

<http://dx.doi.org/10.17703/JCCT.2024.10.2.447>

JCCT 2024-3-51

인공지능 통제 가능성 고찰과 글로벌 규제 현황 연구

Study on Controllability of Artificial Intelligence and Status of Global Regulations

장미경*

MiKyung Chang*

요약 생성형 인공지능 기술의 놀라운 성과가 점차 가시화됨에 따라, 기계의 인간 지배 가능성 등 잠재적인 실존 위협이 제기 되는 현시점에서 인공지능에 대한 ‘통제 가능성’이 첨예한 글로벌 키워드로 주목받고 있다. 이에 따라 이 연구는 인공지능 기술을 중심으로 펼쳐질 미래 사회의 혁신적 변화에 대응하기 위하여 인공지능에 대한 통제 개념과 현주소, 글로벌 현황을 면밀하게 탐색함으로써 사회적 공론장 형성의 토대를 마련하고자 하는 데 목적이 있다. 이를 통해 인공지능 기술 진화에 따라 야기될 사회문제와 예측 불가능한 변수에 대해 대응책을 마련하기 위한 시사점을 모색하고, 정부 규제 수립에 대한 가이드라인과 전략적 통찰력을 제시하는 한편, 사회적 공개 담론 형성을 위한 함의를 찾아 보고자 한다.

주요어 : 인공지능, 통제 가능성, 통제, 규제, 법안, 인공지능 법안, 국회

Abstract As the remarkable achievements of generative artificial intelligence technology become increasingly visible, the issue of ‘controllability’ in artificial intelligence is emerging as a prominent global keyword. This comes at a time when existential threats, such as the possibility of machines dominating humans, are being raised. Accordingly, this study aims to establish the groundwork for shaping a social public sphere by closely examining the concept of control, the current status, and the global landscape of artificial intelligence. It seeks to address the innovative changes anticipated in future society, with artificial intelligence technology at its core. The study aims to derive implications for preparing countermeasures against social problems and unpredictable variables that may arise from the evolution of artificial intelligence technology. It also aims to present guidelines and strategic insights for the establishment of government regulations. Furthermore, the study seeks to uncover implications for the formation of social public discourse.

Key words : Artificial intelligence, controllability, control, regulation, legislation, AI act, the National Assembly

1. 서론

최근 첨예한 글로벌 이슈로 떠오른 인공지능 기술 관련 연구는 광범위한 분야에서 활발하게 진행되고 있다.

기술 자체에 주안점을 두고 있는 연구는 물론, 기술의 진화에 따라 급변할 사회 제반 환경에 대한 사회과학적 관점의 연구, 인간에게 가해지는 실존적 위협에 따른 철학적 시각의 연구 또한 주목받고 있다. 생성형 인공

*고려대학교 과학기술협동과정 과학언론학 박사 (저자)
접수일: 2024년 1월 2일, 수정완료일: 2024년 1월 21일
게재확정일: 2024년 2월 1일

Received: January 2, 2024 / Revised: January 21, 2024
Accepted: February 1, 2024

*Corresponding Author: rosem@korea.ac.kr
Science and Technology Studies, Korea Univ., Korea

지능 챗봇 챗GPT가 보여준 첨단 과학기술의 놀라운 위력은 인공지능에 대한 사람들의 관심과 흥미를 자극하는 촉매 역할을 했고, 향후 펼쳐질 성과 역시 상상을 초월하는 속도로 가시화됨에 따라 인공지능에 대한 통제 가능성을 주요 화두로 삼은 세계적 석학들의 목소리도 더욱 커지고 있다. 구체적으로 살펴보면, 저서 ‘사피엔스’로 일반인에게 친숙한 유발 하라리(Yuval Harari)는 2023년 3월부터 세계 유수의 언론매체 기고와 포럼 참여 등을 통해 인공지능 통제의 중요성을 경고했다. 그는 인간이 인공지능을 통제할 수 없는 시점이 오기 전에 인공지능 능력에 상응하는 책임과 통제의 조화를 이루어야 한다고 촉구했다. 인공지능과 딥러닝의 대부로 알려진 제프리 힌턴(Geoffrey Hinton) 역시 인공지능의 긍정적 측면보다 부정적 측면이 훨씬 더 크다고 우려했다. 그는 인공지능의 경우 핵무기 개발과 달리 국가가 비밀리에 기술을 개발하는지 파악할 수 없다면서, 세계 과학자들이 인공지능에 대한 통제 방법을 찾기 위해 협력해야 한다고 강조했다. 저서 ‘슈퍼인텔리전스’로 화제를 모았던 영국 옥스퍼드대 교수 닉 보스트롬(Nick Bostrom)도 인공지능이 인류에게 존재론적 위협까지 불러올 수 있다고 언급하면서, 인공지능이 출현하기 전에 인공지능 통제 방안에 대해 고민하고 연구를 서둘러 추진해야 한다고 주장해 왔다[1].

