램을 설치하여 형법적으로 중대한 결과를 초래한 경우, 또는 사용자가 의도적으로 특정한 명령을 내려 인공지능으로 하여금 범죄적 결과를 낳게 한 경우, — 그 알고리즘 입증의 어려움은 일단 제쳐두고 — 생산자나 관리자, 또는 사용자에게 형사책임을 귀속시킬 수 있음에는 별다른 논란의 여지가 없다. 이처럼 인공지능 배후의 사람이 고의를 갖고 있는 경우에는 오늘날의 간접정범론과 유사한 논리를 구성하거나,[58] 또는 보다 단순하게 인공지능을 행위수단에 불과한 것으로 보아도 될 것이다. 그에 반해 인공지능 배후의 사람에게 불법결과에 대한 고의가 없는 경우, 예컨대 알고리즘상 예측하지 못했던 오류나 하자로 인해 불법적인 결과가 발생한 경우에는 과실에 의한 형사책임 귀속문제가 발생하게 된다.[59] 자율적으로 판단하고 행동하는 인공지능의 특성상 인공지능의 "독자적 의지에 따른 고유한 판단"이 생산자나 운영자, 또는 사용자의 행위와 형법적으로 중대한 결과발생 사이의 인과관계를 차단한다고 보는 경우라면 책임귀속이 부정될 가능성도 있다. 다만 아직까지는 인공지능이 독자적 의지형성능력을 가진다고 볼 수는 없기 때문에, 관련자 각각의 예측가능성과 주의의무에 따라 어떠한 "사람"에게 책임이 귀속될 것인가가 판가름 날 것이다. 그리고 이는 사용자나 관리자, 생산자, 또는 설계자가 각각 불법결과를 예측하고 예방했어야 하는 지위와 권한, 능력을 갖고 있었는지를 입증하는 문제로 귀결될 것이다.

---

57) 같은 취지로 민사법적 책임귀속에 관해서는 Schuhr, *Neudefinition tradierter Begriffe*, 2012, S. 22. 이 견해는 이른바 인공지능의 "행위"를 인공지능 뒤에 있는 사람의 "부작위"로 보고 있으며(S. 23), 인공지능에 대한 "책임비난"은 사람에 대한 "과실책임비난"으로 귀결된다고 한다(S. 25).

58) Gleß/Weigend, 앞의 글, 579-580.

59) 윤지영 외, 앞의 보고서, 162쪽.

Ⅴ. 정당화상황의 사전적 프로그래밍

인공지능에게 형법상 행위능력이나 책임능력, 형벌능력을 인정하지 않고, 따라서 설계자나 제조자, 공급자 등에게 행위결과와 책임을 귀속시킨다 하더라도 인공지능의 활용은 정당화 영역에서, 즉 위법성조각사유와 관련하여 새로운 문제를 발생시킬 것으로 보인다.[60] 지금의 형법도그마틱에서 위법성조각사유는 어떠한 행위가 구성요건에 해당하지만 상황의 특수성으로 인해 예외적·사후적으로 위법하지 않은 것으로 평가하여 허용해주는 작용을 한다.[61] 그에 비해 인공지능 활용으로 인해 마주하게 될 상황은 일반적·사전적 예측과 그러한 예측을 토대로 한 프로그래밍을 요구하게 될 것이다.

경비기능을 수행하던 인공지능이 침입자를 발견한 경우를 대비한 알고리즘을 구축한다고 가정해보자. 단순침입자인 경우와 무장침입자인 경우를 구분하여 대응하도록 설계하는 방법을 생각해 볼 수 있겠으나, 단순침입자와 무장침입자를 어떠한 기준으로 구분할 것인지 문제될 것이며, 실제로 발생할 수 있는 상황은 이보다 훨씬 다양할 것이다. 무인자동차가 자율주행 중 지나가던 어른 5명을 치거나 어린아이 1명을 치게 될 수밖에 없는 양자택일 상황에 놓인 경우도 정교한 알고리즘 구축이 문제되는 사례라 할 수 있다. 앞의 사례는 불법행위에 대한 대응이기 때문에, 오상방위상황이 아니고 그 대응도 과도하지 않은 이상 가치판단의 문제는 그리 크지 않을 수 있다. 그러나 후자는 정당한 이해의 충돌이 예견된 상황이기 때문에 보다 첨예한 가치판단의 문제가 발생한다. 인공지능의 성능에 따라 이와 같은 정당화상황을 어느 정도 예측하여 학습알고리즘과 가치판단 기준, 각 상황에 맞는 세세한 규준 등을 사전에 프로그래밍해야 할 상황이 발생하게 될 것이다.

