국회토론회

EU와 미국은 왜 인공지능을 규제하려는가?

일시 | 2023년 7월 20일(목) 오후2시

장소 | 국회의원회관 제6간담회실

주최 | 더불어민주당 국회의원 장경태, 국회의원 황운하, 정의당 국회의원 장혜영,
건강권 실현을 위한 보건의료단체연합, 광주인권지기 활짝, 무상의료운동본부,
민주사회를위한변호사모임 디지털정보위원회, 사단법인 정보인권연구소,
서울YMCA 시민중계실, 소비자시민모임, 언론개혁시민연대,
연구공동체 건강과대안, 전국민주노동조합총연맹, 전국사무금융서비스노동조합,
전북평화와인권연대, 진보네트워크센터, 참여연대, 천주교인권위원회, 홈리스행동

순 서

2:00 ~ 2:10	인사말	**장경태** (더불어민주당 국회의원, 국회 과학기술정보방송통신위원회) **황운하** (더불어민주당 국회의원, 국회 정무위원회, 국회운영위원회) **장혜영** (정의당 국회의원, 국회 기획재정위원회)
	사회	**김병욱** (민주사회를위한변호사모임 디지털정보위원회, 변호사)
2:10 ~ 2:50	발제	인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점 | **유승익** (한동대학교 연구교수)
		유럽연합 인공지능법의 내용 및 시사점 | **허진민** (참여연대 공익법센터 소장, 변호사)
2:50 ~ 4:20	토론	미국의 인공지능 규제 동향 | **장여경** (사단법인 정보인권연구소 이사)
		인공지능과 소비자 권리구제 | **허유경** (소비자시민모임 이사, 변호사)
		건강권과 인공지능 규제 | **전진한** (보건의료단체연합 정책국장)
		과학기술정보통신부 | **최동원** (인공지능기반정책국 과장)
		개인정보보호원회 | **이병남** (정책국 과장)
		공정거래위원회 | **강승빈** (시장감시정책과 서기관)
		국회입법조사처 | **박소현** (입법조사관)
		국가인권위원회 | **김재석** (인권정책과 과장)
		인터넷기업협회 | **김영규** (정책1실장)
4:20 ~ 4:30	플로어 토론 및 참석자 전체 토론	

인사말

장경태 의원

더불어민주당

국회 과학기술정보방송통신위원회

안녕하세요. 더불어민주당 국회의원 장경태입니다.

“EU와 미국은 왜 인공지능을 규제하려는가? 소비자 안전과 인권 보호를 위한 인공지능에 대한 각국의 노력이 주는 함의” 토론회 개최를 진심으로 축하합니다.

인공지능 알파고와 이세돌 9단이 벌인 ‘세기의 대결’이 있었던지도 벌써 7년이 흘렀습니다.

인공지능의 발전을 전세계가 눈으로 확인한 순간이 지나고 우리 실생활 속에서도 하나둘씩 그 영역이 넓어지고 있습니다. 이제는 단순한 학습을 넘어 추론의 영역까지 넘보는 초거대 인공지능의 시대가 다가오고 있습니다.

이러한 인공지능 기술이 인간의 존엄성과 도덕성을 위협할 수 있다는 두려움은 언제나 존재해왔습니다. 기술은 인간을 위해서 존재해야 합니다. 다른 모든 기술과 마찬가지로 인공지능 역시 가치가 사람을 넘어서는 일이 있어서는 안 될 것입니다. 기술의 편의성을 추구하는 것을 넘어 사람들을 감시하고 차별하거나 인류의 불평등을 조장하는 수단이 되어서는 안 됩니다.

그렇기 때문에 우리는 고민해야 합니다. 더 많은 시간과 노력을 들이더라도 인공지능 기술의 윤리적 활용과 예상치 못하게 발생할 수 있는 사회적 갈등을 최대한 점검해야 합니다. 유럽연합과 미국의 입법례는 물론 더 다양한 가능성을 충분히 검토해 나가야 합니다. 그것이 무지성적 기술의 추종을 거부하고 합리적이고 안정적인 기술발전을 담보할 수 있는 방법이라 생각합니다.

끝으로 오늘 토론회를 공동주최해주신 시민사회단체 여러분과 장혜영 의원님, 황운하 의원님께 감사의 인사를 드립니다. 토론회 발제를 맡아주신 유승익 교수님과 허진민 변호사님, 사회를 맡아주신 김병욱 변호사님께도 존경과 감사를 전합니다.

과학기술정보방송통신위원회의 소속 위원으로서, 한 사람의 국회의원으로서 오늘의 토론회에서 모아주신 의견을 경청하고 깊이 숙고하도록 하겠습니다. 감사합니다. ▢

인사말

장혜영 의원

정의당

국회 기획재정위원회

안녕하세요. 정의당 국회의원 장혜영입니다.

“소비자안전과 인권보호를 위한 인공지능에 대한 각국의 노력이 주는 함의: 미국과 유럽연합 등의 입법사례와 한국 인공지능법안 비교” 토론회 개최를 진심으로 환영하고 축하합니다.

2020년 12월 출시된 인공지능 챗봇 ‘이루다’는 성희롱과 소수자 혐오로 인해 잠정 중단되기도 하면서 우리 사회에 인공지능과 윤리라는 큰 화두와 과제를 남겼습니다. 이루다는 100억개의 사적인 SNS 대화내용을 데이터로 활용해 딥러닝시켜 출시하면서 개인정보 유출 문제도 불거졌습니다. 혹자는 인공지능의 상업화 측면에서 기술적 실수이자 결함이라며 옹호했지만, 개발자의 윤리 문제와 사적 정보의 오남용 및 유출 의혹, 그리고 무엇보다 약자 혐오와 차별이라는 측면에서 비판을 피할 수 없었습니다.

기술의 발전을 거듭하고 있는 인공지능은 인간의 편리성 추구에 발맞춰 우리 사회 곳곳에서 찾아볼 수 있습니다. 그러나 굉장히 복잡하고 폐쇄적인 딥러닝 기법으로 인해 소수자 혐오 등 문제가 발생할 경우 ‘사람이 아니라 기술이 낸 결과’라는 논리로 개발자는 물론 그 누구도 책임지지 않는 문제가 발생할 수 있습니다. 무엇보다 인공지능이 초래한 인권 침해와 차별에 대해 기술 자체 또는 개발자에게는 ‘의도가 없었다’는 식으로 무마할 가능성이 있으며, ‘이루다’ 사태 당시 개발자 측의 해명이 이를 증명하고 있습니다.

시민들의 안전과 소수자 인권 보호를 위해 인공지능의 윤리적 개발과 사용에 필요한 보호장치와 규제 마련이 필요합니다. 인공지능의 기술 발전과 편익이 시민들의 안전과 인권보다 우선해서는 안 됩니다. 미국과 유럽연합 등 해외 각국의 입법사례를 검토하며 한국의 인공지능 규제입법의 과제를 모색하는 오늘 토론회가 매우 의미 있는 이유입니다.

오늘 토론회를 공동주최해주신 시민사회단체와 동료의원님들께 감사의 인사를 드립니다. 그리고 오늘 토론회 발제를 맡아주신 유승익 교수님과 허진민 변호사님, 사회를 맡아주신 김병욱 변호사님께 존경과 감사의 인사를 드립니다. 지정토론을 맡아주신 허유경 변호사님과 전진한 국장님, 장여경 상임이사님과 정부 관계자 여러분들께도 감사의 인사를 드립니다.

시민 모두의 존엄한 삶을 위해 저와 정의당은 최선을 다 하겠다는 약속을 드립니다.
감사합니다. ▢

인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점*

유승익

(한동대학교 연구교수)

< 차 례 >

Ⅰ. 문제설정

인공지능 기술의 기하급수적 발전과 함께 그 잠재적 위험성을 경고하는 목소리도 높아지고 있다. 대표적인 예가 최근 제프리 힌튼(Geoffrey Hinton) 교수의 행보다.[1] 그는

* 이 글은 민주주의법학연구회 2023년 봄 정기학술대회 “과학기술과 인권, 그리고 민주주의”(2023. 5. 19.)에서 발표한 “인공지능이 인권과 민주주의에 미치는 영향과 규제거버넌스의 필요성”을 수정하여, 민주법학 제82호(2023. 7. 1. 발간)에 투고한 논문임.

1) New York Times, 2023. 5. 1., “‘The godfather of AI’ Leaves Google and warns of danger ahead”,<https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>, 검색일: 2023. 5. 3. 힌튼 교수는 인공신경망을 개발한 인공지능의 개척자 중 하나로 알려진

자신이 일생을 바친 연구를 후회하며 인공지능의 잠재적 위험성을 경고한다.

그가 지적하는 위험성은 다음과 같다. 통제되지 않는 빅테크에 의해 개발되는 인공지능 기술로 인해 가짜 사진, 동영상, 텍스트가 인터넷에 넘쳐나 일반적인 사람들은 더 이상 무엇이 진실인지 알 수 없게 될 것이며, 고용시장에서 반복적 업무를 처리하는 사람들(법률보조원, 개인비서, 번역가 등)은 기계에 의해 대체된다는 것이다. 또한 인공지능이 분석하는 방대한 양의 데이터에서 예상치 못한 행동을 학습하는 경우가 많다는 점, 인공지능 시스템이 자체 컴퓨터 코드를 생성하는 것을 넘어 스스로 그러한 코드를 실행하게 될 수 있다는 점, 킬러 로봇과 같은 자율살상무기가 현실화될 수 있다는 점 등도 지적하고 있다.

하지만 디지털 기술은 한때 '해방의 기술'(Liberation Technology)로 여겨지기도 했다.[2] 튀니지, 이집트, 레바논 등 아랍의 봄에서 트위터, 페이스북 등 소셜 미디어는 일종의 공론장이었고 봉기의 매개였다.[3] 디지털 기술은 대중들의 소통을 촉진하여 정치적 변혁을 이끌 수 있는 원동력이 되기도 했다.[4]

기술 발전은 긍정적으로든 또는 부정적으로든 사회적 가치와 원칙에 영향을 미치며 새로운 국면의 변화를 여는 전환점이 되어 왔다.[5] 최근 혁신을 거듭하고 있는 인공지능 기술도 이러한 새로운 국면을 여는 사회적 변곡점이 될 것이라는 예상이 지배적이다.[6] 인공지능으로 통칭되는 알고리즘과 데이터 기술은 방대한 양의 데이터를 처리하는

인물이지만, 최근 10년 이상 근무했던 구글에서 퇴사하여 화제가 되었다. 한편, 최근 미국의 생명미래연구소는 공개서한을 통해 인공지능 개발을 6개월 이상 중단해야 한다고 주장하기도 했다. 인공지능을 개발하는 연구자들조차 통제할 수 없는 인공지능 기술이 경쟁적으로 개발되고 있다는 것이다. 규제 없는 인공지능 개발이 계속되면 통제 불가능한 상황에 직면할 수 있다는 경고이다. 이 서한에는 일론 머스크, 요슈아 벤지오, 유발 하라리 등 유명 인사들이 참여하여 화제가 되기도 했다. 힌튼 교수는 이 서명에 참여하지 않았지만, 이제 인공지능의 위험성을 비판하는 흐름에 합류하게 되었다.

2) Larry Diamond, "Liberation Technology", *Journal of Democracy*, vol. 21, no. 3(2010), 69-83쪽.

3) 인남식, "아랍 민주화 운동과 미국의 대중동정책 변화 연구", 서정민/인남식 엮음, 중동 민주화의 대내외 정치역학(대외경제정책연구원, 2011), 110쪽. "페이스북, 트위터, 유튜브 등 소위 소셜네트워크서비스(SNS)가 확산되면서 아랍 대중들끼리 소통이 가능해졌고, 시위로 이어지는 촉매제가 된 것이다. SNS의 파급효과는 상당히 컸다. 특히 튀니지의 채소 행상 모하메드 부아지지의 분신 장면이 유튜브에 실리고 페이스북으로 옮겨가면서 아랍권 대부분의 젊은이들을 자극하게 된다. 실업 상태의 확산과 곡물가의 상승 그리고 정치적 패배주의 등 구조적 요인이 이러한 촉진 요인을 통해 사회를 숙성시키고 결국 변혁의 단계에 이르게 된다."

4) 2010년 이집트 혁명에서 와엘 고님(당시 구글 중동 지역 마케팅 담당 이사)은 "만약 사회를 해방시키고 싶다면, 인터넷만 있으면 된다"며 혁명의 단초를 제공하는 활동을 하였다. 그러나 5년 이후, 정치적 양극화를 초래하며 인터넷 그 자체가 권력이 되어 가는 현상을 겪으며, "사회를 해방시키고 싶다면, 우리는 먼저 인터넷을 해방시켜야 한다"는 말을 남겼다. 박승일, 기계, 권력, 사회: 인터넷은 어떻게 권력이 되었는가(사월의책, 2021), 37쪽 이하 참조.

5) 제프리 삭스는 경제체제는 어느 시간, 어느 장소가 되었든 지리, 기술, 제도라는 세 가지 조건의 상호작용에 의존한다고 하면서, "지리의 경제적 중요성은 변화하는 지식과 기술에 의해 끊임없이 다르게 규정되어왔다"라고 지적한다. 제프리 삭스, 이종인 옮김, 지리, 기술, 제도(21세기북스, 2021), 49, 52쪽.

6) "우리[는] 디지털 기술에 힘입어 경이로운 발전을 거듭하는 시대에 살고 있다. … 한 마디로, 우리는

새로운 방식이다.[7] 정치권에서도 '디지털 시대의 쌀은 데이터'라며 산업적 활용방안을 앞다투어 제시하고 있다. 이른바 디지털 경제에서 대량의 데이터와 정보는 알고리즘 등의 가공을 통해 가치를 창출할 수 있는 자산으로 인식된다.[8] 인공지능은 인간이 세계를 경험하는 방식을 근본적으로 변화시키고 있으며, 정치적·경제적·사회적 구조를 획기적으로 전환할 것이다.

인공지능은 시민들의 권리와 자유에 관한 역량을 높인다는 측면에서 긍정적 영향을 미치기도 하지만, 새로운 헌법적·법률적 문제를 제기하기도 한다. 최근의 인공지능은 전대미문의 기술이다. 대량의 정보처리, 고도의 계산과 예측, 변화하는 상황에 대한 학습과 조정적 반응, 사물 인식 및 분류와 같은 복잡한 작업을 수행할 수 있는 자율 시스템을 인류 역사상 처음으로 다루게 된 것이다. '알고리즘 지배'(Algocracy),[9] '알고리즘 사회'(Algorithmic Society)[10]라는 조어들이 의미하는 바처럼, 인공지능은 정치, 경제, 사회, 법의 새로운 문제계이다. 다양한 문제들이 등장하고 있다. 알고리즘 의사 결정 시스템(Algorithmic Decision Systems: ADS)을 통해 아동복지, 형사사법, 학교 배정, 교사 평가, 화재 위험평가, 노숙자 주거 우선순위 지정, 건강보험, 출입국 관리 및 입국시 위험평가, 예측 치안 등 광범위한 분야에서 인공지능이 의사 결정에 관여하게 될 것이다.[11]

이 글은 정치적·법적 문제의식에 한정하여, 현재 기술 수준에서 관찰되고 예측되는 범위 안에서 인공지능이 인권과 민주주의에 미치는 영향을 살펴보고, 이에 대응할 수

변곡점에, 컴퓨터 때문에 궤도가 크게 구부러지는 지점에 와 있다. 제2의 기계 시대로 진입하고 있는 것이다." 에릭 브린올프슨, 앤드루 맥아피, 이한음 옮김, 제2의 기계시대(청림출판, 2014), 15쪽. 최근 헨리 키신저 등이 펴낸 저서에도 인공지능의 근본적 특성이 지적되고 있다. "인류의 역사는 기술의 변천사이기도 하다. 그러나 기술로 인해 사회정치적 구조가 근본적으로 바뀐 사례는 거의 없었다. … 하지만 AI는 인간 경험의 모든 영역에서 변화를 예고한다. 그 변화의 중심에는 인간이 현실을 이해하는 방식, 그리고 그 안에서 자신이 맡은 역할을 이해하는 방식을 바꿔놓는 철학적 전환이 있을 것이다". 헨리 키신저/에릭 슈밋/대니얼 허튼로커, 김고명 옮김, AI 이후의 세계(윌북, 2023), 53쪽.

7) 대표적으로 Sue Newell/Marco Marabelli, "Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'", *Journal of Strategic Information Systems*, vol. 24, no. 1(2015), 3쪽.

8) 빅토르 마이어 쇤버거/케니스 쿠키어, 이지연 옮김, 빅 데이터가 만드는 세상(21세기북스, 2013), 17쪽. "사람들은 더 이상 데이터를 유통기한이 지난 고정물로 생각하지 않게 되었다. 이전에는 비행기가 착륙하고 나면(혹은 구글에서 검색어가 처리되고 나면), 수집된 데이터는 애초의 목적을 달성했으므로 그 유용성이 끝났다고 생각했다. 하지만 이제 데이터는 비즈니스의 원자재가 되었다. 빼놓을 수 없는 경제 인풋(Input)으로서 새로운 형태의 경제적 가치를 창출하는 원료가 된 것이다".

9) John Danaher, "The Threat of Algocracy: Reality, Resistance and Accommodation", *Philosophy & Technology*, vol. 29, no. 3(2016), 245쪽.

10) Agnieszka M. Walorska, "The Algorithmic Society", Denise Feldner 엮음, *Redesigning Organizations Concepts for the Connected Society*(Springer, 2020), 149-160쪽.

11) Céline Castets-Renard, "Human Rights and Algorithmic Impact Assessment for Predictive Policing", Hans-W. Micklitz 외 엮음, *Constitutional Challenges In The Algorithmic Society*(Cambridge University Press, 2022), 93쪽.

있는 제도적 장치를 규제 거버넌스의 관점에서 논의하고자 한다. 특히 최근 국회에서 논의되고 있는 인공지능 법안의 내용과 문제점을 살펴보고자 한다.

Ⅱ. 인공지능이 인권과 민주주의에 미치는 영향

1. 인공지능의 의미와 영향

인공지능은 매우 다양하게 정의된다.[12] 공학기술의 관점에서 인공지능이란, "지능적인 기계를 만드는 과학과 공학기술"(John McCathy), "인간이 수행한다면 지능이 필요한 일을 하는 기계를 만드는 과학"(Marvin Minsky), "주어진 상황에서 가능한 최선의 행동을 취하는 지능적 에이전트를 구축하는 문제"[13]라 할 수 있다. 우리 법제에서 인공지능을 적시하여 정의하고 있는 경우는 아직 없으며,[14] 최근 논의되고 있는 인공지능 법안에서는 인공지능을 "학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것"으로 정의하고 있다.[15]

인공지능이 인간과 사회에 미치는 영향은 양가적이다. 긍정적 영향과 부정적 영향 모두 가지고 있을 것이다. 반복적 노동을 대체하고, 증거에 기반한 신속한 의사 결정을 지원하며, 예측 시스템을 통해 각종 사회적 위험을 예방할 수 있다는 점은 긍정적 측면일 것이다.

하지만 인공지능 사회에서 인공지능 시스템이 갖는 부정적 영향과 잠재적 위험성은 인권과 민주주의에도 치명적이다. 이하에서는 인권과 민주주의의 측면에서 인공지능이 갖는 부정적 영향과 위험성을 살펴본다.

2. 인공지능 시스템에 의한 인권의 침해 가능성과 부정적 영향

인공지능 시스템은 인권과 기본권 목록의 거의 모든 권리에 침해가능성을 갖는다. 인공지능은 완벽한 기술이 아니다. 품질이 낮은 데이터셋을 학습한 예측 모델은 편향된 결과나 부정확한 결과를 도출하여 차별적 결과를 초래한다. 예를 들어, 챗봇 이루다의

---

12) 김진우, 나의 첫 인공지능 수업(메이트북스, 2022), 18쪽.

13) 스튜어트 러셀/피터 노빅, 류광 옮김, 인공지능: 현대적 접근방식, 제3판(제이펍, 2016), 2쪽 이하.

14) 다만, 행정기본법 제20조는 자동적 처분과 관련하여, 인공지능 기술을 적용한 시스템을 포함한 '완전히 자동화된 시스템'으로 처분을 할 수 있다고 규정하고 있다.

15) 인공지능산업 육성 및 신뢰 확보에 관한 법률안(윤두현의원 대표발의, 의안번호: 18726) 제2조 제1호. 이 법안은 "인공지능 기술"을 '인공지능을 구현하기 위하여 필요한 하드웨어 기술 또는 그것을 시스템적으로 지원하는 소프트웨어 기술 또는 그 활용 기술'로 정의하며, 그 외에도 "고위험 영역에서 활용되는 인공지능", "인공지능 윤리", "인공지능 산업" 등을 정의하고 있다.

혐오 발언 사건, 경비로봇의 유아 공격 사건, 챗GPT의 답변 오류나 환각(Hallucination) 문제 등을 떠올려 볼 수 있다.

아래 [표]는 인공지능 시스템에 따라 침해 가능성이 있는 인권 목록을 예시한 것이다.[16)]

[표] 인공지능 시스템에 의해 침해 가능성이 있는 인권

인공지능 시스템 예시	침해될 가능성이 있는 인권
노인과 장애인 등 대상자의 맥박, 혈당, 활동 등을 감지하고 말벗, 인지기능을 지원하는 돌봄로봇	개인정보자기결정권 침해
얼굴인식에 기반한 출입국 자동화 시스템	인종, 국가 등에 따른 차별 개인정보자기결정권 침해
지역별로 범죄 발생 확률을 예측하여 순찰 인력을 배치하는 인공지능 범죄 예측 시스템	인종, 지역 등에 따른 차별
인공지능 채용(면접) 시스템	성별, 연령, 장애, 용모, 출신지역 등에 따른 차별
공공 장소에서의 행인의 얼굴을 인식하여 용의자와 대조하는 원격 얼굴인식 시스템	이동의 자유 침해, 집회 및 결사의 자유 침해, 자의적 체포
아동이 사용하는 소셜네트워크서비스에서 선정적이고 자극적인 콘텐츠가 우선 노출되도록 하는 알고리즘	아동의 권리 침해 개인정보자기결정권 침해
소셜네트워크 플랫폼에서의 알고리즘 기반 콘텐츠 관리 시스템	표현의 자유 침해 정보접근권 침해
소셜네트워크 플랫폼에서 개인의 정치 성향에 기반한 정치광고 노출 시스템	자유로운 정치참여 제한 선거권 침해
고등학교의 기존 성적에 기반한 인공지능 대학입학 시스템	지역에 따른 차별 교육권 침해
사업장 내에 설치된 생체인식, 위치추적 시스템	노동자 개인정보자기결정권 및 노동3권 침해
인공지능 판결 지원 시스템	공정한 재판을 받을 권리 침해
인공지능을 통한 사회보장급여 부정수급 탐지시스템	사회보장수급권, 장애인권리 침해, 인종 및 장애 등에 따른 차별

16) 아래 [표]는 유승익 외, 인공지능 인권영향평가 도입 방안 연구(국가인권위원회, 2022), 214쪽. 이 표에서 해당 인공지능 시스템이 반드시 인권 침해적이라는 의미는 아니며, 해당 인권과의 관련성만을 나타낼 뿐이다. 또한 해당 목록은 인공지능 시스템에 의한 인권 침해 가능성의 극히 일부 사례일 뿐이다.

이상에서 살펴본 바와 같이, 알고리즘의 편향되고 잘못된 결정에 따라 표현의 자유, 평등권, 정치적 기본권, 개인정보자기결정권 등이 침해될 수 있다. 이 중에서도 표현의 자유와 개인정보자기결정권은 정치적 의사형성과정에서 중요한 기본권이라는 점에서 특별히 강조될 필요가 있다. 예를 들어 인공지능 시스템이 적용된 소셜 미디어 등에서 의견을 자유롭게 표현할 수 없거나 '사적 검열'과 같은 형태의 관리가 별다른 통제 없이 이루어진다면 민주적 의사형성에 장애가 된다.[17] 또한 개인정보자기결정권이 제약되어 개인의 데이터 처리를 규율하는 규칙이 마련되어 있지 않다면 개인은 책임성과 투명성이라는 보호장치 없이 사적 감시체제에 노출될 수밖에 없다.[18]

이러한 의미에서 '콘텐츠 관리'와 '사용자 프로파일링'은 인공지능 시스템이 기본권과 민주주의에 미치는 위험성을 예시해 주는 사례라 할 수 있다.

콘텐츠 관리(Content moderation)는 혐오 표현이나 허위 정보, 거짓 주장들에 대응하기 위해 개별 기술 기업들이 "내부 '관리자(moderator)'를 고용하여 개별 콘텐츠를 확인하고 플랫폼이 정한 규칙을 위반하는 게시물을 삭제"[19]하는 관리 방식을 말한다.[20]

사용자 프로파일링(Users' profiling)이란, "다양한 방법으로 수집된 데이터를 분석하여 개인 또는 개인 그룹에 대한 새로운 특성 또는 행태 정보를 생성하고 적용하는 등의 작업 일체를 의미하는 것"을 말한다.[21] 주요 사례로 이용자 행동 정보 분석을 통해 맞춤형 온라인 광고를 제공하는 것, 검색 엔진 사업자들이 이용자의 위치, 과거 검색 기록, 다른 이용자와의 관계 등의 분석을 통해 검색 결과를 개인화(personalization)함으로써 이용자의 검색 의도에 부합하고 연관성 있는 결과를 보여주는 것, 개인 신용평가 결과를 이자율 결정 등 금융거래에 활용하는 것 등을 들 수 있다.

특히 콘텐츠 관리는 다양한 관점에서 문제를 지적할 수 있다. ① 빅테크 등 일반 사기업이 디지털 환경에서 '사적' 보호 기준을 설정함으로써 기본권 보호 여부를 결정한다. 사적 '검열' 권한을 인정하는 꼴이다. ② 사적 판단을 통해 콘텐츠를 관리함으로써 공적 기준과 사적 기준의 경계를 모호하게 하고, 예측가능성이나 법적 안정성의 문제를

---

17) 홍남희, "디지털 플랫폼에 의한 '사적 검열(private censorship)'", 미디어와 인격권 통권 제6호(2018), 135쪽 이하.

18) Oreste Pollicino/Giovanni De Gregorio, "Constitutional Law in the Algorithmic Society", Hans-W. Micklitz 외 엮음, *Constitutional Challenges In The Algorithmic Society*(Cambridge University Press, 2022), 7쪽.

19) Melissa Heikkilä, "사회를 오염시키는 소셜미디어, 콘텐츠 관리만으로는 해결할 수 없다", 〈https://www.technologyreview.kr/social-media-polluting-society-moderation-alone-wont-fix-the-problem/〉, 검색일: 2023. 5. 1.

20) 콘텐츠 관리와 반대성명의 한계 및 설계의 문제를 지적하는 다음 글을 참조. Bits of Freedom, "'혐오표현'에 대항하기: '설계'는 '콘텐츠 관리'로부터 우리를 구할 수 있을까?", 〈https://act.jinbo.net/wp/39075/〉, 검색일: 2023. 5. 1. 이에 따르면, 콘텐츠 관리는 디지털 플랫폼에게 과도한 검열 권한을 부여하고, 기업들로 하여금 무엇이 옳고 그른지 결정하게 하는 민영화된 법집행으로 실질적인 법 집행을 대체하여, 표현의 자유를 더욱 취약하게 할 것이다.

21) 방송통신위원회/한국인터넷진흥원, 개인정보 처리에서의 프로파일링 사례집(2020), 3쪽.

제기한다. ③ 디지털 플랫폼의 콘텐츠 관리는 투명성과 책임성이 결여될 수밖에 없다. 표현물의 삭제 여부를 결정하는 디지털 플랫폼 기업은 정부나 공공기관처럼 헌법상 표현의 자유에 요구되는 엄격한 심사요건을 따를 의무가 없으며, 사업 목적에 따라(대개는 상업적 목적) 자유롭게 평가하여 삭제할 수 있다.[22] 표현의 자유의 보장 여부가 디지털 플랫폼의 사적 기준에 의존하게 되는 것이다.

같은 맥락에서 온라인상의 허위 정보(disinformation) 문제는 법적·정치적으로 매우 민감한 문제를 제기하기도 한다. 미국 대선 당시의 '피자게이트'나 영국의 브렉시트 국민투표의 사례에서 볼 수 있었던 것처럼, 정치적 의사형성에서 허위 정보의 파괴적 영향력은 개인의 프라이버시 침해에 한정되지 않는다. 이와 관련하여 허위 정보에 대한 규제수단의 헌법적 한계를 획정하는 문제, 허위 정보의 한계에 관한 기준을 설정하고 이를 관리하는 데 있어서 인공지능 시스템을 활용하는 문제와 같은 새로운 쟁점들이 등장한다. 아직 이러한 쟁점에 대한 헌법적 한계가 불분명하다.

현재의 수준에서 보자면, 디지털 플랫폼 기업은 (허위)정보의 양을 사적 기준에 의해 '조절'할 수 있고, 이를 통해 정치적 지형을 형성할 수도 있다. 전 세계적으로 관찰되는 포퓰리즘은 상업적 콘텐츠 관리의 '완화되고 조절된' 허위 정보를 유통시키기 위해 디지털 플랫폼을 적극 활용한다. 이탈리아의 오성운동(M5S)의 성공이 대표적인 예일 것이다.[23]

이렇듯 통제되지 않는 인공지능 시스템은 개인의 인권을 취약하게 할 뿐만 아니라 정치적 공동체의 운명을 민간기업의 처분에 맡길 수 있는 위험도 안고 있다.

---

22) 이 문제를 자세히 서술하고 있는 롭 라이히 외, 이영래 옮김, 시스템 에러: 빅테크 시대의 윤리학(어크로스, 2022), 307쪽 이하.

23) 정병기, "오성운동(M5S)의 직접 의회주의와 사이버크래틱 집중주의: 포스트포퓰리스트 정치 운동의 성공과 한계", 한국정치연구 제29권 제2호(2020), 91-116쪽. 오성운동은 정치풍자로 공영TV 출연이 금지됐던 코미디언 그릴로(Beppe Grillo)가 웹 전략가인 카사레죠(Gianroberto Casaleggio)와 함께 2009년 창당한 정당이다. 웹 2.0 기반으로 활동하면서 인터넷 민주주의를 추구하는 정당으로 성장했고, 2013년 총선에서 25%의 지지율을 얻어 제1당이 된 후, 2018년 선거에서는 집권까지 하였다. 오성운동은 그릴로의 블로그, 온라인 플랫폼 모임인 밋업(Meetup), 당원들의 최종 의사 결정을 위해 활동하는 공간인 '루소' 등 디지털 플랫폼을 적극 활용한다. 이 플랫폼들은 카사레죠의 웹 회사가 독자적으로 제작 관리하고 있는데, 그 불투명성, 당원 제한, 리더십 집중 현상 등의 문제가 지적되고 있다.

3. 감시사회의 고도화와 분화 : 전통적 감시 권력에서 빅데이터 감시 권력으로

인공지능 사회의 출현을 통해 등장하는 가장 민감하고 중요한 문제 중의 하나는 감시사회의 성격이 고도화, 심화, 확장되고 있다는 점이다.[24] 인공지능이 보편화된 사회는 과거 '정보사회'로 지칭된 사회가 진화된 형태이며,[25] 새로운 형태의 '대중감시'(mass-surveillance)가 가능해지는 사회이다.[26] 과거 푸코식의 판옵티콘 규율 체계와 일방향·전방위 시각감시 중심의 감시사회에서 "유동형(liquidity-driven) 감시와 네트워크형(network-driven) 알고리즘 감시가 공모하는 형태로 진화"하고 있다는 것이다.[27]

사실 감시사회는 새로운 주제가 아니다. 한국 현대사에서 만연했던 사찰과 정보정치부터 CCTV와 전자주민증까지 다양한 감시기법과 기술의 위험성에 대한 인식은 꾸준히 높아져 왔다.[28] 하지만 팬데믹을 거치며 사회적 거리두기나 감염병 의심자에 대한 통제 정책에 관한 효과적인 도구로서 인공지능 기술은 한층 더 각광받고 있다.[29] 이제 인공지능 기술은 자유의 적이라기보다 보건과 건강을 지키는 안전막처럼 인식되고 있다.[30] 감시기술과 감시문화는 공공과 민간을 가로지르며 별다른 저항 없이 수용되고 있다. 이를 "감시 자본주의"라 부르기도 한다.[31]

그런데 이렇게 고도화된 감시사회는 인권의 관점에서 미묘한 쟁점을 제기한다. 팬데믹 시대를 거치며 확인한 것처럼 고도화된 감시기술이 사회에 순기능을 하며 시민들의 삶을 개선하는 측면도 분명 존재한다는 것이다. 순응과 자발적 참여를 통해 감시 권력이 인권 주체의 대립적 객체로 인식되기보다 생명, 건강, 재산, 자유의 인프라 구조처럼 자리매김한다. 안전, 보호, 편의성, 효율성 등의 혜택을 위해 개인 데이터가 기록, 저장, 복구, 이동, 교차, 교환되는 것에 동의하는 것을 당연하게 받아들인다. "숨길 이유가 뭔가"("nothing to hide")하는 감시 이데올로기 자체가 뉴노멀로 되고, 감시에 의문을 제

---

24) 이광석, "포스트-판옵티콘 시대 감시 연구, 새로운 지형", 김동욱 외 엮음, 스마트 시대의 위험과 대응방안(나남출판, 2015), 192쪽 이하.

25) 프랭크 웹스터, 조동기 옮김, 현대 정보사회 이론(나남출판, 2016).

26) Neil M. Richards, "The Dangers of Surveillance", *Harvard Law Review*, vol. 126, no. 7(2013), 1934쪽.

27) 이광석, "포스트-판옵티콘 시대 감시 연구, 새로운 지형", 192쪽. 그리고 이 글에서 주로 참고하고 있는 지그문트 바우만/데이비드 라이언, 한길석 옮김, 친애하는 빅브라더(오월의봄, 2014)도 참조.

28) 대표적으로 한홍구 외, 감시사회(철수와영희, 2021), 11쪽 이하.

29) 코로나 확진자 동선 파악은 카드 사용내역, 스마트폰 GPS, 주변 CCTV를 통해 이루어졌다. 이 중에서 CCTV는 단순히 영상녹화 및 저장기능에 머무르지 않고, 얼굴인식, 자동차 번호판 인식, 영상분석 추적 등이 가능한 '지능형 CCTV'로 발전했다. 다른 CCTV에 촬영된 영상 속 사람들과 비교해 동선을 파악할 수 있다. ICT 표준에 의한 CCTV호환이 가능해졌고, 딥러닝, 빅데이터 기술이 이를 뒷받침한다. CCTV 관제센터, 경찰서, 소방서가 연결되어 통합·관리된다. 한국정보통신기술협회, 스마트 시티 정보의 통합 관리 및 운영을 위한 플랫폼 소프트웨어 요구사항(TTAK.KO-10.1118) 참조.

30) Natalie Ram/David Gray, "Mass Surveillance in the Age of COVID-19", *Journal of Law and the Biosciences*, vol. 7, no. 1(2020), 1쪽.

31) 쇼샤나 주보프, 김보영 옮김, 감시 자본주의 시대(문학사상, 2021).

기할 역량 자체가 무력화된다.[32]

시민들의 통제에서 벗어난 기술을 통해 구현되는 이러한 새로운 감시사회에서 감시질서는 푸코적 의미에서 정상화된다.[33] '정상적' 감시질서하에서 장기적으로 시민들은 새로운 형태의 감시를 준수하고 순응하며 당연시하게 된다. 이때 개인이 갖는 프라이버시권과 개인정보자기결정권은 "데이터 조각"으로 파편화되어 비인격화되고,[34] 개인(Individual)으로서 법적 권능은 저하하고 분할되어 무화된다.