전 세계 석학들이 인공지능의 ‘통제 가능성’을 매우 중요한 키워드로 빈번하게 언급하고 있다는 점을 고려할 때, 인공지능 통제에 대한 현주소와 글로벌 규제 현황을 탐색해 보는 연구는, 첨단 과학기술 관련 국가 경쟁력을 높이고 인공지능이 주도할 미래 사회 변화에 적극적으로 대응할 시사점을 찾을 수 있다는 데 의미가 있을 것으로 판단된다. 이 연구를 통해 인공지능 기술 발전과 진화에 따라 야기될 사회문제와 예측 불가능한 변수에 대해 대응책을 마련하기 위한 시사점을 모색하고, 정부의 인공지능 규제 마련에 대한 가이드라인을 제시하는 한편, 시민의 관심과 사회적 공개 담론 형성을 위한 함의를 찾아보고자 한다.

II. 인공지능 통제 가능성 논의

일반적으로 논의되는 인공지능 통제 문제의 핵심은 ‘첨단기술에 대한 제어력을 상실할 수 있다는 것’이다[2]. 닉 보스트롬(Nick Bostrom)은 인류가 포스트휴먼 단계에 도달하기 전에 멸종될 가능성이 높고, 현재 인류는 거의

확실하게 컴퓨터 시뮬레이션 속에 살고 있다고 주장하면서, 인간이 인공지능에 대한 통제력을 잃는 상황에 대해 경고해 왔다[3]. 인공지능 기술 진화에 대한 논의가 진행될 때마다 학자들은 통제 가능성 문제를 끊임없이 제기해 왔다. 예를 들어, 인공지능 기술을 핵심 두뇌로 안착한 ‘자율주행차’의 상용화와 관련해 사람들이 가장 우려하는 문제는 “인간이 자율주행차를 충분히 통제할 수 있을지” 여부다. 특히 이러한 상황에서 통제에 대한 우려는 잠재적 책임 소재와 관련된 경우가 많다. 인공지능과 같이 첨단기술 시스템에 대한 통제가 불가능한 상태에서 사고가 발생해 누군가 다친다거나 사망하게 된다면, 누가 책임을 져야 하는지 불분명할 수 있기 때문이다[4]. 이 같은 논의는 킬러 로봇, 군사용 드론 등 자율 무기 시스템에 대한 통제 문제 제기에서도 같은 맥락으로 적용될 수 있다. 또한 인공지능 기술 개발 단계에서는 자율주행차 사고나 알고리즘의 편향된 의사결정 등 인공지능 통제 가능성과 관련된 예기치 못한 사회적 문제가 유발될 수 있다.

1. 인공지능 통제의 개념과 흐름

인공지능과 관련된 통제에 대한 우려는, 인공지능이 무엇인지 또는 인공지능을 무엇으로 간주해야 하는지에 대한 일반적 고찰에서 비롯된다고 할 수 있다[5]. 흔히 ‘인공지능’ 개념의 근간을 만들었다고 손꼽히는 앨런 튜링(Alan Turing)은 1950년 ‘인공지능’이라는 용어를 직접 정의해 사용한 것이 아니라, “기계가 생각할 수 있는지”에 대한 연구인 ‘튜링 테스트’(Turing test)에 집중했다[6]. 튜링 테스트에 따르면, 만약 기계가 제시하는 메시지와 사람이 작성한 메시지를 구별할 수 없을 정도로 인간을 모방할 수 있는 기계가 발명된다면, 지능적이고 생각하는 기계, 즉 ‘인공지능’이 현실로 구현되었다고 할 수 있다. 이러한 연구에 영감을 둔 인공지능에 대한 정의는 “생각하는 인간을 성공적으로 모방할 수 있는 기계를 만들 때” 완성된다고 할 수 있다[7].