인공지능의 수준이 높아질수록 그 독자적 판단영역의 범위는 넓어질 것이며, 위법성조각이 문제되는 상황에 직면하게 될 수 있음은 어렵지 않게 상상해 볼 수 있

60) 어쨌거나 위법성조각이 성립하는 상황이라면 굳이 결과나 책임의 귀속을 따질 필요는 없겠으나, 위법성이 완전히 제거되지 않는 오상방위나 과잉방위 등의 사례에서는 인공지능의 판단과 행동에 대한 최종책임을 누가 져야하는가 하는 문제가 발생하는데 이는 앞에서 서술한 바와 마찬가지로 각 관련자의 지위와 주의의무정도를 고려해 판단해야 할 것이다.

61) 배종대, 앞의 책, 57/5.

다. 그리고 인공지능은 그 수준여하를 떠나 근본적으로 사람이 만든 인위적 창조물이기 때문에, 이와 같은 상황에서 취해야 할 행동은 정도의 차이는 있겠으나 사전적 프로그램을 필요로 한다. 요컨대 인공지능이 활용되는 시대에는 오늘날 형법도그마틱의 사후적 정당화 평가뿐만 아니라 사전적 정당화 프로그래밍도 필요하게 될 것이다. 그러나 아무리 프로그램을 정밀하게 구성한다 하더라도 구체적 상황에서 인공지능이 어떻게 작동할지 정확하게 예측하기란 매우 어려울 것이다.[62]

Ⅵ. 맺는 글

인공지능의 출현 내지 활용 자체에 대해서도 여전히 찬반 견해가 대립하고 있으며, 인공지능에 대한 법적 논의는 아직 걸음마 단계에 불과하다. 그리고 본 논문도 몇 차례 걸음마 가운데 한 걸음 정도의 의미 정도밖에는 없을지도 모르겠다.

물론 인공지능 활용과 관련한 이와 같은 쟁점들은 오늘날의 법이론과 법해석으로도 충분히 해결할 수 있을 것이라는 견해도 있다. 새로운 법이론과 법정책이 필요하다는 입장도 아직까지는 구체화된 내용을 제시하지는 못하고 있다. 현실 내지 존재의 변화가 곧바로 규범의 변화로 이어질 수 없는 규범학의 내재적 한계가 있기 때문일 것이다.[63] 그럼에도 인공지능의 급속한 발전과 눈앞으로 다가온 상용화 가능성으로 인해 규범학도 이제는 어느 정도는 미래를 대비하지 않을 수 없게 되었음은 부인할 수 없을 것이다.

이러한 차원에서 다소 거칠게나마 형법적 쟁점을 찾아보면, 우선 인공지능 자체의 행위능력, 형사책임, 형벌능력 등으로 요컨대 인공지능에게 형법적 주체성을 인정할 수 있는가 하는 점 등이 발견된다. 그런데 오늘날의 확고한 인본주의적 형법관, 형법규정, 기술수준 등으로 인해 인공지능의 형법적 주체성을 인정하기에는 다소 무리수가 있어 보인다. 인공지능의 형법적 주체성을 인정하지 못한다 하더라도

62) Neumann, “기술의 발전에 따른 법학적·법철학적 전망”, 법과학을 적용한 형사사법의 선진화방안 연구(VI) - 지능형 로봇기술과 형사법적 논의의 지평 모색 – 워크숍 자료집, 2015.10.6., 29쪽.
63) 변종필, “위험사회, 위험 그리고 형법의 대응”, 비교형사법연구 제14권 제2호, 2012, 535쪽.