이러한 새로운 감시사회에서 정부와 민간 기업 모두에 의해 대규모 감시가 이루어지게 된다. 경우에 따라서 개인 상호 간 감시도 일상화된다. 이 모든 시나리오를 종합할 때, 개인의 사생활은 지속적이고 침습적인 감시의 대상으로 격하되고, 데이터에 대한 권리주체의 통제력은 상실된다.

4. 자유의 의미변화: 축소와 왜곡

인공지능 사회는 자유의 핵심적 내용도 변형시킨다. 인공지능은 특정 개인이나 집단을 표적으로 자유를 직접적으로 실현하거나 제한하는 것이 아니라 대량의 데이터를 모집단으로 패턴을 인식하여 자율적으로 결정을 내릴 뿐이다. 개인이나 집단이 설정한 목적이나 내용, 개인정보의 개별적 의미나 인과적 관계(causal link)는 중요성을 상실하고 상관관계(correlation)와 예측성으로 대체된다. 기계는 인간을 위해 또는 인간을 대신하여 데이터의 상관관계에 기초하여 예측적으로 결정을 내린다. 여기에서 인간의 자유는 침해된다고 말할 수도 있지만, 엄밀히 말해 자유 그 자체가 기계적으로 환원되고 포획된다는 표현이 더 정확할 것이다.

인간 자유를 포획(capture)하는 방식은 두 가지이다. 하나는 인공지능 시스템으로 하여금 인간을 대신해서 직접 결정하도록 위임하는 경우이다. 마치 대의제하에서 자유위임의 원리와 같이 인간은 의사 결정 시스템이 자율적으로 결정할 수 있도록 위임한다. 인간 에이전트로서 이러한 위임결정의 모멘트에만 자유를 행사하고 이후 자기 구속된다.

다른 하나는 인간이 특정한 행위나 결정을 할 때 필요한 정보를 인공지능 시스템이 제공하도록 하는 경우이다. 배경지식을 자동화된 기계가 독점 또는 과점하도록 하는 방식이다. 인간은 지식과 정보를 통해 자유롭게 결정하고 행동한다. 정보원은 인간의 결

---

32) Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*(W. W. Norton & Company, 2015).

33) David Lyon, *Surveillance After September 11*(Polity, 2003), 150쪽.

34) 이광석, "포스트-판옵티콘 시대 감시 연구, 새로운 지형", 197쪽. "개인데이터의 경우에는, 가치가 추출될 수 있도록 인터넷 이용자들이 뒤에 남기는 무수한 클릭과 네트상의 동선과 흔적들, '데이터 배출'(data exhaust)이 빅데이터의 핵심이 된다. 이는 이용자들이 남긴 데이터 부스러기, 즉 '데이터 조각'으로 불리기도 한다."

정에 영향을 미치는 핵심 요소 중의 하나이다. 어떤 정보를 취득하느냐에 따라 어떤 결정과 자유를 행사하는지 달라진다. 예를 들어, 국회의원 선거에서 특정 후보 선택은 해당 후보에 대한 이력, 정책, 여론, 평가 등의 정보에 따라 달라진다. 오늘날 직접 경험 외에 정보를 얻기 위해 이용하는 채널은 압도적으로 스마트폰이다.[35]

또한 자동화된 기술 시스템을 통해 대량으로 유통되는 정보를 일괄적으로 소비하게 됨으로써 개인이나 집단이 자신의 정체성을 과거 대면공동체나 소속 집단의 공유된 정보에 의해 구획적으로 형성하던 방식에서 탈피하게 되었다. 대규모로 집적된 데이터에서 알고리즘에 의해 상관관계적으로 큐레이트된 지식의 흐름에 따라 개인이나 집단의 정체성이 유동적으로 "유도"된다.[36]

이러한 인공지능 사회에서 전통적인 의미에서의 자유의 의미는 축소되고 변형된다. 첫째, 인공지능 시스템으로 하여금 대신 결정하도록 '동의'한 경우, 시스템이 인간의 자유를 제한하는 결정을 내렸을 때 그것을 권리 주체의 자유에 대한 제한으로 평가할 수 없게 된다. 이미 인간 에이전트는 해당 시스템이 인간의 결정보다 더 증거에 기반하고 있으며 과학적, 효과적, 중립적일 것이라는 가정하에 결정을 위임했다. 명시적으로 또는 묵시적으로 인공지능 시스템에게 의사 결정을 위임했다는 점에서, 인간의 의지나 자기 결정에 반하여 자유가 침해되었다고 말하기 힘들어진다.[37] 인공지능 시스템을 활용한 통치는 시스템 도입 단계에서 사용자가 이용약관을 동의함으로써 저항 없이 완성된다.

둘째, 인공지능 기술이 제공하는 정보를 통해 결정을 내릴 때, 자유에 대한 제약은 더 이상 외재적인 것이 아니게 된다. 인간은 스스로 인공지능에 의해 제공된 정보를 신뢰하고 이용할 뿐이다. 여기에 외부적 강제나 폭력이 개입되지 않는다. 자신이 가진 스마트폰이나 PC를 통해 자유롭게 뉴스 플랫폼을 선택하여 이용할 뿐이다. 이러한 정보들을 믿을 만하고 유효한 근거로 스스로 받아들일 뿐이다. 물론 이러한 정보 수용은 인터넷 플랫폼에서 제공하는 웹상의 정보들이 과학적으로 정확하거나 신뢰할 수 있다는 가정에 기초한다.

그러나 이러한 정보에 대한 신뢰는 인과성에 기초한 과학이나 객관적 검증과는 전혀 관계가 없다. 플랫폼은 다만 기존 파라미터에서 선호를 추출함으로써 행동 패턴을 학습하는 알고리즘을 사용하여 개인과 집단이 갖는 기존의 신념을 반복적으로 강화할 뿐이다. 알고리즘이 객관성이나 정확성을 보증한다는 것은 확증편향에 의해 생성되는 기존

---

35) 2022년 기준 신문구독 이용매체 현황에 따르면, 스마트폰 애플리케이션 80.2%, 스마트폰 웹브라우저 11.2%로 91.5%가 스마트폰으로 신문을 구독하고 있다. 그 외에 PC 4.8%, 종이신문 3.8% 순이다. 미디어통계포털(Kisdistat), 〈https://stat.kisdi.re.kr/〉, 검색일: 2023. 5. 1.

36) Holger Pötzsch, "Archives and Identity in the Context of Social Media and Algorithmic Analytics: Towards an Understanding of iArchive and Predictive Retention", *New Media & Society*, vol. 20, no. 9(2018), 3304쪽.

37) Andrea Simoncini/Erik Longo, "Fundamental Rights and the Rule of Law in the Algorithmic Society", in Hans-W. Micklitz 외 엮음, *Constitutional Challenges In The Algorithmic Society*(Cambridge University Press, 2022), 40쪽.

의 신념을 확률적으로 반복함을 의미할 뿐이다. 사용자는 관심 있는 커뮤니티별로 모여 해당 커뮤니티의 기존 선호도를 학습한 알고리즘에 의해 추출된 정보를 '객관적' 정보로 습득하게 된다. 전체적으로 보면, 집단양극화와 분리, 통약불가능성을 초래한다. 결국 자유는 외재적으로 제약되는 것이 아니라 내재적으로 왜곡된다.

5. 사적 권력의 문제

인공지능 사회에서 빅테크 기업이나 디지털 플랫폼 기업 등 사적 행위자의 지위와 역할이 국가나 정부의 역할을 초월하면서 공사의 경계가 희석된다. 빅테크 기업은 대중들의 데이터를 통해 막대한 수익을 얻고 있지만, 별다른 대안 없이 빅테크가 구축한 디지털 플랫폼 서비스에 더 의존하게 된다(Lock-in 효과). 이러한 플랫폼은 애초 귀찮은 일에서 해방해주고 보다 창의적인 활동을 할 수 있도록 약속했지만, 역설적으로 디지털 플랫폼 자체가 불가결의 존재로 되어 가고 있다.

개인들은 점점 더 자신들의 의사 결정에 영향을 미치는 기술 시스템에 둘러싸이게 되지만, 이러한 현상을 이해하거나 통제할 가능성은 점점 더 줄어들게 된다. 결과적으로 민주적 의사 결정에 의식적으로 참여할 가능성도 축소될 수밖에 없다.[38] 이러한 현상은 알고리즘 자체가 불투명성을 가지고 있는 것이기도 하지만, 다른 한편으로는 인공지능 시스템이 사적으로 개발되기 때문이기도 하다.

인공지능 사회에서 혁신의 주도세력은 기하급수적 성장을 거듭하면서 인터넷 관련 상품 및 서비스 시장을 주도하는 거대 빅테크 기업이다.[39] 감시 자본주의 또는 "플랫폼 자본주의"는 이러한 빅테크 기업들이 압도적 기술과 자본으로 사회적 권력을 행사하는 새로운 시스템이다.[40]

시가총액의 약 25%를 차지하는 5대 빅테크 기업(애플, 아마존, 마이크로소프트, 페이스북, 구글(알파벳))을 '몰리고폴리'(moligopoly)라는 합성어로 정의하기도 한다.[41] 이러

---

38) Oreste Pollicino/Giovanni De Gregorio, "Constitutional Law in the Algorithmic Society", 5쪽.

39) 예를 들어, 스마트 기기(Apple, Samsung, Huawei, Xiaomi), 웹 검색 엔진(Google), 소셜 미디어 기업(Facebook, Instagram, Twitter), 클라우드 서비스 제공업체(Amazon, Microsoft, Google), 전자상거래 기업(Amazon, Netflix), 소셜 플랫폼(Zoom, Cisco Webex) 등이다. Andrea Simoncini/Erik Longo, "Fundamental Rights and the Rule of Law in the Algorithmic Society", 32쪽.

40) 쇼샤나 주보프, 김보영 옮김, 감시 자본주의 시대, 420쪽. "역사적으로 다른 어떤 시기에도 지금처럼 전례 없는 부와 권력을 지닌 민간 기업들이 돈으로 살 수 있는 모든 첨단 과학의 노하우로 전 지구적인 유비쿼터스 컴퓨팅 지식 및 통제 아키텍처를 구축하고 유지하며, 이를 토대로 행위의 경제에 입각한 사업 활동을 이처럼 자유롭게 영위한 적은 없었다." 또한 주보프는 페이스북 등을 감시 자본가로 지칭하며, "사회적 권력과 통제권을 위한 투쟁이 상대해야 할 눈에 보이지 않는 적은 이제 계급이나 생산관계가 아니라 자동화된 행동 수정이다"라고 강한 어조로 비판한다.

41) Nicolas Petit, *Big Tech and the Digital Economy: The Moligopoly Scenario*(Oxford University Press, 2020); Oreste Pollicino/Giovanni De Gregorio, "Constitutional Law in the Algorithmic Society", 32쪽 이하.

한 몰리고폴리 기업들은 사용자들의 거래를 통해 발생하는 네트워크 효과로부터 수익을 얻을 뿐만 아니라, 사실상의 정치적 영향력까지도 획득하고 있다. 그 가장 극단적인 예는 케임브리지 애널리티카(Cambridge Analytica)가 유권자 타겟팅을 통해 2016년 미국 대선과 영국의 브렉시트 국민투표에 개입한 사건일 것이다.[42)]

인공지능 사회는 사적·사회적 기업 권력이 정치 권력과 병존하고 교차하는 사회이다. 이러한 새로운 권력의 출현은 헌법 전통에 도전하고 있으며, 이에 대응하여 헌법은 어떻게 진화해야 하는지에 관해 새롭고도 중대한 문제를 제기한다.

Ⅲ. 우리나라 인공지능 법안의 내용과 문제점

인공지능 사회는 인권과 민주주의의 가치에 다양한 위협요인이 되지만, 이를 예방하거나 완화한 제도적 장치는 충분히 마련되지 않은 실정이다.

인공지능 산업은 아직 혁신이 진행 중인 민간 주도 분야라는 점, 인공지능 기술에 대한 무정부적 탈규제는 인권과 민주주의의 본질적 내용과 근본원칙을 방기할 수 있다는 점, 새로운 인공지능 기술이 등장했을 때 법제도가 개입하면 너무 늦은 경우가 많다는 점 등이 종합적으로 고려되어야 한다.

따라서 사후적 책임 분배와 구제수단을 강구하는 사법적 조치만으로는 인공지능 시스템의 위험성을 예방하거나 완화할 수 없다. 인공지능 기술의 안전성을 확보하고 위험성을 완화하기 위해서는 기술의 형성과정을 역추적하여 적절한 시점에 개입하는 것이 중요하다. 따라서 보다 두터운 보호를 위해서는 규제를 통한 보호, '설계'에 의한 보호, '기본값'(default setting)에 의한 보호까지도 필요하다.[43)]

세계 여러 나라는 인공지능 기술에 잠재하는 위험에 대처하기 위한 입법적 준비에 한창이다. 신뢰가능한 인공지능 기술의 발전을 위해 공정성, 윤리성, 투명성과 설명가능성, 안전성 등의 차원에서 통제와 규율을 강화하고 있다.

우리 국회도 인공지능 기술에 대응하는 입법에 속도를 내고 있다. 지난 2월 14일 국회 과학기술정보방송통신위원회의 법안소위는 '인공지능산업 육성 및 신뢰 기반 조성에 관한 법률안'을 통과시켰다. 이하에서는 이 법안을 둘러싼 주요 쟁점과 문제점을 검토하고 앞으로의 과제를 살펴본다.

---

42) 이에 대하여는 시바 바이디야나단, 홍권희 옮김, 페이스북은 어떻게 우리를 단절시키고 민주주의를 훼손하는가(아라크네, 2020), 207쪽 이하. 또한 브리태니 카이저, 고영태 옮김, 타겟티드(한빛비즈, 2020); 로저 맥나미, 김상현 옮김, 마크 저커버그의 배신(에이콘출판, 2020) 참조.

43) Andrea Simoncini/Erik Longo, "Fundamental Rights and the Rule of Law in the Algorithmic Society", 41쪽. 여기에서 더 나아가 설명가능성과 의사 결정에 대한 인간의 개입(Human-in-the-loop for decisions) 원칙 등의 구체적인 의미를 전달한다는 의미에서 "교육에 의한 보호"까지도 요구된다.

1. 법안의 주요 내용

해당 인공지능법안은 2020년 7월부터 과방위에 발의된 인공지능 관련 법안 7개를 통합한 법안이다(대표발의 의원은 이상민, 양향자, 민형배, 정필모, 이용빈, 윤영찬, 윤두현 의원이다. 통합된 수정안은 아직 공개되어 있지 않다).[44)]

이 법안은 인공지능산업을 진흥하고 인공지능사회의 신뢰기반 조성에 필요한 사항을 규정함으로써 국민의 권익과 존엄성을 보호하고 국민의 삶의 질 향상과 국가경쟁력을 강화하는 데 이바지함을 목적으로 한다고 밝히고 있다(윤두현의원안 제1조). 인공지능산업 육성 도모와 신뢰기반 조성으로 요약된다.

법안의 주요 내용은 다음과 같다.

① 인공지능산업 육성 및 신뢰 확보를 위한 추진체계로, 과학기술정보통신부(이하, '과기정통부'라 함) 장관은 3년마다 인공지능기술 및 인공지능산업의 진흥과 국가경쟁력 강화를 위하여 인공지능 기본계획을 수립·시행, 국무총리 소속으로 심의·의결기관인 인공지능위원회와 그 산하에 전문위원회(인공지능 신뢰성 전문위원회 포함) 설치, 지능정보사회진흥원 산하 국가인공지능센터 설치가 포함된다.

② 인공지능 기술개발 및 산업 육성을 위하여, 인공지능기술 및 알고리즘의 연구·개발 및 인공지능제품 및 서비스 출시 등에 대한 '우선허용·사후규제 원칙', 인공지능기술 개발 및 안전한 이용 지원 사업 실시, 인공지능기술의 표준화, 인공지능 학습용데이터 관련 시책의 수립, 기업의 인공지능기술 도입·활용 지원, 창업 활성화, 인공지능 융합의 촉진, 과기정통부장관의 법령정비 등 제도개선 노력의무, 전문인력의 확보 시책 추진, 국제협력 및 해외시장 진출의 지원, 인공지능집적단지 지정, 대한인공지능협회의 설립 등이 포함된다.

③ 인공지능윤리 및 신뢰성 확보를 위하여, 정부는 인공지능사업자 및 이용자가 인공지능의 개발·이용과정에서 지켜야 할 인공지능 윤리원칙 제정·공표, 신뢰할 수 있는 인공지능 기반조성을 위한 시책 마련(과기정통부장관), 인공지능 신뢰성 검·인증 지원사업 추진(과기정통부장관), 고위험 영역 인공지능의 확인, (제품 또는 서비스 제공자의 이용자에 대한) 고위험 영역 인공지능 고지 의무 등을 규정하였다.[45)]

④ 기타 보칙으로 인공지능산업의 진흥을 위한 재원의 확충 방안 마련, 실태조사, 통계 및 지표의 작성·관리 및 공표, 비밀누설 등에 대한 벌칙 조항, 국가인공지능센터 또는 이와 유사한 명칭 사용한 자에 대한 과태료 부과 조항을 규정하고 있다.

---

44) 심사의견 중, 해당 법안의 구체적인 조문은 적시하고 있지 않다는 의견이 있었다. 해당 법안은 2023. 2. 14. 과방위 법안심사소위를 통과하였다고 알려졌지만, 아직 병합·수정된 이른바 '소위원장 대안'을 공개하지 않고 있다(2023. 6. 기준). 국회는 본회의 상정 직전까지 이를 공개하지 않는데, 이 법안의 사례처럼 상당한 기간 입법정보의 공백을 방치하고 있는 꼴이다. 제도적 개선이 필요한 부분이다.

45) 기타 고위험 영역 인공지능과 관련한 사업자의 책무, 민간자율인공지능윤리위원회의 설치 등도 논의되고 있는 것으로 알려져 있다.

2. 법안의 문제점

이 법안을 둘러싼 논의의 주요 쟁점은 이 법안이 과연 신뢰가능한 인공지능을 위한 제도 설계로 적합한 것인지, 인공지능 기술에 내재되어 있는 위험성을 통제하거나 규제할 수 있는 원칙과 수단을 제공하고 있는지에 모아진다.

첫째, 법안이 인공지능 기술 규제방식으로 채택하고 있는 "우선허용 · 사후규제 원칙"의 문제이다. 이른바 '포괄적 네거티브 규제방식'이다.[46] 인공지능은 신생 기술이므로 산업기반 조성을 위해 규제의 강도를 완화하자는 취지로 이해된다. 이 법안은 인공지능 기술, 제품 또는 서비스가 '국민의 생명 · 안전 · 권익에 위해가 되거나 공공의 안전 보장, 질서 유지 및 복리 증진을 현저히 저해할 우려가 있는 경우'가 아니면 제한할 수 없도록 함으로써 사전규제를 엄격히 제한한다. 현저한 위험이 사전에 확인되지 않는다면, 사전규제는 사실상 불가능하다.

그러나 이러한 "우선허용 · 사후규제 원칙"은 모든 산업에 적용할 수 있는 것이 아니다. 특히, 국민의 안전·생명·건강에 위해가 되거나 환경에 중대한 위협을 초래할 우려가 있는 경우, 개인정보의 안전한 보호 및 처리 여부가 문제되는 경우 네거티브 규제방식의 도입은 부적합하다.[47] 인공지능 기술은 기본적 인권에 직접적이면서도 광범위하게 부정적 영향을 미칠 수 있다. 개인정보자기결정권을 비롯하여 표현의 자유, 프라이버시권, 평등권, 생명권, 안전권 등 거의 모든 인권 목록이 문제된다.[48] 챗GPT와 같은 인공지능 기술이 막대한 전력을 소모하여 기후환경을 파괴한다는 지적도 있다.[49] 인공지능 기술에 잠재한 위험성을 고려한다면 이러한 사후규제방식은 적절한 규제수단이 아니다.

특기할 만한 점은 이러한 규제방식이 기존 ICT 법제의 규제 프레임을 따르고 있다는 것이다. 인공지능법안의 해당 규정은 정보통신융합법이나 지능정보화기본법 등의 내용을 거의 그대로 답습하고 있다.[50] 인공지능 정책은 민간기업 간 경쟁 촉진보다는 공정

---

46) 행정규제기본법 제5조의2에 따르면, 국가나 지방자치단체가 신기술을 활용한 새로운 서비스 또는 제품과 관련된 규제를 법령등이나 조례·규칙에 규정할 때, 네거티브리스트, 포괄적 개념정의, 유연한 분류체계, 그리고 사후규제를 규정한다. 포괄적 네거티브 규제방식에 대하여는 정관선/박균성, "네거티브 규제의 재검토", 법제 제699권(2022), 189쪽 이하.

47) 행정규제기본법 제19조의3 참조.

48) 자율주행차의 보행자 사망 사고, 경비로봇의 유아 공격, 인공지능 스피커의 오주문, 챗봇 이루다의 혐오 발언, 법무부 출입국 인공지능 식별추적 시스템의 얼굴정보 무단 이용, 그리고 챗GPT의 개인정보 및 보안 이슈 등 인공지능의 위험성은 최근까지도 수차례 보고되고 있다.

49) 중앙일보, 2023. 2. 14., "'AI 더러운 비밀'…구글보다 챗GPT'가 지구에 더 나쁜 이유", <https://www.joongang.co.kr/article/25140421>, 검색일: 2023. 5. 1.

50) 정보통신융합법 "제3조의2(우선허용·사후규제 원칙) ① 누구든지 신규 정보통신융합등 기술·서비스를 활용하여 사업을 할 수 있으며, 국가와 지방자치단체는 신규 정보통신융합등 기술·서비스를 활용하는 과정에서 국민의 생명과 안전을 저해하는 경우에 이를 제한할 수 있다." 지능정보화기본법도 유사한 규정을 두고 있다. "제31조(규제 개선 등) ① 누구든지 지능정보기술, 지능정보서비스 및 지능정보기술 제품을 개발·제공·활용할 수 있으며, 정부는 지능정보기술, 지능정보서비스 및 지능정보기술 제품을 개발·제공·활용하는 과정에서 사람의 생명과 안전을 저해하는 경우 등에 한정하여 이를 제한할 수 있다."

성, 윤리성, 투명성, 설명가능성, 견고성, 안전성과 같은 인공지능의 신뢰성을 확보하기 위한 인프라 구축에 중점을 두어야 하며,[51] 이를 위해서는 기술의 전 형성과정에서 규제 거버넌스가 요구된다. 사후규제만으로는 새로운 데이터 인프라 구조 구축이나 인공지능 기술개발 단계 등에서 제기되는 윤리문제 및 위험성에 적절히 대처할 수 없다. 인공지능 정책에는 사전규제에 의한 보호, 설계에 의한 보호, 기본값(default setting)에 보호, 인공지능 리터러시 교육 등 다양한 정책패키지가 각 단계에 합리적으로 배치되어야 한다. 사후규제는 단기적으로 신기술의 시장진출을 용이하게 할 수는 있으나, 규제에 대한 예측가능성이 없어 산업육성에 반드시 유리한 것도 아니다.[52]

둘째, 고위험 영역 인공지능의 정의와 분류 문제이다. 법안은 "사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 영역에서 활용되는 인공지능"을 고위험 영역 인공지능으로 정의한다. 고위험 외에 다른 분류규정은 없다. 그러나 이는 국제적 표준으로 인식되는 유럽연합의 「인공지능법(AI Act)」(안)에서 인공지능이 생성할 수 있는 위험 수준을 네 단계로 구분하고 각 수준에 따라 공급자와 사용자에게 다른 의무를 설정하고 있는 것과 대조를 이루고 있다.

유럽연합의 법안은 인공지능의 위험을 수용불가한 위험, 고위험, 저위험, 최소위험으로 분류하면서, 특히 수용불가한 위험도를 가진 인공지능에 대해서는 그 개발 및 활용을 엄격히 금지한다. 여기에는 잠재의식기술을 사용하는 인공지능 시스템, 사람의 취약성을 공격하는 시스템, 사회적 점수평가(social scoring)에 활용되는 시스템, 공개장소에서 실시간 원격 생체 인식 시스템, 성별, 인종 등 민감한 특성을 사용하는 생체 인식 분류 시스템, 예측 치안 시스템 등이 포함된다. 반면, 우리 법안은 금지되는 인공지능 시스템에 관한 규정 자체가 없다.

유럽연합의 법안은 건강, 안전, 기본권 또는 환경에 대한 위험성을 포함하여 고위험 영역을 설정하고 있으며, 최근 논의를 통해 선거운동 등 정치캠페인에서 유권자에 영향을 미치는 인공지능 시스템과 4,500만 명 이상의 사용자를 보유한 소셜 미디어 플랫폼에서 사용하는 추천 시스템도 고위험군 목록에 추가하였다. 우리 인공지능법안의 경우, 에너지, 먹는물 등의 공급, 보건의료의 제공 및 의료체계, 의료기기, 핵물질과 원자력시설 등에서 사용하는 인공지능을 고위험 영역으로 분류하고 있다.

하지만 사람의 생명, 안전, 기본권에 미치는 중대한 위험성이 충분히 예상되는 영역이 다수 고위험 분류에서 제외되었다는 비판이 제기되고 있다.[53] 특히 생체정보의 분석·활용의 경우 범죄수사나 체포업무라는 일부 업무에 한정하여 고위험 영역으로 분류하고 있어, 이를 금지하는 유럽연합의 입법태도와 대조를 이룬다. 법률상 고위험군으로

---

51) 한상기, 신뢰할 수 있는 인공지능(클라우드나인, 2021) 참조.

52) 정관선/박균성, "네거티브 규제의 재검토", 208쪽.

53) 예를 들어, 유럽연합 법안은 ① 실시간 또는 사후적으로 사람의 생체정보를 활용하여 신원확인 작업을 수행하는 인공지능, ② 지원서 선별, 후보자 평가, 승진결정, 작업 할당, 업무 성과 모니터링 등 인사 관리 업무에 사용되는 인공지능, ③ 직업 훈련 기관의 선정 및 지원 결정, 교육생 및 훈련생 평가에 사용되는 인공지능을 포함하고 있지만, 우리 인공지능법안에서는 제외되었다.

명시되지 않았더라도 대통령령을 통해 사후 고위험으로 규율할 수 있다는 반론도 가능하지만, 충분히 예상되는 위험에 대한 사전의 법률적 규율과 사후의 위임입법적 규율은 규제강도를 달리한다.

또한 유럽연합의 법안은 허용할 수 없는 위험이나 고위험 인공지능시스템이 제재사항을 준수하지 않을 경우 3천만 유로 내지 직전회계연도의 전 세계 연간 총매출액의 6% 중 높은 금액을 과태료로 부과한다. 반면 우리 법안에 따르면, 고위험 영역 인공지능에 대한 사업자의 이용자에 대한 사전고지의무, 사업자의 신뢰성 확보조치와 그에 대한 준수 '권고'만을 규정하여, 고위험을 완화하거나 방지할 수 있는 실효적 장치를 찾아볼 수 없다.

셋째, 인공지능 정책의 거버넌스 문제이다. 법안은 과기정통부를 인공지능 정책의 주관기관으로 설정하고 있다. 인공지능에 관한 폭넓은 정책권한을 부여하고 있는 것이다.

인공지능 기본계획 수립, 인공지능위원회의 간사위원 참여, 국가인공지능센터 설치, 인공지능 기술 개발 활성화 사업, 학습용 데이터 관련 시책 수립, 전문인력 확보, 인공지능 윤리원칙 및 그 실천방안에 대한 권고, 고위험 영역 인공지능의 확인 등이 과기정통부의 권한에 포함된다. 그러나 이 영역 중 상당 부분은 공정거래위원회, 고용노동부, 국가인권위원회, 개인정보보호위원회, 방송통신위원회, 산업통상자원부, 행정안전부, 각 지자체 등 다른 기관의 업무와 중첩된다. 산업진흥에 본연의 가치를 두고 있는 과기정통부가 개인정보보호, 소비자보호, 인권보장, 차별금지, 제품안전 등 광범위한 영역의 가치를 두루 고려하면서 기본계획, 시책, 윤리원칙을 마련할 수 있는지 의문이다.

인공지능법안은 인공지능 등에 관하여 '다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법이 정하는 바에 따른다'(윤두현의원안 제4조)고 규정하는데, 이를 통해 형식적으로 다른 규제기관의 작용을 무력화시키지는 않겠지만, 해당 인공지능법안이 인공지능에 관한 최초의 법률로서 갖는 위상, 이 법에 따라 정립될 과기정통부와 인공지능위원회의 지위와 권한 및 영향력으로 볼 때, 이 인공지능법안이 통과된다면 인공지능 정책에 관한 거버넌스는 과기정통부 중심으로 형성될 것이다. 인공지능 기술의 다면적 성격과 내재적 위험성에 대한 선제적·종합적 대응을 요구하는 환경하에서 기술관료 중심의 거버넌스는 산업편향의 결과를 산출할 가능성이 크다.

3. 향후 과제

우선, 인공지능 기술, 제품, 서비스가 갖는 잠재적 위험성을 예방하고 완화할 수 있는 규제방식을 채택해야 한다. 이와 관련하여 국제기구와 세계 각국은 인공지능에 대한 영향평가제도를 도입하고 있다. 특히 인권실사(HRDD)와 그 핵심도구인 인권영향평가를 인공지능 규제정책에 포함할 필요가 있다. 인공지능 인권영향평가는 인공지능 기술의 잠재적·현실적 위험성을 예방, 완화하기 위한 제도적 대안으로 주목받고 있다. 법안이 추상적으로 선언하는 우선허용·사후규제 원칙은 영향평가제도의 도입을 봉쇄하는 조항

이다.

둘째, 인공지능 사용 영역의 위험도와 심각도에 대한 사회적 합의와 이에 기초한 입법이 필요하다. 유럽연합이 제시한 금지/고위험/제한적/최소의 위험도 구분을 기준으로 할 때, 먼저 우리 사회에서 금지되어야 할 인공지능 활용영역을 규범적으로 확인해야 한다. 헌법적 기준에 따른 위험도 평가가 필요하다. 유럽연합의 법안처럼 잠재의식기술을 활용한 광고행위도 금지되어야 한다. 고위험 영역에 대해서는 광범위한 사회적 합의가 요구된다. 예를 들어, 앞서 언급한 공공장소 원격 실시간 생체 인식(얼굴인식)의 위험도를 어떻게 분류할 것인지, 입학시험, 성적평가, 노동자 모니터링, 플랫폼 노동 등에 활용되는 인공지능의 위험도를 어떻게 평가할 것인지 등 다양한 이슈는 사회적 합의가 필요하며 입법은 이를 반영해야 한다. 공론조사방식도 고려될 수 있다.

셋째, 인공지능 규제와 정책을 주도할 독립적인 기관 설치를 고려해야 한다. 지금 법안이 상정하고 있는 것처럼 과기정통부가 인공지능 정책을 주도한다면, 인공지능 정책은 균형을 잃고 산업편향에 빠질 것이다. 세계적 입법경향에 부합하는 인공지능 거버넌스를 구축하기에 과기정통부는 좁은 틀이다. 2022년 5월 국가인권위원회는 「인공지능 개발과 활용에 관한 인권 가이드라인」을 발표하면서 “인공지능을 독립적이고 효과적으로 감독할 수 있는 체계 수립”을 권고한 바 있다.

4. 소결

인공지능의 위험성에 대한 입법적 대응은 신약에 대한 검증 통제에서 배울 필요가 있다. ‘우선허용·사후규제 원칙’으로는 인공지능의 잠재적 위험성에 대비할 수 없다. 인공지능은 인간 정신의 깊숙한 영역까지 조정할 수 있는 침습적 기술로 활용될 수 있으므로 장단기적 안전성을 검증하기 위한 제도적 장치가 마련되어야 한다.

2020년 유엔사무총장은 인공지능 사용에 대한 책임성 보장을 위한 적절한 법률체계와 감독체계 수립, 피해구제수단 마련을 권고한 바 있고, 2021년 유엔인권최고대표도 인공지능 사용의 부정적 인권영향을 방지·완화하는 인권실사와 규제체계 도입을 권고한 바 있다. 국가인권위원회, 개보위, 금융위, 방통위도 관련 가이드라인이나 규제방안을 제시한 바 있다. 하지만 국회 과방위와 과기정통부가 주도하고 있는 인공지능법안은 이러한 국내외의 입법 경향과 동떨어진 내용을 담고 있다. 이대로 입법되면 국내 인공지능 산업은 갈라파고스화되고, 시민들은 보호장치 없이 인공지능의 위험성에 그대로 노출될 것이다.

Ⅳ. 나오며

인공지능 시스템이 제기하는 문제는 프라이버시나 데이터권에 국한되지 않는다. 인간의 고유한 능력으로 여겨왔던 활동이나 판단능력을 기계적으로 구현할 수 있는 시대가 다가온 것이다. 자동화된 의사 결정 시스템은 민주적 가치에 총체적인 영향을 미칠 수 있다는 점에 주목해야 한다. 민주적 가치를 추구하는 국가와 정치단체는 인공지능 시스템의 한계와 오용 가능성을 파악하고 이에 대비하는 제도와 전략을 준비해야 한다.

인공지능 시스템을 자율규제라는 이름으로 안전장치 없이 방임한다면, 이 시스템을 지배하는 행위자들(테크노크라트와 빅테크의 소유자)이 감시 자본주의와 디지털 자본주의의 논리에 따라 초국가적 차원에서 권리와 자유의 보호 기준을 자의적으로 결정하게 될 것이다. 인권과 민주주의는 기술결정론으로 대체될 것이다. “기술에 의존하는 세상에서는 그 기술에 대해 정통한 정치꾼이 곧 왕이다”.[54]

고도화된 인공지능 시스템을 지속가능한 방식으로 운영하기 위해서는 규제 거버넌스의 역할을 이해하는 것이 중요하다. 공공부문과 민간부문, 기술혁신과 위험 사이에서 균형 있는 방향잡기가 필요하다. 통제되지 않은 인공지능 기술을 활용하여 확증편향으로 점철된 차별적 편견을 조장하고, 유권자의 특정 성향을 프로파일링하여 가짜 뉴스를 세뇌될 때까지 발신할 수도 있다. 인간의 특정한 반응을 ‘유도’하는 기술도 날로 발전하고 있다. 오늘날 인공지능 기술의 지배자인 빅테크나 디지털 기업이 헌법적 한계를 초월하는 권력을 획득하도록 방치할 수는 없다. 인공지능 시대의 빅테크 기업에 대한 규제 거버넌스는 일순위 의제로 논의되어야 한다.

과거 산업 비즈니스 모델에 입각한 디지털 정책의 관행적 시행에서 벗어나 인공지능 기술이 생성하는 제도적 환경에 적응하는 인공지능 정책을 형성해야 한다. 현재 국회에서 계류 중인 인공지능법안은 기존 ICT 법제의 규제방식을 그대로 채용하고 있으며, 인공지능 법제가 풀어야 할 주요 쟁점인 인공지능 위험도 분류 문제를 별다른 사회적 합의나 뚜렷한 기준 없이 ‘고위험 영역’이라는 단일 척도로 제도화하고자 한다. 고위험 인공지능을 규제하고 금지할 실효성 있는 제도적 장치는 몰각한 채 과기정통부라는 산업부처 주도로 입법절차가 진행되고 있다. 현재 해당 인공지능법안은 소관위 전체회의 의결을 앞두고 있다.

