인권기반 인공지능 개발 모델연구

-혐오차별 예방을 중심으로-

2022년 5월

국가인권위원회

조홍래

차 례

국외훈련개요

1. 훈 련 국 : 미국

2. 훈련기관명 : 일리노이시카고대학교
(University of Illinois at Chicago)

3. 훈련분야 : 인권정책

4. 훈련기간 : 2020년 8월 12일 ~ 2022년 6월 11일

훈련기관개요

1. 일리노이시카고대학교 (University of Illinois at Chicago)

2. 주소: 412 S. Peoria St, Chicago, IL 60607

3. 전화번호: (+1) 312 - 413 - 8088

4. 기능 및 조직

일리노이주립대학교(University of Illinois System) 안에는 어바나삼페인(UIUC), 시카고(UIC), 스프링필드(UIS)의 3개의 대학교가 속해 있다. UIC는 University of Illinois System 안에 있지만 독립적으로 운영되고 있는 연구종합대학이다.

UIC의 행정학과(Department of Public Administration)는 도시계획 및 행정대학(College of Urban Planning and Public Affairs)에 속해 있다. 대학원에는 3개의 석사과정과 1개의 박사과정이 있다. 3개의 석사과정은 MPA(Master of Public Administration), MPP(Master of Public Policy), MSCA(Master of Science in Civic Analytics)로 MPA는 관리 영역, MPP는 정책 분석 영역, MSCA는 통계 분석 영역에 특화되어 있다. 박사과정으로는 PhD in Public Administration 과정을 보유하고 있다.

< 훈련결과보고서 요약서 >

성 명	조홍래	직 급	행정주사
훈 련 국	미국	훈련기간	2020.8.12. ~ 2022.6.11.
훈련기관	일리노이시카고대학교	보고서매수	103 매
훈련과제	인권기반 인공지능 개발 모델연구		
보고서 제목	인권기반 인공지능 개발 모델연구		
내용요약	인공지능의 영향력은 전 세계적이지만 빅데이터 수집과 인공지능에 관련 기기가 필요하므로 주로 선진국을 중심으로 인공지능의 활용이 확산되고 있다. 따라서 인공지능의 역기능으로 인한 문제 발생 및 해결방안 논의 역시 이와 관련된 사례를 가지고 있는 국가를 중심으로 이루어진다. 우리나라는 인공지능 관련 기술과 기기를 보유하고 있으며 사회적으로도 인공지능에 대한 관심이 높아 인공지능으로 인한 문제해결 방안을 논의하기에 좋은 환경을 가지고 있다. 따라서 본 연구의 목적은 인공지능의 역기능을 예방하기 위하여 어떠한 제도적 기반을 갖추어야 하는지 알아보고, 인권에 기반한 인공지능 개발을 위하여 우리 사회가 어떠한 방향으로 나아가야 할지 고민하고자 한다. PEW 연구소(PEW Research Center)는 인터넷 사용자의 인종에 따라 다른 뉴스에 노출된다는 연구결과를 발표했다. 백인의 경우 35%가 인종과 관련한		

	게시물에 노출된 반면, 흑인은 68%가 인종과 관련된 게시물에 노출되었다. 그러므로 흑인은 약 2배 더 높은 비율로 인종과 관련된 게시물에 노출되는 것으로 나타났다. 인종 관련 소셜미디어 게시물이 인종에 따른 불이익, 억울한 상황 등을 다룬다는 것을 고려했을 때, 흑인은 백인보다 사회의 인종문제를 더 심각하게 받아들일 가능성이 높으며 이는 흑인과 백인의 견해 차이를 더 크게 만들 수 있다. 이는 곧 사회의 양극화 및 갈등을 심화시키고 사회 통합을 저해하는 요소로 작용할 수 있다. 2017년 사이언스지(Science)에 게재된 논문 “Semantics derived automatically from language corpora contain human-like biases“에 따르면 번역 인공지능이 성차별적 결과물을 산출했다고 한다. 연구자들은 터키어의 성 중립적 3인칭 대명사인 “o”를 의사와 간호사를 번갈아가며 함께 입력한 결과, 의사와 함께 “o”를 입력하는 경우 “o”를 남성으로 표시했으며, 간호사와 함께 “o”를 입력하는 경우 “o”를 여성으로 표시했다. 즉, 인공지능에게 성별을 입력 값으로 주지 않았음에도 불구하고 인공지능이 기존에 학습한 데이터를 바탕으로 의사는 남성으로, 간호사는 여성으로 번역하는 결과를 산출한 것이다. 인공지능은 자동화된 결정 시스템을 이루는 여러 기술을 통칭하는 개념이다. 여기에는 알고리즘, 신경망, 머신러닝, 딥러닝, 로보틱스 등을 포함시킬 수 있다. 이렇게 인공지능이라고 통칭되는 여러 기술들은 서로 연결되어 운영되고 있다. 예를 들어, 머신러닝은 신경망 기술에 기반하여 운영되고 있고