이후 ‘인공지능’이라는 용어는 1956년 개최된 미국 ‘다트머스 컨퍼런스’(Dartmouth Conference) 학술회의에서 연구진의 제안에 따라 처음 사용되었다[8]. 인공지능 연구의 효시가 된 이 컨퍼런스에서 인공지능의 아버지로 불리는 존 매커시(John McCarthy)를 중심으로 “인간의 사고와 학습, 지능적 측면을 시뮬레이션하는 기계”에 대한 학자들의 논의가 시작된 것이다. 이 정의에

따르면 인공지능은 “인간의 지능을 시뮬레이션하는 기계”라고 할 수 있다. 시간이 흐르면서 현대적 접근 방식에 따라 인공지능은 지능형 에이전트 생성 개념으로 논의되기 시작했다. 여기서 ‘에이전트’는 환경을 ‘인식’하고 특정 목표를 달성하기 위해 ‘행동’할 수 있는 시스템을 의미한다. ‘지능’은 철학계에서 ‘도구적 합리성’이라고 표현하는 것, 즉 ‘목표를 효율적으로 달성하는 능력’으로 정의할 수 있다[9]. 결국 목표를 잘 달성할 수 있고, 다른 지능형 에이전트와 협력할 수 있으며, 어려운 작업 분야에서 인간보다 뛰어난 성능을 발휘할 에이전트의 출현에 대한 아이디어 구상은, 인공지능과 관련한 통제 상실 문제가 왜 발생할 수 있는지, 인간이 인공지능에 대한 통제 불가능 상황을 왜 두려워해야 하는지 짐작하게 한다. 예를 들어 닉 보스트롬은 미래의 어느 시점에 인간이 목표를 달성하는 데 있어 매우 효율적인 ‘초지능’ 시스템에 도달할 수 있다고 우려하면서 대중과 사회의 관심을 이끌어왔다. 여기서 우리가 주목해야 할 문제는, 이러한 상황이 인류에 대한 ‘실존적 위협’으로 표현되는 매우 중대한 위협으로 이어질 수 있다는 점이다[10]. 이에 따라 생성형 인공지능 챗GPT의 출현 이후, 인공지능 통제 관련 문제에 대한 우려가 더욱 커지게 되었고, 세계 석학들은 인공지능 기술의 놀라운 발전에 대한 찬사를 뛰어넘어 기술 진화가 낳게 될 예측할 수 없는 부작용과 사회문제에 대해 강한 두려움과 공포를 표현하기 시작했다.

2. 통제 가능성에 대한 우려와 쟁점

앞선 논의에서 살펴본 바와 같이 인공지능 기술은 향후 어떤 양상으로 사회 변화를 주도하게 될지 가늠할 수 없는 ‘불확실성’과 ‘예측 불가능성’을 주요 특징으로 하는 기술이다. 무엇보다 현시점에 인공지능 기술 관련 통제와 위협 논의, 인공지능 규제에 현주소를 살펴보는 것이 중요한 이유는, 인공지능 기술이 기존에 경험할 수 없었던 변화를 주도할 ‘파괴적 기술’(Disruptive Technology)의 대표 사례라고 할 수 있기 때문이다. ‘파괴적 기술’이라는 용어는 ‘파괴적 혁신’(Disruptive Innovation)이라는 개념에서 그 유래를 찾을 수 있다. 미국 하버드 경영대학원 클레이튼 크리스텐슨(Clayton Christensen) 교수는 저서 ‘혁신기업의 딜레마’(The Innovator’s Dilemma)를 통해 기업이 기술 변화를 인지하지 못하거나 간과했을 경우 발생할 수 있는 격차와 실패를 강조하면서 ‘파괴적 혁신’이라는 개념을

사용했다. 그동안 ‘파괴적 기술’ 관련 논의는 컴퓨터, 인터넷, 차세대 유전체학, 신소재, 빅데이터 등 혁신적 기술에 대해 논할 때마다 진행되었는데, 최근 인공지능 이슈를 중심으로 다시 그 중요성이 강조되고 있다.