세계적으로 인공지능에 대한 규제정책은 일반 산업이나 디지털 및 인터넷 경제에 대응하는 정책적 경향과 차별화되고 있다. 인공지능 기술, 제품, 서비스가 갖는 잠재적 위험성을 경고하면서 인공지능 개발자와 기업이 스스로 사전 규제의 필요성을 주장하는 역설적 상황도 펼쳐지고 있다. 고도 기술로서 인공지능의 지속가능한 운영을 위해서는 세계적 입법경향에 부합하는 규제 거버넌스가 구축되어야 한다. ▢

54) 데이비드 런시먼, 최이현 옮김, 쿠데타, 대재앙, 정보권력(아날로그, 2020), 170쪽.

<참고문헌>

이광석, “포스트-판옵티콘 시대 감시 연구, 새로운 지형”, 김동욱 외 엮음, 스마트 시대의 위험과 대응방안, 나남출판, 2015, 189-207쪽.

인남식, “아랍 민주화 운동과 미국의 대중동정책 변화 연구”, 서정민/인남식 엮음, 중동 민주화의 대내외 정치역학, 대외경제정책연구원, 2011, 73-148쪽.

정관선/박균성, “네거티브 규제의 재검토”, 법제 제699권, 2022, 189-213쪽.

정병기, “오성운동(M5S)의 직접 의회주의와 사이버크래틱 집중주의: 포스트포퓰리스트 정치 운동의 성공과 한계”, 한국정치연구 제29권 제2호, 2020, 91-116쪽.

홍남희, “디지털 플랫폼에 의한 ‘사적 검열(private censorship)’”, 미디어와 인격권 통권 제6호, 2018, 135-172쪽.

김진우, 나의 첫 인공지능 수업, 메이트북스, 2022.

박승일, 기계, 권력, 사회: 인터넷은 어떻게 권력이 되었는가, 사월의책, 2021.

방송통신위원회/한국인터넷진흥원, 개인정보 처리에서의 프로파일링 사례집, 2020.

유승익 외, 인공지능 인권영향평가 도입 방안 연구, 국가인권위원회, 2022.

한홍구 외, 감시사회, 철수와영희, 2021.

데이비드 런시먼, 최이현 옮김, 쿠데타, 대재앙, 정보권력, 아날로그, 2020.

로저 맥나미, 김상현 옮김, 마크 저커버그의 배신, 에이콘출판, 2020.

롭 라이히 외, 이영래 옮김, 시스템 에러: 빅테크 시대의 윤리학, 어크로스, 2022.

브리태니 카이저, 고영태 옮김, 타겟티드, 한빛비즈, 2020.

빅토르 마이어 쇤버거/케니스 쿠키어, 이지연 옮김, 빅 데이터가 만드는 세상, 21세기북스, 2013.

쇼샤나 주보프, 김보영 옮김, 감시 자본주의 시대, 문학사상, 2021.

스튜어트 러셀/피터 노빅, 류광 옮김, 인공지능: 현대적 접근방식, 제3판, 제이펍, 2016.

시바 바이디야나단, 홍권희 옮김, 페이스북은 어떻게 우리를 단절시키고 민주주의를 훼손하는가, 아라크네, 2020.

제프리 삭스, 이종인 옮김, 지리, 기술, 제도, 21세기북스, 2021.

지그문트 바우만/데이비드 라이언, 한길석 옮김, 친애하는 빅브라더, 오월의봄, 2014.

프랭크 웹스터, 조동기 옮김, 현대 정보사회 이론, 나남출판, 2016.

헨리 키신저/에릭 슈밋/대니얼 허튼로커, 김고명 옮김, AI 이후의 세계, 윌북, 2023.

Castets-Renard, Céline, “Human Rights and Algorithmic Impact Assessment for Predictive Policing”, Hans-W. Micklitz 외 엮음, *Constitutional Challenges In The Algorithmic Society*, Cambridge University Press, 2022, 93-110쪽.

Danaher, John, “The Threat of Algocracy: Reality, Resistance and Accommodation”, *Philosophy & Technology*, vol. 29, no. 3, 2016, 245-268쪽.

Diamond, Larry, “Liberation Technology”, *Journal of Democracy*, vol. 21, no. 3, 2010, 69-83쪽.

Newell, Sue/Marabelli, Marco, “Strategic Opportunities (and Challenges) of Algorithmic

Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'", *Journal of Strategic Information Systems*, vol. 24, no. 1, 2015, 3-14쪽.

Pollicino, Oreste/De Gregorio, Giovanni, "Constitutional Law in the Algorithmic Society", Hans-W. Micklitz 외 엮음, *Constitutional Challenges In The Algorithmic Society*, Cambridge University Press, 2022, 3-24쪽.

Pötzsch, Holger, "Archives and Identity in the Context of Social Media and Algorithmic Analytics: Towards an Understanding of iArchive and Predictive Retention", *New Media & Society*, vol. 20, no. 9, 2018, 3304-3322쪽.

Ram, Natalie/Gray, David, "Mass Surveillance in the Age of COVID-19", *Journal of Law and the Biosciences*, vol. 7, no. 1, 2020, 1-17쪽.

Richards, Neil M., "The Dangers of Surveillance", *Harvard Law Review*, vol. 126, no. 7, 2013, 1934-1965쪽.

Simoncini, Andrea/Longo, Erik, "Fundamental Rights and the Rule of Law in the Algorithmic Society", in Hans-W. Micklitz 외 엮음, *Constitutional Challenges In The Algorithmic Society*, Cambridge University Press, 2022, 27-41쪽.

Walorska, Agnieszka M., "The Algorithmic Society", Denise Feldner 엮음, *Redesigning Organizations Concepts for the Connected Society,* Springer, 2020, 149-160쪽.

Lyon, David, *Surveillance After September 11*, Polity, 2003.

Petit, Nicolas, *Big Tech and the Digital Economy: The Moligopoly Scenario*, Oxford University Press, 2020.

Schneier, Bruce, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*, W. W. Norton & Company, 2015.

[보론] 인공지능 규제정책과 인공지능법안의 쟁점*

1. 들어가며

최근 미국의 생명미래연구소는 인공지능 개발을 6개월 이상 중단해야 한다는 성명을 발표한 바 있다. 인공신경망 연구로 인공지능 기술을 개척한 제프리 힌튼(Geoffrey Hinton) 교수도 최근 구글을 나오며 인공지능에 잠재한 위험성을 경고하였다. 국내외에서 인공지능의 위험성과 그 규제에 대한 요구가 커지고 있다.

주요 국가와 국제기관에서 인공지능 기술을 규제하기 위해 여러 제도적 장치를 마련하고 있다. 유럽연합 집행위원회 산하 주요 의원 위원회는 2023. 5. 11.「인공지능법(AI Act)」(안)을 통과시켰으며, 연내 시행을 목표로 입법절차를 진행하고 있다. 이 법안은 인공지능에 내재한 위험에 따른 사전 규제 및 생성AI가 만든 컨텐츠에 대한 표기의무 등을 규정하고 있다. 미국 의회에서도 보안, 책임, 투명성을 높여 인간 통제를 벗어나는 인공지능 기술의 오용 가능성을 낮추는 것을 목표로 사전 규제안을 마련하기 위한 협의체가 구성되었다.[1)]

우리 국회는 지난 2. 14. 과학기술정보방송통신위원회의 법안소위에서「인공지능산업 육성 및 신뢰 기반 조성에 관한 법률안」(이하 '인공지능법안'이라 함)을 통과시켰다. 이 법안이 통과된다면 국내 최초의 인공지능에 관한 단일 법안이 될 것이다. 그러나 이 인공지능법안이 포함하는 규제의 형식과 내용에 대해서는 여러 비판이 제기되고 있다. 과기정통부 주관의 인공지능 기술에 대한 규제정책이 산업진흥에 편중되어 인공지능 기술에 내재하는 규제적 특성을 간과한다는 것이다.

이하에서는 인공지능 기술의 발전에 따른 규제정책의 함의를 기존 신경제 하의 디지털 정책과 비교하여 제시하고, 현재 우리 국회에서 논의되고 있는 인공지능법안의 쟁점과 향후 과제를 살펴보고자 한다.

2. 인공지능 기술 발전에 따른 정책 패러다임 변화 : 디지털 정책에서 인공지능 정책으로

인공지능 시스템의 비약적 발전은 관련 정책패러다임(규제정책 포함)에도 큰 변화를 가져올 것으로 예상된다. 1990년대 후반 인터넷의 확산이 가져온 경제적 환경 변화에 기업과 사회가 일정하게 적응한 것처럼, 인공지능의 발전도 새로운 정책적 과제를 제기

---

* 유승익, "인공지능 규제정책과 인공지능법안의 쟁점", KIPA 규제동향 통권 제42호(한국행정연구원, 2023), 28-32쪽을 전제한 글임.

1) 아시아경제, "美 첫 AI 규제법 연내 채택되나…대선 앞두고 의회 법률화 속도" 2023. 6. 5. 보도

하고 있다. 챗GPT, 바드(Bard) 등 생성형 AI 기술이 보여주고 있는 바와 같이, 이번 인공지능 기술혁신의 물결은 과거 '신경제'(New Economy)와는 다른 정책환경을 조성하고 있다.[2)]

첫째, 과거 디지털 정책이 디지털 인프라구조 구축(민간 기업의 혁신, 경쟁 촉진과 지원 등)에 주된 관심을 두었다면, 인공지능 정책은 '데이터 인프라구조 구축' 그리고 '기술개발 과 활용에서의 윤리 문제'에 초점을 둔다. 인공지능 시스템의 관건은 대량의 고품질 데이터('데이터는 인공지능의 연료')라는 점에서 새로운 데이터 인프라구조의 구축이 요구된다. 또한 인공지능의 도입과 활용에서 인권의 문제를 포함한 법적·윤리적 문제가 선행적으로 고려되어야 한다. 데이터 거버넌스와 알고리즘에 의한 자동화된 의사결정은 필연적으로 법적·윤리적 문제를 제기하기 때문이다. 따라서 인공지능 정책은 선제적·사전적 규제방식과 친화적이다.

둘째, 디지털 혁신에서 주된 정책적 이슈가 '생산성 역설'(productivity paradox)[3)]이라면, 인공지능 정책에서는 인공지능의 '분류'(AI classification)가 주로 문제된다. 디지털 혁신 초기, 무형자산을 통계상 집계하지 못하는 문제가 디지털 정책 수립에 장애로 작용했다(예산 수립, 투자정책 등). 이에 반해, 인공지능 정책에서는 전통적인 척도를 통해 인공지능 부문을 분류하는 것이 불가능하다는 점이 주로 문제된다. 인공지능은 하나의 산업이 아니라, 여러 산업에 걸쳐 있는 범용기술이기 때문이다. 따라서 인공지능 정책에서는 산업별 분류보다 기술별 분류가 주된 분류방법으로 활용된다. 예를 들어, 인공지능 기술은 머신러닝, 데이터분석, 챗봇, AI 비서, 언어 처리, 음성인식, 예측분석, 얼굴 및 감정 인식 등으로 분류될 수 있다. 유럽연합의 인공지능법안은 위험기반접근법(risk-based approach)에 따라 4단계로 인공지능을 분류하고 있다(수용불가/고위험/제한/최소). 이는 인공지능 정책의 거버넌스가 하나의 산업부처에 국한될 수 없음을 말해주기도 한다.

셋째, 디지털 정책의 주요 수용자는 e-비즈니스 등 민간 부문이지만(기존 출판, 방송, 미디어 기업과 인터넷 서비스 제공업체의 경쟁), 인공지능 정책은 공적 목적으로 활용되는 분야(교육, 보건, 국가안보, 환경 등)를 대상으로 우선적으로 수립되는 경향을 보인다.

넷째, 디지털 정책은 수요견인형(demand-pull) 혁신을 대상으로 한다면, 인공지능 정책은 기술주도형(tech-push) 혁신을 대상으로 한다. 전자의 경우, 최종 사용자와 소비자의 적극적이고 광범위한 참여가 기술혁신을 이끌어내고, 산업 비즈니스 모델에 따른 디지털 정책이 후속하여 이러한 혁신을 지원하는 과정을 거친다면, 후자의 경우, 일정한 거버넌스 하에서 빅테크나 혁신형 창업기업(start-up) 등에 속한 연구자의 연구개발

---

2) Birgitte Andersen, Public Policy and Government, C. Kerrigan(ed.), Artificial Intelligence: Law and Regulation, Edward Elgar, 2022, p. 442 이하 참조.

3) ICT에 대한 투자가 증가함에도 불구하고 기업, 산업 및 국가 수준의 생산성이 비례해서 증가하지 않거나 오히려 감소하는 현상을 말한다. 정선영, 디지털 혁신과 우리나라의 생산성 역설, BOK 이슈노트 No.2021-20, 한국은행, 2021, 2쪽 이하.

을 통해 공급자 중심으로 선도적 기술혁신이 일어나고 이에 후속하여 산업계와 최종사용자의 도입 및 채택이 이루어지는 구조를 갖는다. 따라서 인공지능 정책에서는 디지털 정책과 다르게 윤리 및 데이터 거버넌스가 기술개발의 단계에서부터 강조될 수밖에 없다.

<표> 인터넷 정책과 인공지능 정책의 차이[4)]

	디지털 정책	인공지능 정책
정책 포커스	디지털 인프라구조 - 기업 혁신, 경쟁 등을 지원하기 위한 디지털 인프라 구축	데이터 인프라와 윤리 - 새로운 데이터 인프라구조(데이터는 인공지능의 '연료') - 인공지능의 도입과 활용에 대한 윤리적 고려
정책적 이슈	생산성 역설 - 기술발전과 생산성의 괴리현상	인공지능의 분류 - 범용기술로서 인공지능 - 정책적 결정과 경제적 분석에 적합한 최선의 분류방법
주요 수용자	민간 분야 - e-비즈니스 분야(출판, 방송, 미디어 기업과 인터넷 서비스 제공업체(ISP))	공공 분야 - 교육, 보건, 국가안보, 환경 서비스와 같은 공적 목적을 위한 부문
수용의 확산	최종사용자/소비자 중심 - 수요 견인형(산업 비즈니스 모델 및 정책이 후속됨)	공급자/과학 중심 - 기술 견인형(산업계의 도입 및 최종 사용자/소비자 채택이 후속됨)

3. 「인공지능법안」의 쟁점과 문제점

이상에서 살펴본 바와 같이, 인공지능 기술은 규제정책에도 일정한 변화를 요구한다. 따라서 현재 국회에 계류 중인 인공지능법안이 변화하는 정책환경을 고려하여 적합하게 설계되어 있는지 문제 된다. 과거 디지털 정책의 제도적 관성을 넘어 인공지능의 특성을 반영하는 규제정책을 채택하고 있는지 살펴본다.

첫째, 인공지능법안은 '우선허용·사후규제 원칙'을 규제방식으로 채택하고 있다(법안 제11조).[5)] 인공지능기술의 연구·개발 및 서비스 출시 등을 초기 시장으로 보고 규제의 강도를 대폭 축소, 완화하고 있다. 해당 규정에 따르면, 인공지능 기술, 제품 또는 서비

4) 이 표는 Birgitte Andersen, 위의 글, p. 444를 변형한 것이다.

5) 이른바 '포괄적 네거티브 규제방식'을 채택한 것으로 이해된다. 정관선, 박균성, 네거티브 규제의 재검토, 법제 제699권, 2022, 189쪽 이하 참조.

스가 국민의 생명·안전·권익에 위해가 되거나 공익 등을 현저히 저해할 우려가 있는 경우가 아니라면 이를 제한할 수 없도록 함으로써 기술개발, 제품출시, 서비스의 전 과정에서 사전규제를 엄격히 제한하고 있다.

특기할만한 점은 이러한 규제방식이 기존 ICT 법제의 규제 프레임을 따르고 있다는 것이다. 인공지능법안의 해당 규정은 정보통신융합법이나 지능정보화기본법 등의 내용을 거의 그대로 답습하고 있다.[6] 인공지능 정책은 민간기업 간 경쟁 촉진보다는 공정성, 윤리성, 투명성, 설명가능성, 견고성, 안전성과 같은 인공지능의 신뢰성을 확보하기 위한 인프라 구축에 중점을 두어야 하며,[7] 이를 위해서는 기술의 전 형성과정에서 규제 거버넌스가 요구된다. 사후규제만으로는 새로운 데이터 인프라구조 구축이나 인공지능 기술개발 단계 등에서 제기되는 윤리문제 및 위험성에 적절히 대처할 수 없다. 인공지능 정책에는 사전규제에 의한 보호, 설계에 의한 보호, 기본값(default setting)에 보호, 인공지능 리터러시 교육 등 다양한 정책패키지가 각 단계에 합리적으로 배치되어야 한다. 우선허용·사후규제 방식은 규제에 대한 예측가능성을 떨어뜨려 산업진흥에 반드시 유리한 방식도 아니다.[8]

둘째, 인공지능 분류 문제이다. 인공지능법안은 "사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 영역에서 활용되는 인공지능"을 '고위험영역 인공지능'으로 정의하여 분류하고 있다. 이 외에 다른 분류 규정은 존재하지 않는다. 이는 유럽연합의 「인공지능법(안)」에서 인공지능이 생성할 수 있는 위험 수준을 네 단계로 구분하고 각 수준에 따라 공급자와 사용자에게 다른 의무를 설정하고 있는 것과 대조를 이루고 있다.

유럽연합의 법안은 인공지능의 위험을 수용불가한 위험, 고위험, 저위험, 최소위험으로 분류하면서, 특히 수용불가한 위험도를 가진 인공지능에 대해서는 그 개발 및 활용을 엄격히 금지한다. 여기에는 잠재의식기술을 사용하는 인공지능 시스템, 사람의 취약성을 공격하는 시스템, 사회적 점수평가(social scoring)에 활용되는 시스템, 공개장소에서 실시간 원격 생체 인식 시스템, 성별, 인종 등 민감한 특성을 사용하는 생체 인식 분류 시스템, 예측 치안 시스템 등이 포함된다. 반면, 우리 인공지능법안은 '금지되는 영역'에 대한 범주 자체를 상정하고 있지 않다.

유럽연합의 법안은 건강, 안전, 기본권 또는 환경에 대한 위험성을 포함하여 고위험 영역을 설정하고 있으며, 최근 논의를 통해 선거운동 등 정치캠페인에서 유권자에 영향

---

6) "제3조의2(우선허용·사후규제 원칙) ① 누구든지 신규 정보통신융합등 기술·서비스를 활용하여 사업을 할 수 있으며, 국가와 지방자치단체는 신규 정보통신융합등 기술·서비스를 활용하는 과정에서 국민의 생명과 안전을 저해하는 경우에 이를 제한할 수 있다." 지능정보화기본법도 유사한 규정을 포함하고 있다.
"제31조(규제 개선 등) ① 누구든지 지능정보기술, 지능정보서비스 및 지능정보기술 제품을 개발·제공·활용할 수 있으며, 정부는 지능정보기술, 지능정보서비스 및 지능정보기술 제품을 개발·제공·활용하는 과정에서 사람의 생명과 안전을 저해하는 경우 등에 한정하여 이를 제한할 수 있다."

7) 한상기, 신뢰할 수 있는 인공지능, 클라우드나인, 2021 참조.

8) 정관선, 박균성, 위의 논문, 208쪽.

을 미치는 인공지능 시스템과 4,500만 명 이상의 사용자를 보유한 소셜 미디어 플랫폼에서 사용하는 추천 시스템도 고위험군 목록에 추가되었다. 우리 인공지능법안의 경우, 에너지, 먹는물 등의 공급, 보건의료의 제공 및 의료체계, 의료기기, 핵물질과 원자력시설 등에서 사용하는 인공지능을 고위험영역으로 분류하고 있다.

하지만 사람의 생명, 안전, 기본권에 미치는 중대한 위험성이 충분히 예상되는 영역이 다수 고위험 분류에서 제외되었다는 비판이 제기되고 있다.[9] 법률상 고위험군으로 명시되지 않았더라도 대통령령을 통해 사후 고위험으로 규율할 수 있다는 반론도 가능하지만, 충분히 예상되는 위험에 대한 사전의 법률적 규율과 사후의 위임입법적 규율은 규제강도를 달리한다.

셋째, 규제의 실효성 확보 문제이다. 유럽연합의 법안은 수용불가한 위험이나 고위험 인공지능에 관한 의무사항을 준수하지 않을 경우 3천만 유로 내지 직전사업연도의 전 세계 연간 총매출액 6% 중 높은 금액을 과징금으로 부과할 수 있다. 반면, 우리 인공지능법안에 따르면, 고위험영역 인공지능에 대한 사업자의 이용자에 대한 사전고지의무, 사업자의 신뢰성 확보조치와 그에 대한 준수 '권고' 등 미약한 수준의 의무를 부과하여, 실효적 제재를 위한 근거규정을 마련하고 있지 않다.

넷째, 인공지능 정책을 소관하는 거버넌스의 문제이다. 인공지능법안은 과기정통부에 인공지능 정책에 관한 폭넓은 권한을 부여하고 있다. 이 인공지능법안에 따르면, 인공지능 기본계획 수립, 국무총리 산하 인공지능위원회의 간사위원 참여, 국가인공지능센터 설치, 인공지능기술 개발 활성화 사업, 학습용 데이터 관련 시책 수립, 전문인력 확보, 인공지능 윤리원칙 및 그 실천방안에 대한 권고, 고위험영역 인공지능의 확인 등이 과기정통부의 권한에 속한다.

그러나 앞서 살펴본 바와 같이, 인공지능 정책은 디지털 정책과 다른 거버넌스를 요구한다. 기술관료 중심의 과기정통부가 인권, 윤리, 공정성, 신뢰가능성, 헌법원칙 등의 가치들을 두루 고려하면서 설계, 기술개발을 포함한 기술의 전체 형성과정, 출시, 서비스 등을 관장할 수 있는지 검토가 필요하다. 해당 업무 중 상당 부분이 공정거래위, 고용노동부, 국가인권위, 개인정보보호위, 방송통신위, 산업부, 행안부, 각 지자체 등 다른 기관의 기능과 중첩된다. 해당 인공지능법안은 인공지능 등에 관하여 '다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법이 정하는 바에 따른다'(법안 제4조)고 규정하는데, 이를 통해 형식적으로 다른 규제기관의 작용을 무력화시키지는 않겠지만, 해당 인공지능법안이 인공지능에 관한 최초의 법률로써 갖는 위상, 이 법에 따라 정립될 과기정통부와 인공지능위원회의 지위와 권한 및 영향력으로 볼 때, 이 인공지능법안이 통과된다면 인공지능 정책에 관한 거버넌스는 과기정통부 중심으로 형성될 것이다. 인공

9) 예를 들어, 유럽연합 법안은 ① 실시간 또는 사후적으로 사람의 생체정보를 활용하여 신원확인 작업을 수행하는 인공지능, ② 지원서 선별, 후보자 평가, 승진결정, 작업 할당, 업무 성과 모니터링 등 인사 관리 업무에 사용되는 인공지능, ③ 직업 훈련 기관의 선정 및 지원 결정, 교육생 및 훈련생 평가에 사용되는 인공지능을 포함하고 있지만, 우리 인공지능법안에서는 제외되었다.

지능 기술의 다면적 성격과 내재적 위험성에 대한 선제적·종합적 대응을 요구하는 환경 하에서 기술관료 중심의 거버넌스는 산업편향의 결과를 산출할 가능성이 크다.

4. 나오며

우리나라 인공지능 기술이 세계적 흐름에 뒤처지지 않도록 제도적 장치를 시급히 마련할 필요가 있다. 이를 위해서는 과거 산업 비즈니스 모델에 입각한 디지털 정책의 관행적 시행에서 벗어나 인공지능 기술이 생성하는 제도적 환경에 적응하는 인공지능 정책을 형성해야 한다. 그러나 현재 국회에서 계류 중인 인공지능법안은 기존 ICT 법제의 규제방식을 그대로 채용하고 있으며, 인공지능 법제가 풀어야 할 주요 쟁점인 인공지능 위험도 분류 문제를 별다른 사회적 합의나 뚜렷한 기준 없이 '고위험영역'이라는 단일 척도로 제도화하고자 한다. 또한 고위험 인공지능을 규제하고 금지할 실효성 있는 제도적 장치는 몰각한 채 과기정통부라는 산업부처 주도로 입법절차가 진행되고 있다. 현재 해당 인공지능법안은 소관위 전체회의 의결을 앞두고 있다.

세계적으로 인공지능에 대한 규제정책은 일반 산업이나 디지털 및 인터넷 경제에 대응하는 정책적 경향과 차별화되고 있다. 인공지능 기술, 제품, 서비스가 갖는 잠재적 위험성을 경고하면서 인공지능 개발자와 기업이 스스로 사전 규제의 필요성을 주장하는 역설적 상황도 펼쳐지고 있다. 유엔사무총장과 유엔인권최고대표도 인권실사나 영향평가를 포함한 적절한 감독체계와 규제방안 도입을 권고하고 있다. 고도 기술로서 인공지능의 지속가능한 운영을 위해서는 세계적 입법경향에 부합하는 규제 거버넌스가 구축되어야 한다. ▢

유럽연합 인공지능법의 내용 및 시사점

허진민

(참여연대 공익법센터 소장, 변호사)

Ⅰ. 인공지능 규제의 필요성

인공지능 기술의 지속적인 발전과 동시에 인공지능기술의 위험성도 현실화되고 있습니다. 신문기사에서 인공지능기술이 양극화, 편향성, 불확실성, 사생활 침해 등 정치·사회·경제적으로 부정적 영향들을 미칠 수 있음을 쉽게 접할 수 있습니다. 인공지능 기술은 개발된 상태에 머무르지 않고 지속적으로 변화하는 성질을 가지고 있는 점에서 신뢰성 확보와 안전성에 대한 검증은 지속적으로 요구될 수밖에 없고, 인공지능기술의 발전 속도와 사회 대부분의 영역에서 그 적용이 확대될 것이 예견되는 점을 고려하면 기존의 법률 규제만으로는 적정하게 대응하기 어려운 측면이 있습니다. 그리고 인공지능기술이 향후 사회경제적으로 미치는 파급효를 고려할 때 산업진흥의 측면에서 기술개발 촉진을 위한 규제 완화 등의 논의가 필요함과 동시에 인공지능기술의 잠재적 위험성을 고려한 규제에 대한 논의도 필요합니다. 또한 인공지능과 관련된 문제는 동시대적 문제로 세계 각 국가가 동일하게 직면하는 문제라는 점에서 다른 나라의 입법 논의는 국내에서의 인공지능 관련 법안을 입안함에 있어 충분히 고려되어야 합니다.

그러므로 인공지능 관련 법안은 인공지능기술의 발전 및 관련 산업의 육성이라는 목적과 인공지능기술의 잠재적 위험성을 최소화하고 신뢰성 확보라는 목적을 조화롭게 달성할 수 있는 내용이 포함되어야 합니다. 이러한 점에서 최근 EU 의회에서 채택된 EU 인공지능법의 내용은 우리가 인공지능 관련 법안을 제정함에 있어 반드시 참고해야 할

것으로 생각됩니다.

발표문에서는 EU 인공지능법의 내용을 살펴보고 현재 국회에서 논의되고 있는 인공지능법안의 문제점을 살펴보고자 합니다.

Ⅱ. EU 인공지능법의 제정까지의 경과와 EU 인공지능법의 내용

1. EU 인공지능법의 제정까지의 경과

EU는 장클로드 융커의 EU 집행위원회회가 2015년 5월 발표한 '디지털 단일시장 전략(Digital Single Market Strategy)'을 발표한 이후, 인공지능과 관련하여 EU집행위원회는 2018년 4월 유럽의 인공지능 전략(AI for Europe) 수립하고, 2018년 12월 인공지능 합동계획(Coordinated plan on AI) 발표하였으며, 2019년 4월, 신뢰할 수 있는 인공지능 윤리 가이드라인을 발표하였고, 2020년 2월 인공지능 백서(White paper on AI) 발표한 이후 회원국, 전문가, 기업, 학계, 시민사회 등 각계의 의견 수렴을 거쳐 2021년 4월 EU 인공지능법안을 발의하였습니다.

아래에서는 인공지능법안의 토대가 된 신뢰할 수 있는 인공지능 윤리기준 가이드라인과 인공지능백서의 주요 내용을 살펴보고자 합니다.

가. 신뢰할 수 있는 인공지능 윤리 가이드라인의 주요내용

EU 집행위원회는 2018년 유럽 인공지능 전략(Artificial Intelligence for Europe)을 채택하고, 후속 조치로 2018년 6월 '학계, 시민사회, 산업계를 대표하는 52인의 전문가로 구성된 인공지능 고위 전문가 그룹(High-Level Expert Group)을 발족하였으며, 인공지능 고위 전문가 그룹은 2019년 4월 신뢰할 수 있는 인공지능 윤리 가이드라인(Ethics Guidelines for Trustworthy Artificial Intelligence)을 발표하였습니다.

신뢰할 수 있는 인공지능은 합법적이고, 윤리적이며, 견고하여야 한다는 3가지 구성요소를 갖추어야 하며, 3가지 구성 요소는 시스템의 전주기를 통해 만족되어야 하는 것으로 정의하였습니다. 즉, 합법적이라 함은 기본권을 근거로 모든 적용 가능한 법과 규칙을 준수해야 함을, 윤리적이라 함은 윤리적 원칙과 가치를 존중해야 함을, 견고함은 인공지능 시스템이 예측하지 못한 피해를 일으킬 수 있음을 전제로 기술적·사회적 관점에서 강건하고 견고해야 함을 의미합니다.

또한 신뢰할 수 있는 인공지능의 구현 및 실현에 대한 7가지 요구사항을 제시하였는데, 그 내용은 ① 인간의 개입과 감독, ② 기술적 견고성과 안전성(공격 및 보안에 대한 복원력, 만일에 대한 계획 및 일반 안전, 정확성, 신뢰성 및 재현성), ③ 개인 정보

보호 및 데이터 거버넌스(프라이버시 존중, 데이터의 품질 및 무결성, 데이터 접근), ④ 투명성(추적성, 설명성 및 커뮤니케이션), ⑤ 다양성, 차별 금지 및 공정성(불공정한 편견 방지, 접근성 및 유니버설 디자인, 이해 관계자 참여), ⑥ 사회적·환경적 복지(지속 가능성 및 환경 친화성, 사회적 영향), ⑦ 책무(감사 가능성, 부정적인 영향 최소화 및 보고, 균형 및 보상) 입니다.

비록 인공지능 윤리 가이드라인은 법적 효력은 없지만 합법적 인공지능의 실질이나 인공지능 규제 정립을 위한 구체적 지침보다는 사회공학적 체계 내에서 윤리원칙들이 작동할 수 있는 지침을 제공하여 당시 인공지능에 관한 규제가 거의 부재한 현실에서 그 기반이 될 수 있는 사회적 규범을 형성하는 길잡이가 되었다는 점에서 의미를 가지고 있습니다.[1)]

나. 인공지능 백서의 주요내용

EU 집행위원회는 2020년 2월 19일 5년간 유럽연합 디지털 정책의 방향과 대원칙을 밝히는 세 가지 정책 문서, “인공지능 백서”, “유럽 데이터 전략”, “유럽의 디지털 미래”를 발표하였습니다. 인공지능 백서는 인공지능 정책에 관한 제안을 담은 것으로 법적 효력은 없지만 EU 집행위원회의 향후 정책 방향을 제시하는 성격을 가집니다.[2)]

인공지능 백서는 인공지능 이용에 대한 대중, 이해 관계자, 유럽 의회, 유럽 이사회 등의 토론을 위한 자료로, 인공지능의 도입을 촉진하는 동시에 인공지능 활용에 따르는 위험을 해결하기 위한 다양한 정책 방안을 제시하고 있는데, 인공지능의 "수월성(秀越性)[3)] 생태계(ecosystem of excellence)" 구축과 관련하여 민간-공공 파트너십을 촉진하기 위한 정치적·기술적 측면을 제시하고, 인공지능의 "신뢰성 생태계(ecosystem of trust)" 구축을 위한 향후 규제 프레임워크의 핵심 요소를 제안하였습니다.[4)]

수월성 생태계 확보 방안과 관련하여 다양한 인공지능 산업 지원 방안들을 제시하고 있으며, 특히 디지털 혁신 허브(Digital Innovation Hubs)를 통한 중소기업의 인공지능 활용 지원 및 중소기업 간의 협력을 촉진하는 방안을 제안하였습니다.

신뢰성 확보를 위한 생태계 구축과 관련하여 인공지능의 잠재적 리스크를 두 가지로 나누었는데, 첫째, 인공지능 시스템의 설계상 결함이나 편향된 데이터 사용으로 인한 데이터 프라이버시를 비롯하여 표현의 자유, 집회의 자유, 인간 존엄, 성·인종·종교·장

---

1) 윤혜선, 인공지능 규제 정책에 관한 연구: 주요국의 규제 정책 사례를 중심으로, 정보통신정책연구 제26권 제4호(2019. 12) 160쪽

2) 이상윤, 유럽연합 디지털 정책의 동향과 전망: “유럽의 디지털 미래” · “유럽 데이터 전략” ·“인공지능 백서”의 주요 내용과 의의, 고려법학 제97호(2020. 6.) 193~194쪽

3) 우수한 생태계라는 표현으로 번역되기도 합니다.

4) EU 인공지능 백서(White Paper on Artificial Intelligence) 주요 내용 분석, 한국인터넷진흥원(2020. 4.) 3쪽

애·연령·성적 지향에 따른 차별 금지, 효과적인 사법 구제와 공정한 재판을 받을 권리, 소비자 권리 등 인간의 기본적 권리들이 침해당할 수 있는 불확실성을 의미하는 무형의 리스크, 둘째, 인공지능 기술의 결함으로 말미암아 인명, 재산상에 직접적인 피해가 발생할 불확실성을 의미하는 유형의 리스트로 구분하고 있습니다.[5] 그리고 이러한 인공지능 리스크들은 현재 수준의 규제만으로는 한계가 있으므로[6] 새로운 규제 체계의 핵심요소를 제안하였습니다.

인공지능 리스크에 대한 새로운 규제 체계는 비례성 원칙에 따라 리스크 기반 접근 방식(risk-based approach)에 기초한 새로운 규제 체계 도입의 필요성을 제시하면서 위험의 정도에 따른 선별적인 대응 방안이 필요하고, 특히 고위험군 인공지능(high-risks ai applications)의 기준을 제시하여 고위험군 인공지능의 대응은 새로운 법률 제정이 필요하다고 제안하였습니다.

그리고 고위험군 인공지능의 판단 기준, 고위험군 인공지능의 규제를 위한 법률 제정 시 요구될 수 있는 사항, 고위험군 인공지능의 규제를 위한 법률의 수범자, 고위험군 인공지능을 적용하는 모든 공급자에 대한 사전 적합성 평가제도 등을 제안하고 있습니다.