	딥러닝은 머신러닝의 학습방법 중 하나이다. 인공지능은 사람을 모방하여 사람과 비슷한 의사결정을 내리기 위해 개발된 컴퓨터 프로그램이다. 인공지능은 강력한 연산능력, 언제 어디서나 운영할 수 있는 존재감 등 사람보다 강점이 많은 것은 사실이다. 그러나 인공지능은 인간을 이롭게 하기 위하여 개발된 것이다. 따라서 인공지능의 우월성에 사로잡혀 인간의 능력을 평가절하하는 등 사람의 가치가 왜곡되는 일은 없어야 한다. 인공지능은 복잡하고 연속적인 의사결정 과정 때문에 결과에 대한 설명이 어렵고, 이로 인하여 투명성이 부족하다는 지적을 받는다. 머신러닝을 사용하지 않는 컴퓨터 프로그램은 사람이 입력한 프로그램 명령어에 의하여 작동하므로 어떤 명령어(코드)에 따라 컴퓨터가 결과 값을 냈는지 추적이 가능하다. 하지만 머신러닝이 작업한 결과 값은 컴퓨터 프로그램이 스스로 수많은 명령을 내리므로 결과 값에 영향을 미친 입력값과 관련된 연산과정을 찾기 어렵다. 마찬가지로 인공지능의 차별적 의사결정의 원인을 찾아내는 것 역시 굉장히 힘든 작업이다. 차별을 찾기 위해서는 인공지능이 내린 수많은 결정 가운데에서 어떠한 연산 작업에서 문제를 발생시켰는지 알아내야 하지만 이와 같이 복잡한 과정 속에서 원인을 특정하는 것은 굉장히 어려운 일이기 때문이다. 대안으로 제시할 수 있는 것은 인공지능에게 같은 조건에서 다른 정보를 입력한 후 산출된 결과 값을 비교해보는 것인데, 이 또한 중간 연산 과정을 설명할 수 없으므로 결과론적인 논의에서 머무는 한

	계를 가지고 있다. 이러한 문제를 해결하기 위하여 인공지능 관련 주요 기업들이 인공지능을 설명해주는 서비스를 제공하기도 한다. 최근 들어 구글은 Explainable AI, IBM은 Explainable artificial intelligence(XAI)를 통해 인공지능의 결정에 대한 설명을 제공하는 소프트웨어 서비스를 제공하기 시작했다. 인간에게 내재되어 있는 편향으로 인해 실제 데이터는 항상 편향을 내포하고 있다. 훈련용 데이터를 별도 구축하여 인공지능의 훈련에 사용한다 해도 이는 실제 데이터에 기반한 것으로 이미 편향된 구조 안에 머무는 한계가 있다. 즉, 머신러닝에 사용되는 데이터는 모두 직간접적으로 편향되어 있다. 이러한 편향이 잘못되었다는 것은 아니다. 인간과 사회가 각자의 특성을 가지는 것은 너무나 당연한 일이다. 다만 여기서 말하고자 하는 내용은 인공지능이 사용하는 데이터에는 인간의 특성이 이미 반영되어 있고 이것이 편향을 불러일으킨다는 것이다. 역사적으로 인간이 쌓아온 차별과 모순이 우리 사회에 자연스럽게 자리 잡고 있으므로, 사회를 반영한 데이터에도 역시 편향이 내재되어 있다. 미국의 대표적인 인공지능 규제 동향을 살펴볼 수 있는 것은 2022년에 발의된 알고리즘책임법안(The Algorithmic Accountability Act of 2022)이다. 주요 내용을 살펴보면 미국 연방거래위원회(Federal Trade Commission: FTC)에 알고리즘을 모니터링하는 담당부서 신설, 일정 규모 이상의 인공지능 회사에 대한 인공지능 영향평가 의무화, 연방거래위원회 인공지능 담당부서의 동향 보고서 작성 및 공