2023년 11월 2일, 미국 테슬라와 스페이스X의 CEO 일론 머스크(Elon Musk)는 영국 런던에서 개최된 인공지능 안전성 정상회의(AI Safety Summit) 이후 영국 총리 리시 수낙(Rishi Sunak)과 대화를 갖고, 인공지능의 미래와 관련하여 정부의 역할, 일자리 변화, 규제 등 인공지능의 위험과 기회에 대해 논의했다. 일론 머스크는 인공지능에 대해 ‘역사상 가장 파괴적인 힘’이라고 언급하면서, 인공지능이 인간의 일자리를 필요 없게 만들 것으로 예측했다. 머스크에 따르면 인공지능 기술이 지닌 ‘파괴적인 힘’은 인공지능 기술이 기존 산업의 양상을 근본적으로 변화시키거나 대체할 수 있는 혁신적 잠재력을 가지고 있음을 의미한다. 그는 이날 인공지능의 파괴적인 힘을 언급하는 동시에 통제와 규제의 필요성을 강조했다. 아울러 인공지능이 결과적으로는 대부분 ‘선’을 위해 쓰이게 될 것이라는 낙관적인 미래를 전망했지만, 위협이 전혀 없는 것은 아니라고 주의를 당부했다.

한편 2023년 11월 17일, 인공지능 통제 가능성과 관련해 기술 패권 경쟁의 선두에 있는 주도국은 물론, 전 세계 사람들의 촉각을 곤두세우게 하는 대형 사건이 발생했다. 챗GPT 개발사인 오픈AI 이사회가 창업자인 샘 올트먼(Sam Altman)을 전격 해임했다고 발표한 것이다. 당시 명확한 해임 사유를 밝히지 않았지만, ‘인공지능 통제 이슈’가 원인으로 지목된 바 있다. 오픈AI는 2015년 비영리기관으로 설립되면서 ‘인류에게 안전하고 유익한 인간 수준의 인공지능 기술 개발과 구축’을 목표로 삼았는데, 올트먼이 안전한 인공지능 개발을 위한 통제와 규제에 초점을 맞추기보다, 상업성과 수익성 추구에 더 주안점을 두면서 글로벌 빅테크기업의 투자를 끌어내기 위해 활동해 왔다는 것이다. 좀 더 구체적으로 살펴보면, 오픈AI는 2015년 출범 당시 생성형 인공지능 챗GPT 개발에 있어 외부 영향력을 배제하기 위하여 자본과 경영을 분리해 자본 없는 6인 이사회 체제를 구성했는데, 이후 기술 개발과 발전을 추진하기 위한 자금 조달 목적의 영리 자회사를 출범시키면서 각기 다른 가치의 경쟁 구도가 시작되었다. 전문가들은 오픈AI 조직 내에서 ‘통제할 수 있는 안전한 인공지능 개발’과 ‘인공지능 개발을 통한 수익 모델의 실현’이라는 상반된 가치가 충돌했고, 인공지능 개발 간 노선 갈등이

이번 사태의 원인이 되었다고 관측한다. 즉 ‘인공지능 개발론자’와 ‘인공지능 규제론자’의 갈등이 ‘올트먼 해임’이라는 충격적 사건을 불러일으켰다는 것이다.