고위험군 인공지능의 판단기준으로 중대한 위험이 발생할 것으로 예상되는 분야에서 인공지능 제품 및 서비스가 사용되고, 인공지능 제품 또는 서비스가 중대한 위험을 야기할 수 있는 방식으로 사용되면 고위험 인공지능에 해당하는 것으로 보았으며, 판단기준의 충족 여부와 상관없이 채용 과정에서 활용되거나, 노동자 또는 소비자 권리에 영향을 미치는 인공지능 기술의 활용, 원격 생체 인식(remote biometric identification) 감시 등에 인공지능 기술이 적용되는 경우에는 언제나 고위험 인공지능으로 분류하도록 제안하였습니다.[7]

고위험군 인공지능의 규제를 위한 법률 제정 시 요구될 수 있는 사항을 아래와 같이 제시하면서 이들을 바탕으로 향후 세부적인 법적 기준을 마련하겠다는 계획을 밝혔습니다.[8]

① 인공지능 시스템을 트레이닝하기 위한 데이터 셋은 안전을 위해 충분하게 광범위

---

5) 이상윤, 위의 논문, 223~224쪽

6) 현재 수준의 규제만으로는 대응할 수 없는 문제점에 대하여 예시한 내용으로는 ① 인공지능 기술의 불투명성(opaqueness)으로 인한 기본권 보호, 법적 책임 부담, 손해배상 요건 등에 관한 기존 법규들의 적용이 어려운 점, ② 현재 제품 안전에 관한 유럽연합의 법규는 순수 소프트웨어에도 적용될 수 있는지 여부가 불분명하고 인공지능 기반 서비스의 문제가 다루어지지 않는 문제점, ③ 인공지능 시스템·제품의 경우 잦은 소프트웨어 업데이트나 머신러닝 등으로 인해 시간이 지나면서 시장 출시 시점에서의 상태와는 전혀 다른 시스템·제품으로 바뀌어버리는데, 기존 법은 이런 특성을 반영하지 못하고 시장 출시 시점에서의 안전 리스크에만 집중하는 한계가 있는 점, ④ 인공지능 적용 상품·서비스 공급망의 복잡성으로 인해 제품 안전 책임과 제조물 책임의 소재를 밝히는 데 불확실성이 있는 점, ⑤ 기존의 법체계가 상정하고 있지 않던 새로운 안전 리스크들이 등장하고 있는 점입니다(이상윤, 위의 논문, 225~226쪽).

7) 이상윤, 위의 논문, 227~228쪽

8) 이상윤, 위의 논문, 228~229쪽

하고 모든 관련된 시나리오를 반영하는 것이어야 하고, 비차별을 보장하기 위해 충분하게 대표성을 지니는 것이어야 하며, 현행 법규가 제공하는 프라이버시와 개인 데이터의 보호 수준에 부합하는 것이어야 한다,

② 인공지능 기술의 복잡성과 불투명성 등으로 발생하는 법적 문제들을 해소하기 위하여 인공지능 시스템을 트레이닝 하는데 사용된 데이터에 관한 기록들, 필요한 경우에는 데이터 셋 자체, 그리고 알고리듬 프로그래밍에 관한 기록들을 보관하도록 해야 한다.

③ 인공지능 시스템의 목적, 작동 조건, 정확도 등에 관한 정보가 투명하게 제공되어야 하며, 이와 별개로 사용자가 사람이 아닌 인공지능 시스템을 상대하고 있는 경우 그에 관한 정보를 객관적이고 간결하면서도 이해하기 쉬운 형태로 제공받을 수 있어야 한다.

④ 인공지능 시스템은 그 기술을 신뢰할 수 있을 만큼 생애 주기 내내 견고하고 정밀해야 하고, 시스템의 처리 결과는 재현 가능한 것이어야 하며, 시스템 작동 중 발생할 수 있는 오류와 비일관성을 적절한 수준에서 통제할 수 있어야 하고, 외부의 공격이나 부적절한 개입을 받았을 때 원상태로 회복하는 탄성력이 있어야 한다.

⑤ 사전 또는 사후에 인간 개입을 보장하는 것처럼 적절한 형식과 수준으로 인간에 의한 감독이 이뤄질 수 있도록 해야 한다.

고위험군 인공지능의 규제를 위한 법률의 수범자와 관련하여 인공지능 시스템의 전체 수명주기 동안 개발자·생산자·배포자·최종 사용자 등 다양한 참여자들이 시스템에 관여하게 되므로, 수명주기 각 단계별 위험에 대해 참여자들 사이에 적절한 책임 소재를 결정하고 배치할 것과 법적 요구사항의 지리적 적용 범위와 관련하여 법률에 명시되는 요구사항들은 기업 또는 조직의 소재지와 무관하게 EU 시장에 인공지능 기반의 제품 또는 서비스를 제공하는 모든 기업 또는 조직에 적용할 수 있도록 검토되어야 한다는 의견을 제시하였습니다.[9)]

고위험군 인공지능을 적용하는 모든 공급자에 대한 사전 적합성 평가제도는 고위험군 인공지능의 규제를 위한 법률 제정 시 요구될 수 있는 사항들이 제대로 준수될 수 있도록 테스트, 조사, 인증 관련 절차들로 구성된 사전 적합성 평가가 이루어져야 함을 제안하였습니다.

인공지능백서의 주요 내용들은 이후 EU 인공지능법에서 구체화되어 규정되었습니다.

9) EU 인공지능 백서(White Paper on Artificial Intelligence) 주요 내용 분석, 한국인터넷진흥원(2020. 4.) 11쪽

2. EU 인공지능법의 주요 내용

EU 집행위원회는 2021. 4. 21. EU 의회에 인공지능에 관한 일괄규제 법안의 성격을 가진 EU 인공지능 법안(Proposal for a Regulation laying down harmonized rules on artificial intelligence, Artificial Intelligence Act)[10]을 발의하였고, 2023년 6월 14일 EU 의회는 EU 인공지능 법안을 채택하였으며, 향후 EU 집행위원회에서 통과되면 2026년부터 시행됩니다.

EU 인공지능법은 인공지능기술의 급속한 발전에 비하여 기존 법률에 따른 규제만으로는 한계가 있어 인공지능 시스템이 초래할 리스크의 정도에 따른, 즉 리스크 기반 접근 방식에 근거한 규제를 규정하고 있습니다. 또한 EU 인공지능법은 신뢰할 수 있는 인공지능 윤리 가이드라인, 인공지능 백서 및 EU 의회의 인공지능 관련 각종 후속 결의들과 그 동안 정부, 지방자치단체, 상업/비상업기구, 전문가 집단, 학계, 시민 등 공공부문과 민간부문을 망라한 이해관계자 전원을 대상으로 한 협의를 반영하여 제정된 것이기도 합니다.

가. EU 인공지능법의 목표 및 적용대상

EU 인공지능법의 목표는 ① EU 역내시장에서 사용되는 인공지능 시스템이 안전하고, 기본권 및 유럽의 가치에 관한 기존 법률을 준수하도록 보장, ② 인공지능에 대한 투자와 혁신 지원을 위해 법적 불확실성 제거, ③ 인공지능 시스템에 적용 가능한 기본권 및 안전 요구사항에 대한 기존 법률의 거버넌스 및 효과적인 집행 강화, ④ 합법적이고 안전하며 신뢰할 수 있는 인공지능 애플리케이션을 위한 단일시장 구축 지원 및 시장파편화의 방지입니다.[11]

이에 따라 EU 인공지능법은 제1조에서 (a) 유럽 연합에서 인공 지능 시스템('AI 시스템)의 출시, 서비스 개시 및 사용을 위한조화 규칙, (a) 특정한 인공 지능 관행의 금지, (b) 고위험 인공지능 시스템에 대한 요구사항 및 동 시스템의 운영자에게 부과되는 의무, (c) 자연인, 감정 인식 시스템 및 생체 인식 분류 시스템, 그리고 이미지, 오디오 또는 비디오 콘텐츠를 생성하거나 조작하는 데 사용되는 인공지능 시스템과 상호 작용하는 인공지능 시스템에 대한 조화 투명성 규칙, (d) 시장 감시 및 모니터링에 관한 규칙을 목적으로 규정하였습니다.

그리고 EU 인공지능법의 적용대상은 EU 내에서 인공지능 시스템을 출시하거니 서비

---

10) 본 발제문은 한국법제연구원이 번역한 EU 인공지능 법안(글로벌법제전략 자료 21-17-②)의 번역본에 기초하여 작성 되였습니다.

11) 김현정, 인공지능 기반 사회에 대비한 EU의 전략과 정책 : EU의 AI 규제안을 중심으로, 한국과 국제사회(제5권 4호), 263쪽

스하는 제공자(제공자의 소재지가 EU든 제3국이든 불문), EU 내에 위치한 인공지능 시스템의 사용자, 인공지능 시스템의 산출물이 EU에서 이용될 경우에 한정하여 제3국에 위치한 인공지능 시스템의 제공자와 사용자(제2조 1항)로 명시하고 있으며, EU 인공지능법의 시행일은 일부 규정[12]을 제외하고 EU 인공지능법 발효 후 24개월이 지난 후입니다(제85조).

나. 인공지능의 정의

EU 인공지능법은 "인공 지능 시스템(AI 시스템)은 부속서 I[13]에 명시된 기법과 방식 중 한 가지 이상을 적용하여 개발된 것으로, 인간이 정한 목적을 위해 콘텐츠, 예측, 추천, 또는 주변환경에 영향을 미치는 결정 등의 산출물을 만들어 내는 소프트웨어"로 정의하였습니다(제3조 제1항).

다. 리스크 기반에 기초한 규제 내용

EU 인공지능법은 인공지증 시스템이 초래할 리스크의 정도에 따라 용인할 수 없는 위험(an unacceptable risk), 고위험(a high risk), 낮은 위험 혹은 최소 위험(low or minimal risk)으로 분류하고 각각의 리스크에 대응하는 규제를 하고 있으며, 그 내용은 아래 표[14]와 같습니다.

---

12) 고위험 인공지능 시스템에 대한 적합성평가기구, 인증기구와 관련한 규정(제30조 내지 제39조)과 특정한 인공지능 시스템에 대한 투명성의무 규정(제52조)는 EU 인공지능법 발효 후 3개월이 지난 이후에 시행되면, 회원국의 제재처분에 관한 규칙 제정 등과 관련된 규정(제71조)은 EU 인공지능법 발효 후 12개월이 지난 이후에 시행됩니다.

13) 부속서 I에서는 딥 러닝을 포함한 다양한 방법을 사용하는 지도형, 비지도형 및 강화형 학습을 포함한 기계 학습 접근 방식. 지식 표현, 귀납적(논리) 프로그래밍, 지식 기반, 추론 및 연역적 엔진, (기호) 추론 및 전문가 시스템을 포함한 논리 및 지식 기반 접근 방식. 통계적 접근 방식, 베이즈(Bayesian) 추정, 검색 및 최적화 방법을 명시하고 있습니다.

14) 이경선, EU 인공지능 규제안의 주요 내용과 시사점, 정보통신정책연구원(2021. 5. 4.) 4쪽

용인할 수 없는 위험 (unacceptable risk)	▶ 기본권 침해 등 EU 가치에 위배되는 다음과 같은 목적하에서의 AI 시스템 활용은 금지 (a) 잠재의식에 영향을 미치는 기술(subliminal technique)을 통해 사람들의 행동을 왜곡/조작 (b) 나이, 신체적 또는 정신적 장애 등에 기반한 특정 그룹의 취약성 악용 (c) 공공기관이 AI 기반 사회적 점수화(social scoring)를 통해 자연인의 신뢰도를 평가 및 분류 (d) 법 집행을 위한 목적으로 공개적으로 접근가능한 공간에서 '실시간' 원격 생체 인식 시스템 사용. 단, 범죄 피해자 표적수색, 임박한 위협방지 등 일부 상황에서는 예외적으로 허용
고위험 (high risk)	▶ 자연인의 건강/안전/기본권에 고위험을 야기할 수 있는 AI 시스템은 요구사항 준수, 사전 적합성 평가 수행이 요구됨 • 고위험 AI 시스템은 (a) 제품의 안전요소로써 사용되어 사전 제 3자 적합성 평가가 요구되는 AI 시스템, (b) 생체 인식 및 분류, 교육, 고용, 법집행 등 기본권에 영향을 미칠 수 있는 환경에서 독자적으로 활용되는 AI 시스템 • 요구사항은 ① 위험관리 시스템 구축, ② 데이터 거버넌스 수행, ③ 기술 문서화, ④ 기록, ⑤ 이용자에게 투명성 및 정보 제공, ⑥ 사람에 의한 감독, ⑦ 정확성/견고성/사이버보안 ▶ 위원회는 건강/안전/기본권에 악영향을 미칠 수 있는 고위험 AI 시스템의 추가, 채택의 권한을 가짐
낮은 위험 (non-high risk; low or minimal risk)	▶ 낮은 위험(non-high risk) AI 시스템에 대해서는 고위험 AI 시스템에 부여된 요구사항들이 강제되지는 않으나 자발적 준수를 위한 행동강령(code of conducts)의 수립이 권장됨 • 추가적으로 환경을 위한 지속가능성, 장애를 가진 사람들에 대한 접근권, AI 시스템의 설계/개발 시 이해관계자들의 참여, 개발팀의 다양성 보장 등을 위한 자발적 노력이 권장됨 ▶ ① 챗봇처럼 사람과 상호작용하거나, ② 감정인식, 생체데이터에 기반한 (사회적) 분류에 사용되거나, ③ 딥페이크처럼 진짜처럼 보이는 콘텐츠의 생성·조작에 사용되는 특정 AI 시스템에 대해서는 투명성 의무를 부여 • 투명성 의무 : 해당 AI시스템의 제공자, 사용자는 사람들이 충분한 정보에 기반하여 시스템 이용여부를 결정할 수 있도록 시스템 구동방식 등을 고지할 필요

이처럼 용인할 수 없는 위험에 해당하는 인공지능시스템은 활용이 금지되며, 낮은 위험에 해당하는 인공지능시스템은 시민의 권리나 안전에 영향이 거의 없어 추가적인 법적 의무 없이 기존 법규에 따라 개발 및 사용이 가능하도록 규정하고 있습니다, 그러므로 EU 인공지능법의 주된 규제 내용을 이루는 고위험 인공지능시스템에 대하여 자세하게 살펴 볼 필요가 있습니다. 고위험 인공지능 시스템에 대하여는 부속서Ⅲ에서 구체적으로 언급하고 있으며 그 내용은 아래 표와 같습니다.[15)]

15) 이경선, 위의 논문 5쪽

구 분	내 용
1. 자연인의 생체 인식 및 분류:	(a) 자연인의 'real time', 'post' 원격 생체 인식 식별에 사용되는 AI 시스템
2. 중요 인프라의 관리 및 운영:	(a) 도로 교통 및 물, 가스, 난방 및 전기 공급의 관리 및 운영에서 안전 구성 요소로써 사용되도록 의도된 AI 시스템
3. 교육 및 직업훈련:	(a) 교육 및 직업 훈련 기관에 대한 자연인의 접근을 결정하거나 할당할 목적으로 사용되는 AI 시스템 (b) 교육 및 직업 훈련 기관의 학생을 평가하고 교육 기관 입학에 일반적으로 필요한 시험에서 참여자를 평가할 목적으로 사용되는 AI 시스템
4. 고용, 근로자 관리 및 자영업에 대한 접근:	(a) 자연인의 채용 및 선택, 특히 공석 광고, 심사 또는 지원서류 필터링, 인터뷰 또는 테스트 과정에서의 후보자 평가 등을 위해 사용되는 AI 시스템 (b) 승진, 업무 관련 계약 관계의 종료, 작업 할당, 이러한 관계에서의 사람의 성과 및 행동을 모니터링, 평가하기 위해 사용되는 AI
5. 필수 개인 서비스 및 공공 서비스, 혜택에 대한 접근 및 향유:	(a) 공공 지원 혜택 및 서비스에 대한 자연인의 적격성을 평가하고 그러한 혜택 및 서비스를 부여, 축소, 취소 또는 회수하기 위해 공공기관 또는 공공기관을 대신하여 사용하도록 의도된 AI 시스템 (b) 소규모 공급자가 자체 사용을 위해 서비스하는 AI 시스템을 제외한, 자연인의 신용도를 평가하거나 신용 점수를 정하는 데 사용되는 AI 시스템 (c) 소방관 및 의료 지원을 포함, 긴급 first response 서비스의 파견 또는 파견에 우선 순위를 설정하는 데 사용되는 AI 시스템
6. 법 집행:	(a) 법 집행 당국이 자연인의 공격 또는 재범 위험, 또는 형사 범죄의 잠재적 피해자에 대한 위험을 평가하기 위해 사용하는 AI 시스템 (b) 법 집행 기관에서 거짓말 탐지기 및 유사한 도구로 사용하거나 자연인의 감정 상태를 감지하기 위해 사용하는 AI 시스템 (c) 제 52조(3)에 언급 된 바와 같이 법 집행 기관이 딥페이크를 탐지하기 위해 사용하는 AI 시스템 (d) 형사 범죄 수사 또는 기소 과정에서 증거의 신뢰성을 평가하기 위해 법 집행 기관에서 사용하는 AI 시스템 (e) 법 집행 기관이 자연인의 프로파일링을 기반으로 실제 또는 잠재적 범죄 행위의 발생 또는 재발을 예측하거나 또는 자연인 또는 그룹의 성격적 특성 및 특징, 또는 과거 범죄 행위를 평가하기 위해 사용하는 AI 시스템 (f) 형사 범죄의 탐지, 조사 또는 기소 과정에서 자연인의 프로파일링을 위해 법 집행 기관에서 사용하는 AI 시스템 (g) 법 집행 당국이 데이터에서 알려지지 않은 패턴을 식별하거나 숨겨진 관계를 발견하기 위해 다양한 데이터 소스 또는 다양한 데이터 형식의 대규모 데이터 세트를 조사하는 방식으로 자연인에 대한 범죄 분석에 사용되는 AI 시스템
7. 이주, 망명 및 국경 통제 관리:	(a) 관할 공공기관에서 거짓말 탐지기 및 유사한 도구로 사용하거나 자연인의 감정 상태를 감지하기위한 AI 시스템 (b) 관할 공공기관이 회원국의 입국을 원하거나 입국한 자연인의 보안, 비정규 이민, 건강 위험 등의 위험을 평가하기 위해 사용하는 AI 시스템 (c) 관할 공공 기관에서 자연인의 여행 및 관련 문서의 진위를 확인하는데 사용하는 AI 시스템 (d) 관할 공공기관에서 망명, 비자 및 거주 허가를 신청하는 자연인의 적격성 평가, 관련 불만을 조사하기 위해 사용하는 AI 시스템.
8. 정의와 민주적 과정의 관리:	(a) 사법 당국이 사실과 법을 조사 및 해석하고 구체적인 사실에 법을 적용하는 데 도움을 주기위해 사용하는 AI 시스템.

EU 인공지능법은 제8조에서부터 제15조까지 고위험 인공지능 시스템이 갖추어야 하는 요구사항을 상세하게 규정하고 있습니다. 즉, 고위험 인공지능 시스템은 인공지능 시스템의 수명주기 전체를 대상으로 연속적이고 반복적인 프로세스가 포함된 리스크관리 시스템을 도입하여 시행, 기록, 유지할 수 있을 것(제9조), 데이터를 활용해 모델을 훈련시키는 기술을 이용하는 고위험 인공지능 시스템은 제10조 제2항 내지 제5항에 명시된 품질기준에 부합하는 교육과 검증, 시험데이트[16] 세트를 이용하여 개발할 것(제10조), 고위험 인공지능 시스템을 출시하거나 서비스 개시하기 전에 EU 인공지능법이 명시하고 있는 고위험 인공지능 시스템의 요구사항을 준수하고 있음을 입증하는 기술문서를 작성하고 최신상태를 유지할 것(제11조), 고위험 인공지능 시스템을 설계하고 개발할 때에는 공인된 표준이나 공통규격에 부합하는 고위험 인공지능 시스템이 작동하는 동안의 사건을 자동으로 기록('로그')하는 기능을 포함시킬 것(제12조), 고위험 인공지능 시스템은 작동방식이 충분히 투명하여 사용자가 시스템의 산출물을 해석하고 적절히 이용할 수 있도록 설계 및 개발할 것(제13조), 고위험 인공지능 시스템을 설계하고 개발할 때 사람-기계간 인터페이스 도구를 삽입하는 등 시스템이 사용되는 동안 자연인이 그 시스템을 효과적으로 감독할 수 있게 해야 할 것(제14조), 고위험 인공지능 시스템은 시스템이 의도한 설계와 목적에 비추어 적정한 수준으로 정확도와 견고성, 사이버보안을 갖추어야 하고, 수명이 다할 때까지 이러한 성능을 유지하도록 설계하고 개발되어야 할 것(제15조)입니다.

한편 EU 인공지능법은 제8조에서부터 제16조에서 제29조까지 고위험 인공지능 시스템의 제공자[17], 제품제조업자, 수입업자[18], 유통업자[19], 사용자[20] 및 기타 당사자의 의무를 규정하고 있습니다.

제공자는 고위험 인공지능 시스템이 갖추어야 하는 요구사항을 준수할 의무, 품질관리시스템 구축 의무(제17조), 기술문서 작성의무(제18조), 고위험 인공지능 시스템의 시장 출시나 서비스 개시 전 관련 적합성평가[21]를 실시할 의무(제19조), 고위험 인공지능

16) "시험 데이터"란 훈련과 검증을 마친 인공지능 시스템을 출시하거나 서비스를 개시하기 전, 그 시스템의 기대 성능을 확인하기 위한 독립 평가에 제공되는 데이터를 의미한다(제3조 (31)호).

17) '제공자(provider)'는 유료, 무료로 불문하고 자기 명의 또는 상표로 출시하거나 서비스 할 의도로 인공지능 시스템을 직접 또는 위탁하여 개발하거나 자연인, 법인, 공공 기관, 관청, 기타 기관 등을 의미한다(제3조 (2)호).

18) '수입업자(importer)'는 EU 내에 소재하는 자연인 또는 법인으로 EU 이외에 소재하는 자연인 또는 법인의 명의나 상표를 지닌 인공지능 시스템을 출시하거나 서비스하는 자를 의미한다(제3조 (6)호).

19) '유통업자(distributor)'는 공급망에 속한 개인이나 법인 중 제공자나 수입업자가 아닌 자로서, 인공지능 시스템에 변형을 가하지 않고 EU시장 내에서 인공지능 시스템을 이용할 수 있게 하는 자를 의미한다(제3조 (7)호).

20) '사용자(user)'는 자신의 권한으로 인공지능 시스템을 이용하는 자연인, 법인, 공공기관, 관청, 기타 기관을 의미하며, 다만 직무와 무관하게 개인 활동으로 인공지능 시스템을 이용하는 경우는 사용자에서 제외한다(제3조 (4)호).

21) "적합성 평가"란 인공지능 시스템과 관련된 EU 인공지능법 Ⅲ편 2장에 명시된 요건의 이행 여부를 검증하는 과정을 의미한다(제3조 (20)호).

시스템을 관리하는 동안 시스템에서 자동으로 생성되는 로그를 보관할 의무(제20조), 제공자가 고위험 인공지능 시스템이 요구사항에 부합하지 않거나 그렇게 볼만한 사유가 있다고 판단할 경우 시스템의 적합성 확보를 위해 즉시 필요한 시정조치를 취하거나 시스템을 퇴출[22]시키거나 리콜[23] 조치해야 할 의무(제21조), 고위험 인공지능 시스템이 EU 인공지능법 제65조 제1항에 규정된 리스크가 존재하고, 제공자가 이를 알게 된 경우 즉시 회원국 관할기관 또는 인증기구에 위반사실을 보고할 의무 및 위반사실을 인지하고 시정조치를 취했다면 그 사실을 보고할 의무(제21조), 회원국 관할기관 요청 시 고위험 인공지능 시스템이 EU 인공지능법에 규정된 요구사항을 준수하였음을 입증하는데 필요한 정보와 문서를 제출할 의무(제23조), EU 인공지능법 제51조에 규정된 등록 의무를 준수할 의무(제16조 제1항 (f)호), EU 인공지능법 제49조에 따라 고위험 인공지능 시스템에 규정준수를 알 수 있도록 고위험 인공지능 시스템에 CE 마크[24] 부착의무 및 EU 인공지능법에 부합한다는 사실을 표시할 의무(제16조 제1항 (i)호)를 부담합니다. 또한 제공자는 인공지능 기술의 성격과 고위험 인공지능 시스템의 리스크에 비례하여 사후 모니터링 시스템을 구축하고 문서화해야 하며(제61조), EU 법률에서 정한 의무의 위반에 해당하는 중대한 사고나 오작동이 있으면 해당 사건 또는 위반이 발생한 회원국의 시장감시 기관에 보고할 의무가 있습니다(제62조).

제품제조업자에게는 부속서 II의 A절에 열거된 법률의 적용을 받는 제품과 관련된 고위험 인공지능 시스템이 해당 법률에 따라 제조되고, 제품제조업자의 상호로 출시되거나 서비스되는 경우, 해당 제품의 제조업자는 그 인공지능 시스템이 본 규정을 준수하도록 해야 할 의무와 그 인공지능 시스템에 관한 한 제공자에게 부과하는 의무와 동일한 의무를 부과하고 있습니다(제24조).

수입업자는 고위험 인공지능 시스템을 출시하기 전에 적합성평가 절차를 제대로 실시하였는지 여부, 기술문서 작성 여부, 적합성CE마크가 부착되었는지 여부를 확인하고 보장할 의무를 부담하며, 자신의 판단에 따라 고위험 인공지능 시스템이 EU 인공지능법에 부합하지 않거나 그렇게 판단할만한 사유가 있을 경우 적합성을 충족할 때까지 출시하지 않을 의무 등을 부담합니다(제26조).

유통업자는 고위험 인공지능 시스템을 출시하기 전에 적합성CE마크가 부착되었는지 여부, 필수 문서와 사용설명서[25]가 구비되어 있는지 여부, 시스템의 제공자와 수입업자

---

22) “인공지능 시스템의 퇴출”이란 사용자가 이용할 수 있는 상태에 놓인 인공지능 시스템을 해당 제공자에게 되돌려 보내기 위해 취하는 조치를 통칭한다(제3조 (17)호).

23) “인공지능 시스템의 리콜”이란 인공지능 시스템의 유통과 전시, 판매 제안을 막기 위해 취하는 조치를 통칭한다(제3조 (16)호).

24) “적합성 CE 마크”란 제공자가 인공지능 시스템에 부착하는 것으로 해당 시스템이 EU 인공지능법 III편 2장과 그 외 제품의 판매 조건통일에 관한 EU 법률에 명시된 요건을 준수하였다는 것을 나타내는 표시를 의미한다(제3조 (24)호).

25) “사용설명서”란 제공자가 사용자에게 인공지능 시스템의 설계 목적과 적합한 용도를 말려주기 위해 제공하는 정보를 의미하며, 그 내용에는 고위험 인공지능 시스템이 쓰일 지리적, 행동적, 기능적 환경이 포함된다(제3조 (15)호).

가 EU 인공지능법에 명시된 의우를 준수했는지를 검증할 의무, 자신의 판단에 따라 고위험 인공지능 시스템이 EU 인공지능법에 부합하지 않거나 그렇게 판단할만한 사유가 있을 경우 적합성을 충족할 때까지 출시하지 않을 의무 등을 부담합니다(제27조).

사용자는 고위험 인공지능 시스템에 첨부된 사용설명서에 따라 그 시스템을 사용할 의무, 사용설명서에 따라 고위험 인공지능 시스템의 작동을 모니터링하고, 사용설명서에 따라 사용했을 때 해당 시스템이 EU 인공지능법 제65조 제1항에 규정된 리스크가 발생할 가능성이 있다면 제공자나 유통업자에게 통지하고 시스템의 사용을 중단할 의무 등이 있습니다(제29조).

한편 유통업자, 수입업자, 사용자 또는 기타 제3자는 자신의 상호 또는 상표로 고위험 인공지능 시스템을 출시하거나 서비스시하는 경우, 이미 출시되었거나 서비스가 개시된 고위험 인공지능 시스템의 설계목적[26]을 변경하는 경우, 고위험 인공지능 시스템에 중대한 변경을 가한 경우 중 하나에 해당하면 제16조에 따른 제공자의 의무를 부담하게 됩니다(제28조).

그리고 고위험 인공지능 시스템과는 별개로 제52조에서 특정 인공지능 시스템에 대하여 제공자와 사용자에게 투명성 의무를 부과하고 있습니다. 제공자에게는 자연인과 상호 작용하는 인공지능 시스템을 설계하고 개발할 때 자연인에게 인공지능 시스템과 교류하고 있음을 알리는 기능을 넣을 의무를 부과하고 있습니다. 그리고 감정인식 시스템[27] 또는 생체분류 시스템[28]의 사용자는 그 시스템에 노출되는 자연인에게 시스템이 운용되고 있다는 사실을 고지할 의무가 있으며, 기존 인물, 사물, 장소, 기타 실체 또는 사건과 현저히 유사하여 보는 이로 하여금 마치 진자 또는 사실처럼 느끼게 만드는 이미지, 청각 또는 시각 콘텐츠를 생성하거나 조작하는(“딥페이크”) 인공지능 시스템의 사용자는 해당 콘텐츠가 인공적으로 생성되었거나 조작되었음을 공개할 의무가 있습니다.

한편 제5조에 명시된 인공지능 활용 관련 금지행위를 위반한 경우와 인공지능 시스템이 제10조에 명시된 요건을 위반한 경우 최고 3,000만 유로와 직전 회계연도 전세계 연간 총매출의 6%에 해당하는 금액 중 큰 금액을 상한으로 한 과태료를 부과하는 규정과 인공지능시스템이 EU 인공지능법 규정에 따른 요건이나 의무를 위반하면 2,000만 유로와 직전 회계연도 전세계 연간 총매출의 4%에 해당하는 금액 중 큰 금액을 상한으로 한 과태료를 부과하는 규정을 통해 규정의 실효성을 확보하고 있습니다(제71조).

---

26) “설계목적”이란 제공자가 사용설명서나 홍보 또는 영업자료/명세서, 기술문서에 명시한 인공지능 시스템의 본래 용도, 예를 들면 환경과 이용 조건 등을 의미한다(제3조 (12)호).

27) “감정인식 시스템”이란 생체데이터(자연인의 신체적, 생리학적, 행동적 특성을 기술적으로 처리하여 얻은 개인 정보 중에서 얼굴 이미지나 지문 정보 등과 같이 그 자연인의 고유한 신원을 확인할 수 있는 정보)를 토대로 자연인의 감정이나 의도를 식별하거나 추정하는 목적을 지닌 인공자능 시스템을 의미한다(제3조 (34)호).

28) “생체분류 시스템”이란 생체데이터를 기초로 성별, 연령, 머리색, 눈동자색, 문신, 인종, 성적 지향, 정치성향 등 특정 범주에 따라 자연인을 분류하는 인공자능 시스템을 의미한다(제3조 (35)호).

Ⅲ. 한국의 인공지능법안 발의 이전까지의 경과 및 인공지능법안의 주요 내용

1. 한국의 인공지능법안 발의까지의 경과

과학기술정보통신부와 한국정보화진흥원은 2018년 9월 지능정보사회 윤리 가이드라인 및 윤리헌장을 발표하였습니다. 지능정보사회 윤리 가이드라인을 발표하게 된 배경으로는 4차 산업혁명은 사회, 경제, 산업구조, 노동환경 및 개인 삶의 방식까지 근본적으로 변화시킬 것으로 예측되어 지능정보기술의 산업적, 사회적 활용에 따른 부작용에 대한 우려가 높아짐에 따라 주체별 권리와 책임의 범위에 대한 규범적 논의의 전개를 위해 사회적 공론화가 필요하고, 국내외 정책적 대응이 관련 산업의 진흥을 위한 부작용 예방 차원에서 윤리규범 등을 논의하는 수준에 머물러 있어 지능정보화사회의 행위주체에 따른 세부적 규범 내용을 포함한 종합적 접근이 필요하다는 것입니다.

지능정보사회 윤리 가이드라인의 목적은 인간중심의 지능정보사회를 위해, 지능정보기술 및 서비스 개발자, 공급자의 책임윤리 강화 및 이용자의 오남용 방지를 위한 지침 제공이며, 기본방향은 지능정보기술의 잠재적 위험으로부터 사회시스템을 보호하는 사전 예방원칙, 윤리적 규율이 필요한 분야에 대해 구체적 행위지침을 제시하는 것입니다. 이에 따라라 지능정보사회, 지븐정보기술, 지능정보서비스, 개발자, 공급자, 이용자에 대한 정의를 하고, 지능정보기술을 개발 · 활용하거나 지능정보서비스를 제공 · 이용할 때의 공통원칙으로 ① 공공성, ② 책무성, ③ 통제성, ④ 투명성을 제시하면서 각 원칙별로 개발자, 공급자, 사용자가 준수해야 할 세부적 내용을 제시하였습니다.

이후 과학기술정보통신부와 정보통신정책연구원은 2020년 12월 23일 대통령 직속 4차산업혁명위원회 전체회의에서 인공지능 시대 바람직한 인공지능 개발·활용 방향을 제시하기 위한 사람이 중심이 되는 인공지능(AI) 윤리기준을 마련하였는데, 이는 윤리적 인공지능을 실현하기 위해 정부·공공기관, 기업, 이용자 등 모든 사회구성원이 인공지능 개발~활용 전 단계에서 함께 지켜야 할 주요 원칙과 핵심 요건을 제시한 것이었습니다.

‘인간성을 위한 인공지능(AI for Humanity)’을 구현하기 위해 인공지능의 개발 및 활용 과정에서 고려되어야 하는 3대 기본원칙으로 ① 인간의 존엄성 원칙(인공지능은 인간의 생명은 물론 정신적 및 신체적 건강에 해가 되지 않는 범위에서 개발 및 활용되어야 한다), ② 사회의 공공선 원칙, ③ 기술의 합목적성 원칙을 제시하였습니다.

그리고 3대 기본원칙을 실천하고 이행할 수 있도록 인공지능 전체 생명 주기에 걸쳐 충족되어야 하는 10가지 핵심 요건을 제시하였는데 이는 ① 인권 보장, ② 프라이버시 보호, ③ 다양성 존중, ④ 침해금지(인공지능을 인간에게 직간접적인 해를 입히는 목적으로 활용해서는 안 된다), ⑤ 공공성, ⑥ 연대성(윤리적 인공지능의 개발 및 활용에

국제사회가 협력하도록 노력해야 한다), ⑦ 데이터 관리, ⑧ 책임성(인공지능 설계 및 개발자, 서비스 제공자, 사용자 간의 책임소재를 명확히 해야 한다), ⑨ 안전성(인공지능 개발 및 활용 전 과정에 걸쳐 잠재적 위험을 방지하고 안전을 보장할 수 있도록 노력해야 한다), ⑩ 투명성(인공지능 활용 상황에 적합한 수준의 투명성과 설명 가능성을 높이려는 노력을 기울여야 한다)입니다.

2. 인공지능법안[29]의 주요 내용과 문제점

인공지능법안은 정부의 인공지능 기본계획 수립, 인공지능사회의 구현, 인공지능산업의 진흥 및 신뢰 확보와 관련된 사항을 심의·의결하기 인공지능위원회 설립, 인공지능산업 기반 조성을 위한 규제 원칙, 인공지능기술의 표준화, 기업에 대한 지원, 국제 협력 등을 주요 내용으로 하여 인공지능산업육성에 초점이 맞추어져 있습니다. 아래에서는 인공지능법안의 주요 내용 및 문제점에 대하여 살펴보겠습니다.

인공지능법안은 제1조에서 인공지능산업을 진흥하고 인공지능사회의 신뢰 기반 조성에 필요한 기본적인 사항을 규정함으로써 국민의 권익과 존엄성을 보호하고 국민의 삶의 질 향상과 국가경쟁력을 강화하는 데 이바지함을 목적으로 한다고 규정하여 법안의 목적이 인공지능산업 육성에 있음을 밝히고 있습니다.