	개 등이다. 이 법안에 따르면 알고리즘이 자동적으로 중요한 의사결정을 내리는 시스템을 사용하는 회사는 알고리즘의 의사결정에 의한 영향을 사전에 평가해야 한다. 이 영향평가는 법 이후에 개발되어 활용되는 알고리즘뿐만 아니라 이전에 활용되고 있는 알고리즘까지 모두 포함한다. 각 회사에서 알고리즘의 영향을 자체적으로 평가할 때 측정 내용이 불명확하거나 회사마다 기준이 다르다면 평가의 실효성이 떨어질 수 있다. 따라서 이 법은 알고리즘에 대한 회사 자체의 평가가 표준화될 수 있도록 미국 연방거래위원회(FTC)가 이러한 평가의 기준을 안내하는 가이드라인을 마련하도록 정하고 있다. 해당 가이드라인은 영향평가 진행 및 보고 방법에 대한 구체적인 기준을 제시한다. 이 가이드라인은 본 법안에서 중요한 위치를 차지하고 있는데 그 이유는 영향평가가 향후 알고리즘 규제에 있어 시작점과 같은 역할을 하기 때문이다. 먼저 알고리즘을 활용하는 회사가 영향평가 결과에 대한 책임을 지도록 명시하고 있다. 이렇게 작성된 영향평가 자료는 연방거래위원회에 제출되며 연방거래위원회는 아카이브를 구축하여 이를 보관한다. 연방거래위원회는 향후 알고리즘에 문제가 발생하면 알고리즘 운영사에서 절차에 따라 영향평가를 진행하고 문제 예방을 위한 노력을 기울였는지 확인할 수 있다. 이러한 접근 방식은 자율통제를 통하여 연방정부의 업무 부담을 줄이는 한편, 기업의 자발적인 올바른 인공지능 사용문화 확산을 의도한 것으로 보인다. 민간의 인공지능 개발인력 규모와 기술발전 속도를 정부에서 따라잡기

	란 거의 불가능하다. 따라서 월등히 앞서가고 있는 민간을 규제하기 위해서는 이를 효율적으로 관리하기 위한 권한과 자원이 필요하다. 의료분야는 인공지능이 광범위하게 활용되고 있는 대표적인 분야 중 하나이다. 미국 식품의약국(US Food and Drug Administration: FDA)은 향후 의료분야에서 사용되는 인공지능 규제모델을 개발하기 위해 Digital Health Software Precertification (Pre-Cert) Program을 마련하고 시범적으로 인공지능 소프트웨어를 규제하고 있다. Pre-Cert 프로그램은 이미 개발되어 출시된 소프트웨어를 점검하는 것이 아니라 개발 단계에 있는 소프트웨어를 테스트해보고 문제가 있다면 개발 단계에서 수정하는 것이 특징이다. FDA는 전통적으로 의약품 사용에 따른 사고 발생을 예방하기 위해 제품 출시 전에 안전성을 확보한 후 허가를 내준다. FDA의 Pre-Cert 프로그램 역시 이러한 기존의 접근 방식을 사용한 것으로 보인다. 기존 의약품 인허가 방식은 새로운 약의 위험성을 진단해보고 일반 국민들이 사용해도 안전하다고 판단되면 시장에 출시할 수 있도록 허가를 내주는 형식인데, 여기서 의약품을 소프트웨어로 바꾼 것이 Pre-Cert 프로그램이라고 볼 수 있다. 우리의 일상생활 깊숙이 활용되고 있는 인공지능이 인권을 침해하지 않도록 예방하고, 인권침해 발생 시 이에 대응할 수 있는 제도를 마련하는 것이 필요하다는 것은 누구나 동의하고 있다. 하지만 이용자의 알 권리와 인공지능 개발사의 영업상 비밀을 보장받을 권리가 충돌하므로 이를 조정하기 위한

	제도가 필요하다. 인공지능이 인권을 침해하지 않기 위한 가장 효율적인 방법은 인공지능 설계 단계에서부터 인권침해를 예방하기 위한 노력을 기울이는 것이다. 이를 위하여 인권에 기반한 인공지능 개발 가이드라인이 필요하다. 인공지능 개발자는 인공지능이 편향된 데이터로 훈련을 받는다는 것을 고려하여, 이를 줄이도록 노력해야 한다. 그러나 개발자에게만 이러한 의무를 지우는 것은 바람직하지 않다. 기업의 경영자, 인공지능을 사용하는 이용자 모두 인공지능이 인권에 기반하여 개발되고 운영되어야 함을 최우선 가치로 여겨야 한다. 인권에 기반한 인공지능이 만들어지게 하기 위해서는 인공지능의 인권침해 예방과 관련된 사회적 논의 확산 및 지속이 필요하다. 국회의 입법과 정부의 정책 등 실질적인 대응 조치는 모두 사회적 논의에서부터 비롯된다. 이러한 논의 확산을 위해 사회 내에 다양한 주체를 세워서 인공지능에 의한 권리침해를 예방하는 문화를 확산해야 한다. ‘민간자율인공지능윤리위원회’는 사회 곳곳에서 인권기반 인공지능 개발 문화를 확산하는 데에 큰 역할을 할 수 있을 것이다. 이러한 민간 차원의 위원회가 사회 곳곳에 설치 및 운영되면 인공지능의 인권침해 예방 문화를 확산시킬 수 있다. 이를 통해 인공지능의 역기능에 대한 감시 주체가 크게 증가하고 해결방안 논의가 더욱 활성화될 수 있을 것으로 본다.