올트먼 해임이 발표된 직후, 마이크로소프트는 올트먼을 자사의 연구팀에 합류시킬 예정이라고 밝혔고, 오픈AI의 직원 770명 중 702명이 올트먼을 따라가겠다고 선언함에 따라, 5일 만에 올트먼이 다시 복귀하면서 해임 사건은 마무리되었다. 더불어 올트먼 해임을 주도한 이사진 3명은 사임했고, 이에 따라 전문가들은 오픈AI가 인공지능 기술 개발과 상업화를 위해 더 절주하게 될 것으로 전망했다. 결국 올트먼 해임 사건은 인공지능 기술 개발 및 규제와 관련된 상반된 관점에 따라 ‘인공지능 통제’를 둘러싼 기술계의 깊어지는 갈등이 얼마나 심각한지를 보여주는 한 편의 드라마로 남게 되었다. 이러한 기술계의 사건은 인공지능 기술이 인류에게 위협이 되지 않도록 통제하고 규제하는 책임성이 정부와 사회 구성원 몫으로 넘어오고 있음을 시사한다. 인공지능을 윤리적으로 사용할 수 있도록 이끌어야 할 정부 역할은 물론, 전혀 예측할 수 없는 사회 패러다임 변화를 불러올 새로운 기술 이슈에 대해 활발하게 논의하고 토론하는 사회적 공론장의 형성이 더욱 중요해졌다는 의미다. 그렇다면 글로벌 쟁점으로 부상하는 인공지능 법안 제정의 현주소는 어떤 모습일지 좀 더 구체적으로 살펴볼 필요가 있다.

III. 인공지능 통제 및 규제 동향

1. 국외 규제 동향

인공지능 기술 선진국은 기술 개발의 효율성을 끌어 올려 인공지능 관련 산업의 발전과 진흥을 도모하면서도, 인공지능 기술로 인해 야기될 사회문제 등 우려와 혼란을 잠재우기 위한 제도적 기반과 통제 방안 마련에 고심하고 있다. 가장 앞서나가는 사례로는 유럽연합의 ‘인공지능 법안’(Artificial Intelligence Act)을 손꼽을 수 있다[11]. 유럽연합은 2021년 4월 인공지능 법안을 세계 최초로 제안했고, 2023년 6월에는 생성형 인공지능 관련 내용을 반영한 수정법안을 유럽의회에서 통과시켰다. 이후 2023년 12월 8일, 27개 EU 회원국이 모여 법안에 잠정 합의하는 등 가장 빠른 속도로 인공지능 규제 방안 마련에 앞장서고 있다. 유럽연합은 법안 운영을 통해 인공지능 기술을 활용하는 과정에서 시민의 기본권을 지키고 민주주의를 보호하며, 개인의 건강과 안전을 보장하는 것을 목표로 삼고 있다.

특히 관련 법규를 위반할 경우, 기업의 연간 글로벌 매출 7%에 해당하는 벌금을 부과하는 등 인공지능이 충분히 신뢰할 수 있고 인간 중심으로 활용될 수 있도록 지원하면서 법안 실효성을 강화할 예정이다. 유럽연합은 2018년 인공지능 윤리에 관한 연구 전문가그룹을 구성해 2019년 인공지능이 가져올 사회문제 등을 대비하기 위한 규범 등 연성법 형태의 ‘인공지능 윤리 가이드라인’을 발표했고, 이후 2021년 강제력을 지닌 법적 규제 프레임으로 인공지능 법안을 제안했다는 점에서 주목할 만한 법안 내용 구성과 절차를 적용하고 있다는 평가를 받는다. 즉 사회적으로 파급력이 클 수 있는 인공지능 통제 문제에 대해 지침이나 가이드라인을 제시하는 자율적 규제에서 출발해, 강제력 있는 법적 규제에 다다를 때까지 인공지능 발전의 단계에 맞추어 서서히 사회에 안착시키고 노력했음을 유추할 수 있다. 이에 따라 유럽연합은 법안에서 강제하는 엄격한 규제는 물론, 윤리 문제 해결 지침 등의 대응까지 탄력적으로 적용할 수 있다. 이러한 측면에서 유럽연합은 인공지능 관련 산업의 성장동력을 저해하지 않음과 동시에, 각종 부작용과 사회문제에 선제적으로 대응할 수 있는 전략적 방안을 만들어 내고 있다고 해석된다. 우리 정부는 물론, 인공지능 지침과 법안을 준비하는 국제사회가 주의 깊게 바라보아야 할 부분이다. 또 하나 주목할 만한 점은, 유럽연합의 인공지능 법안의 경우 위험 기반 접근에 따라 인공지능 규제 대상을 최소 위험(minimal risk), 제한된 위험(limited risk), 고위험(high risk), 수용 불가 위험(unacceptable risk) 등 4단계로 구분하고 있다는 사실이다. 인공지능은 하나의 산업으로 묶을 수 없는 범용기술이기 때문에, 기술 분야별·위험 수준별 분류에 따른 기술 정책이 적용되어야 한다는 점을 고려한 것으로 판단된다.