제2조에서 인공지능에 대하여 학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것으로 정의 내리고(제2조 제1호), 인공지능기술은 인공지능을 구현하기 위하여 필요한 하드웨어 기술 또는 그것을 시스템적으로 지원하는 소프트웨어 기술 또는 그 활용 기술로(제2조 제2호), “고위험영역 인공지능”에 대하여 열거된 개별 법령의 규율 영역 중 하나에 해당하는 인공지능으로서 사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 영역에서 활용되는 인공지능으로 정의하고 있습니다(제2조 제3호). 반면 인공지능사업자에 대하여는 인공지능산업(인공지능기술 또는 인공지능기술을 활용한 제품)을 개발 · 제조 · 생산 또는 유통하거나 이와 관련한 서비스를 제공하는 산업)과 관련된 경제활동을 영위하는 자로 포괄적으로 규정하고 있어(제2조 제6호), 인공지능기술의 개발과 활용을 포함한 모든 단계에서 각 당사자의 책무가 구체적이고 명확하게 규정되기 어려운 문제점이 있습니다.

제4조 제2항에서는 타법과의 관계에 대하여 인공지능 등에 관하여 다른 법률을 제정하거나 개정하는 경우에는 이 법의 목적에 부합되도록 하여야 한다는 규정을 두었는데, 이는 이미 시행되고 있는 지능정보화기본법[30]이나 학습용 데이터 관련 시책의 수립(제

---

29) 2020년 7월부터 과방위에 발의된 인공지능 관련 법안 7개를 통합한 법안을 의미합니다.

30) 지능정보화 기본법은 제2조 제4호에서 지능정보기술을 아래와 같이 정의하고 있는데 이에 의하면 인공지능기술을 포함하는 것으로 해석됩니다.

“지능정보기술” 이란 다음 각 목의 어느 하나에 해당하는 기술 또는 그 결합 및 활용 기술을 말한다.

14조)과 관련된 개인정보보호법의 내용과 상충될 수 있을 뿐만 아니라 산업육성법의 성격을 가지고 있음에도 기본법으로서 지위를 부과하고 있다는 점에서 문제가 있습니다.

제5조에서 과학기술정보통신부장관으로 하여금 관계 중앙행정기관의 장 및 지방자치단체의 장의 의견을 들어 3년마다 인공지능기술 및 인공지능산업의 진흥과 국가경쟁력 강화를 위하여 인공지능 기본계획을 수립하고 시행하도록 규정하고 있는데, 인공지능산업의 광범위성 및 이해당사자가 다양한 점을 고려하면 과학기술정보통신부가 독립성을 가지고 부처간 이해관계를 조율할 수 있는 인공지능 기본계획의 수립 및 시행의 주체로 적정한지도 의문입니다.

제11조에서는 우선허용·사후규제 원칙하에 인공지능기술, 인공지능제품 또는 인공지능서비스가 국민의 생명 · 안전 · 권익에 위해가 되거나 공공의 안전 보장, 질서 유지 및 복리 증진을 현저히 저해할 우려가 있는 경우가 아니라면 이를 제한하여서는 아니 된다는 원칙적 규정만을 두어 인공지능기술의 잠재적 위험에 대비하기 위한 구체적 기준이 없어 법적 안정성이 확보되지 않아 법규범으로의 기능을 할 수 없는 문제점이 있으며, 또한 인공지능산업의 이해당사자로 하여금 제한되는 인공지능기술 등의 구체적인 기준에 대한 예측가능성을 제공하지 않아 사실상 인공지능기술의 개발자, 제공자, 사용자 등의 자의에 맡겨두었다는 점에서도 심각한 문제점이 있습니다.

제20조에서는 정부로 하여금 인공지능 · 인공지능기술의 개발, 인공지능사회 및 인공지능윤리에 관한 국제적 동향을 파악하고 국제협력을 추진하여야 하는 책무를 부과하고 있지만 인공지능법안의 내용은 세계 각국의 인꽁지능기술의 잠재적 위험성에 대한 규제 논의에 대한 내용들이 포함되어 있지 않아 향후 인공지능기술을 토대로 국내 기업들의 해외 진출에 있어 심각한 문제점을 야기할 우려가 있으며, 이는 EU 인공지능법이 적합성 CE 마크의 부착을 의무화 한 점만 보아도 산업육성법으로서의 기능을 제대로 할 수 있는지도 의문입니다.

제26조에서는 제품 또는 서비스를 개발·활용·제공하려는 자로 하여금 고위험영역 인공지능에 해당하는지에 대한 확인을 과학기술정보통신부장관에게 요청하도록 하여 고위험영역 인공지능에 대하여 제품 또는 서비스를 개발·활용·제공하려는 자의 자의에 맡겨두고 있는 점, 고위험영역 인공지능을 이용하여 제품 또는 서비스를 제공하려는 자의 의무로는 해당 제품 또는 서비스가 고위험영역 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지할 의무(제27조)와 고위험 인공지능의 신뢰성과 안전성을 확보하기 위한 조치 의무(제28)만을 부과하고 있지만 이러한 의무 위반에 대하여 어떠한 제재 조치도 없어 실효성이 있을지 의문입니다.

가. 전자적 방법으로 학습 · 추론 · 판단 등을 구현하는 기술

나. 데이터(부호, 문자, 음성, 음향 및 영상 등으로 표현된 모든 종류의 자료 또는 지식을 말한다)를 전자적 방법으로 수집 · 분석 · 가공 등 처리하는 기술

다. 물건 상호간 또는 사람과 물건 사이에 데이터를 처리하거나 물건을 이용 · 제어 또는 관리할 수 있도록 하는 기술

이처럼 인공지능법안은 인공지능산업육성 방안에 대하여는 비교적 구체적인 내용들을 담고 있지만 인공지능기술의 신뢰성, 안전성 확보를 위한 구체적인 기준과 방안에 대한 규범적 내용은 거의 없다는 점에서 수정 없이 입법되는 것은 적절하지 않다고 생각됩니다.

Ⅳ. 인공지능법안의 수정 필요성

유럽연합은 인공지능 윤리기준 가이드라인의 내용과 인공지능백서의 내용을 토대로 공동체 구성원의 논의를 통하여 위 내용들을 규범으로 구체화한 인공지능법을 만들었습니다. 반면 우리나라도 비슷한 시기에 지능정보사회 윤리 가이드라인을 통하여 인공지능기술의 윤리규범 차원에서의 논의를 벗어나 인공지능기술의 개발과 활용에 있어 행위주체에 대한 권리와 책임의 범위에 대한 규범적 논의를 위해 공통원칙으로 ① 공공성, ② 책무성, ③ 통제성, ④ 투명성을 제시하면서 각 원칙별로 개발자, 공급자, 사용자가 준수해야 할 세부적 내용을 제시하였고, 이후 인간성을 위한 인공지능을 구현하기 위한 3대 기본원칙과 이를 실천하고 이행할 수 있도록 인공지능 전체 생명 주기에 걸쳐 충족되어야 하는 10가지 핵심 요건을 제시한 인공지능 윤리기준을 발표하였습니다. 그러나 현재 국회에서 논의되고 있는 인공지능법안의 내용에는 지능정보사회 윤리 가이드라인과 인공지능 윤리기준에서 제시한 내용들이 규범으로서 구체화되지 않았음을 알 수 있습니다. 특히 인공지능법안은 지능정보사회 윤리 가이드라인에서 제시한 개발자, 공급자, 사용자가 준수해야 할 세부적 내용보다 구체적이지 않습니다. 또한 인공지능 윤리기준에서 제시한 책임성, 안정성, 투명성의 핵심 요건들이 반영되지 않았습니다.

이처럼 인공지능법안은 그 동안 우리사회에서 규범화를 전제로 논의해왔던 인공지능기술의 신회성과 잠재적 위험을 최소화하기 위한 규제의 내용들이 충분히 반영되지 않았다는 점에서 대폭적인 법안의 수정이 필요한 것으로 생각됩니다. 인공지능법안의 수정을 논의함에 있어 그동안 우리나라에서 논의되었던 내용들과 유럽연합의 리스크 기반 방식의 새로운 규제 체계에 대한 수용 여부가 반드시 검토되어야 할 것으로 생각됩니다. 인공지능기술의 산업정책적 차원에서의 중요성과 인공지능기술의 신속한 발전 속도를 고려할 때 이를 규제하는 인공지능법의 도입이 시급하지만 특히 고위험 인공지능기술의 경우, 그 위험성이 확인되지 않은 상태에서 산업 육성의 목적에 치우쳐 실효성 없는 원칙적 규정만으로는 예측불가능한 위험을 관리하기는 어려우며, 향후 위험성이 관리되지 않고 현실화 되었을 경우의 파급효를 고려하여 구채적이고 명확한 규제의 내용이 포함되어야 할 것입니다. ▢

<참고문헌>

김민수, 4차 산업혁명의 담론과 인공지능 기술에 대한 철학적 분석과 비판, 인공지능인문학연구 7권(2021)

김창화, 인공지능(AI) 윤리 가이드라인 연구, 한국인터넷진흥원(2020)

김현정, 인공지능 기반 사회에 대비한 EU의 전략과 정책 : EU의 AI 규제안을 중심으로, 한국과 국제사회(제5권 4호)

박도현, 인간 편향성과 인공지능의 교차, 서울대학교 법학 제63권 제1호(2022)

손영화, 인공지능(AI) 시대의 법적 과제, 법과 정책연구 제16권 제4호(2016. 12.)

윤혜선, 인공지능 규제 정책에 관한 연구: 주요국의 규제 정책 사례를 중심으로, 정보통신정책연구 제26권 제4호(2019. 12.)

이경선, EU 인공지능 규제안의 주요 내용과 시사점, 정보통신정책연구(2021. 5. 4)

이상윤, 유럽연합 디지털 정책의 동향과 전망: "유럽의 디지털 미래" · "유럽 데이터 전략" ·"인공지능 백서"의 주요 내용과 의의, 고려법학 제97호(2020. 6)

이숙연, 인공지능 관련 규범 수립의 국내외 현황과 과제, 법조 제72권 제1호(2023. 2.)

정원준 외 2인, 인공지능 시대의 법제정비 방안, 정보통신정책연구원(2019)

한국인터넷진흥원(2020. 4.), EU 인공지능 백서(White Paper on Artificial Intelligence) 주요 내용 분석

미국의 인공지능 규제 동향

장여경
(사단법인 정보인권연구소 상임이사)

❍ 인공지능(AI) 규제에 대한 미국의 접근법은 다음과 같은 자료들을 통해 살펴볼 수 있음.

I. 백악관 AI 권리장전 청사진[1)]

❍ 백악관 과학기술정책국(OSTP)은 2022. 10. 5. ‘AI 권리장전(AI Bill of Rights)’ 청사진을 공개함.

❍ 청사진은 ▲안전한 시스템 ▲차별 방지 ▲데이터 사생활 보호 ▲사전 고지와 설명 ▲인적 대안 및 대비책 등 기업과 정부 기관들이 지켜야 할 5가지 기본 원칙을 제시함.

❍ 모든 원칙에서 위험을 사전에 식별하는 독립적인 평가, 평가 공개, 위험 완화 조치를 요구하고 있다는 점이 특기할만 함.

1) The White House (2022). Blueprint for an AI Bill of Rights : Making Automated Systems Work For The American People.
<https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>; 조선일보 보도(2022. 10. 6). 백악관, ‘AI 권리장전’ 발표...인종·성차별 도구 사용 안 돼.

원칙	주요 내용
안전하고 효과적인 시스템	여러분은 안전하지 않거나 효과적이지 못한 시스템으로부터 보호되어야 합니다. … 시스템이 안전하고 효과적이라는 사실을 확인하는 독립적인 평가와 보고가 수행되어야 하며, 여기에는 잠재적 위해를 완화하기 위해 취한 조치에 대한 보고가 포함되어야 하고, 그 결과는 가능할 한 공개되어야 합니다.
알고리즘 차별로부터 보호	여러분은 알고리즘에 의해 차별받지 않아야 하며, 시스템은 공평한 방식으로 사용되고 설계되어야 합니다. … 이에 대한 보호 조치로서, 시스템 설계, 대표 데이터 사용 및 인구집단 특성별 대리변수 보호에 대한 사전 공평성 평가를 수행하고, 설계 및 개발에 대한 장애인에 대한 접근성을 보장하며, 배치 전 및 지속적으로 편향성 테스트 및 완화조치를 수행하고, 조직적 감독을 명확하게 수행하여야 합니다. 알고리즘 영향평가 형식의 독립적인 점검과 쉬운 용어로 이루어진 보고가 이루어져야 하며, 여기에는 편향성 테스트 결과 및 완화 조치에 대한 정보가 포함되어야 하고, 이러한 보호 조치 여부를 확인하기 위해 보고서가 가능한 한 공개되어야 합니다.
개인정보 보호	여러분은 설치형 보호(Built-In Protection)를 통해 개인정보를 오남용하는 관행으로부터 보호받아야 하고, 여러분에 관한 개인정보가 어떻게 이용되는지에 대해 여러분이 권한을 가져야 합니다. … 감시 기술이 사생활과 시민권에 미치는 잠재적 위해와 이를 제한하는 범위를 최소한 배치 전에 평가하는 등 감시 기술에 대한 감독을 강화하여야 합니다… 여러분은 가능한 한 여러분의 개인정보 결정권이 존중되었음을 확인하고 여러분의 권리, 기회 또는 접근에 미치는 감시 기술의 잠재적 영향에 대해 평가하는 보고서에 접근할 수 있어야 합니다.
통지 및 설명	여러분은 자동화된 시스템이 사용되고 있다는 사실과 이 시스템이 여러분에게 영향을 미치는 결과물에 작동하는 방법과 이유를 이해할 수 있어야 합니다. … 자동화된 시스템에 대한 요약 정보가 쉬운 용어로 포함된 보고서와 더불어, 이들 통지 및 설명의 명확성과 품질에 대한 평가 내용이 가능한 한 공개되어야 합니다.
인간 대안, 검토 및 대체	적절한 경우 여러분은 제외(opt-out)될 수 있어야 하며 여러분이 직면한 문제를 신속하게 검토하고 구제할 수 있는 사람에게 연락할 수 있어야 합니다. … 인간 거버넌스 절차에 대한 설명과 적시성, 접근성, 결과물 및 효과성에 대한 평가 보고서는 가능한 한 공개되어야 합니다.

II. 미국 의회 알고리즘책무성법안 (2022. 2. 3. 발의)[2)]

❍ 일정 규모 이상의 대기업이 배치하는 자동화된 의사결정 시스템 및 AI에 기반한(증강된) 중요한 의사결정 프로세스를 사용하는 경우 FTC가 감독하는 사전적인 위험 영향평가를 수행하고 위험을 완화할 의무를 부과함.

제2절 정의

(8) 중요한 의사결정 - "중요한 의사결정"이란 용어는 다음에 대한 접근 또는 비용, 조건 및 가용성과 관련하여 소비자의 삶에 법적으로나 중요한 또는 유사하게 중대한 영향을 미치는 의사결정 또는 판단을 의미한다.

(A) 평가, 인정, 인증을 포함한 교육 및 직업 훈련

(B) 고용, 근로자 관리, 자영업

(C) 전기, 난방, 수도, 인터넷·통신 접근, 교통과 같은 필수 설비

(D) 입양 서비스, 생식 서비스를 포함한 가족 계획

(E) 모기지 회사, 모기지 브로커, 채권자가 제공하는 금융 서비스를 포함한 모든 금융 서비스

(F) 정신건강의학과, 치과, 안과를 포함한 모든 보건의료

(G) 주택 임대, 단기 주택 임대, 숙박 서비스를 포함한 모든 주택 및 숙박

(H) 사적 중재 또는 조정을 포함한 법률 서비스

(I) 위원회가 규칙 제정을 통해 소비자의 삶에 비교적 법적, 물질적 또는 유사하게 중대한 영향을 미친다고 판단한 서비스, 프로그램 또는 기회 결정

(12) 영향 평가 - "영향평가"란 용어는 자동화된 의사결정 시스템 또는 증강된 중요 의사결정 프로세스 및 이들이 소비자에게 미치는 영향에 대하여 지속적으로 연구·점검하는 것을 말한다.

제3절 자동화된 의사결정 시스템과 증강된 중요 의사결정 프로세스의 영향평가

(b) 규정

(1) 총칙

(D) 각 대상 기업에 대하여, 배치된 자동 의사결정 시스템 또는 증강된 중요 의사결정 프로세스의 지속적인 영향 평가에 대한 요약 보고서를 매년 위원회에 제출할 것을 요구한다.

(E) 각 대상 기업에 대하여, 새로운 자동화된 의사결정 시스템 또는 증강된 중요 의사결정 프로세스에 대한 초기 요약 보고서를 대상 기업이 이를 배치하기 전에 위원회에 제출할 것을 요구한다.

(G) 각 대상 기업에 대하여, (A)호에 기술된 영향 평가를 수행함에 있어 가능한 한, 각 대상 기업이 관련 내부 이해관계자(직원, 윤리 팀 및 담당 기술팀 등) 및 독립적인 외부 이해관계자(영향을 받는 집단의 옹호자나 대표, 시민 사회 및 인권단체, 기술 전문가 등)와 필요에 따라 수시로 의미 있는 협의(참여 설계, 독립 감사 또는 피드백 요청 및 통합)를 할

2) Algorithmic Accountability Act of 2022 (알고리즘 책무성법안), H.R.6580 및 S.3572, SEC.4. 〈https://www.congress.gov/bill/117th-congress/house-bill/6580/text〉.

것을 요구한다.

(H) 각 대상 기업에 대하여, 소비자의 삶에 법적 또는 유사하게 중대한 영향을 미치는 물질적·부정적 영향이 나타나는 경우 증강된 중요 의사결정 프로세스에 의해 발생하는 모든 영향을 시기적절한 방식으로 제거하거나 완화하기 위해 노력할 것을 요구한다.

제4절 대상 기업 영향평가에 대한 요구사항

(a) 영향평가 요구사항

(4) 다음 사항에 대한 문서화를 포함하여 기준(benchmarking) 데이터셋, 대상 기업의 과거 데이터의 대표 예시, 기타 표준과 같은 측정을 사용하여, 자동화된 의사결정 시스템 또는 증강된 중요 의사결정 프로세스의 현재 및 과거 성능에 대하여 지속적인 테스트 및 검토를 수행한다.

(A) 대상 기업이 성공적인 수행과 기법으로 간주하는 사항과, 성능을 평가하기 위해 사용한 기술 및 사업 측정기준(metrics)에 대한 설명

(B) 테스트 조건에서 해당 시스템 또는 프로세스의 성능에 대한 검토, 또는 이러한 성능 테스트가 수행되지 않은 이유에 대한 설명

(C) 배치 조건에서 해당 시스템 또는 프로세스의 성능에 대한 검토, 또는 배치 조건에서 성능이 검토되지 않은 이유에 대한 설명

(D) 배치 조건에서 해당 시스템 또는 프로세스의 성능을 테스트 조건과 비교하거나, 이러한 비교가 불가능한 이유에 대한 설명

(E) 소비자의 인종, 피부색, 성별, 성정체성, 연령, 장애, 종교, 가족 상태, 사회경제적 상태, 병역 퇴역 상태, 또는 위원회가 적절하다고 판단하는 기타 특성(해당 특성의 조합을 포함)과 관련된 모든 차별적인 수행에 대한 검토. 대상 기업이 정보를 갖고 있는 경우, 이러한 검토를 위한 방법론에 대한 설명, 데이터에서 해당 특성을 식별하는 데 사용한 방법(우편번호를 비롯한 대리적인(proxy) 데이터를 사용하는 등)에 대한 정보 및 문서화

(F) 테스트 및 검토에 하위 집단이 사용된 경우, 사용된 하위 집단과 해당 하위집단이 테스트 및 검토와 관련이 있다고 판단된 방법 및 이유에 대한 설명

(7) 자동화된 의사결정 시스템 또는 증강된 중요 결정 프로세스를 개발, 테스트, 유지 관리, 갱신하는 데 사용되는 데이터 및 기타 입력 정보의 최신 문서를 다음과 같이 유지 관리하고 보관한다

(A) 다음 사항을 포함하여 해당 데이터 및 기타 입력 정보를 취득한 시기 및 방법, 해당되는 경우 라이선스가 부여되었는지 여부

(i) 파일 유형, 파일 생성·수정일, 데이터 필드 설명 등 데이터 및 기타 입력 정보의 구조와 유형에 대한 메타데이터

(ii) 대상 기업이 데이터 및 기타 입력 정보를 수집, 추론, 획득한 방법론. 해당되는 경우 대상 기업이 수집, 추론, 획득하기 전 해당 데이터 및 기타 입력 정보가 라벨링, 분류, 정렬, 군집화되었는지 여부를 포함하여 해당 데이터 및 기타 입력 정보에 라벨링, 분류, 정렬, 군집화를 적용한 방법론에 대한 설명

(iii) 소비자가 자신에 관한 데이터 및 기타 입력 정보의 포함 및 추가 사용에 대하여 설명에 입각한 동의를 제공했는지 여부 및 그 방법, 이 포함 및 추가 사용 제한에 대한 규정 사항에 대하여 제공받았는지 여부 및 그 방법

(B) 해당 데이터 및 기타 입력 정보가 사용된 이유 및 다른 대안이 탐색되었는지 여부

(C) 데이터 및 기타 입력 정보에 대한 다음의 기타 정보

(i) 데이터셋의 대표성 및 증강된 중요 의사결정 프로세스가 배치되는 인구집단 분포에 대한 가설 등 해당 요소가 측정된 방법

(ii) 데이터 품질, 그 품질의 검토 방법, 데이터를 정규화, 수정 또는 정제하기 위해 취한 조치토되지 않은 이유에 대한 설명

(8) 소비자의 다음 권리를 검토한다.

(A) 대상 기업이 소비자에게 다음 사항을 제공하는 정도를 평가한다.

(i) 해당 시스템 또는 프로세스가 사용될 것이라는 명확한 통지

(ii) 이러한 사용에서 제외(opt-out)될 수 있는 방법

(B) 다음 사항을 포함하여 해당 시스템 또는 프로세스의 투명성과 설명 가능성을 평가하고, 소비자가 결정에 대해 이의제기·정정·재심을 청구하거나 해당 시스템 또는 프로세스에서 제외될 수 있는 정도를 평가한다.

(i) 그 변경 시 시스템 또는 프로세스가 다른 결정을 내리도록 하는 기여 요인 설명 등 특정 결정에 대한 기여 요인을 포함해 시스템 또는 프로세스에 대해 소비자 또는 소비자의 대표자 및 대리인이 이용할 수 있는 정보. 해당 소비자, 대표자 및 대리인이 해당 정보에 접근할 수 있는 방법

(ii) 해당 시스템 또는 프로세스와 관련하여 소비자가 대상 기업에 제출한 불만, 분쟁, 정정, 재심, 제외 요청에 대한 문서

(iii) 소비자의 우려나 피해를 해결하기 위해 대상 기업이 취한 모든 시정 조치의 과정 및 결과

(C) 제3절(b)(1)(I)(iii)에서 위원회가 정의한 바에 따라, 제3자 결정 수취인이 해당 시스템 또는 프로세스의 결과에 대한 사본을 받거나 이에 접근할 수 있는 범위를 기술하고, 해당 제3자 결정 수취인의 범주를 기술한다.

(9) 자동화된 의사결정 시스템 또는 증강된 중요 의사결정 프로세스가 소비자에게 미치는 중대한 부정적 영향 가능성을 식별하고 적용가능한 완화 전략을 다음과 같이 평가한다.

(A) 해당 영향을 식별하고 측정하기 위해 취한 조치를 문서화하는 등 소비자에게 미치는 시스템 또는 프로세스의 중대한 부정적 영향 가능성을 식별하고 측정한다.

(B) 시장에서 시스템이나 프로세스를 철수하거나 개발을 종료하는 등의 조치를 포함하여 식별된 중대한 부정적 영향 가능성을 제거하거나 합리적으로 완화하기 위해 취한 조치를 문서화한다.

(C) 식별된 중대한 부정적 영향 가능성과 관련하여, 해당 영향이 완화되지 않은 상태로 남아 있는 것과 조치를 취하지 않은 이유를 문서화하고, 이해관계가 비차별적이면서 설득력 있음을 설명하고 해당 이해관계가 다른 수단으로 충족될 수 없는 이유(2명 이상의 소비자에 대한 영향 간에 동등한 제로섬 균형이 있는 경우 또는 필요한 완화 조치가 민권 또는 기타 법률을 위반하는 경우 등)를 자세히 서술한다.

(D) 소비자에 미치는 중대한 부정적 영향 가능성을 식별, 측정, 완화 또는 제거하는 데 사용되는 표준 프로토콜 및 관행을 문서화하고, 관련 부서와 직원이 해당 프로토콜 및 관행에 대해 정보를 제공받고 교육을 받는 방법을 문서화한다.

(12) (1)~(11)호에 서술된 사항 중 시도되었으나 실행이 불가능하여 준수할 수 없었던 영향평가 요구사항을 문서화하고, 해당 요구사항을 준수할 수 없었던 다음의 근거에 대하여 문서화한다.

(A) 다른 사람, 제휴자, 기업이 개발한 자동화된 의사결정 시스템의 특정 정보가 부재함

(B) 대상, 고객, 라이선스 사용권자, 파트너 및 기타 개인, 제휴자 또는 기업이 증강된 중요 의사결정 프로세스에 자동화된 의사결정 시스템을 배치하는 방법에 대한 특정 정보가 부재함

(C) 해당 데이터가 수집, 추론 또는 저장하기에는 너무 민감하기 때문에 차별적 수행을 평가하는 데 필요한 인구집단 및 기타 데이터가 부족함

(D) 기술 혁신을 포함하여 해당 요구사항을 수행하는 데 필요한 특정 기능이 부족함

제9절 집행.

(a) 위원회에 의한 집행.

(1) 불공정하거나 기만적인 행위 또는 관행 - 본 법 또는 이에 따라 공포된 규정의 위반은 연방거래위원회법 제18조(a)(1)(B)에 따라 불공정하거나 기만적인 행위 또는 관행을 정의하는 규칙의 위반으로 간주된다(15 U.S.C. 57a(a)(1)(B)).

(2) 위원회의 권한.

(A) 일반적 사항 - 위원회는 연방거래위원회법(15 U.S.C. 41 이하)의 모든 해당 약관 및 조항이 본 법에 통합되어 본 법의 일부가 된 것처럼 동일한 방식, 동일한 수단, 동일한 관할권, 권한 및 의무로 본 법과 본 법에 따라 공표된 규정을 집행한다.

(B) 특권 및 면책 - 본 법 또는 이에 따라 공표된 규정을 위반하는 사람은 처벌을 받으며 연방거래위원회법(15 U.S.C. 41 이하)에 규정된 특권 및 면책을 받을 자격이 있다.

(C) 권한 보존 - 본 법의 어떠한 조항도 다른 법률 조항에 따른 위원회의 권한을 제한하는 것으로 해석되어서는 안 된다.

III. 유럽연합 AI법안과의 비교 평가[3)]

❍ 미국 법안은 회사가 (i) 자동화된 의사결정 시스템을 배치하기 전 이 시스템과 (ii) 자동화된 의사결정 시스템을 배치한 후 증강된 의사결정 프로세스에 대하여 영향 평가를 수행하도록 규정함. 이는 사전과 사후 평가를 모두 포함하며 유럽연합 법안에서 사전적합성 평가 및 시판 후 모니터링 계획과 대응함

- 다만 영향평가에는 한계가 있음. 특정 위험을 식별 및 완화하지 못하거나 단순 체크박스 절차로 축소될 위험이 있음. 그러나 강력한 제도적 지원이 뒷받침 되어 절차적

3) Mökander, J., Juneja, P., Watson, D.S. et al. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other?. Minds & Machines 32, 751-758 (2022). https://doi.org/10.1007/s11023-022-09612-y.

이고 투명하게 이루어지는 영향평가는 규제기관의 추적 가능성을 보장하고 회사의 규칙 준수를 촉발함.

❍ 미국 법안의 강점

- 유럽연합 법안처럼 '고위험 AI 시스템'이 아닌 '중요한 의사결정 프로세스'를 규제하는 데 초점을 맞춤으로써 AI 시스템이 무엇인지에 대한 존재론적 질문을 피하고의사결정 프로세스에서 자동화 수준은 스펙트럼의 정도 차이로 취급될 수 있도록 함. 인공지능에 대한 특정 기술 종속성을 벗어날수 있어 미래 기술에 대해 확장성이 있음.

- '중요한 의사결정'은 교육, 고용, 금융 서비스 등 소비자의 삶에 중대한 법적 또는 실질적 영향을 미치는 모든 결정에 대하여 (위험성 여부와 무관하게) 적용되기 때문에 열거된 '고위험'에 대해서만 평가를 요구하는 경우보다 유럽연합 법안보다 요구 범위가 넓음. 위험은 인공지능 시스템 단독으로 발생하는 것이 아니라 그로 인해 증강된 의사결정 상황에서 발생할 수 있음.

- 미국 법안은 기존의 의사결정 절차를 자동화된 의사결정 시스템으로 대체하는 이유와 그 이점 자체를 평가 대상으로 삼고 있음. 자동화된 의사결정 시스템이 개인정보 침해 및 차별의 위험이 있다면, 인간의 판단 역시 많은 인지 편향, 편견, 상황적 요인의 영향을 받음.

❍ 미국 법안의 한계

- 미국 법안은 (a) 연간 매출액이 5천만 달러 이상이거나, (b)자본 가치가 2억 5천만 달러 이상이거나, (c) 100만 명 이상의 사용자 정보를 처리하는 '대기업'에만 적용됨. 대기업이 형식적 아웃소싱으로 의무를 회피할 수 있으며, 무엇보다 공공복지기관이나 법집행 기관 등 공공부문이 제외됨.

- 미국 법안이 추구하는 동등성의 가치들이 상호 충돌될 수 있음. 예를 들어 의사 결정 프로세스의 전반적인 정확성을 향상시키는 한편으로 모집단의 특정 하위 그룹을 차별할 위험이 있음. 개인에 대한 동등한 대우, 인구통계학적 동등성, 기회의 평등 역시 상호 배타적일 수 있음. 특정 지표를 평가에서 제외하여 차별적인 결과를 모호하게 만들 수도 있음(배경이 다른 개인들에게 동일한 비율로 대출서비스를 제공하지만, 대출 규모에서는 차별할 수 있음). 대상의 속성에 대한 정보를 수집하지 않는 방식으로 차별금지 의무를 회피할 수 있음. 여러 우회 기술과 지표의 모호성으로 인해 모니터링과 집행상 한계가 있을 수 있음.

- 미국 법안은 중요하고 구체적인 정책 설계를 FTC에 위임함. '가능한 한', '가능한 범위'와 같은 문구는 법안의 취지를 약화시킬 우려가 있음.

- 집행위원회 차원에서 강력하게 추진하는 유럽연합에 비하여 입법 동력이 약한 편임.

IV. FTC, 챗GPT 진정 사건에 조사 착수

❍ 2023. 3. 30. 미국 비영리 연구소 CAIDP(인공지능 및 디지털 정책 센터), FTC에 “오픈AI 회사와 그 제품 챗GPT에 대해 조사를 요청하는” 진정을 제기함. 진정인은 오픈AI에 대해 조사가 이루어지고 필요한 보호장치가 마련될 때까지 추가 모델 출시를 금지해 줄 것을 요청함.

❍ 진정인은 “오픈AI의 사업 관행은 불공정하고 기만적이며, FTC의 AI 관행에 대한 성명, 보고서, 지침은 물론 AI 거버넌스에 대한 최근의 법 규범을 위반”하였다고 주장하며, 크게 ▲편향성, ▲아동 안전, ▲소비자 보호, ▲사이버 보안, ▲기만, ▲개인정보 보호, ▲투명성, ▲공공 안전 측면에서 위험이 있거나 이를 알고 있음에도 오픈AI가 이를 방지하거나 완화하지 않고 일반에 상업적 용도로 챗GPT 서비스를 공개한 것이 위법하다고 지적함.

- 오픈AI는 편향성 위험에 대하여 “우리는 이 모델이 특정 소외 집단에 대해 유해한 고정관념과 비하적인 연상 등 특정한 편견과 세계관을 강화하고 재생산할 가능성이 있음을 발견했습니다.”고 인정함.

- 오픈AI는 사이버 보안 위험에 대하여 챗GPT가 “사회 공학이나 기존 보안 도구의 강화 등을 통해 성공한 사이버 공격에서 특정 단계 비용을 감소시키는 추세를 이어가고 있습니다. 안전 조치가 없다면 GPT-4는 유해하거나 불법적인 활동을 수행하는 방법에 대하여 상세하게 안내할 수 있습니다.”고 인정함.

- 오픈AI는 현재 모델인 GPT-4가 “더 신뢰할 수 있고 더 설득력이 있기 때문에” 기만 위험을 증가시킬 수 있다고 인정함. 또한 “GPT-4는 뉴스 기사, 트윗, 대화 및 이메일 등 그럴듯하게 사실적이고 맞춤한 콘텐츠를 생성할 수 있습니다.”고 인정함. 더불어 GPT-4가 “사실을 만들어 내고, 잘못된 정보를 두 배 늘리고, 잘못된 작업을 수행하는 경향이 있습니다. 또 이러한 경향을 이전 GPT 모델보다 더 설득력 있고 믿을 법한 방식(예: 권위 있는 어조 또는 정밀하고 매우 상세한 정보의 맥락 제시)으로 드러내곤 해서 과의존 위험을 증가”시킬 뿐 아니라 “나쁜 행위자가 GPT-4를 사용하여 오도시키는 콘텐츠를 만들고 따라서 미래 사회의 인식론 일부를 설득력 있는 LLM이 형성할 수 있는 위험을 증가”시킨다고 인정함.

- 진정서는 특히 생성형 AI 모델이 “판매용으로 이를 출시한 회사에서 사전에 식별하지 못한 동작을 하기 때문에 특이한 소비자 제품”이라고 지적하며, 오픈AI가 “‘긴급한 위험 행동’이 발생할 위험성을 인정했지만 그럼에도 불구하고 GPT-4의 상용 출시를 진행하기로 결정”하였다고 비판함.

❍ 결론적으로 진정인들은 FTC에 다음을 요청함.

- 오픈AI가 개발한 상용 GPT의 추가 배포를 중단시켜 줄 것
- GPT 제품의 향후 배포에 앞서 독립적인 평가를 실시할 것을 요구하여 줄 것
- GPT의 추가 배포에 앞서 FTC AI 지침을 준수할 것을 요구하여 줄 것
- GPT AI 수명 주기 전반에 대해 독립적인 평가를 요구하여 줄 것
- FTC의 소비자 사기 신고 체계와 유사하고 공개적으로 이용할 수 있는 GPT-4 사고 신고 체계를 구축하여 줄 것
- 생성형 AI 시장 부문에서 제품 기본 표준을 수립하는 규칙 제정을 추진하여 줄 것
- 기타 위원회가 필요하고 적절하다고 판단하는 구제 조치를 시행하여 줄 것

❍ 2023. 7. 13. FTC가 챗GPT의 소비자보호법 위반 여부를 조사하기 시작했다고 보도됨[4].

V. 미국 인공지능 규제의 향후 전망

❍ 민주당 싱크탱크 〈브루킹스 연구소〉가 2023. 6. 15. "AI 규제의 세 가지 과제"라는 제목의 해설을 발표함.

- 필자는 톰 휠러(2013-2017 오바마 정부에서 FCC 위원 역임)[5]

❍ 2023. 5. 16. 미 상원 법사위원회 AI 청문회에서 OpenAI의 CEO 샘 알트만은 “특정 규모 이상의 모든 노력에 라이선스를 부여하고 해당 라이선스를 박탈하고 안전 표준 준수를 보장할 수 있는 새로운 기관”이 필요하다고 밝힘.