1. 서 론

1.1. 연구배경 및 목적

2019년 6월, 유엔인권이사회(UN Human Rights Council)는 이 사회의 자문위원회(Advisory Committee)에 인공지능을 포함한 신기술이 인권에 미치는 영향에 대한 보고서를 작성할 것을 요청했다.[1] 해당 자문위원회는 각종 인권 사안에 대해 인권이사회가 최선의 결정을 내릴 수 있도록 자문하는 기구로서 다수의 인권 전문가로 구성되어 있다. 자문위원회는 초안을 작성할 그룹을 별도로 구성하여 보고서를 준비하기 시작하였고 해당 초안 작성 그룹은 우리나라 출신의 전 유엔인권이사회 자문위원이 이끌었다.

유엔인권이사회가 이러한 움직임을 보인 이면에는 4차 산업혁명과 이로 인한 신기술의 영향력이 유엔 회원국에게 미치는 영향이 확대되었기 때문이었다. 인공지능 다양한 산업과 사회 분야는 물론 국경을 넘어 영향력을 급속하게 확대하였으며, 이는 유엔 기구의 활동에도 역시 영향을 미쳤다. 2018년 11월 유엔총회는 결의안(73/17)을 통해 급격한 기술변화가 지속가능발전목표 달성에 어떠한 영향을 미치는지에 대해 언급하며 신기술(인공지능)을 주목했다.[2]

2018년 유엔표현의자유특별보고관(UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression) 데이비드 케이(David Kaye)는 보고서를 통해 인공지능에 의한 인권침해 위험을 경고했다. 특별보고관은 인공지능 회사들이 보편적 인권기준에 따라 내부 방침을 정하고 시스템을 디자인하는 것이

1) A/HRC/41/L.14 (9 July 2019). New and emerging digital technologies and human rights.

2) A/RES/73/17 UN General Assembly resolution 73/17 (26 November 2018). 본 유엔총회 결의문에서 유엔은 급격한 기술변화가 지속가능발전목표 달성에 미치는 영향을 검토했다.

필요하다고 권고했다.[3] 특별보고관은 또한, 인공지능 회사들이 인권적 책임성을 고려하여 기술을 설계하고 운영할 수 있는 전문적 기준을 개발할 것을 권고했다. 이러한 유엔의 논의는 필자가 인권에 기반한 인공지능 개발이 어떻게 이루어질 수 있을지 고민하는 계기가 되었다.

당시 우리나라는 유엔인권이사회에서 관련 논의에 주도적으로 참여하는 등 해당 이슈를 선도하기 위하여 많은 활동을 펼쳤다. 유엔뿐 아니라 OECD 등 다른 국제무대에서도 관련 논의에 적극적으로 나서고 있었다. 하지만 국내적으로는 인공지능의 역기능으로 인한 문제를 다룰 수 있는 제도적 기반을 구축해야 하는 과제를 안고 있었다. 이는 비단 우리나라만 가지고 있는 과제가 아니었다. 인공지능 기술을 선도하고 있는 미국을 비롯한 주요 선진국 역시 이러한 과제를 가지고 있었다. 해당 이슈를 선도하고 있는 유럽연합(EU)의 경우 다른 나라보다 몇 년 일찍 논의를 시작하였고, 활발한 논의를 통해 제도 마련의 기초를 다지고 있었다.

인공지능의 영향력은 전 세계적이지만 빅데이터 수집과 인공지능에 관련 기기가 필요하므로 주로 선진국을 중심으로 인공지능의 활용이 확산되고 있다. 따라서 인공지능의 역기능으로 인한 문제 발생 및 해결방안 논의 역시 이와 관련된 사례를 가지고 있는 국가를 중심으로 이루어진다. 우리나라는 인공지능 관련 기술과 기기를 보유하고 있으며 사회적으로도 인공지능에 대한 관심이 높아 인공지능으로 인한 문제해결 방안을 논의하기에 좋은 환경을 가지고 있다.