인공지능으로 인해 발생할 수 있는 불법결과는 누구의 책임인가 하는 문제는 여전히 남는다. 이는 결국 인공지능의 생산에서부터 사용단계 사이에 관련된 자들 가운데 누구에게 결과발생에 대한 예측가능성 및 회피의무가 있는지를 찾아내는 문제로 귀결될 것이며, 형사상 제조물책임에 관한 논의와도 접점을 형성한다. 그 밖에 사람과 인공지능 사이에서도 책임귀속 문제가 발생하게 될 것이다. 인공지능의 합리적 판단과 명령에 따라 사람은 단순히 기계적으로 행동하는 경우를 배제할 수 없기 때문이다. 이러한 상황은 다시 인공지능에게 형법적 주체성을 인정할 수 없기 때문에, 그 인공지능의 배후에 있는 어떤 사람에게 형사책임을 물어야 하는지의 문제로 귀결된다. 그리고 누구에게 형사책임을 귀속시켜야 하는가 하는 실체법상의 문제는 다시 결과발생의 근본적·결정적 원인이 무엇인가라는 절차법상 인과관계 입증의 문제로 옮겨질 것이다.[64] 끝으로 인공지능이 이른바 정당화상황 내지 이익충돌 상황에 직면했을 때 어떻게 판단하고 행동해야 할지를 사전에 프로그래밍해야 하는 상황이 발생할 수도 있다.

인공지능에 호의적인 입장이든 부정적인 입장이든, 인공지능 자체 및 인공지능으로부터 야기될 수 있는 모든 위험을 완벽하게 통제할 수는 없을 것이라는 점에는 이견이 없어 보인다.[65] 그리고 인공지능이 점차 다양한 기능을 탑재하고 똑똑해질수록 인공지능은 사람의 통제에서 벗어나기 쉬우며 따라서 종래의 이론이나 해석으로는 해결할 수 없는 문제상황이 발생할 가능성은 점차 높아질 것이다.[66] 궁극적으로는 “인간이란 무엇인가?”라는 물음과 함께 인간과 인간 아닌 것 사이의 이분법을 계속해서 고수해야 하는지에 대한 (법)철학적 재론까지도 필요하게 될 것이다.

64) 윤지영 외, 앞의 보고서, 166쪽.
65) 윤지영 외, 앞의 보고서, 165쪽.
66) Gleß/Weigend, 앞의 글, 590.

참고문헌

김일수, 한국형법 I [총론 상] 개정판, 박영사, 1996.

배종대, 형법총론 제12판, 홍문사, 2016.

윤지영·윤정숙·임석순·김대식·김영환·오영근, 법과학을 적용한 형사사법의 선진화 방안(VI), 연구총서 15-B-16, 한국형사정책연구원, 2015.

이재상, 형법총론 제7판, 박영사, 2011.

고인석, "로봇윤리의 기본 원칙: 로봇 존재론으로부터", 범한철학 제75집, 범한철학회, 2014.

관계부처 합동, 제2차 지능형 로봇 기본계획(안), 2014.

김영환 "위험사회에서의 책임구조: 자연 재해에 대한 법적 담론", 홍익법학 제14권 제1호, 홍익대 법학연구소, 2013.

김호기, "개발위험의 항변과 형법적 제조물책임 - 가습기 살균제 등 대량생산되어 사용되는 일상생활용품의 사례를 중심으로", 형사정책연구 제27권 제1호, 한국형사정책연구원, 2016.

박현섭, "지능형 로봇 산업발전을 위한 정부정책", TTA 저널 제158호, 한국정보통신기술협회, 2015.

변종필, "위험사회, 위험 그리고 형법의 대응", 비교형사법연구 제14권 제2호, 한국비교형사법학회, 2012.

신현우, "진화론적 지능형 서비스 로봇에 대한 실천윤리학적 고찰", 윤리연구 제79호, 한국윤리학회, 2010.

윤상석·박성기·최종석, "자폐 범주성 장애 아동의 사회적 기능향상을 용이하게 하는 상호작용로봇", 한국HCI학회 학술대회 2014, 한국HCI학회, 2014.

전지연, "형벌론적 관점에서 본 형법적 제조물책임의 필요성", 형사정책 제13권 제1호, 한국형사정책학회, 2001.

정보통신기술진흥센터, "미국의 인공지능(AI) 기술 R&D 추진 동향", 해외 ICT R&D 정책동향 2015년 3호, 2015.

하태훈, “缺陷製造物로 인한 법익침해와 그 형사책임”, 형사법연구 제17호, 한국형사법학회, 2002.

Neumann, Ulfrid, “기술의 발전에 따른 법학적·법철학적 전망”, 법과학을 적용한 형사사법의 선진화방안 연구(VI) - 지능형 로봇기술과 형사법적 논의의 지평 모색 – 워크숍 자료집, 2015.10.6.