- 마이크로소프트 브래드 스미스 사장도 유사한 입장을 표명함.
- 그러나 AI 규제 논의가 일반 논의에서 유럽연합 AI법안 등 구체적인 규제 구현으로 옮겨가자 알트만은 유럽연합 AI법안에 반대하는 입장을 발표하였음(이틀뒤 트위터에서 이를 번복함).
- 결국 AI 규제법의 세부사항이 중요함(The details really matter).

---

4) 조선일보 보도(2023. 7. 14). 美FTC, 챗GPT 개인정보 유출 등 소비자보호법 위반 여부 조사. 〈https://n.news.naver.com/mnews/article/001/0014065852?sid=104〉.

5) Wheeler, Tom (2023). The three challenges of AI regulation. Brookings Commemtary (2023. 6. 15). 〈https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/〉.

❍ 해설은 AI 규제의 세부 쟁점이 AI 감독의 세 가지 문제를 다루어야 한다고 지적함.

- 이는 (1) AI 개발 속도 대응 (2) 규제 대상 분석 (3) 규제 방법 결정.

❍ 해설은 (1) AI 개발 속도에 다음과 같이 대응해야 한다고 봄.

- AI는 한동안 무대 뒤에서 조용히 진화해 왔음. Google이 검색어를 자동 완성하거나 Amazon이 책을 추천할 때 AI가 작동하고 있음.

- 그러나 2022년 11월 ChatGPT-3가 출시되면서 AI는 소프트웨어 엔지니어를 위한 그림자 도구에서 소비자 상품으로 폭발적인 관심 대상이 되었음.

- 마이크로소프트, 구글, 메타 등 빅테크들 간에 AI 경주가 시작됨.

- 관건은 심판 없이 진행되는 역사상 가장 빠른 경주에서 공익을 어떻게 보호하느냐임. 기업 AI 경쟁이 무모해지는 것을 막으려면 규칙의 수립과 개발, 법적 가드레일의 시행이 필요함.

- 그러나 AI 변화 속도가 기존 연방 정부의 전문성과 권한을 능가할 수 있음. 현재 정부 규제 법령 및 규제 구조는 산업 시대를 가정하고 구축되어 있음. 즉, 기존 규칙은 AI 개발 속도를 처리하기에 민첩성이 부족함.

- 반면 자율규제 접근법은 지난 20년 간 디지털플랫폼 빅테크 기업들이 반복해 온 '내버려둬' 주장의 반복임. 그 결과 우리 사회는 전례 없는 개인 사생활 침해, 시장 집중, 사용자 조작, 혐오, 허위, 조작 정보의 유포 등 온라인 위해에 속수무책인 상황. 이윤 추구 경향이 실효성 있는 가드레일 구축을 앞지를 수 있다는 사실을 알기 때문에, AI에 대해서는 기업의 자율 규제보다 더 나은 것을 필요함.

- 빅테크가 사실상 사이버 정부가 되어 AI에 대한 자체 규칙을 만들도록 방관한다면 과거 온라인 플랫폼에 대한 정책적 실패를 반복하게 됨. 리처드 블루먼솔 상원의원(민주당)은 “의회는 소셜 미디어의 시간을 맞추지 못했습니다. 이제 우리에게는 AI의 위협과 위험이 현실화되기 전에 그 시간을 맞추어야 할 의무가 있습니다.”라고 밝힘.

- 속도 문제에 대처하는 것은 집중력과 민첩성의 문제임. 규제기관이 AI 문제를 전면에서 집중적으로 다룰 수 있어야 하고, 미세규제관리 방식보다 기술 발전 속도에 민첩하게(agile) 대응할 수 있어야 함.

❍ 해설은 (2) 규제 대상을 다음과 같이 분석함

- AI는 다면적인 기능을 가지고 있기 때문에 "일률적인" 규제는 어떤 경우에는 과잉규제, 어떤 경우에는 과소 규제가 될 수 있음. 예를 들어 온라인게임에서 사용되는 AI는 중요 인프라의 보안을 위협하거나 인간을 위험에 빠뜨릴 수 있는 AI와 다른 영향을 미치며 다르게 취급되어야 함.

- 따라서 AI 규제는 위험 기반 접근법을 취해야 하고 표적화되어야 함. 이는 (a)전통적인 오남용 문제, (b)디지털 남용 문제, (c)AI 그 자체를 다루는 문제로 나누어 접근할 수 있음.

(a) 전통적인 오남용 문제는 AI를 전통적인 불법 활동에 적용하는 경우임.

- AI로 인해 소비자 기만과 범죄가 전례없는 규모와 정교한 수준에 이를 수 있음.

- FTC는 다음과 같이 사실적인 대화형 AI의 기만 문제에 대해 경고하고 대응하기 시작함(3초 오디오클립을 AI로 가공하여 가족을 사칭함).

- EEOC(미국 고용평등기회위원회)는 고용관계에서 " 채용, 업무 모니터링, 급여 또는 승진 결정"에 AI 모델을 사용하는 것이 연방법을 위반하는 차별적인 결과를 초래할 수 있다고 경고함.

- 법무부는 주택 임대에서 AI에 기반하여 세입자를 심사하고 선정하는 것이 불법적이라고 경고함.

- 이와 같은 전통적인 오남용 문제에는 전통적인 규제 수단을 적용하는 것이 가능함.

- 바이든 행정부는 FTC, EEOC, 법무부, CFPB(소비자 금융 보호 위원회) 등 미국 소비자 대면 규제기관 4곳을 모아 AI 문제에 기존 법령을 적용하는 집중 이니셔티브를 발표함.

- 리나 칸 FTC 위원장은 "현행법상 AI에 대한 면책 조항은 없습니다(There is no AI exemption to the laws on the books.)"고 밝힘.

(b) 디지털 남용 문제는 빅테크의 침해적 관행이 AI로 악화되는 문제임.

- 빅테크 독과점 디지털 기업이 개인 정보를 광범위하게 수집하고, 이를 기업 자산화하여 시장을 통제하고, 시장 지배력을 사용하여 소비자가 보는 정보를 통제하는 상황이 악화될 수 있음.

- 또한 AI의 현재와 미래 역시 유사한 데이터의 광범위한 수집과 활용 경로를 밟고 있음.

- 2023년 봄 현재, GPT-4는 GPT-3보다 6배 많은 1조 개의 파라미터를 가지고 있음. 학습을 위해 온라인 사용자의 글, 동영상, 발언 등 방대한 양의 데이터가 수집되었음. 구글, 메타 등 기존에 개인정보 권력을 남용해온 빅테크가 AI에 뛰어들면서 우리의 영상 및 음성 등 새로운 개인 정보 침해의 기반이 되고 있음.

- 방대한 데이터 비축량을 가진 기업들의 독과점 이점이 커지고 있음.

- 빅테크가 책임을 회피하는 허위, 조작 정보가 거짓 이미지, 오디오, 텍스트를 생성하는 AI로 기하급수적으로 확대될 수 있음.

(c) AI 그 자체 문제는 AI의 주의 의무, 투명성, 안전성, 책임성을 확보하는 문제임.

- 기업은 주의 의무를 이행할 책임이 있음. 이는 상품 또는 서비스 제공자가 잠재적인 악영향을 식별하고 완화할 책임이 있음을 의미하며, 이러한 의무를 이행하지 않으면 규제를 포함한 법적 조치가 취해질 수 있음.

- 투명성은 AI의 지속적인 위험을 식별하고 완화하기 위해 필요함. 모델 작동 방식이 불가해하여 AI 공급자조차도 자신이 만든 것이 무엇인지 정확히 알 수 없는 경우에 대응할 수 있어야 함. 학계, 정부, (AI의 대상이 되는) 시민들이 모델에 접근하여 AI가 야기하는 새로운 위협을 추적할 수 있어야 함. 소비자가 AI와 상호작용하고 있는 사실과 학습 데이터 소스를 공개하고 영상 및 음성이 AI 산출물이라는 사실을 밝히는 것은 소비자 혼란을 방지할 수 있음. 투명성은 알고리즘 편향을 완화하는데 도움이 될 수 있음. 뉴욕시는 회사가 채용 지원자에게 AI 사용 사실을 밝히고 제3자 감사를 받도록 신규 규제를 시행함.

- 안전성은 투명성의 결과이자 그 자체로 추구해야 할 원칙임. AI 안전성 기준은 NIST(미 국립표준원)가 〈AI 위험 관리 프레임워크〉에서 밝힌 바 있음.

- 책임성은 백악관 〈AI 권리장전〉 청사진의 핵심임.

- [미국에서] 주의 의무는 집행가능해 보임. 투명성, 안전성 및 책임성은 규제가 없다면 이상에 그치게 됨.

❍ 해설은 (3) 규제 방법을 다음과 같이 제시함.

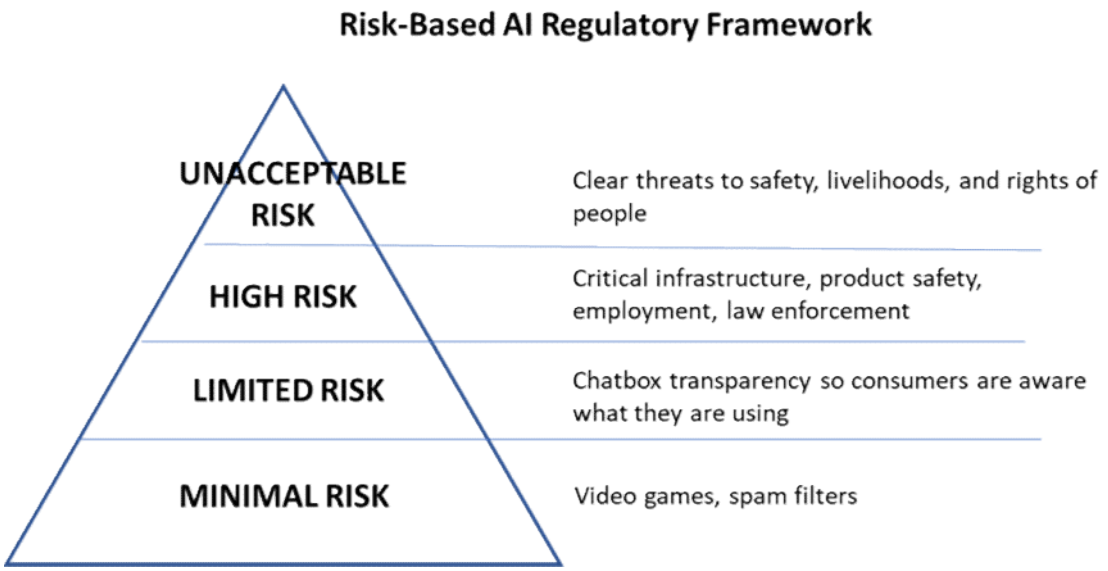
- 유럽연합은 GDPR, 디지털시장법, 디지털서비스법에 이어 AI법에도 세계적인 주도권을 발휘하고 있음. 6월 14일 유럽의회에서 협상안이 통과됨. 미국도 패스트 팔로워가 되기 위해 준비하여야 함.

- 규제 주체와 관련하여, 오픈AI, 마이크로소프트, 구글은 AI 감독이나 규제를 전담하는 연방기구를 지지함. 상원 청문회 후 마이클 베넷과 피터 웰치 상원의원은 신규 디지털 플랫폼 위원회(DPC)를 설립하는 법안을 제출함. 린지 그레이엄과 엘리자베스 워런 상원의원도 디지털 기구에 대한 법안을 준비 중임.

- 규제 방법과 관련하여, 오픈AI와 마이크로소프트는 허가(라이선스) 방식을 지지함. 미국은 방송, 통신, 핵물질, 원자로, 시추, 어업, 항공 등의 영역에서 라이선스 방식을 운영하여 왔음. 그러나 라이선스를 규제 감독 수단으로 사용하면 라이선스를 받는 사

람들의 지위를 강화할 우려가 있음. 따라서 AI 분야 지배적인 기업으로서는 라이선스 방식을 선호할 수 있음. 이는 독과점 상업 시장의 진입장벽 문제를 라이선스 권한 문제로 치환함.

- 무엇보다 위험에 기반한 민첩한 규제 체제를 갖출 필요가 있음.



- 민첩한 규제 기관은 투명하고 반응적이며 민첩한 접근 방식으로 행동 권장 표준을 개발할 필요가 있음. 이는 금융산업규제당국(FINRA)이 SEC(증권거래위원회) 감독 하에 금융시장을 규제하거나 북미전력안정성기구(NERC)가 FERC(연방에너지규제위원회) 감독 하에 정전방지에 대처하는 방식임. ▢

인공지능과 소비자 권리구제

허유경

(소비자시민모임 이사, 변호사)

1. 서론

현재 국회에서 논의 중인 인공지능법안은 “우선허용·사후규제 원칙”을 채택하여(법안 제11조), 인공지능 기술, 제품 또는 서비스가 국민의 생명·안전·권익에 위해가 되거나 공익 등을 현저히 저해할 우려가 있는 경우가 아니면 이를 제한할 수 없도록 함으로써 **사전규제를 엄격히 제한**하고 있다. 한편, 동 법안은 “사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 영역에서 활용되는 인공지능”을 “고위험영역 인공지능”으로 정의·분류하는 외에 다른 분류 규정을 두고 있지는 않다.

위와 같은 규정들을 종합해보면, “고위험영역 인공지능”에 해당하지 않는 영역에서는 소비자 피해를 **사전적으로 예방하거나 관리하기 어렵고,** 인공지능 시스템이 소비자에게 제공된 후에 발생한 피해에 대해서 **사후적인 피해구제만 가능하다**는 결론에 이른다.

그러나 인공지능의 블랙박스와 같은 특징과 현행 법제도의 불완전성 등을 고려할 때, “우선허용 원칙"에 따라 이미 개발·제공된 인공지능으로 인하여 발생한 손해를 입증하고 구제받는 것은 사실상 불가능하지 않을까 우려된다. 인공지능이 소비자 편익 및 효용 증대 등 소비자 권익에 기여하는 측면도 있지만, **인공지능 시스템이 불가역적이고 광범위한 소비자 피해를 발생시킬 가능성**이 있는 점을 고려하면 **소비자 보호 및 소비자 권익을 반영하는 보완 입법이 필요**하다.

이하 본 토론문에서는 해외 소비자 및 인권 단체에서 제기한 인공지능으로 인한 소비자 피해 유형을 검토한 후, 현행 법제도 하에서는 그에 대한 소비자 피해구제 및 입증이 어렵다는 점을 설명하고자 한다.

2. 편향성 및 차별적 취급의 문제

(1) 인공지능의 편향성 및 차별적 취급 유형 및 사례

소비자 보호의 핵심은 모든 소비자를 공정하고 평등하게 대우하는 것이다. 인공지능을 비롯한 기술 자체는 인식 및 고의가 없으므로 중립적이라는 견해도 있지만, 인공지능의 영역에서도 학습데이터 및 알고리즘의 편향성, 조작자의 오류 등으로 인해 불공정성, 편향성이 존재할 수 있다는 점은 이미 국내외 다수 연구를 통하여 밝혀진 바 있다. 즉, 인공지능은 학습데이터에 기반하여 작동되므로 데이터가 편향되면 편향된 결과가 생성될 수 있고, 개발자가 데이터를 분류·선택하는 과정에서도 편향성과 차별이 발생할 수 있다.

소비자 영역에서 알고리즘 및 인공지능의 적용으로 발생한 차별 및 편향성에 관한 구체적인 연구 사례는 다음과 같다.[1)]

- **글로벌 온라인 데이팅 애플리케이션 틴더의 가격차별 정책**: 2022년 한국의 소비자시민모임과 해외 소비자단체들이 국제소비자기구(Consumers International, CI) 및 미국의 모질라재단과 공동으로 조사한 결과, 글로벌 온라인 데이팅 애플리케이션 틴더(Tinder)의 개인별(나이 등)·국가별 '가격 차별'이 심각하다는 사실을 밝혀냈다. 예컨대, 한국의 연령별 틴더 평균 가격을 비교하면 18~29세 연령층의 평균 가격(9.03달러)은 30~49세 평균 가격(18.11달러)이나 50세 이상 평균 가격(18.85달러)의 절반 정도였다. 위 조사결과가 공표된 후 틴더 측은 연령 기반 가격정책을 중단키로 결정하였다.
  - 위 조사 결과에 대하여 국제소비자기구는 "전 세계적으로 소비자들은 알고리즘을 사용한 개인화된 가격 책정의 '잠재적 위험'에 대해 우려하고 있다"며 "기업의 개인정보 활용에 대한 투명성을 높이고, 소비자의 알 권리를 보장하며, 기업의 비합리적 차별을 방지하기 위한 조치가 마련돼야 한다"고 지적하였고, 소비자시민모임은 "기업은 소비자가 정확한 정보에 입각한 결정을 내릴 수 있도록, 개인화된 가격 책정 알고리즘에 대해 투명하게 공개해야 한다"면서 "소비자단체와 감독기관은 가격이 공정한지 확인하기 위해, 알고리즘에 대한 접근 권한을 부여해야 한다"고 강

1) 일반적인 내용으로는 정원섭, “인공지능 알고리즘의 편향성과 공정성”, 인간·환경·미래 제25호 (2020).

조한 바 있다.

- **네이버의 검색 알고리즘 조작 사례**: 2020년 10월 6일 공정거래위원회는 네이버(주)가 자사의 검색 알고리즘을 인위적으로 조정·변경한 것은 독점규제 및 공정거래에 관한 법률(이하 공정거래법)상 시장지배적지위 남용행위 중 거래조건 차별행위 및 불공정거래행위 중 부당한 차별취급행위 , 부당한 고객유인행위를 한 것으로 보고, 시정 명령과 함께 약 267억의 과징금을 부과한바 있고, 2022년 12월 법원에서 동 과징금 부과가 정당하다는 판결을 내렸다(서울고등법원 2021누36129 판결, 현재 상고심 계류중[2]). 위 사건의 특이점은 공정거래위원회 측에서 네이버의 알고리즘 조작을 밝히기 위하여 "네이버 직원들의 이메일과 각종 회의자료"는 물론 검색 알고리즘을 분석해, 네이버가 제휴 업체에만 가중치를 부여하는 등 알고리즘을 조작한 사실을 확인하였다는 점이다.[3]

- **Chat GPT와 같은 생성적 인공지능에서 발생하는 차별적이고 편향적인 내용**: 워싱턴 포스트가 조사한 바에 의하면 구글 및 메타의 언어모델(LM)에 활용되는 구글의 C4의 학습 데이터에는 위키피디아, 레딧, 언론기사, 정부 웹사이트 등 다량의 인터넷 공개 정보가 포함되어 있는바, 이는 인종차별, 음란, 혐오, 광고 등의 내용이 여과없이 "학습"되어 생성적 인공지능 모델에 반영된다는 것을 의미한다. 예컨대, 생성적 인공지능을 통해 만든 그림은 유색인종 여성을 성적인 대상으로 표현하거나, "아프리카 노동자"는 육체 근로자로, "유럽 노동자"는 사무직 근로자로 표현하였고, 구글 및 메타의 이미지 인식 알고리즘은 피부색이 어두운 사람을 고릴라로 인식하여 비판받기도 하였다.[4]

- **미국 연방 공정거래위원회의(FTC) 인공지능 알고리즘에 대한 경고**: 미국 FTC는 2021년 4월 기업이 소비자에게 실질적인 영향을 미칠 수 있는 방식으로 인공지능 기술을 사용하는 것을 주시하고 있다고 밝혔다. FTC에 의하면, 의료 분야에서 적용된 알고리즘이 보건의료 분야에 존재하는 흑백 인종차별을 고착화하는 사례를 들며, 사업자들은 인공지능에 있어서 진실성, 공정성 및 평등을 추구해야 한다고 강조하였다.[5]

---

2) 법률신문, 비교쇼핑 검색 알고리즘 조작 혐의' 네이버에 266억 과징금 부과 정당, 2022. 12. 15.
3) 조선비즈, [로펌의기술](96) "네이버, 스마트스토어 밀어주기 위해 알고리즘 조작" … 공정위 '266억 과징금' 부과 정당 입증한 지음 법률사무소, 2023. 1. 11.
4) Forbrukerradet, Ghost in the Machine, (2023).
5) FTC, Aiming for truth, fairness, and equity in your company's use of AI | Federal Trade Commission, (2021).

(2) 관련 법제의 미비 및 입증책임의 어려움

공정거래법은 제23조에서 불공정거래행위를 금지하면서, 그 유형의 하나로 “부당하게 거래를 거절하거나 거래의 상대방을 차별하여 취급하는 행위”를 규정하고 있다 (제1호, 차별적 취급금지). 공정거래관련 법령은 소비자와 연관성이 높은 차별적 취급금지의 유형을 “불공정거래행위의 유형 및 기준"에서 세분화하고 있는데, 이에 따르면 “부당하게 거래지역 또는 거래상대방에 따라 현저하게 유리하거나 불리한 가격으로 거래하는 행위"는 금지된다(동법 시행령 [별표 2]).

이러한 불공정거래행위가 성립하기 위해서는 **“부당성" 및 “현저성" 요건이 요구된다**. 즉, 인공지능 시스템을 이용한 소비자 차별행위가 불공정거래행위에 해당하려면 단순히 윤리적 비난가능성을 넘어 법이 금지하는 정도의 “부당성"이 입증되어야 한다. 우리 대법원은 부당성과 관련하여, “행위로서 그에 대한 **의도와 목적**이 있었다는 점을 입증하여야” 하고,[6] “부당성 유무를 판단할 때에는 당사자의 거래상 지위 내지 법률관계, 상대방의 선택 가능성·사업규모 등의 시장상황, 그 행위의 목적·효과, 관련 법규의 특성 및 내용 등 여러 사정을 고려하여 그 행위가 공정하고 자유로운 경쟁을 저해할 우려가 있는지 여부에 따라야 한다”고 판시 한 바 있다.[7]

따라서, 현행법상 “부당성” 및 “현저성"의 입증책임은 원칙적으로 공정거래위원회에 있는데, 인공지능의 “블랙박스”와 같은 특징, 인공지능은 법인격을 가지지 않은 점 등을 고려할 때 이를 입증하는 것은 사실상 불가능에 가까울 것으로 예상된다.[8] 해외 인공지능 연구자들도 인공지능이 대규모 데이터셋을 학습하는 과정을 기록, 추적할 방법이 없을 경우에는 그에 대한 책임을 지울 방법도 없다고 지적하고 있다.[9]

한편, 국가인권위원회법도 차별적 취급을 금지하면서 “차별금지사유”[10] 및 “차별

---

6) 대법원 2007. 11. 22. 선고 2002두8626 판결.

7) 대법원 2010. 8. 26. 선고 2010다28185 판결.

8) 서완석, “인공지능에 의한 소비자권익 침해에 관한 유형과 법적 과제" 상사법연구, 제37권 제98호(2018), 363면.

9) "웹에서 수집된 대규모 선별되지 않은 정적 데이터셋으로 학습한 LM[언어 모델]은 소외된 사람들에게 해로운 지배적 관점을 내포하고 있다. 이에 우리는 LM 학습 데이터를 선별하고 문서화하는 데 상당한 자원을 투자해야 할 필요성이 있음을 강조한다. 큰 데이터셋에 의존하면 할수록 문서화 부족이 발생할 위험이 있다. 데이터셋이 문서화되지 않거나 사후에 문서화하기에는 너무 커진 상태일 수 있는 것이다. 문서화는 책임을 부과할수 있게끔 하는 반면, 문서화되지 않은 학습 데이터는 아무런 보상 없이 피해를 지속시킨다. 문서화가 이루어지지 않는다면, 확인된 문제 중 일부 또는 채 알려지지 않은 문제를 완화하기 위해 학습 데이터의 특성을 이해하려는 시도조차 할 수 없다." (CAIDP-FTC-Complaint-OpenAI-GPT, 제43항)

10) 국가인권위원회법 제2조 제3호는 구체적인 “차별금지사유”로 성별, 종교, 장애, 나이, 사 회적 신분, 출신지역(출생지, 등록기준지, 성년이 되기 전의 주된 거주지역 등을 말한다), 출신국가, 출신민족, 용모 등 신체조건, 기혼·미혼·별거·이 혼·사별·재혼·사실혼 등 혼인 여부, 임신 또는 출산, 가족형태 또는 가족 상황, 인종, 피부색, 사상 또는 정치적 의견, 형의 효력이 실효된 전과, 성적 (성적) 지향, 학력, 병력(병력) 등 19가지를 규정하고 있다.

금지영역”에 대하여 규정하고 있다. 차별금지영역의 경우, 고용(모집, 채용, 교육, 배치, 승진, 임금 및 임금 외의 금품 지급, 자금의 융자, 정년, 퇴직, 해고 등을 포함한다)과 관련하여 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위, 재화·용역·교통수단·상업시설·토지·주거시설의 공급이나 이용과 관련하여 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위, 교육시설이나 직업훈련기관에서의 교육·훈련이나 그 이용과 관련하여 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위, 성희롱 행위 등 4개의 영역을 규정하고 있다(동법 제2조 제3호). 그에 따르면 사업자와 소비자 간의 재화·용역의 공급 및 이용도 “차별금지영역”에 해당한다. 그러나 이러한 소비자 차별 관련 규정들은 선언적 규정에 불과하고, 실제로 차별행위를 이유로 사업자를 규제하기 위해서는 국가기관 또는 피해자가 사업자들의 개별행위를 조사·분석해야 하는바, 앞서 살펴본 인공지능 시스템의 특징 등을 고려할 때 이 또한 현실적으로 불가능해 보인다.[11)]

미국의 소비자보호법령도 금융, 교육, 고용, 주택, 여행을 비롯하여 미국 경제의 주요 부문에서 편향 및 차별적인 취급을 금지하고 있는데, 동법은 “행위중심” 및 “귀책중심” 규정을 채택하고 있어 인공지능으로 인한 차별적인 취급과 그로 인한 손해의 입증이 어렵다. 그와 관련하여 인과관계, 합리적 예측가능성, 리스크 관리체계의 유무에 따라 손해배상책임을 부과하는 것은 소송에서 인공지능의 편향성을 입증하기 어렵게 하므로 관련 법제의 정비가 필요하다는 견해도 있다.[12)]

소비자 및 감독당국 등이 인공지능에 의한 행위의 인과관계 등을 증명하기 어렵다는 점을 고려하면, 인공지능에 의한 불공정거래행위에 대해서는 사업자가 위법성이 없다는 것을 증명하도록 하는 입증책임의 전환이 필요하다.[13)]

나아가 소비자 및 사업자 간의 거래관계에 인공지능이 활용될 경우, 이는 인공지능법안이 정한 “고위험 영역"에 해당하지 않은 경우가 많을 것이다. 그러나 앞서 살펴본 것처럼 가격차별 행위 등 “고위험영역 인공지능” 아닌 영역에서도 인공지능으로 인한 소비자 피해가 발생할 수 있음에도 현행법 체계로는 규율이 어려운 상황이므로, 인공지능법안에 대응 방안 등을 반영될 필요가 있다고 판단된다.

---

11) 서완석, 363면.

12) Jason Jia-Xi Wu, “Algorithmic Fairness In Consumer Credit Underwriting: Towards a “Harm-Based” Framework for AI Fair Lending,” Berkeley Business Law Journal (2023).

13) 서완석, 363면.

3. 인공지능의 조작적(manipulative) 내용 문제

(1) 인공지능의 조작적 내용, 오류, 안전성 등이 문제된 사례

최근에 생성적 인공지능이 등장한 이후, 인공지능의 조작·오인·오류 가능성과 그로 인한 소비자의 신체·생명·재산상 위해 등에 관한 논의가 활발하다. 이미 국내외에서 사업자와 소비자 간의 상거래 전반에 챗봇의 도입이 보편화되었는데, 특히 해외에서는 챗GPT(Chat GPT) 등 인공지능을 사용한 챗봇이 고령자, 청소년 등 취약한 소비자 계층에 대한 위해 가능성이 높다고 보고되고 있으며, 심지어 "병적인 거짓말쟁이(pathological liar)" 또는 "자신감 있게 거짓말을 하는 사람(confident bullshitter)"이라는 비판을 받을 정도로 허위의 내용을 생성하는 경우도 있다.[14)]

생성형 인공지능은 기술이 발달될수록 보다 세련되고 권위있는 언어적 기법을 사용할 수 있게 되어, 전혀 사실에 기반하지 않는 내용으로도 소비자를 오인케 할 수 있다. 또한 인공지능을 활용한 챗봇은 인간과 유사한 태도 및 감정을 가진 것으로 오인될 수 있어, 허위 내용의 유포, 기만적인 판매 행위 등으로 악용될 소지가 있다.[15)]

이러한 이유로 유럽 32개국 46개 독립 소비자단체의 연합체인 BEUC는 챗GPT가 소비자에게 미치는 영향이 커지고 있다고 경고하면서 생성형 인공지능 시스템에 대한 적절한 규제가 필요하다고 강조하였다. 나아가 BEUC는 기업이 자진하여 적절한 조치를 취하지 않을 경우 관련 당국이 이에 대하여 조사 및 통제를 할 것을 요구하였다. 즉, 소비자들이 생성형 인공지능의 현실적 위협을 당면하고 있으므로, 소비자 피해 방지를 위한 즉각적·심층적 평가가 필요하다고 강조하고 있다.

BEUC 등 해외 소비자 단체에서 소개한 알고리즘 및 인공지능의 조작적 내용 사례는 다음과 같다.

- **금융부문에서 로보어드바이저 등 잘못된 투자 조언의 가능성**: BEUC은 일련의 트윗에서 잘못된 금융 자문으로 인한 위험에 대해 설명했는데, 예를 들어, "챗GPT가 금융 부문에 출시되어 소비자에게 투자 또는 부채 관리에 대해 조언하기 시작할 때… 나쁜 조언으로 소비자에게 재정적으로 부정적인 결과가 초래되는 것을 막을 방법이 있습니까?"라고 지적하였다.

- **신용평가, 보험 인수 등에서 인공지능의 오류 또는 차별적 취급**: BEUC은 또한 "챗GPT가 소비자 금융 또는 보험 평가에 사용되는 경우, 불공정하고 편향적인 결과를 생성하여 특정 유형의 소비자에 대해 금융서비스 접근을 차단하거나 건강보험이나 생명보험 가격을 인상하는 일을 방지할 방법이 있습니까?"라고 물어 금융부문에서 인공

14) Forbrukerradet, Ghost in the Machine (2023), 2.2.1 항.
15) Forbrukerradet, Ghost in the Machine (2023), 2.2.1 항, 2.2.2 항.

지능의 오류 및 편향성의 가능성을 지적하였다. [16]

- **기만적인 광고 및 판매 행위의 가능성**: BEUC은 "챗GPT가 기존의 챗봇을 대체한다면 그 판매 포인트는 더 '인간적'이고 신뢰할 수 있을 것처럼 보인다는 데 있습니다. 그러나 챗GPT는 소비자를 속이고 하지 않았을 구입을 하도록 소비자를 압박한다는 점을 지적하여 무엇을 살지 조언을 구하는 사람에게 미치는 영향력을 생각해 보십시오"[17] 라고 하여 인공지능을 활용한 기만적인 광고 및 판매행위의 위험성을 지적한 바 있다.

- **"인간적인" 챗봇과의 대화하여 자살한 사례**: BEUC는 벨기에에서 **30대 남성이 인공지능 챗봇의 "부추김" 을 받아 스스로 목숨을 끊었다며**, 챗본이 공공의 안전에 영향을 미칠 수 있다고 지적한바 있다.[18] 워싱턴 포스트는 Replika와 같은 "친구형 챗봇(Companionship bots)"이 인간적 연결과 유사한 감정 또는 동반자적 관계를 만들어 이용자에게 위로를 준다고 보도하였고,[19] 생성형 인공지능의 사용자들이 '인공지능 답변'과 '사람의 답변'을 구별할 수 없었다는 연구결과도 있다.[20] 이러한 이유로 인공지능 챗봇이 1인칭 대화법 또는 인간과 유사한 이모티콘을 사용하는 것을 금지시켜야 한다는 견해도 있다.

- **미성년자에 대한 부적절한 컨텐츠가 노출된 사례**: 생성형 인공지능은 미성년자와 같은 취약 계층에 더욱 심각한 위해를 초래할 가능성이 있다. 이에 이탈리아에서는 대화형 인공지능 챗봇인 Replica에 미성년자인지 여부를 확인하는 '연령 확인 기능'이 없고 미성년자에게 부적절한 내용을 노출할 수 있다는 이유로 해당 앱이 금지된 바 있고,[21] 미국 등 해외에서 널리 사용 중인 메신저 Snapchat에서 출시한 챗봇은 13세 미성년자에게 성행위, 주류 및 마약에 대한 조언을 하여 크게 비판받은 바 있다.[22]

---

16) CAIDP-FTC-Complaint-OpenAI-GPT, 제56항.
17) CAIDP-FTC-Complaint-OpenAI-GPT, 제57항.
18) BEUC, 2023. 4. 12. 진정서, https://www.beuc.eu/sites/default/files/publications/BEUC-X-2023-046_BEUC_concerns_over_AI_and_mental_health_%20Ms_Pinuccia_Contino.pdf .
19) Washington Post, They fell in love with AI bots. A software update broke their hearts, 2023. 3. 30.
20) Forbrukerradet, Ghost in the Machine, 2.2.2 항.
21) https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9852506#english
22) Forbrukerradet, Ghost in the Machine, 2.2.2 항.

(2) 관련 법제의 미비 및 입증책임의 어려움

사업자가 인공지능을 활용하여 조작적 내용, 허위 또는 기만적 내용, 소비자의 안전성에 위해가 되는 행위 등을 하면 기본적으로 **표시광고의 공정화에 관한 법률(이하 표시광고법) 및 제조물책임법의 규율 대상**이 될 것이지만, 관련 법규의 입법취지 및 판례의 태도 등에 비추어 볼 때 **현행 법체계 하에서는 소비자가 인공지능으로 인한 손해를 배상(피해구제)받기는 어려울 것으로 예상**된다.

제조물책임법의 적용 여부에 관하여 먼저 살펴보면, 인공지능을 제조물책임법 제2조 제1호에서 말하는 "제조물"(제조되거나 가공된 동산(다른 동산이나 부동산의 일부를 구성하는 경우를 포함한다)로 볼 수 있는지 여부가 문제된다. 이에 대해서는 국내 연구자들 사이에 견해의 대립이 있는데, 인공지능을 '유체물'이 아니라 '서비스'로 볼 경우 현행법상 인공지능의 제조물성이 인정되기는 어려워 보인다.[23]

나아가, 제조물책임법상 손해배상 책임이 인정되기 위해서는 동법 제2조 제2호의 "결함" (제조·설계 또는 표시상의 결함이 있거나 그 밖에 통상적으로 기대할 수 있는 안정성이 결여된 것)이 존재해야 하는데, 우리 대법원 판례는 "물품을 제조·판매하는 제조업자는 그 제품의 구조·품질·성능 등에 있어서 **그 유통 당시의 기술수준과 경제성에 비추어 기대 가능한 범위 내의 안전성과 내구성을 갖춘 제품**을 제조·판매하여야 할 책임이 있고, 이러한 안전성과 내구성을 갖추지 못한 결함으로 인하여 소비자에게 손해가 발생한 경우에는 불법행위로 인한 손해배상의무를 부담"한다고 판시하여 결함의 범위를 제한하고 있다.[24] 생성형 인공지능이 현재 개발 중인 신기술인 점을 고려하면, 그로 인하여 소비자에게 손해가 발생하여도 제조물책임법상 "결함"이 인정되기는 어려워 보인다.