한편 미국의 경우에는 2022년 10월, 바이든 행정부의 과학기술정책실(OSTP, Office of Science and Technology Policy)이 인공지능 기술과 자동화시스템 설계 및 사용, 기술 배포 과정에서 시민의 권리를 보호하기 위한 지침인 ‘인공지능 권리장전 청사진’(Blueprint for an AI Bill of Rights)을 발표했다. 이는 미국 정부 차원에서 인공지능 관련 인권 보호 원칙을 처음으로 공식화했다는 측면에서 의의가 있으며, 총 5가지 기본원칙, 안전하고 효율적인 시스템의 구축, 알고리즘을 통한 차별 방지, 데이터 관련 사생활 보호, 자동화시스템의 활용에 대한 고지와 설명, 인간 대안 마련을 제시했다. 미국은 이 원칙을 실현하기

위해 자동화된 시스템을 설계할 때 고려해야 할 가이드 라인을 별도로 마련해 제공하고 있다.

영국은 2022년 7월, '디지털 규제 원칙'을 통해 디지털 전환에 따른 기술혁신을 촉진하고 규제를 완화하기 위한 원칙을 제시했다. 이 원칙은 영국이 유럽연합 탈퇴 후 디지털 규제 체계를 구축하기 위해 3가지 목표를 설정해 범정부 차원에서 마련한 것이다. 특히 영국은 디지털 전환에 따른 기술혁신 촉진과 안전성을 중점 원칙으로 삼은 '산업적 측면의 규제'라는 차원에서 차별성을 갖고 있다. 목표 사항 3가지는, 성장 주도를 위한 디지털 부문 전반의 경쟁과 혁신 촉진, 성장과 혁신이 시민과 기업에 해가 되지 않도록 온라인에서 안전하게 보호, 기본권과 자유를 보호하기 위해 반영하는 민주적 사회를 촉진하는 디지털 경제 형성이다. 목표 실현을 위한 규제 원칙으로는 적극적 혁신의 촉진, 미래지향적이고 일관된 성과 달성, 국제적 기회와 도전에 대한 대처를 제시하고 있다.

국제연합(UN)은 2020년, 디지털 협력에 관한 고위급 패널의 라운드테이블 활동에서 도출된 내용을 바탕으로 '디지털 협력을 위한 로드맵'을 제시했다. 이 로드맵에서 UN은 2030년까지 지속 가능한 발전 목표(SDG)를 달성하기 위해 디지털 기술의 기회를 활용하는 동시에 위험을 최소화하는 데 초점을 맞추고 있다. 이를 위해 UN은 총 다섯 가지, 포용적 디지털 경제와 사회 구축, 인적·제도적 역량 강화, 인권과 인간 자율성 보호, 디지털 신뢰, 안보, 안정성 촉진, 글로벌 디지털 협력 증진 등을 제시했다.

2. 우리나라 규제 동향

인공지능 기술에 대한 통제와 규제 방향은 인공지능 기술만이 갖고 있는 특성을 올바르게 이해하고 적용해서 설정해야 한다. 첫째, 인공지능 기술은 방향성이 명확하게 설정되지 않았고 산업 전반에 불러올 변화를 확인할 수 없다는 점, 둘째, 인공지능에 대해 적절한 통제가 이행되지 않는다면 인권과 민주주의 근간을 흔들 수 있는 사건이 발생 가능하다는 점, 셋째, 인공지능 관련 산업 분야의 패러다임 변화가 완성된 후에 법과 제도를 도입한다면 이미 돌이킬 수 없는 상황이 된다는 점 등을 고려해야 할 것이다. 이에 따라 인공지능 관련 사회문제에 대한 책임 부여 또는 규제 방안 적용 등을 사후에 반영하는 형태는, 인공지능 기술이 불러올 위험과 부작용 우려를 줄일 수도, 예방하기도 어렵다는 점을 명확하게 인지해야 한다.