따라서 본 연구의 목적은 인공지능의 역기능을 예방하기 위하여 어떠한 제도적 기반을 갖추어야 하는지 알아보고, 인권에 기반한 인공지능 개발을 위하여 우리 사회가 어떠한 방향으로 나아가야 할지 고민하고자 한다.

---

3) A/73/348, p. 18 참고. https://digitallibrary.un.org/record/1643488?ln=en

1.2. 연구내용 및 범위

인공지능이 인권에 어떠한 영향을 미치는지 구체적으로 파악하기 위해 언론에서 그동안 다루어졌던 인공지능의 인권침해 사례를 검토하고자 한다. 특히, 인공지능이 공정하지 못하고 편향된 결정을 하는 사례가 다수 보고되었으므로 차별과 관련된 이슈에 중점을 두고 내용을 파악해보고자 한다.

다음으로는 인공지능이 왜 이러한 인권침해 사례를 발생시켰는지에 대한 원인을 파악하기 위해 먼저 인공지능에 관련된 개념을 알아보고자 한다. 인공지능은 알고리즘, 신경망, 머신러닝, 딥러닝 등 여러 컴퓨터 기술을 통칭하는 단어이므로 인공지능과 연관된 개념을 살펴보며 인공지능의 운영원리를 파악해보고자 한다.

또한, 인공지능 개발 및 운영에 필수적인 빅데이터에 대해서도 알아보고자 한다. 빅데이터는 인간의 경험과 비슷한 것으로서 인공지능이 의사결정을 내릴 때 근거가 되는 요소이다. 따라서 인공지능의 운영원리를 파악하려면 빅데이터에 대한 이해 역시 필수적이므로 빅데이터에 대한 개념과 함께 빅데이터가 인권에 미치는 영향도 알아보고자 한다.

다음으로는 미국의 인공지능 규제 동향을 파악해보고자 한다. 미국은 인공지능 개발과 활용을 주도하고 있는 국가이며 기업은 물론 정부까지 인공지능 관련 제품을 폭넓게 활용하고 있다. 미국은 인공지능 활용 사례가 많은 만큼 문제점도 다수 발생하였으며 사회적으로 이에 대한 논의가 활발하게 이루어지고 있다. 따라서 미국의 인공지능 규제 사례를 통해 참고할 점을 알아보고자 한다.

마지막으로는 인공지능 개발 시 어떻게 인권을 존중하고 보호할 수 있을지에 대해 고민하고자 한다. 먼저 인권기반 인공지능

개발 가이드라인을 구상해보고 이러한 노력이 인공지능을 개발하는 개발자와 기업에게 어떠한 영향을 미칠 수 있을지 고민해본다. 이와 함께 국내의 인공지능 관련 논의 동향, 법률안 검토 등을 통해 인권에 기반한 인공지능 기술 개발을 위하여 우리나라에 어떠한 시스템이 필요한지 고민하고자 한다.

1.3. 연구의 한계

시의성 부족은 본 연구의 한계점이다. 본 연구 주제를 확정한 시점은 2019년으로 아직 국내에서 인공지능의 문제점과 해결방안에 관련된 사회적인 논의가 비교적 활발하지 않은 시기였다. 당시에는 국내에 이와 관련한 법률안이 제안되지 않았으므로 본 연구가 국내의 인공지능 규제 법안 마련에 있어 기초적인 참고자료로 활용되는 것을 목적으로 했다.

하지만 2020년부터 2021년 동안 인공지능과 관련된 법안이 9개나 발의되었으며 그 중 인공지능에 대한 규제내용을 담고 있는 법안도 다수이다. 또한, 2021년 11월 발의된 ‘알고리즘 및 인공지능에 관한 법률안’의 경우 본 연구에서 검토한 인공지능 규제방법의 많은 내용을 이미 담고 있었다. 따라서 본 연구에서는 제안된 법률안의 내용을 중심으로 실효성 제고 방안을 제시하고 인권기반 인공지능 개발을 위해 어떠한 요소가 필요한지 제시하고자 한다.

마찬가지로 2019년에 본 연구 기획 당시 구상했던 인권기반 인공지능개발 가이드라인의 경우 2020년에서 2021년 사이에 다수의 정부기관에서 인공지능 윤리 관련 가이드라인을 발표하였으므로 독창적인 내용을 만들기에는 어려운 상황이 되었다. 그럼에도 불구하고 개인정보보호 등 다른 요소보다 인권 존중을 중심으로 가이드라인을 만들기 위해 노력했다.