한편, 현행 표시광고법상 금지되는 "거짓·과장의 표시·광고" 또는 "기만적인 광고"는 허위 또는 기만적인 내용만으로 그 요건이 충족되고 허위의 인식이라는 주관적인 의도는 필요하지 않으므로, 인공지능으로 허위 표시광고 내용을 적시할 경우에도 동법이 적용된다고 볼 여지가 있다.

그러나 동법이 적용되기 위해서는 동법 제2조 제1호 또는 제2호의 "표시" 또는 "광고" (상품의 내용을 소비자에게 널리 알리거나 제시하는 것)에 해당되어야 하는데, 생성형 인공지능을 적용한 챗봇의 대화 내용을 광고로 의율하기는 어려울 것이므로, 결국 표시광고법으로 위와 같은 소비자 피해 사례를 규율하기는 어려워 보인다.

---

23) 서완석, 355-357면.

24) 대법원 2004. 3. 12. 선고 2003다16771 판결.

4. 인공지능의 반경쟁적 효과

(1) 인공지능의 반경쟁적 사례

인공지능이 소비자 후생에 악영향을 미칠 수 있는 또 다른 유형으로 인공지능을 활용한 디지털 카르텔 및 생성적 인공지능의 반경쟁 효과를 들 수 있는바, OECD[25] 및 미국 FTC에서도 **인공지능의 반경쟁적인 위험성에 대하여 빈번하게 지적**해왔다.

- **인공지능을 활용한 카르텔**: 인공지능을 활용하여 담합을 할 경우, 경영자들은 담합의 시작 등을 위한 명시적인 의사소통을 피할 수 있고, **반경쟁적 가격이 지속될 수 있는 경우에도 자기학습 알고리즘을 통해 공동이익을 최대화하면서 소비자 피해를 극대화하는 가격을 더욱 쉽게 결정할 수 있게 된다.**[26] 즉, 기업들은 **인공지능의 광범위한 이용을 통해 공식적 계약이나 인적 상호작용 없이 담합을 하고 이를 유지하는 것이 용이**해진다.
- **생성적 인공지능의 반경쟁 효과**: 최근 미국 FTC는 생성형 인공지능이 다량의 학습 데이터를 활용하고, 전문인력의 쏠림 현상이 나타나고 있으며, 컴퓨팅 기술의 확보 등이 제한적일 수 있으므로 신규 사업자의 진입장벽이 존재한다고 지적하였다.[27]

(2) 관련 법제의 미비 및 입증책임의 어려움

예컨대, 인공지능의 발전으로 자기학습 알고리즘이 적용될 경우, 사업자의 개입 없이 인공지능에 의하여 담합의 결과가 발생할 수 있고, 그에 사업자 간의 명시적·묵시적 합의가 있었다고 판단하기는 어렵다. 자기학습 알고리즘은 AI를 활용하는 사업자도 결과 도출 과정을 알 수 없기 때문에 사업자가 담합 발생을 정확히 예상하기도 어렵다.[28]

그러나 구글의 딥 마인드 사례 등을 통해 AI 간의 협력, 즉 담합의 가능성이 증명되기도 하였으므로, 이에 대한 지속적인 연구와 입법적 대응이 긴요한 시점이다.

---

25) https://www.oecd.org/competition/algorithms-and-collusion.htm.
26) 서완석, 346면.
27) FTC, Generative AI Raises Competition Concerns, 2023. 6. 29. https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.
28) 서완석, 360면.

5. 결론

현재 입법 논의 중인 인공지능법안은 인공지능의 긍정적 활용 가능성에 주목한 결과, 인공지능의 발전 및 육성에 무게 중심을 두고 있다. 그러나 최근 인공지능에 의해 발생한 국내외 소비자 피해사례와 현행 국내법 체계의 미비점을 고려할 때, 인공지능으로 인하여 시장질서가 왜곡되고 소비자의 권익이 침해될 가능성이 있다는 점을 충분히 고려한 입법적 대응이 필요하다. ▢

건강권과 인공지능 규제

- 은밀한 살인자(unnoticed killer) 인공지능 규제, 포기할 것인가?

전진한

(보건의료단체연합 정책국장)

2023년 2월 14일 국회 과학기술정보방송통신위원회 심사소위에서 통과된 것으로 알려진 「인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안」 은 인공지능(AI)에 “우선허용·사후규제 원칙”을 채택해 사전규제를 사실상 불가능하게 만드는 규제완화가 핵심이다. 그런데 반드시 AI에 국한하지 않더라도 우리는 충분치 않은 검증과 규제가 안전과 생명, 인권을 위협할 수 있다는 사실을 수없이 경험한 바 있다. 가습기살균제 사태를 보자. 이 일련의 사건과 이후 경과는 사후규제와 피해보상은 이미 발생한 심각한 위해를 돌이킬 수 없고, 그런 피해의 입증과 사후규제도 어렵다는 사실을 보여주었다. 여러 사례에서 드러났듯이 기업들은 자사 제품의 문제점과 부작용을 알면서도 이를 은폐하고 출시하는 경우가 허다하다. 이런 기업 자율에 모든 걸 맡기는 건 국가의 존재 이유를 부정하는 것이다.

보건의료 분야에서 부실한 규제가 일으킨 대표적 참사는 인보사 사태다. 식약처가 연골세포를 이용한 관절염 치료제라는 기업 말만 믿고 수백만원 짜리 치료제를 허가했는데 사실은 연골세포가 아니라 종양을 유발하는 신장세포 유래로 밝혀진 사건이었다. 오늘날 '규제가 혁신을 막는다'는 신자유주의 기업규제완화 논리가 사회를 지배하고 있지만, 오히려 사람들의 안전과 생명을 지키기 위해 국가가 엄격한 규제와 검증을 해야 할 필요는 점점 더 분명해지고 있다.

검증되지 않은 인공지능은 임상시험을 거치지 않은 신약보다 더 위험하다

인공지능도 다르지 않다. 오늘날 보건의료 영역에서 산업계와 규제완화론자들은 AI같은 신기술은 다르다고 주장한다. 그들은 기술의 발전 속도가 매우 빠르고, 해당 기술의 복잡성이 매우 높기 때문에, 기존의 규제는 효과가 없거나 심지어 불필요한 것이라고 한다. 디지털 예외주의(digital exceptionalism)라는 신조어가 만들어졌을 정도다[1]. 하지만 새로운 기술이라고 근거에 기반한 검증을 회피할 명분이 되지 않는다. 하버드대학 의생명정보학과 초대 의장 아이작 코헤인(Isaac Kohane) 박사는 “검증되지 않은 인공지능을 허용하는 건 신약을 테스트하지 않고 환자에게 투여하는 것과 같다”고 말한 바 있다[2].

그런데 검증되지 않은 인공지능은 검증되지 않은 의약품보다 훨씬 더 위험한 결과를 초래할 수도 있다. AI는 단지 개인에게 생의학적 문제를 일으키는 데 그치지 않고 불특정 다수에게 생명권, 건강권을 포함한 광범한 안전침해를 일으킬 수 있기 때문이다. 이것이 시민사회가 인공지능에 대한 규제완화를 특히 더 우려하는 이유다.

세계보건기구(WHO)는 최근인 5월 성명을 내 챗GPT를 비롯한 생성형 인공지능을 의료 분야에 활용할 때는 엄격한 검증이 필요하다고 지적했다[3]. WHO는 AI의 위험성을 매우 신중하게 검토하지 않으면 안 되는 이유를 다음과 같이 들었다. 첫째, AI를 학습시키는 데 사용되는 데이터가 편향되어 건강을 위협하거나, 불평등을 확대하는 부정확한 정보를 생성할 수 있다. 둘째, AI는 사용자에게 권위 있고 그럴듯하게 보일 수 있는 답변을 생성하지만, 특히 건강 관련 정보의 경우 완전히 부정확하거나 심각한 오류를 포함할 수 있다. 셋째, AI는 사용자에게 적절한 동의를 구하지 않은 데이터로 학습할 수 있고, 사용자가 대답을 구하는 과정에서 제공한 민감한 정보를 보호하지 않을 수 있다. 넷째, 텍스트, 오디오 또는 비디오 콘텐츠의 형태로 매우 설득력 있어 보이는 가짜뉴스를 생성하고 유포하는 데 오용될 수 있다. WHO는 이런 위험이 큰데도 불구하고 AI에 대한 열광만 높을 뿐, 일반적으로 신기술에 적용해온 엄격한 검증이 AI에 대해서는 똑같이 적용되지 않고 있는 게 큰 우려점이라고 지적했다. 이는 오늘날 한국에서 장미빛 환상을 제시하며 '인공지능법'을 밀어붙이는 이들을 겨냥하는 듯하다.

상임위 법안심사를 통과한 ‘인공지능법’은 「보건의료기본법」 제3조제1호에 따른 보건의료의 제공 및 이용체계 등에 사용되는 인공지능과 「의료기기법」 제2조제1항에 따른 의료기기에 사용되는 인공지능에 대해서는 “고위험영역 인공지능”이라고 분류는 하지만 규제라고는 '고위험 인공지능을 활용한다'고 홈페이지에 게시하거나 제품·서비스

1) People's Health Movement Digital Health Working Group, 포스트 코로나 세계에서의 디지털 정의 회복 : 디지털 헬스의 현황과 과제, 2022.5.
2) Jeremy Hsu, How using a medical startup's AI “symptom checker” could go very wrong, QUARTZ, 2019.12.13.
3) WHO, WHO calls for safe and ethical AI for health, 2023.5.16.

설명서에 적는 것으로 고지하는 것 외에는 의무를 두지 않는다. 인공지능법은 보건의료 인공지능에 대해서도 우선허용·사후규제를 할 뿐 아무런 안전장치가 없다.

검증되지 않은 AI가 낳을 문제 1 : 부정확한 진단과 치료로 개인의 건강 위협

검증되지 않은 인공지능은 지난 몇 년 사이에도 전 세계에서 문제를 일으켜 왔다. 이것이 오늘날 EU 등 세계 각국이 인공지능에 대한 규제를 강화하려는 이유일 것이다. 먼저 개인에게 영향을 미치는 진단·치료기기의 성능 문제에 대해서 살펴보자. IBM '왓슨'은 의료 인공지능 중 최근까지 가장 각광 받는 존재였다. 왓슨은 2011년 유명 퀴즈쇼에서 인간 챔피언을 이기며 화려하게 등장, 의료에 진출했다. IBM은 왓슨을 '암 치료의 혁명'이라고 홍보했다. 하지만 제대로 검증한 바는 없었다. 결과는 심각했다. 왓슨은 세계 암환자들에게 안전하지 않은 잘못된 치료를 권장했다. 진단 정확도는 폐암의 경우 17.8%에 그쳤다. IBM은 이런 문제를 알면서도 홍보와 마케팅에 열을 올렸다. 많은 병원도 왓슨이 환자에게 도움이 되지 않는다는 걸 알고도 경쟁적으로 도입했다. 소위 의료계 군비경쟁의 결과였다. 의료기관들이 인공지능에 대한 환상을 이용해 환자를 유치하고 높은 치료비를 청구할 수 있기 때문이었다. 일부 의사들이 "왓슨은 쓰레기"라고 경영진에 항의해도 소용이 없었다[4]. 한국 병원들도 마찬가지였다. 길병원, 부산대, 건양대, 대구가톨릭대 병원 등이 왓슨을 도입해 '인공지능 암센터'를 운영한다며 환자를 끌어들였다. IBM은 최근 왓슨을 헐값 매각하고 시장에서 철수했는데 이미 왓슨의 진단과 치료 대상이 된 환자들이 받은 영향과 불필요하게 지출한 의료비에 대해서는 제대로 평가되거나 보상된 바가 없다.

검증되지 않은 AI가 낳을 문제 2 : 대규모 의사결정 시스템이 낳는 집단적 건강 문제

인공지능 의료기술은 개별 환자를 넘어 더 많은 사람에게 영향을 미치기도 한다. 물론 개별 의료진도 환자에게 오진하고 실수할 수 있지만, 체계화된 인공지능 시스템에 오류가 있으면 단기간에 수천·수만의 사람들을 위험에 빠뜨릴 수 있다. 코로나19 시기 영국 정부는 감염자와 접촉한 사람에 자가격리를 안내하는 앱을 도입했는데, 정확하지 않았다. 앱 사용자는 정부 지침을 따르는 것보다 5배나 더 오래 감염자 곁에 머물렀다. 1900만명이 앱을 사용해 극히 적은 이들만 격리됐고, 나머지는 감염에 노출됐다. 세계보건기구가 검증되지 않은 인공지능이 수많은 사람의 목숨을 앗아가는 '은밀한 살인자(unnoticed killer)'가 될 수 있다고 우려한 이유 중 하나다[5].

---

4) Casey Ross, Ike Swetlitz, IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show, STAT+, 2018.7.

바빌론 챗봇도 유사한 사례다. 영국은 최근 국가 의료비용을 절감한다며 '바빌론'이라는 AI 챗봇을 도입했다. 바빌론은 AI를 사용하여 의학적 치료가 필요한 환자만 필터링함으로써 국가 의료비용을 절감시켜주겠다고 약속했다. 규제 당국은 충분한 검증을 거치지 않고 이 서비스를 출시했다. 전문가들은 효과가 있다는 충분한 증거 없이 서둘러 출시했다는 우려를 초기부터 제기했다. 아니나 다를까, 결과적으로 이 부정확한 챗봇은 도움이 필요한 많은 환자의 치료를 지연시키는 결과를 낳았다. 2018년 란셋에 게재된 논문에서 이들은 바빌론이 "현실적인 상황에서 의사보다 더 나은 성능을 발휘할 수 있다는 설득력 있는 증거를 제시하지 못하며, 오히려 성능이 현저히 떨어질 가능성도 있다"고 결론냈다[6].

검증되지 않은 AI가 낳을 문제 3 : 의료 불평등과 차별 강화·영속화[7]

앞서 예로 든 바빌론은 성별 편향 문제도 있었다. 흉통과 메스꺼움을 호소하는 59세 여성에게는 우울증과 공황발작을, 같은 증상을 호소하는 같은 프로필의 남성에게는 심각한 심장질환 가능성을 제시하며 구급차 호출을 권했다[8]. 차별과 편견으로 가득한 사회의 데이터로 학습한다는 사실 때문에, 데이터의 질 자체에 문제가 없더라도 불평등을 심화시키는 결과가 나타나는 것이다. 미국에서 추가 자원이 필요한 환자를 선별하기 위해 널리 사용되는 AI는 흑인보다 백인 환자에게 더 많은 의료자원을 쓰게한 것으로 나타난 바도 있다. 문제는 시스템 개발자가 현재의 병적 상태가 얼마나 심각한지를 확인하는 데 과거 대상자들이 사용한 의료비용을 대리지표로 활용했다는 점이다. 아프리카계 미국인들이 동일한 질병에도 과거에 더 적은 의료자원을 활용했다는 사실로부터, 이들이 더 적은 의료자원을 필요로 한다는 잘못된 결론을 이끌어 낸 것이다. 이와 비슷하게 소수민족 환자들이 자동 1차 진료 예약시스템에서 불참률이 높았다는 사실이 AI가 이들을 더 나쁜 시간대의 진료에 배정하는 근거가 되기도 했다. 이처럼 인공지능을 잘 설계하지 않으면 이 기술은 이미 사회에 존재하는 불평등을 강화하고 영속화할 수가 있다.

인공지능은 흔히 의료 접근성을 높일 것이라고 예측되는데, 그러나 이런 높은 접근성은 디지털 의료에 쉽게 더 많이 접근하는 '걱정 많은 건강한 이들(worried well)'이 의료체계를 압도하고, 자원이 부족한 이들은 오히려 의료서비스에서 더 멀어지는 결과를 초래할 수 있다는 우려도 제기되고 있다. 게다가 사회경제적 접근보다 생의학적 접근을 우선시하는 기술의 도입은 보편적 건강향상보다는 이런 기술이용의 의미를 파악하고 거

5) WHO, Ethics and governance of artificial intelligence for health: WHO guidance, 2021.
6) Jeremy Hsu, 같은 글
7) Alexander d'Elia etc, Artificial intelligence and health inequities in primary care: a systematic scoping review and framework, BMJ, 2022
8) WHO guidance(2021)

기에 시간과 자원을 쏟을 수 있는 사람에게는 유용함을 제공하지만, 그렇지 못한 사람들에게는 서비스가 제한되는 결과를 낳을 가능성이 높다. 이는 건강불평등을 더 심화시킬 수 있다. 특히 인공지능과 관련된 규제가 빠른 기술변화를 반영해야 한다는 압력 때문에 완화되면, 불평등을 심화시키는 인공지능이 의료시스템에 적용되기가 더 쉽다고 지적된다.

검증되지 않은 AI가 낳을 문제 4 : 의료비 증가와 상업화 추세 가속화

AI 기술에 대한 의존은 또한, 건강의 사회적 결정요인에 대한 관심과 역량강화로 이어지기보다는 의료를 상업 부문으로 넘겨주는 결과를 초래할 수도 있다는 문제가 제기되고 있다. 공중보건조치 같은 건강의 사회적 결정요인에 대한 정책적 조치 같은 상류의 개입이 의료서비스의 변화와 새로운 치료옵션 같은 하류의 개입보다 불평등을 줄이는 데 더 효과적이라는 것은 이미 입증된 사실이다. 따라서 불평등 완화에 초점을 맞추지 않고 공중 보건적 개입보다 새로운 치료옵션 제공의 한 수단으로 자리 잡은 AI는 본질적으로 불평등을 더 강화시킬 수 있다[9].

이런 AI는 의료비를 더 상승시킬 가능성도 크다. 벤저민 메이저(Benjamin Mazer) 존스홉킨스대 병리학 교수에 따르면 오늘날 AI가 의사 일자리를 대체할 거라는 예측이 많지만, 의료계는 이를 두려워하기보다는 AI를 통해 수익을 창출할 기회를 노릴 거라고 주장한다. AI 의사는 무료이며 편리하고 효율적 도구가 되기보다는, 불필요한 치료를 유발하는 비용이 많이 드는 존재가 될 가능성이 높다는 것이다. 예컨대 AI는 환자의 사소한 증상과 생체징후들을 더 예리하게 포착해 더 많은 검사를 요구하고 더 많은 청구서를 내밀 것이다. 이는 꼭 필요한 의료행위이기보다는 불필요하고 우려되며 때로 해로운 조치이기 쉬울 것이다[10]. 그는 매우 상업적이고 '행위별수가제'에 기반한 미국 시스템이 이런 경향으로 나아가도록 만들거라고 주장하는데, 이는 95%가 민간의료기관이고 행위별수가제에 미국보다 더 전적으로 의존하는 한국에서는 비슷한 현실화이 될 가능성이 크다.

인공지능은 아예 노골적으로 기업 이윤을 위해 비윤리적으로 활용될 수도 있다. 알려진 최근 사례에 따르면, 미국 민간보험사인 시큐리티 헬스 플랜(Security Health Plan)은 최근 어깨 골절로 입원한 85세 노인의 의료비 지불을 17일 만에 중단해버렸다. 인공지능이 16.6일이면 퇴원할 수 있다고 예측했다는 이유다. 그러나 환자는 여전히 심한 통증을 호소하는 거동 불능상태였다. 이런 식으로 민간보험사들은 3개월 내 사망할 수 있는 환자의 보험금 지급을 거절해 최대 2~3년이 걸리는 이의신청 절차를 밟게 한다고 알려진다[11].

---

9) Alexander d'Elia etc, 같은 글

10) Benjamin Mazer, The AI doctor will charge you now, The Boston Globe, 2023.5.16.

11) Casey Ross, Bob Herman, Denied by AI: How Medicare Advantage plans use algorithms to cut

검증되지 않은 AI가 낳을 문제 5 : 의사결정의 책임소재 문제와 의료현장의 혼란[12)]

인공지능은 의사결정 결과의 책임 문제라는 새로운 도전을 제기한다. 인공지능 기술의 도입은 의료인과 의료기관의 의사결정을 개선할 수 있다는 개념에 기반을 두지만 AI기술의 특성인 불투명성과 확장성, 복잡성 등은 책임의 문제를 어렵게 만든다.

많은 사람들이 AI를 개발한 개발자들에게 책임을 물어야 한다고 생각할 수 있지만, AI 프로그래밍이 갈수록 자동화되면서 AI가 개발자와 독립적으로 작동하고 개발자가 예측할 수 없는 방식으로 진화할 수 있기 때문에 AI 개발자 및 설계자가 책임에서 면제될 수 있다. 또 AI 개발에는 많은 주체들의 기여가 필요하기 때문에 책임이 분산될 수 있고, 개인은 피해를 보상받지 못할 수 있다.

한편으로 AI 기술은 임상적 의사 결정을 대체하는 것이 아니라 이를 보조하거나 개선하는 데 사용되므로, 의료 서비스에서 AI 기술을 사용함으로써 발생하는 모든 피해에 대해 임상의가 책임을 져야 한다는 주장이 제기될 수 있다. 하지만 임상의는 AI에 대한 통제권을 행사하지 않고, AI 기술이 불투명한 '블랙박스' 알고리즘을 사용하는 경우가 많으며, 임상의가 AI 기술을 사용하기로 선택한 것이 아니라 병원 시스템이나 다른 외부 의사 결정권자의 선호도 때문에 AI 기술을 사용하게 될 수도 있기 때문에 임상의에게 전적으로 책임을 묻기 어려운 지점이 존재하게 된다. 이처럼 AI 의사결정이 문제를 발생시킬 경우 책임소재는 매우 불분명할 수 있다.

이에 대해 WHO는 의료 시스템 전반에 걸쳐 AI 기술을 사용하기로 결정한 경우 문제는 해당 기술을 선택하고 검증하고 배포한 정부 기관이나 의료기관에 있을 수 있다고 지적하기도 한다. 이처럼 복잡한 책임소재의 궁극적 당사자인 정부가 AI의 검증에 아무런 책임을 지지 않고 ‘우선허용·사후규제’의 원칙을 도입한다면 의료현장은 매우 큰 혼란에 빠질 수가 있다.

또 AI를 의료현장에 적용할 경우 AI 기계와 의사 사이에 '동료 의견 불일치'가 발생할 가능성이 있는데, 이러한 상황에서 의사는 알고리즘에 접근해 알고리즘이 내린 결정의 이유를 추론할 수 있는 수단이 없고 알고리즘의 생각을 바꾸기 위해 관여할 수가 없다는 문제가 존재한다. 이에 대해서 WHO는 그 해결책으로 의료 인공지능이 투명하고 설명할 수 있어야 한다는 핵심 원칙을 제시하고 있다. 이는 매우 본질적 문제일 수 있다.

즉 의료 AI의 경우에는 정확성을 높여야 한다는 기술적 문제는 물론, 불평등과 차별을 없애야 한다는 윤리적 문제도 고려한 설계와 검증이 있어야 하지만, 투명하고 설명가능해야 한다는 원칙을 갖춰야 하는 것이다. 그러나 한국의 규제 당국은 설명가능성은

---

off care for seniorsin need, STAT, 2023.3.13.

12) WHO guidance(2021)

커녕 최소한의 기계적 정확성 검증의 수준도 포기하고 있는 형편이다.

마치며

앞서 살핀 것처럼 인공지능에 대한 검증을 포기하는 것은 개별 제품이나 심지어 의약품·의료기기의 안전과 효과에 대한 평가를 무력화하는 것과는 차원이 다른 문제일 수가 있다. 정부와 국회가 인공지능을 통한 대규모 의사결정이 낳을 파괴적 영향을 조금이라도 이해한다면 이런 무책임한 법을 내놓지는 못했을 것이다.

이는 그간 역대 정부는 시민사회의 오랜 반대에도 불구하고 '규제샌드박스' 5법과 각종 의료 관련 법령의 개악으로 생명과 안전에 대한 규제를 완화해 기업 이윤추구를 장려해 왔던 것과 일관되는 행보이기도 하다. 시민의 안전과 생명보다는 기업이윤을 앞세워온 태도가 더 한층의 윤리적 무감각으로 발전해 이제 재앙적 결과를 낳을 규제완화를 시도하려 하는 것이다.

시민사회는 발달된 기술을 적용하는 데 반대하지 않는다. 오히려 제대로 된 검증이 있고 그것이 보편적·공공적 이익으로 돌아올 때 기술은 사회적 신뢰를 얻을 수 있다는 점에 주목한다. 그러기 위해서 이 '인공지능법'은 당연히 폐기되거나 재검토되어야 한다. 인공지능법에 대한 재검토는 무분별한 AI 뿐 아니라 사회 전체에 만연한 규제완화에 대한 재검토의 시작이어야 한다. 사실상 모든 영역에서 '우선허용·사후규제'라는 망령이 스멀스멀 자라나고 있다.

마지막으로 앞서 검토한 것처럼, 기술의 검증이란 안전성과 유효성을 넘어선 사회적 논의를 필요로 한다. 특히 이 사회의 우선순위가 개별화된 솔루션을 제공하는 새로운 기술의 적용인가, 아니면 공공적 사회정책인가 하는 사회적 논의도 민주적으로 이뤄진 후에 기술적용을 결정할 수 있어야 한다. 그것이 극심한 불평등 뿐 아니라 디스토피아적 생태위기를 맞닥뜨린 지금 우리에게 절실한 과제일 것이다. ▢

과학기술정보통신부 토론문

최동원

(과학기술정보통신부 인공지능기반정책국 과장)

개인정보보호위원회 토론문

이병남

(개인정보보호위원회 정책국 과장)

공정거래위원회 토론문

강승빈

(공정거래위원회 시장감시정책과 서기관)

인공지능 법안 관련 토론문

박소현

(국회입법조사처 입법조사관)

먼저 이런 자리에서 토론할 기회를 주신 장경태 의원님께 감사드리고, 귀한 발표로 많은 배움과 생각의 기회를 주신 두 분의 발제자분들께도 감사의 말씀을 드립니다. 유승익 교수님께서는 '인공지능이 인권과 민주주의에 미치는 영향과 인공지능법안의 쟁점'이라는 주제로 인공지능이 인권과 민주주의에 미칠 수 있는 부정적인 영향과 위험성을 짚어주시면서, 현재 국회 과학기술정보방송통신위원회의 법안소위를 통과한 「인공지능산업 육성 및 신뢰 기반 조성에 관한 법률안」(이하 「인공지능법안」)의 문제점 지적을 통해 향후 과제를 제시해 주셨습니다. 먼저 동 법률안이 인공지능기술 규제방식으로 채택하고 있는 '우선허용·사후규제 원칙'은 인공지능 기술의 잠재적 위험성을 고려할 때, 적절한 규제수단이 될 수 없다는 점, 둘째 금지되는 인공지능 기술은 규정하지 않고 있으며, 고위험 영역 인공지능도 사람의 생명, 안전, 기본권에 중대한 위험성을 미칠 것으로 충분히 예상되는 다수의 영역이 제외되었으며, 규정된 고위험 인공지능에 대한 준수사항의 이행을 확보할 만한 실효적 장치가 마련되어 있지 않다는 점, 마지막으로 과기부를 중심으로 한 인공지능 정책에 관한 거버넌스의 산업편향적 결과 산출의 우려가 있음을 문제점으로 지적해 주셨습니다. 그리고 향후 과제로 인공지능 기술의 잠재적 위험성을 예방적으로 관리하고 완화시킬 수 있는 규제방식으로 영향평가제도 도입을 제안하시면서, 우리 사회에서 금지되어야 할 인공지능 활용영역을 규정하고, 인공지능 정책의 균형성 확보를 위해 인공지능 규제와 정책을 주도할 독립적 기관 설치할 것을 제안해 주셨습니다. 허진민 소장님 역시 유럽연합의 인공지능법안을 상세하게 소개

해 주시면서, 인공지능기술의 개발과 활용을 포함한 모든 단계에서 각 당사자의 책무가 구체적이고 명확하게 규정되어 있지 않다는 점, 과기부가 인공지능 기본계획의 수립 및 시행의 주체로 적정하지 않다는 점, '우선허용·사후규제 원칙'이라는 원칙적 규정만 두고 있어서 인공지능기술의 잠재적 위험에 대비하기 위한 구체적 기준이 없고, 해외의 인공지능 규제 논의에 대한 내용이 포함되어 있지 않아 국내 기업들의 해외 진출에 있어 장애물로 작용할 수 있다는 점, 고위험 인공지능에 대한 준수사항에 대해 제재조치를 두지 않아 실효성을 확보하기 어렵다는 점 등을 우리나라 「인공지능법안」의 구체적인 문제로 지적해 주셨습니다.

인공지능 산업의 경쟁력을 높이기 위한 산업 육성 및 지원 그리고 인공지능 기술의 신뢰성과 안전성 확보를 통한 국민의 인권 및 기본권 보호라는 두 가치 사이에서 적절한 조화점을 찾는 것이 쉽지 않은 일임은 분명하지만, 인공지능법을 제정함에 있어서 향후 반드시 달성하여야 할 과제라고 생각합니다. 동 법률안이 과방위 법안심사소위를 통과할 때의 회의록을 보면, 구체적으로 미비한 점들이 많이 있긴 하지만, 산업 육성 및 인공지능 윤리 측면에서 신뢰 기반을 다지기 위해서 법안 통과가 시급하다는 데에는 위원님들도 모두 동의하고 계셨던 것 역시 같은 취지인 것으로 보입니다.`

사후규제라는 방식이 인공지능의 잠재적 위험성과 파급력 관리라는 관점에서 매우 부적절하며, 현 상태에서 결단코 허용될 수 없는 인공지능시스템을 규정하고, 특별한 의무가 부과되는 고위험 인공지능 시스템의 개발·활용에 있어서 의무이행을 강제할 수 있는 실효적 제재수단이 마련되어야 한다는 점에서 두 발제자분들의 문제의식과 향후 과제에 대하여 전적으로 동의하고 있기 때문에, 저는 몇 가지 궁금한 점을 여쭙는 것으로 토론을 갈음하고자 합니다.

먼저, 법이 아무리 잘 만들어진다 하더라도 법률이 적용될 대상이 특정되지 않거나, 목적한 대상을 넘어서서 너무 광범위한 범위까지 모두 포섭되는 경우 입법취지에 부합하지 않을 수 있기 때문에, 적용대상의 개념정의는 입법과정에서 매우 중요한 부분이라고 생각합니다. 그래서 인공지능의 개념정의에 관하여 먼저 여쭙고자 합니다.

인공지능에 대한 OECD 이사회 권고에 따르면, 인공지능시스템을 "인간이 정의한 일련의 주어진 목적을 위하여 실제 또는 가상 환경에 대해 영향을 미치는 예측, 추천 또는 의사결정을 할 수 있는 기계 기반 시스템"으로 정의하고 있습니다.[1] 유럽의 인공지능법안은 이러한 인공지능에 대한 OECD 이사회 권고에 따라 "AI 시스템을 인간이 설정해 둔 일련의 목적과 관련하여 AI 시스템과 상호작용을 하는 환경에 영향을 미칠 수 있는 내용이나 예측, 권고 또는 의사결정과 같은 결과값을 생성할 수 있는 소프트웨어"라 정의하면서, 이 때 사용되는 기술[2]은 부속서에서 별도로 규정하고 있습니다. 동 법

1) OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, June 2019.
2) Annex I (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning

안이 AI 시스템을 정의하면서 AI 기술을 부속서에서 추가로 정의하는 방식을 취한 이유는 가능한 한 기술 중립적이고 미래에 적합하게 규정함으로써 인공지능기술과 인공지능시장에서의 급격한 발전내용들을 고려하기 위함이라고 밝히고 있습니다.[3)]

그런데 윤두현의원이 대표발의한 우리나라의 「인공지능법안」은 인공지능을 학습, 추론, 지각, 판단, 언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현한 것이라고만 개념정의를 하고 있다는 점에서, 인공지능 산업 육성과 규제를 통한 인권 보호라는 두 측면 모두에서 적절하다고 생각하시는지 두 분의 견해를 여쭈어 보고 싶습니다.

두 번째, 유승익 교수님께서는 먼저 금지되어야 할 인공지능 활용영역을 규범적으로 확인해야 한다고 하시면서, 금지되어야 할 유럽연합의 법안처럼 잠재의식기술을 활용한 광고행위를 예로 들고 계십니다. 이와 관련하여 아직은 인공지능을 활용하는 것이 원천적으로 금지되어야 하는 영역으로 어떠한 것들을 염두에 두고 계시는지, 또한 실시간 안면인식기술과 같은 인공지능 기술의 활용에 대하여는 어떻게 생각하시는지 의견을 여쭙고자 합니다.

세 번째, 「인공지능법안」은 인공지능 제품 또는 서비스를 개발·활용·제공하려는 자가 고위험 영역에서 활용되는 인공지능에 해당하는지에 대한 확인을 과기부에 요청한 경우에, 과기부가 고위험 인공지능 해당 여부를 확인하며, 필요한 경우에 전문위원회를 설치하여 자문을 받을 수 있도록 규정하고 있습니다. 물론 고위험 영역에서 활용되는 인공지능의 기준과 예시 등에 관한 가이드라인을 수립하여 보급할 수 있도록 규정하고는 있지만, 인간의 생명, 안전, 기본권에 중대한 영향을 미칠 우려가 있는 인공지능을 인공지능의 기획·개발단계에서부터 체계적으로 관리·감독해야 할 필요성이라는 관점에서, 동 법률안이 제시하고 있는 안에 대한 두 분의 견해와 이에 대하여 혹시 구체적인 개선안을 염두에 두고 계신 것이 있는지 여쭙고 싶습니다.

네 번째로 저는 인공지능 설계 및 개발단계에서부터의 관리·감독을 위해서 설계 및 개발 단계별 문서화, 사생활 및 개인정보자기결정권 보호·데이터 거버넌스 적용·안전성·보안성 등을 점검할 수 있는 영향평가, 알고리즘의 위험성 식별과 이에 대한 적절한 안전장치 설치 여부 평가 시스템과 결함 발견 시 이를 수정할 수 있는 위험관리시스템 구축, 시스템 유통 전의 사전 적합성 평가 등의 의무를 법률에 구체화하고, 알고리즘의 추구목적과 작동방식에 대한 검증체계를 구축해야 하며, 이러한 검증을 위해 별도의 국가검증기관에게 알고리즘의 추구목적과 작동방식 및 문서화된 자료 등을 보고하게 함으로써, 인공지능시스템에 대한 평가와 인증을 통한 신뢰가능성 확보 및 영업비밀의 과도

---

and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods.

3) European Commission, 「Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS」, 2021.4.21., COM/2021/206 final.

한 공개로부터의 보호를 가능하게 하는 방안을 마련해야 한다고 생각합니다. 이에 대하여 규제가 너무 과도하기 때문에 산업 육성에 저해된다고 생각하시는지, 두 분의 생각을 여쭙고자 합니다.