2023년 2월, 국회 과학기술정보방송통신위원회 법안

소위에서 통과된 '인공지능산업 육성 및 신뢰 기반 조성에 관한 법률안'은 인공지능 산업 진흥과 인공지능 사회의 신뢰 기반 조성에 필요한 사항을 규정하고 이를 시행함으로써 국민의 권익과 존엄성을 보호하고 국민 삶의 질 향상과 국가 경쟁력을 강화하는 데 이바지함을 목적으로 한다. 법안의 주요 내용은 다음과 같이 구성되어 있다.

첫째, 인공지능 산업 육성과 신뢰 확보를 위한 추진 체계를 구체화하고 있다. 과학기술정보통신부 장관이 3년마다 인공지능 기술 및 산업 진흥과 국가 경쟁력 강화를 위해 기본계획을 수립·시행하고, 국무총리 소속 심의·의결 기관인 인공지능위원회와 그 산하에 전문위원회를 설치하며, 지능정보사회진흥원 산하 국가 인공지능센터를 설치한다는 내용이다. 둘째, 인공지능 기술 개발과 산업 육성을 위한 방안을 포함하고 있다. 인공지능 기술과 알고리즘 연구·개발, 관련 제품과 서비스 출시에 대한 '우선 허용·사후 규제' 원칙을 제시하고 있고, 인공지능 기술 개발 및 안전한 이용 지원 사업 실시, 인공지능 기술 표준화, 인공지능 학습용 데이터 시책 수립, 기업의 인공지능 기술 도입·활용 지원, 창업 활성화, 인공지능 융합 촉진, 법령 정비 등 제도 개선 노력 의무, 전문인력 확보 시책 추진, 국제 협력 및 해외시장 진출 지원, 인공지능 집적단지 지정, 대한인공지능협회 설립 등을 포함하고 있다. 셋째, 인공지능 윤리와 신뢰성을 확보하기 위한 구체적 방안이 제시되어 있다. 예를 들어 인공지능 개발·이용 과정에서 인공지능 사업자 또는 이용자가 지켜야 할 인공지능 윤리 원칙을 정부 차원에서 제정·공표하고, 신뢰할 수 있는 인공지능 기반 조성을 위한 시책 마련 및 인공지능 신뢰성 감·인증 지원 사업을 추진하며, 고위험 인공지능의 확인은 물론, 제품 또는 서비스 이용자를 대상으로 한 고위험 영역의 인공지능 고지 의무화 등을 규정하고 있다. 넷째, 인공지능 산업의 진흥을 위한 기타 세부 사항을 명시하고 있는데, 예를 들어 재원 확충 방안, 실태조사, 통계 및 지표 관리 공표, 비밀누설 등에 관한 벌칙 조항, 과태료 부과 조항 등을 포함하고 있다.

IV. 시사점과 대응 방안

2023년 12월, 인공지능 법안 시행에 최종적으로 잠정 합의한 유럽연합 등의 선진 법안 사례를 통해 인공지능 관련 규제가 위험성 통제를 위한 제도로서 적절한지 확인하기 위해 먼저 짚어볼 수 있는 부분은, '우선 허용·사후