국내는 물론 국외에서도 인공지능의 역기능을 해결하기 위한 논의가 활발하게 진행되고 있으므로 본 보고서가 논의하고 구상한 내용의 시의성이 부족할 수 있다. 또한, 인공지능 기술의 급격한 향상으로 인해 기술적 논의 역시 시의성이 부족할 수 있다. 본 연구가 제시하는 대안은 현재 알려진 인공지능 기술을 고려한 방법이므로, 연구자가 미처 파악하지 못한 인공지능 기술이나 가까운 미래에 개발되는 인공지능 기술이 문제를 해결할 수도 있다.

2. 인공지능에 의한 차별사례

2.1. 광고 및 사회관계망 서비스(SNS)

광고 산업은 인공지능을 적극적으로 활용하고 있는 분야 중 하나이다. 특히, 인터넷 사용자에게 광고를 전송할 때 광고 수신자 선정 과정에서 편향적 결과가 발생할 수 있다. 광고사가 광고 수신자 선정 시 인적사항을 고려하는데 이러한 인적사항에 성, 연령, 인종, 학력, 지역 등 다양한 정보가 활용되고, 그에 따라 다른 광고를 전송하기 때문이다. 언론에 보도된 관련 사례를 살펴보면 광고사는 광고 수신자의 성, 인종, 정치성향 등에 따라 다른 광고를 전송하였고, 그 결과 광고 수신자의 편향성을 더욱 강화시켜 사회 통합을 저해하는 문제를 발생시켰다는 지적이 있다.

개인정보보호기술 발전동향(Proceedings on Privacy Enhancing Technologies) 학술지에 기고된 논문[4]에 따르면, 구글의 광고 타켓팅 도구가 성별에 따라 다른 광고를 전송했는데 그 광고의 내용이 여성에게 불리했다고 한다.

연구자들은 가상의 개인 프로파일 17,370개를 생성하여 구글 검색을 통해 구인광고 웹페이지를 방문하는 실험을 진행했다. 즉, 연구자들이 가상의 남녀 17,370명을 생성한 뒤 각 인물이 구직을 위해 인터넷 구직사이트를 방문하도록 한 것이다. 구직 사이트 검색을 위해 구글 엔진을 사용했고, 구글 엔진은 구직자의 인적정보를 활용해서 맞춤형 광고를 전송했다.

---

4) Amit Datta, Michael Carl Tschantz, Anupam Datta, "Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination," *Proceedings on Privacy Enhancing Technologies* Volume 2015: Issue 1 (April 18, 2015): https://content.sciendo.com/configurable/contentpage/journals$002fpopets$002f2015$002f1$002farticle-p92.xml

연구자들이 각 개인에게 전송된 광고를 확인한 결과 고소득 경력코치 서비스 광고가 여성 그룹에서는 318회 노출된 반면, 남성 그룹에는 1,852회 노출되었다. 남성이 여성보다 고소득 경력 코칭 서비스 광고에 약 6배 더 많이 노출된 것이다.