Beck, Susanne, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 6/2009.

Fitzi, Gregor, Roboter als “legale Personen” mit begrenzter Haftung. Eine soziologische Sicht, in: Eric Hilgendorf/Jan-Philipp Günther(Hrsg.), Robotik und Gesetzgebung. Beiträge der Tagung vom 7. bis 9. Mai 2012 in Bielefeld, 2013.

Gleß, Sabine/Weigend, Thomas, Intelligente Agenten und das Strafrecht, ZStW 2014.

Gruber, Malte-Christian, Rechtssubjekte und Teilrechtssubjekte des elektronischen Geschäftsverkehrs, in: Susanne Beck(Hrsg.), Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012.

Hilgendorf, Eric, Können Roboter schuldhaft handeln?, in: Susanne Beck(Hrsg.), Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012.

_____, Recht, Maschinen und die Idee des Posthumanen, Telepolis 2014(http://www.heise.de/tp/druck/mb/artikel/41/41777/1.html, 2015.04.02. 검색).

Joerden, Jan C., Strafrechtliche Perspektiven der Robotik, in: Eric Hilgendorf/Jan-Philipp Günther(Hrsg.), Robotik und Gesetzgebung. Beiträge der Tagung vom 7. bis 9. Mai 2012 in Bielefeld, 2013.

Kuhlen, Lothar, Strafrechtliche Produkthaftung, in: Hans Achenbach / Andreas Ransiek(Hrsg.), Handbuch Wirtschaftstrafrecht, 3. Auflage, München, 2012.

Matthias, Andreas, Automaten als Träger von Rechten, Logos Verlag, Berlin, 2008.

Münchener Kommentar zum Strafgesetzbuch, Wolfgang Joecks/Klaus Miebach(Hrsg.), Band 1 §§ 1-37 StGB, 2. Auflage, Müchen, 2011. (인용방법 : 저작자, MK-StGB)

Meyer, Stephan, Der Einsatz von Robotern zur Gefahrenabwehr, in: Eric Hilgendorf(Hrsg.), Robotik im Kontext von Recht und Moral, 2014.

Schuhr, Jan C., Neudefinition tradierter Begriffe (Pseudo-Zurechnung an Roboter), in: Eric Hilgendorf(Hrsg.), Robotik im Kontext von Recht und Moral, 2014.

Ziemann, Sascha, Wesen, Wesen, seid's gewesen? Zur Diskussion über ein Strafrecht für Maschinen, in: Eric Hilgendorf / Jan-Philipp Günther(Hrsg.), Robotik und Gesetzgebung. Beiträge der Tagung vom 7. bis 9. Mai 2012 in Bielefeld, 2013.

Strafrechtliche Verantwortungzurechnung der künstlichen Intelligenz

Im, Seok-Soon*

Aufgrund der dringenden Entwicklung der Künstlichen Intelligenz(KI) und deren Generalisierungsmöglichkeit muss nun die Normwissenschaft einschließlich Rechtswissenschaft die Zukunft vorbereiten. In diesem Aspekt handelt es sich strafrechtswissenschaftlich um die Subjektivität der KI, die Zurechnung des Unrechtserfolgs durch die Verwendung der KI und die Vorprogrammierung der Rechtfertigungsgründe usw. Mindestens jetzt könnte die Subjektivität der KI noch nicht bejaht werden. Die Zurechnung des strafrechtlichen Verantwotung soll endlich die Beweisungsfrage der Kausalität werden.

Unabhängig von den Standpukten über die KI könnte man auf jeden Fall die alle Risiko von/aus der KI nicht vollkommen kontrollieren. Je klügger die KI wird, desto höher wird die Entstehbarkeit der nicht vorgestellten Situation, die mit den bisherigen Theorien und Hermeneutik nicht gelöst wird. Grundlegend notwendig wird die (rechts)philosophische Rediskussion über den Menschen per se.

❖ Keyword: Künstliche Intelligenz, Strafrechtliche Subjektivität der künstlichen Intelligenz, Handlung der künstlichen Intelligenz, Strafrechtliche Verantwortung der künstlichen Intelligenz, Vorprogrammierung der Rechtfertigungsgründe

투고일 : 11월 25일 / 심사일 : 12월 9일/ 게재확정일 : 12월 9일

* Research Fellow in Korean Institute of Criminology, Dr. jur.