규제 원칙'의 타당성 여부이다. 유럽연합의 법안 내용을 살펴보면, 사전 규제 원칙이 비교적 엄격하게 적용된다. 물론 사전 규제가 아닌 선제적 허용과 사후 규제 원칙은, 산업 발전을 위한 불가피한 조치일 수 있다. 인공지능 관련 산업의 특성상 기술 개발 속도에서 뒤처지거나 원천 기술력을 확보하지 못한다면 글로벌 기술 패권 경쟁에서 밀려날 수밖에 없기 때문이다. IT기술 강국인 우리나라가 인공지능 기술 분야에서 앞서나가기 위해서는 기술 발전의 발목을 잡을 수 있는 법적 규제를 최소화해 산업 기반을 탄탄하게 조성할 필요가 있다. 하지만 앞서 살펴본 바와 같이 인공지능은 복잡하고 전문적이며 불확실성을 안고 있어 잠재적 위험을 항상 내포하고 있으므로, 현저한 위험 가능성을 사전에 파악하여 규제를 적용하기란 사실상 불가능하다. 이에 따라 과학기술, 교육, 문화 등 사회의 다양한 영역에서 인공지능 관련 위험성 이슈에 대한 사전 규제 방식을 적절하고 타당하게 설계하여 적용할 필요가 있다. 아울러 유럽연합의 인공지능 법안에서는, 인공지능 규제 대상을 크게 최소 위험, 제한된 위험, 고위험, 수용 불가 위험 등 4단계로 분류하고 있다. 특히 수용 불가 위험은 인공지능 개발과 활용을 전면 금지하는 강력한 규제에 적용된다. 고위험 영역에는 건강, 안전, 기본권, 환경에 대한 위험 시스템, 정치 캠페인 등 선거에 영향을 미칠 수 있는 시스템, 4천5백만 명 이상 사용자를 보유한 SNS 플랫폼에서 사용하는 추천 시스템 등이 포함되어 있다. 따라서 우리나라의 관련 법안에서도 규제 대상을 세분화하는 한편, 고위험 영역을 구체화해 인권은 물론 생명이나 안전에 있어 위험이 예상되는 내용을 세부적으로 포함해야 할 필요가 있다. 이밖에 유럽연합이 인공지능 법안을 위배할 경우 높은 금액의 과태료 부과를 강제하고 있는 것처럼, 우리나라 법안에도 위배에 대한 제재 방안을 명확하게 정의함으로써 글로벌 표준에 부합하는 실질적 조치로 법안 실효성을 높여야 할 것이다.

V. 결론

인공지능 기술 진화로 인해 야기될 위험에 대한 통제 및 규제의 필요성은 일부 국가가 주도적으로 해결해야 할 문제가 아닌 글로벌 협력을 통해 대응해야 할 난제가 되었다. 인공지능 기술 발전에 따른 사회문제는 데이터 오남용, 개인 정보 유출, 사생활 침해, 저작권 논란 등에 국한되지 않는다. 인공지능 알고리즘에 따른 자동화된

의사결정은 과학기술, 교육, 문화 등 사회 시스템 전반에 막대한 영향력을 발휘할 수 있고, 인권은 물론 민주주의 구현 자체에도 영향을 미칠 수 있으므로 이슈의 출발점에서 있는 현시점에서 대응 방안을 모색하는 것이 무엇보다 중요하다. 선진 법안을 선제적으로 면밀하게 검토하고 올바른 방향성을 설정함으로써 효율적인 규제 거버넌스를 구축해야 하는 이유이다. 인공지능 기술 진화가 불러올 부작용이나 위험성에 얼마나 적절하고 효율적으로 대처할 수 있을지, 사회문제 최소화 방안을 얼마나 신속하게 마련할 수 있을지가 인공지능 기술 중심의 미래 전망을 밝게 하는 핵심 요소가 될 것이다.

References

- [1] Bostrom, N. (2003). Are we living in a computer simulation? *The philosophical quarterly*, 53(211), 243-255.
- [2] Nyholm, S. (2023). A new control problem? Humanoid robots, artificial intelligence, and the value of control. *AI and Ethics*, 3(4), 1229-1239.
- [3] Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. New York: Oxford University Press.
- [4] Hevelke, A. & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: an ethical analysis. *Sci. Eng. Ethics* 21(3), 619-630.
- [5] Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin, London.
- [6] Turing, A. (1950). Computing machinery and intelligence. *Mind* LIX, 433-460.
- [7] Dignum, V. (2019). *Responsible Artificial Intelligence*. Springer, Berlin.
- [8] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4), 12-12.
- [9] Russell, S. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, New York.
- [10] Ord, T. (2020). *The Precipice: Existential Risk and the Future of Humanity*. Hachette Books, New York.
- [11] Mosley, T. ryan. (2023). Five Big Takeaways from Europe's AI Act. *MIT Technology Review*.