마지막으로 인공지능 관련 정책 수립을 과기부가 담당하도록 규정하고 있다는 점에서, 관련 정책이 산업 육성으로 편향되지 않도록 하기 위해서는 인공지능위원회의 구성과 역할이 매우 중요할 것이라 생각됩니다. 그런데 동 법안은 국무총리와 인공지능 등에 관한 전문지식과 경험이 풍부한 사람 중 대통령이 위촉하는 1인을 공동위원장으로 하고, 관계 중앙행정기관의 장과 인공지능 등에 관한 전문지식과 경험이 풍부한 사람들이 위원이 되도록 규정하고 있습니다. 이와 관련하여, 인공지능기술의 신뢰성·안전성 확보를 통한 인권 보호가 보장되게 하려면 구체적으로 최소한 어떠한 분야의 전문가들은 반드시 포함되어야 한다고 생각하는지 여쭙고 싶습니다. ▢

인공지능 법률 제정안의 문제점과 개선방안

김재석

(국가인권위원회 인권정책과 과장)

1. 인공지능 법률 제정안

가. 개요

국회 과학기술정보방송통신위원회(이하 ‘과방위’라 함) 정보통신방송법안심사소위원회(이하 ‘법안소위’라 함)는 2022. 12. 15. 당시까지 인공지능과 관련하여 제안된 법률안 7개를 통합하여 대안을 만들기로 의결하였습니다. 법안소위는 이 결정에 따라 대안으로 마련된 ‘인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안(이하 ‘인공지능 법률안’이라 한다.)’을 상정하여 심의한 이후 의원실, 관계기관 등의 협의를 거쳐 최종적으로 수정한 대안을 과방위 본회의에 상정하기로 결정하였습니다.

인공지능 법률안에는 과학기술정보통신부(이하 ‘과기정통부’라고 함)로 하여금 3년마다 인공지능 기본계획을 수립하도록 하고, 기본계획 및 주요 정책, 관련 예산 등을 심의하기 위해 국무총리 소속으로 인공지능위원회를 설치하며, ‘우선 허용, 사후 규제’ 원칙을 명시하는 한편, 고위험 영역에서 인공지능을 활용하는 사업자에게 이용자에 대한 고위험 인공지능 사용 사실의 고지 의무, 신뢰성 확보 조치, 인공지능 도출 최종 결과 등에 대한 설명 의무를 부여하는 등의 내용이 포함된 것으로 알려졌습니다.[1)]

1) 이데일리, 2023. 2. 14. 보도, AI기술 우선 허용-사후 규제...‘인공지능법 제정안’ 법안소위 통과, 김현아 기자

시민사회단체는 2023. 3. 2. 인공지능 법률안이 국제인권규범 및 해외 입법 정책의 예에 비추어 보았을 때 소관 부처의 적절성, 안전과 인권에 미치는 인공지능 위험 규제의 효과 측면에서 중대한 문제를 가지고 있고, 특히 '우선 허용, 사후 규제' 조항은 국민의 개인정보, 소비자 권리 및 인권을 위험에 처하게 할 수 있는 독소조항으로 즉각 폐기하여야 한다는 등의 내용이 담긴 의견서를 국회와 개인정보보호위원회, 인권위 등 관련 기관에 제출하였습니다.[2)]

인권위도 인공지능 법률안의 중심이 된 윤두현 의원의 발의안을 예비 검토한 결과 우선 허용 · 사후 규제 원칙 등 일부 내용이 인권위가 권고한 인공지능 인권 가이드라인과 배치되는 측면이 보았고, 이에 인권위는 인공지능에 관한 기본법의 지위를 가지게 될 인공지능 법률안이 인공지능 개발 · 활용 과정에서 발생할 수 있는 인권침해 및 차별 문제를 방지하기 위한 제도적 기반을 갖추고 있는지 인공지능 인권 가이드라인을 토대로 검토하였습니다.

나. 문제점

1) 이용자 및 정보주체의 권리 보장 미흡

인공지능, 특히 학습기반 인공지능(머신러닝)은 이용자가 제공한 정보뿐만 아니라 온라인 상에서 수집한 방대한 양의 정보를 기반으로 서비스를 제공하므로, 정보 수집 과정에서 이용자와 정보주체의 권리를 침해할 소지가 농후하다. 또한 결과물 산출 과정에서 개인정보 유출, 인격권 침해 및 명예 훼손, 허위정보 생산, 저작권 등 재산권 침해 문제 등을 초래할 위험성이 있습니다.

따라서 인공지능 개발 및 활용 과정에서 이용자와 정보주체의 권리를 명확히 하고, 이용자 등이 권리 침해를 입은 경우 이에 대한 구제 절차를 마련하는 것이 무엇보다 중요합니다. 이와 관련하여 유엔 인권이사회는 2021년 9월 제48차 회기에서 채택한 '디지털 시대 프라이버시권 보고서'에서 인공지능 시스템에서 법률로써 정보주체를 효과적으로 보호해야 할 필요성을 강조하면서 설명에 대한 권리와 완전 자동화된 의사결정에 반대할 권리를 강화하고, 독립적인 개인정보보호 감독기구 등 안전장치를 개인정보 보호 체계에 수립할 필요가 있다는 의견을 밝힌 바 있습니다.

인권위도 인공지능 인권 가이드라인에 알권리, 개인정보자기결정권을 개인이 인공지능에서 보장받아야 할 기본적인 인권으로 명시하고, 알권리와 관련하여 자동화된 의사결정에 의하여 영향 받는 당사자는 결정의 이유에 대하여 설명을 듣고, 당사자 진술을 할 수 있으며, 이의를 제기할 수 있어야 한다고 명시하였습니다. 특히 완전히 자동화된 의사결정만으로 개인에게 법적 효력 또는 생명·신체·정신·재산에 중대한 영향을 미치는

2) 경제정의시민실천연합 등 15개 시민사회단체, 「인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안(소위안)」에 대한 인권시민단체 의견, 2023. 3. 2.

일은 제한하여야 하고, 이러한 의사결정이 이루어진 경우에는 당사자가 해당 방식을 거부하거나 인적 개입을 요구할 수 있는 권리를 보장할 것을 강조하였다. 또한 국가가 인공지능으로 인하여 인권을 침해당하거나 차별을 받은 사람이 진정을 제기하여 권리를 구제받을 수 있는 기회를 보장하고, 이의를 제기할 수 있는 기관과 방법에 대한 정보를 일반에 공개하여야 한다고 밝혔습니다.

그러나 인공지능 법률안은 인공지능의 개발 · 활용 과정에서 인공지능이 안전성, 신뢰성을 확보해야 함을 원칙으로 정하고, 고위험 영역 인공지능을 활용하여 제품 또는 서비스를 제공하려는 자에게 해당 제품 또는 서비스가 고위험 영역 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지할 의무만 부과할 뿐, 이용자와 정보주체의 권리, 권리 침해에 대한 구제 절차를 명확히 규정하지 않았습니다.

2) 고위험 영역 인공지능 관련 문제점

인공지능 기술이 활용되는 분야 및 영역이 점차 확대되고 있는바, 최근에는 인공지능을 기업의 직원 면접 및 채용, 치안 · 수사 업무, 난민심사, 범죄자의 재범 가능성 판단 등과 같이 사람의 삶과 인권에 직접 영향을 미치는 분야에 활용하는 사례가 나타나고 있습니다.

인공지능이 인권에 밀접한 영향을 미치는 영역에 사용되고, 그 영향력이 갈수록 막대해지자 국제사회와 세계 각국은 인공지능 개발 · 활용 과정에서 발생할 수 있는 문제점을 사전에 최소화하기 위한 방안을 마련하고 있습니다. 그 중 가장 선도적인 유럽연합은 2021년 4월 인공지능 기술을 위험성에 따라 '금지되는 인공지능', '고위험 인공지능', '제한된 인공지능', '최소위험 인공지능'으로 구분하고, 각 등급에 따라 법적 의무를 부과하는 '인공지능법안'을 발의하였으며, '인공지능법안'은 2023. 6. 14. 유럽 의회(European Parliament) 전체회의(plenary meeting)를 통과하여 유럽 이사회(European Council) 의결을 앞두고 있습니다.

인권위도 인공지능 인권 가이드라인에서 인공지능을 개인의 인권과 안전에 미치는 위험성의 정도에 따라 구분하는 방법을 제시하였는바, 위험성이 매우 높아 '인공지능이 금지되는 영역', '상당한 제한이 필요한 고위험 영역', '일정 정도의 제한이 필요한 제한된 위험을 가진 영역', '위험성이 거의 없는 영역'으로 구분하고, 각 등급에 맞는 규제와 인적 개입이 이루어져야 한다고 제안하였습니다.

그러나 인공지능 법률안은 인공지능 기술을 활용되는 분야의 특성을 고려하여 위험도에 따라 단계별로 구분하지 않고, 고위험 영역 인공지능에 대해서만 규율하고 있으며, 고위험 영역 인공지능에 해당하는 대상도 유럽연합의 '인공지능법안' 등 해외 사례에서 제시하는 기준에 비교하여 볼 때 상대적으로 협소하게 정의하고 있습니다.

예를 들어 인공지능 법률안은 생체정보를 범죄 수사나 체포 업무에 분석·활용하는 데 사용하는 경우를 고위험 영역 인공지능으로 정의하는 데 반하여, 유럽연합은 '인공지능

법안'에서 형사사법 분야에서 재범 위험 또는 범죄의 잠재적 피해 위험을 평가하기 위하여 사용하는 인공지능, 거짓말 탐지기 및 유사한 도구로 사용하거나 감정 상태를 감지하기 위하여 사용하는 인공지능, 딥페이크 감지를 위하여 사용하는 인공지능, 증거의 신뢰성을 평가하기 위하여 사용하는 인공지능, 프로파일링을 기반으로 범죄 행위의 발생 또는 재발을 예측하거나 과거의 범죄 행위를 평가하는 데 사용하는 인공지능 등을 고위험 영역 인공지능으로 정하여 인공지능 법률안보다 폭넓게 규정하고 있습니다.[3)]

또한 유엔 인권이사회가 채택한 '디지털 시대 프라이버시권 보고서'에서 법 집행, 국가안보, 형사사법, 사회보장, 고용, 보건의료, 교육 및 금융 등 개인의 이해관계가 특히 높은 분야에서 인공지능 기술을 사용하는 경우 법적 요건을 엄격히 하여야 한다고[4)] 밝힌 점에 비추어 보면 인공지능 법률안에서 정의한 고위험 영역 인공지능은 지정 범위가 상대적으로 협소합니다.

물론 인공지능 법률안이 대통령령으로 국민의 안전 · 건강 및 기본권 보호에 중대한 영향을 미치는 인공지능을 고위험 영역 인공지능으로 지정하도록 규정하여 대상을 넓힐 수 있는 여지는 두고 있으나, 법률 위임의 한계를 고려하면 시행령에서 법률이 직접 규정한 대상보다 범위를 폭넓게 지정하는 것은 불가능하므로, 현재 법률안 체계 내에서는 대상을 확대 · 지정하는 데 한계가 있을 것으로 보입니다.

아울러 인공지능 법률안은 인공지능사업자 등이 고위험 영역 인공지능을 개발하여 활용할 경우 사업자 등에게 이용자에 대한 사전 고지, 신뢰성 · 안전성 확보 조치 등의 의무를 부과하는 규정만 두고 있을 뿐이고, 소관 기관인 과학기술정보통신부장관이 고위험 영역 인공지능의 신뢰성 확보 조치의 구체적인 내용을 정하여 사업자에 이를 준수하도록 권고할 수 있다고 규정할 뿐 실효적인 규제 수단은 규정하고 있지 않습니다.

3) 우선허용 · 사후규제 원칙의 위험성

인공지능 법률안은 우선허용 · 사후규제 원칙을 명시하고, "인공지능기술, 인공지능제품 또는 인공지능서비스가 국민의 생명 · 안전 · 권익에 위해가 되거나 공공의 안전 보장, 질서 유지 및 복리 증진을 현저히 저해할 우려가 있는 경우가 아니라면 이를 제한하여서는 아니 된다."라고 규정하고 있습니다.

그러나 인공지능은 개발 과정에서 개발자의 이념이나 사상 등이 무의식 중에 반영될 수 있고, 인공지능이 학습하는 데이터에 이미 사회적 편견이나 차별적 요인이 내재되어 있으므로, 결과물 역시 인권침해 또는 차별적인 요소를 포함할 수 밖에 없으며, 서비스 이용자로부터 개인정보를 과도하게 수집하거나 허위정보를 생산하는 등 각종 문제가 발

3) 국가인권위원회, 인공지능(AI)개발과 활용에서의 인권 가이드라인 연구, 2021년 11월, 49쪽
4) 유엔 문서 A/HRC/48/31(2021), The right to privacy in the digital age, Roport of the United Nations High Commissioner for Human Rights. 45문(출처: 국가인권위원회, 인공지능(AI)개발과 활용에서의 인권 가이드라인 연구, 2021년 11월, 240쪽)

생할 수 있습니다. 이와 같은 위험성은 이미 '이루다'사태에서 여실히 드러났으며[5], 앞서 살펴보았듯이 최근 큰 인기를 끌고 있는 '챗 지피티' 역시 이용자의 질문에 답하면서 허위정보를 생산하여 문제가 되었습니다.[6]

따라서 인공지능 기술 개발 · 활용 과정에서 사전 검증 및 사후 통제가 중요한데, 관련 법률에 우선허용 · 사후규제 원칙을 명시한다면 개발자 및 사업자가 인공지능 기술이 인권에 미치는 영향을 사전에 평가하지 않고 효율성과 편익만을 따져 인공지능 기술을 개발 · 활용함으로써 이 같은 문제점을 간과하게 될 우려가 있습니다.

현재 유럽을 포함한 세계 각국은 인공지능 기술이 초래할 수 있는 여러 가지 부작용과 문제점을 최소화하기 위하여 사전 영향평가를 실시하고 있거나 제도화를 추진하고 있고, 특히 사전 영향평가를 통과하지 못한 고위험 영역 인공지능 기술의 활용을 제한하는 입법을 추진하고 있는 상황입니다.

구체적으로 미국은 2022년 2월 하원과 상원에서 '알고리즘 책무성법안(Algorithmic Accountability Act of 2022)'을 발의하여 연방거래위원회(FTC)가 소관하는 일정 규모 이상의 기업들을 대상으로 자동화된 의사결정 시스템 또는 증강된 중요 의사결정 프로세스 등이 소비자에게 미치는 영향에 대하여 지속적으로 '영향평가'를 실시하도록 의무화하고, 이를 연방거래위원회가 감독하는 방안을 논의하고 있습니다.[7]

영국은 2020년 6월 '인공지능 조달 지침(Guidelines for AI procurement)'을 발표하였는데, 동 지침은 공공기관이 인공지능 기술 및 시스템, 프로세스 등을 공공조달을 통하여 도입하는 경우 10대 원칙을 따르도록 하고, 조달 절차 개시 전에는 데이터 평가를, 개시 단계에서는 인공지능 배치의 편익과 위험에 대한 영향평가를 실시하도록 규정하고 있습니다.[8]

네덜란드는 '기본권 알고리즘 영향평가(Fundamental Rights and Algorithms Impact Assessment)'를 개발하여 알고리즘 시스템 개발 또는 조달을 검토하는 공공기관이 초기 단계에서 4단계에 걸쳐 인권과 관련된 쟁점을 검토하도록 하고 있으며, 캐나다는 2019년부터 '자동화된 의사결정 지침(Directive on Automated Decision-Making)'을 제정하여 시행하고 있는데, 공공기관이 자동화된 의사결정 시스템을 도입 · 활용하기 위해서는

---

5) 이루다는 2020년 12월 출시된 인공지능 챗봇 서비스로, 일부 남성 이용자들이 이루다를 성적 대상으로 취급하고, 발화 내용에 여성 · 성소수자 · 장애인 · 흑인을 혐오하는 내용이 포함되어 인공지능 윤리 문제와 차별 논란을 빚었으며, 개인정보 보호법 위반 소지까지 불거져 개인정보보호위원회가 이루다 개발사에 대하여 조사를 실시하고 확인된 위법 사항에 대하여 총 1억 330만 원의 과징금과 과태료를 부과하였음(출처: 연합뉴스, 2021. 1. 11. 보도, 성희롱 · 혐오 논란에 3주만에 멈춘 '이루다'…AI윤리 숙제 남기다, 이효석 기자/ 개인정보보호위원회, 2021. 4. 28. 보도자료, 개인정보위 '이루다' 개발사 ㈜스캐터랩에 과징금 과태료 등 제재 처분)

6) 챗 지피티를 운영하는 오픈AI사의 CEO 샘 울트먼은 2023. 5. 16. 미국 상원 법사소위원회 청문회에 출석하여 "점점 더 강력해지는 인공지능 모델의 위험을 완화하려면 정부의 규제 개입이 중요하다."라고 하여 인공지능 규제의 필요성에 대해 공감함(출처: 매일경제, 2023. 5. 17. 보도, 오픈AI CEO "인공지능 규제해야"…MS "인공일반지능 진입", 이상덕 특파원)

7) 국가인권위원회, 인공지능 인권영향평가 도입 방안 연구, 2022년 12월, 59쪽

8) 앞의 연구용역 보고서, 48쪽~49쪽

‘캐나다 알고리즘 영향평가 도구(Canadian Algorithmic Impact Assessment Tool’를 활용한 평가를 의무적으로 실시하도록 하고 있습니다.[9)]

위와 같은 국제 흐름을 감안하면 인공지능 법률안처럼 우선허용 · 사후규제 원칙에 따라 산업의 경제성 · 효율성만을 따져 무분별하게 인공지능 기술을 개발 · 활용하게 할 경우, 사전 · 사후 평가 없이 개발 · 활용된 인공지능 기술이 국제 기준에 부합하지 못한 결과를 초래하여 오히려 국제 경쟁력을 저하시키고, 근본적으로 인공지능 기술의 신뢰성을 떨어뜨릴 위험성이 있습니다.

4) 인공지능 감독 · 규제의 실효성 문제

인공지능 법률안은 과학기술정보통신부장관으로 하여금 인공지능 기술 및 알고리즘 개발 활성화를 위한 지원, 기업의 인공지능 기술 도입 · 활용 지원, 창업 활성화, 인공지능 융합 촉진, 인공지능 관련 제도 개선, 전문인력 확보 등 인공지능 기술 및 산업 진흥 · 육성을 담당하도록 할 뿐만 아니라, 인공지능 윤리 원칙 실천 방안 수립 및 홍보, 인공지능 윤리 관련 법령 · 기준 · 지침 등에 관한 권고, 인공지능 신뢰성 검 · 인증 지원, 고위험 영역 인공지능 확인, 고위험 영역 인공지능의 기준과 예시 등에 관한 가이드라인의 보급 등 인공지능 규제에 관한 업무도 담당하도록 하고 있습니다.

그러나 산업 진흥과 규제 업무 모두를 한 기관이 담당할 경우 규제가 실효적으로 이루어질 수 있을지 의문이 제기됩니다. 한 기관이 결정하거나 지원한 행위에 대하여 동일한 기관이 관련 법령 및 기준의 준수 여부, 권리침해 여부를 판단하게 될 경우 자가당착의 모순이 발생할 수도 있습니다. 즉, 자기 자신이 행한 행위가 올바른지 스스로 판단하여야 할 상황이 발생하게 되므로, 규제의 실효성을 담보하기 어려워집니다.

따라서 산업 진흥과 규제를 분리하여 각 기관이 독립적으로 수행하도록 하는 것이 바람직하다. 유엔 인권이사회도 ‘디지털 시대 프라이버시권 보고서’에서 인공지능 시스템에 대한 적정하고 독립적이고 공정한 감독이 필요하고, 이는 행정적 · 사법적 · 준사법적 기관 및 의회 감독 기관의 조합으로 수행될 수 있다고 하면서, 개인정보 보호 기관, 소비자 보호 기관, 부문별 규제 기관, 차별 방지 기구 및 국가인권기구가 감독 시스템의 일부를 구성해야 한다고 밝힌 바 있습니다.[10)]

한편 인공지능 법률안은 소관 기관인 과학기술정보통신부장관으로 하여금 인공지능 윤리 원칙 제정 및 보급, 인공지능 신뢰성 검인증 지원, 인공지능 신뢰성 확보 조치의 구체적 내용 고시 및 준수 권고 등의 권한만 부여하고 있을 뿐, 인공지능으로 인하여 발생한 인권침해, 차별 등의 문제를 규제할 수 있는 적절한 제재 수단은 규정하지 않았

9) 앞의 연구용역 보고서, 39쪽~41쪽

10) 유엔 문서 A/HRC/48/31(2021), The right to privacy in the digital age, Roport of the United Nations High Commissioner for Human Rights. 47문(출처: 국가인권위원회, 인공지능(AI)개발과 활용에서의 인권 가이드라인 연구, 221년 11월, 241쪽)

습니다.

2. 인공지능 법률 개선 방안

가. 이용자 · 정보주체의 권리 명시 및 피해구제 절차 마련

인공지능 개발 · 활용 과정에서 이용자와 정보주체가 보장받아야 할 권리를 명확히 하고 이를 보호하기 위한 체계를 구축하는 것이 무엇보다 중요하므로, 인공지능 인권 가이드라인과 개인정보 보호법 제5장에 명시된 권리 보장에 관한 사항을 참고하여 인공지능 법률안에 정보주체의 권리를 구체적으로 명시하여야 합니다. 더불어 인공지능 기술로 인하여 권리 침해를 당한 이용자와 정보주체를 위한 구체 절차 역시 마련하여야 합니다.

인공지능 기술은 다양한 분야에서 활용되므로, 권리 침해 역시 다양한 형태로 발생할 수 있다. 따라서 인공지능으로 인해 발생할 수 있는 권리 침해 사안을 유형별로 분석 · 분류하고, 이미 제도화되어 있는 구제 절차 중 각 유형별로 이를 처리하기 적합한 구제 절차를 이행할 수 있도록 인공지능 법률안에 근거 규정을 마련하여야 합니다. 관련 법률 역시 이에 맞게 정비하여야 합니다. 또한 기존의 구제 절차로 다룰 수 없는 유형의 권리 침해 사안에도 대응할 수 있도록 인공지능의 특성을 고려한 별도의 구제 절차 역시 마련하여야 합니다.

예를 들어 정보통신망을 통하여 일반에게 공개를 목적으로 제공된 정보로 사생활 침해나 명예훼손 등 타인의 권리가 침해된 경우 그 침해를 받은 사람은 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제44조의2에 따라 정보통신서비스 제공자에게 해당 정보의 삭제 또는 반박 내용의 게재를 요청할 수 있습니다. 챗 지피티와 같이 인공지능 챗봇 서비스를 제공하는 과정에서 유사한 문제가 발생한 경우 이와 같은 구제 절차를 밟을 수 있도록 관련 규정을 정비하거나 별도의 구제 절차를 신설하는 방안을 모두 강구할 필요가 있습니다.

나. 분야별 · 영역별 위험성을 고려한 인공지능 등급 분류

인공지능 기술을 무분별하게 도입하여 활용할 경우 심각한 인권침해, 차별 문제 등을 초래할 우려가 있으므로, 인공지능을 도입 · 활용하고자 하는 분야 및 영역별로 인공지능이 국민의 인권과 안전에 미치는 위험성을 고려하여 등급을 구분하고, 위험도에 따라 규제 수준을 달리 하여야 합니다.

인권위가 인공지능 인권 가이드라인에서 제안한 단계별 구분 방안, 유럽연합이 입법 추진 중인 ‘인공지능법안’의 4단계 분류 방안, 캐나다가 현재 시행하고 있는 ‘자동화된

의사결정 지침'의 분류 방식 등을 참고하여, 인공지능이 도입 · 활용되는 분야 및 영역의 특성, 해당 분야 및 영역에 도입 · 활용될 인공지능이 국민의 인권과 안전에 미치는 영향을 분석하여 등급을 나누고, 각 등급별로 이 같은 위험을 최소화하기 위한 적정한 수준의 규제를 정하여야 합니다.

다. 고위험 영역 인공지능 확대 · 재정의

인공지능 법률안은 고위험 영역 인공지능을 상대적으로 협소하게 정의하여 국제 기준에 부합하지 못하는 측면이 있습니다. 따라서 유엔 인권이사회에서 채택한 '디지털 시대 프라이버시권 보고서'에서 고위험 영역으로 제시한 기준, 유럽연합의 '인공지능법안'에서 제시한 기준 등을 참고하여 고위험 영역 인공지능의 대상을 확대 · 재정의하여야 합니다.

아울러 고위험 영역 인공지능의 경우 공공과 민간 구분 없이 인공지능 감독 · 규제를 담당하는 기관이 사전에 해당 인공지능의 알고리즘이 투명성과 설명가능성을 충족하는지, 개인정보 보호, 인권침해 및 차별 문제를 일으킬 가능성은 없는지 엄격히 점검하도록 하여야 합니다. 또한 활용 과정에서 문제가 발생한 경우 인공지능 감독 · 규제 담당 기관이 해당 인공지능의 알고리즘 점검, 일시 사용 중지 명령 등 적절한 대응 조치를 취할 수 있도록 규제 수단 역시 마련하여야 합니다.

라. 인공지능 인권영향평가 도입

인공지능 기술에 대한 국제 기준이나 세계 각국의 대응을 살펴보면 공통적으로 인공지능의 개발과 활용 과정에서 발생할 수 있는 인권침해, 차별 등의 문제를 예방하기 위한 조치로서 사전적인 품질 관리와 사후적인 모니터링을 강조하고 있으며, 이를 위하여 인권영향평가를 도입하는 추세입니다.

유럽연합은 2016년 제정하여 2018년부터 시행한 '일반 개인정보보호법'에 자연인의 자유와 권리에 고위험을 야기할 수 있는 개인정보 처리에 대하여 공공과 민간을 불문하고 '개인정보보호 영향평가(Data Protection Impact Assessment)'를 실시하도록 규정하였습니다. 또한 2020년 5월 발표한 '인공지능 공공조달 백서'에서 국민의 기본권에 영향을 미치는 공공 부문 인공지능 조달에 있어 위험 기반 · 체계적 접근법에 따라 5단계 실사 절차를 권장하고, '위험영향평가'를 사전적으로 실시하여야 한다고 밝혔습니다.

아울러 위에서 살펴보았듯이 네덜란드, 캐나다는 공공부문에서 인공지능 기술 및 시스템을 조달하는 경우 사전에 영향평가를 실시하도록 하고 있으며, 인권위도 인공지능 인권 가이드라인에서 인공지능 출시 전 인권영향평가를 거치도록 하고, 출시 후 기능이 변경되거나 범위가 조정되는 경우 다시 한번 인권영향평가를 실시하도록 하여야 한다고

제안하였습니다.

따라서 인공지능 법률안도 인공지능 개발 및 출시 전에 인권침해와 차별의 가능성 및 정도, 영향을 받는 당사자의 수, 사용된 데이터의 양 등을 고려하여 인권영향평가를 실시하도록 하고, 인공지능의 기능 또는 활용 범위 변경 시에도 평가를 갱신하도록 관련 규정을 마련하여야 합니다. 아울러 인권영향평가 결과 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 해당 인공지능을 개발 · 활용하는 자가 이를 방지하거나 완화하기 위한 조치를 취하도록 관련 규정을 마련하여야 합니다. 그리고 이에 관한 사항을 인권 전문성과 독립성을 확보한 기관이 담당하도록 하여야 합니다.

마. 인공지능 산업 진흥과 규제의 분리

인공지능으로 인한 인권침해, 차별 등의 문제를 실효성 있게 예방하고 감독하기 위해는 산업 진흥과 규제를 분리하는 것이 바람직하므로, 인공지능 기술 및 산업 진흥에 관한 사항은 과학기술정보통신부 등 소관 부처 및 기관이 담당하도록 하고, 인공지능 감독 · 규제는 이와 분리하여 제3의 기관이 독립적으로 수행하도록 하여야 합니다.

이와 관련하여 인권위는 인공지능 인권 가이드라인에서 인공지능을 독립적이고 효과적으로 감독할 수 있는 체계를 수립하여 개인의 인권과 안전을 보장하고 피해를 구제하여야 하고, 인공지능 국가 감독 체계가 사건을 조사하기에 충분한 자원, 권한 및 전문지식을 구비해야 한다는 의견을 밝힌 바 있습니다.

따라서 개인정보보호위원회, 방송통신위원회 등 유사 업무를 수행하고 있는 기존 기관으로 하여금 인공지능 감독 · 규제, 개별 피해 사례에 대한 조사 및 구제 등에 관한 업무를 담당하도록 하는 것이 효율적인지, 아니면 이를 전담할 독립 기관을 신설하는 방안이 효율적인지 검토하여 가장 적정한 방안을 인공지능 법률안에 규정하여야 합니다.

3. 결어

인공지능은 인간의 단순 노동, 감정 노동을 대체하고, 업무의 효율성과 생산성을 향상시킬 뿐만 아니라 인간의 판단과 결정에 도움을 줌으로써 인간의 삶을 더욱 편리하고, 풍요롭게 해 줄 수 있을 것으로 기대되나, 인공지능 개발 · 활용 과정에서 인권침해, 차별, 사회적 편견의 확대 및 재생산, 개인정보 유출, 허위정보 생산, 저작권 침해 등의 문제가 발생할 수 있습니다.

따라서 인공지능 기술 및 산업의 진흥도 중요하지만, 이러한 문제점이 최소화될 수 있도록 관련 법령·제도를 잘 정비하는 것 역시 무척 중요합니다. 인권위는 2023. 7. 13. 제22차 상임위원회에서 이와 같은 문제점을 지적하고 개선방안을 담은 “인공지능법률

제정안에 대한 의견”을 표명하기로 결정하였으며, 조만간 결정문을 확정하여 국회에 송부하고 공개할 예정입니다. 인공지능법률 입법과정에서 인권위의 의견이 충실히 반영되기를 기대합니다. ▢

인공지능 법안 국회 토론회 토론문

김영규
(한국인터넷기업협회 정책1실장)

현장에서는 AI라는 파도와 글로벌 전쟁이라는 것을 정말 심각하게 받아들이고 있습니다. 신기한 것이 나왔다고 치부해버리기에는 우리의 삶에 너무나도 근본적인 변화가 일어나고 있기 때문입니다. 이 부분에서 우리가 자체적인 기술력, 역량, 성공사례들을 만들어내지 못한다면 많은 부분 사회에서 비효율이 발생할 수 있다는 우려가 있습니다.

이에 온갖 회의적인 시각들 사이에서도 오늘 제가 이 자리에 참석하게 된 것은 실리콘벨리 기업들이 선점 및 비교 우위에 있는 AI 시대에 국내 기업 또는 스타트업에게 어떠한 기회가 될 수 있을까 고민이 필요한 시점인 것 같아 토론회에 나오게 되었습니다.

기존에는 인터넷을 통해서 정보를 보려고 하면 복잡한 과정을 통해서 어디에 있는지 알아보고 복잡하게 작성되어있는 것들을 이해할 수 있어야 했습니다. 과거 '인터넷정보검색사'라는 자격증이 유행했던 적을 생각해 보시면 잘 이해가 되실 겁니다. 그러나 인공지능을 통해서라면 말 한마디에 정보를 가져다주고, 개인적인 이해도에 맞춰서 전달도 가능합니다. 불과 몇 년 만에 큰 변화를 맞이하게 된 것입니다.

그럼 한국에 왜 한국어 초거대 AI가 필요한가? 라는 질문에 답은 'AI 주권'에 있다고 생각합니다. 전 세계적으로 인공지능을 만들 수 있는 나라는 몇 안 됩니다. 미국, 중국, 싱가포르, 영국, 한국 정도이고 프랑스, 이스라엘 정도에서 인공지능 개발에 힘쓰고 있습니다. 만약 우리가 좋은 AI를 가지고 있지 않다면 외국 AI를 쓰게 될 것이고 대부분

ICT의 생산활동이 의존하게 된다면 경제적으로 큰 비용이 수수료로 지불되게 될 것입니다. 우리나라는 다시 조용하게 당하게 될 AI 식민지가 될 가능성이 큽니다.

이에 유럽에서 발의된 인공지능 법안 역시 이러한 상황과 무관하지 않습니다. 이전 플랫폼을 규제하기 위한 DMA나 DSA와 같이 유럽에서는 자국 플랫폼이 없는 상황에서 자국 플랫폼을 육성하기 위해 글로벌 빅테크를 규제하는 법안을 마련하였습니다. 그러나 우리나라는 자국 플랫폼을 가지고 있는 나라이고 이러한 법안이 우리나라 현실에 맞지 않다고 많은 전문가는 지적하고 있습니다.

인공지능 관련 기술은 플랫폼의 플랫폼이라 불리고 있고 앞서 말씀드린 바와 같이 주권을 빼앗기게 되면 경제적 손실이 발생할 수 있습니다. 이제 시작 단계에서 성급한 일반화나 입법보다는 유연한 제도 운용이 필요하다고 생각합니다.

물론 오늘 발제에 나와 있는 인권, 민주주의와 같은 대원칙에 대한 인공지능의 효과적인 관리체계를 위한 감독 체계 수립, 고위험 영역에 대한 사회적 합의 등 유승익 교수님이 말씀하신 부분에 대해 공감하지 않는 바는 아닙니다.

교수님과 다른 의미이지만 「인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안」에 대한 의견을 드리자면,

인공지능의 의미에 관하여 논의되는 바 등에 비추어볼 때 본 대안 제2조의 정의가 적절한 것인지 의문의 여지가 있습니다. 인공지능은 ‘인간이 가진 지적 능력(이하 지능)’이라는 개념을 기반으로 하여 창조된 개념으로, 해당 정의는 해석 주체, 국가별로 다르게 규정되고 있습니다. 그런데 본 대안에서 규정한 인공지능의 정의에 따르면 굉장히 넓은 범위의 소프트웨어 또는 자동화 기술을 포괄하게 되어 적용 대상이 모호해지고, 이에 따라 입법 목적을 달성하기 어려워질 것이 우려됩니다.

특히 현재 전 세계적으로 인공지능 산업의 발전에 관심을 기울이고 있다는 점을 고려하면, 본 대안의 취지인 산업 진흥 및 신뢰 기반 조성을 위해 국제 규범 및 표준을 고려해 정합성 및 상호운용성을 확보할 필요가 있다고 생각합니다.

대안 제2조 제3호는 ‘고위험 영역 인공지능’의 의미를 정의하며, 고위험 영역 인공지능에 해당할 수 있는 영역을 가목부터 자목까지 열거하고 있습니다. 그러나 열거된 내용을 살펴보면, 그 영역 자체가 매우 광범위할 뿐만 아니라 해석 및 실질적인 활용 방법에 따라 그 범위가 모호해질 우려가 있어 인공지능사업자 혹은 이용자의 입장에서 해당 인공지능이 고위험 영역 인공지능에 해당하는지를 판단하기 어려움이 있습니다.

인공지능은 특정 영역에만 이용할 수 있는 기술이 아니며, 다양한 영역에서 이용 가능하기 때문에 영역별 이용 방식, 그에 따른 결과, 영역별 기존의 규율 방식이 모두 다릅니다.

예를 들어, 제2조 제3호 가목의 에너지나 먹는 물, 라목의 핵물질 및 원자력시설 영역에서의 인공지능은 일차적으로 기계 및 설비에 적용되며 지역적인 영향을 미칠 가능성이 우선적인 것으로 예상되는 한편, 마목의 생체정보, 바목의 개인 권리·의무 판단 또는 평가 영역의 인공지능은 사람에 적용되며 개인적인 영향을 미칠 것이므로 그 이용

맥락이 크게 다른 것으로 보입니다.

이처럼 대안은 인공지능의 사례 및 맥락별 이용을 구체적으로 고려하지 않고 모든 인공지능을 획일적으로 규정하고 있어, 오히려 인공지능에 의해 발생할 수 있는 위험을 완화함에 있어 실효성이 낮으며 중복 규제 등으로 인해 불필요한 비용을 발생시킬 것으로 판단됩니다.

앞서 이야기드린바와 같이 인공지능과 고위험 영역이라는 부분에 정부, 전문가 집단, 이해관계자 등 신중한 논의를 통한 사회적 합의에 의한 입법이 진행되기를 바라고 현재 정부에서 이 부분과 관련해서 프로세스를 진행 중인 것으로 알고 있어 보조를 맞추어 진행되기를 희망합니다.

감사합니다. ▢

메모