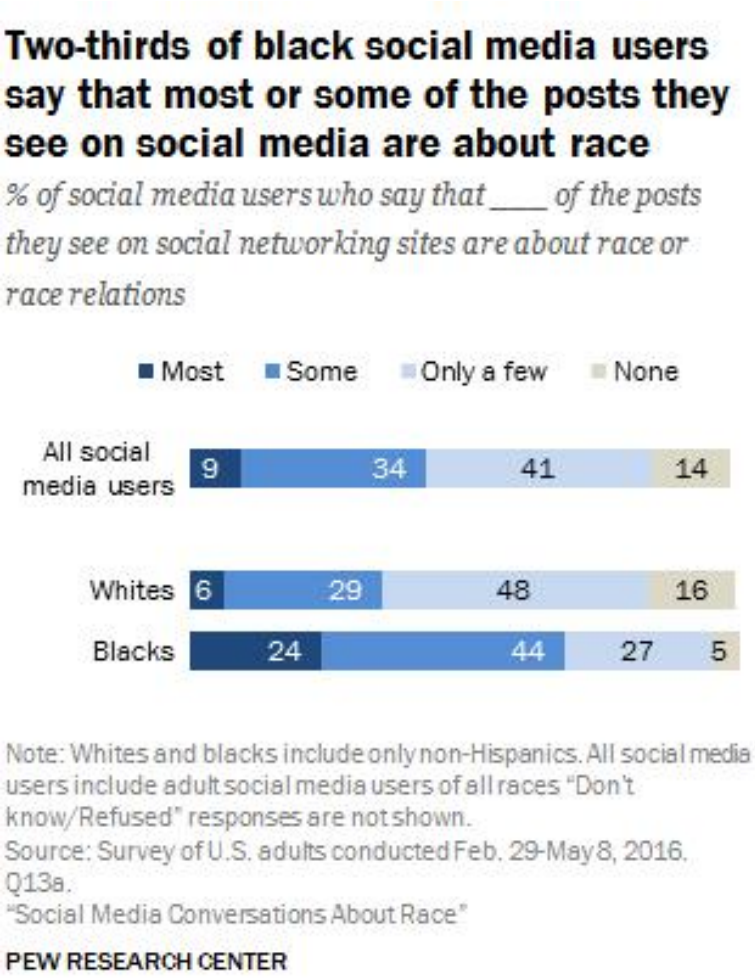
물론 이는 자동화된 광고시스템이 남성에게 고소득 경력 코치 서비스를 6배 더 제공한 것이 아니라 해당 서비스를 안내하는 광고에 노출된 비율이다. 하지만 인터넷이 정보 습득의 주요한 도구임을 감안하면 남성은 고소득의 직업을 가질 수 있는 조언을 들을 수 있는 서비스에 더 쉽게 접근할 수 있음을 의미하는 것이다.

뉴스, 검색엔진 등 웹페이지를 통한 서비스는 광고 수입을 기반으로 운영되는 경우가 대부분이다. 따라서 인터넷 이용자는 인터넷 사용 시마다 많은 양의 광고를 접하고 이에 영향을 받기 마련이다. 위에서 언급한 광고 사례를 단편적으로 본다면 시각에 따라 작게 느껴질 수도 있지만 이러한 차별이 사회에 누적된다면 우리 사회에 미치는 영향이 크다. 게다가 어린이부터 노인까지 모든 연령대가 스마트기기를 사용하고 있는 현실을 감안한다면 인터넷 광고가 얼마나 중요한 역할을 하는지 알 수 있다.

위 사례에서는 성별로만 실험을 진행하였지만 실제 인터넷 광고에서는 연령과 지역 등 다양한 정보를 활용하여 이용자에게 ‘맞춤형’ 서비스를 제공한다. 이러한 맞춤형 서비스 기술은 점차 발전되어 이용자의 거의 모든 프로파일 내용을 활용하고 있다. 그 결과 인터넷 광고업체는 이용자가 관심을 가질만한 정보를 제공하는 데에는 성공했지만 이는 사회 양극화를 불러일으켰다. 다음 사례에서는 소셜미디어가 어떻게 사회에 영향을 미치고 있는지에 대해 알아보도록 한다.

PEW 연구소(PEW Research Center)는 인터넷 사용자의 인종에 따라 다른 뉴스에 노출된다는 연구결과를 발표했다.[5] 연구에서 진행된 설문조사에 따르면 흑인 소셜미디어 사용자의 세명 중 두 명이 인종과 관련된 소셜미디어 게시물에 노출되었다고 응답했다. 한편, 백인 소셜미디어 사용자는 인종과 관련된 소셜미디어 게시물에 대한 노출 비율이 비교적 낮았다.

<표 2-1> 흑인 소셜미디어 사용자의 인종관련 게시물 노출 비율



5) Monica Anderson, "Social Media Conversations About Race," *Pew Research Center* (August 15, 2016),

https://www.pewresearch.org/internet/2016/08/15/social-media-conversations-about-race/

위의 조사 결과를 보면 인종과 관련된 게시물에 노출되지 않았다고 응답한 비율이 전체적으로는 14%이지만 응답자를 인종으로 구분해보면 백인은 16%가 인종관련 게시물에 노출되지 않았다고 응답한 반면, 흑인은 5%만 인종관련 게시물에 노출되지 않았다고 응답했다. 한편 인종과 관련한 게시물에 자주(Most) 노출되었다고 응답한 비율은 흑인 24%, 백인 6%로 약 4배 높게 흑인에게 관련 게시물이 노출된 것으로 나타났다.

또한 가끔씩(Some) 인종과 관련된 게시물에 노출되는 비율을 살펴보면 백인 29%, 흑인 44%로 흑인의 경우 가장 많은 비율이 가끔씩 인종과 관련된 게시물에 노출된다고 응답했다. 이를 자주(Most), 가끔씩(Some)으로 묶어서 살펴보면 백인의 경우 35%가 인종과 관련한 게시물에 노출된 반면 흑인은 68%가 인종과 관련된 게시물에 노출되었다. 그러므로 흑인은 약 2배 더 높게 인종과 관련된 게시물에 노출되는 것이다.

대부분의 경우 인종 관련 소셜미디어 게시물이 인종에 따른 불이익, 억울한 상황 등을 다룬다는 것을 고려했을 때, 흑인은 백인보다 사회의 인종문제를 더 심각하게 받아들일 가능성이 높으며 이는 흑인과 백인의 견해 차이를 더 크게 만들 수 있다. 이는 곧 사회의 양극화 및 갈등을 심화시키고 사회 통합을 저해하는 요소로 작용할 수 있다.

2.2. 대화형 인공지능(챗봇) 및 자동번역 분야

챗봇, 자동번역 등 언어와 관련된 분야는 인공지능 분야에서 많은 성과와 실용화를 이루어낸 분야 중 하나이다. 언어와 관련된 인공지능은 온라인에 존재하는 방대한 텍스트 자료를 인공지능이 학습한 결과 정확도를 높일 수 있었다. 하지만 만약 학습 자료에 잘못된 내용이 포함되어 있다면 인공지능은 그 내용을 그대로 학

습하여 결과를 산출하므로 문제를 재생산할 수 있다. 사람은 컴퓨터가 제시한 산출물이 사회적 맥락에 비추어 잘못되었다고 판단되면 이를 수정할 수 있지만, 인공지능은 잘못된 결과를 산출했을 때에 이를 자정하기 어렵다는 제한점이 있기 때문이다.

챗봇의 인권침해 사례로는 대화형 인공지능 테이(Tay)를 들 수 있다. 이 사례는 인공지능이 인종차별 등 인격을 모독하는 발언을 할 수 있음을 보여주며 인공지능의 역기능을 단적으로 보여주었다. 테이는 히틀러를 옹호하고 홀로코스트(에서의 대규모 인종학살) 사건이 가짜로 조작된 것이라는 등의 잘못된 결과를 산출했다. 결국 테이는 대화서비스를 개시한 지 하루도 되지 않아 운영이 중단되었고, 지금까지도 인공지능 실패의 주요사례로 거론되고 있다.

<그림 2-1> 테이의 인종차별 발언 사례

테이는 개발 시 대화 상대방의 언어를 모방하여 학습 한 후 비슷한 언어를 도출하도록 디자인되었다. 이를 파악한 특정 그룹이 테이에게 인종차별 및 성차별에 대한 내용을 반복적으로 말하여 학습시켰다. 그 결과 테이는 서비스를 시작한지 하루도 되지 않아 인종차별 및 성차별 발언을 시작하였다. 테이를 개발한 해당 회사

(마이크로소프트)는 서비스를 중단하고 공식 사과했다.

해당 회사는 테이에 대한 사례 분석 자료[6]에서 테이는 여러 환경에서 다양한 사용자 그룹을 통해 사전에 위험 테스트를 진행했고 테스트 단계에서는 문제가 발견되지 않았다고 주장했다. 하지만 테이가 온라인에 연결된 이후 특정 그룹의 (집중적인) 공격에 의해 문제가 발생했으며, 이러한 특수한 공격은 개발사가 미처 고려하지 못한 상황이었다고 설명했다.

테이 사례는 인공지능의 순기능이 큰 만큼 역기능도 클 수 있다는 것을 보여주었다. 개발사는 테이가 전 세계 청소년들의 말벗이 되기를 바라며 온라인을 통해 서비스를 개시했지만, 역기능이 발생하는 경우 전 세계 청소년을 대상으로 인종차별 및 성차별 발언을 할 수 있음을 보여주었기 때문이다. 게다가 인공지능은 시간과 공간의 제약 없이 무한히 결과물을 산출할 수 있으므로 대규모의 피해를 발생시킬 수 있었다.

인공지능 개발 시 인권적 요소를 고려했음에도 불구하고 학습 데이터에 따라 의도하지 않은 결과를 산출할 수 있다는 점은 이 사례에서 주목할 부분이다. 어떤 인공지능 개발사라도 자사의 인공지능이 인권을 침해하는 것을 원하지는 않을 것이다. 하지만 인권침해를 예방하도록 인공지능 프로그램을 디자인한 경우라도 현실 세계의 데이터에 의해 잘못 훈련되는 경우 의도했던 것과 다르게 인공지능이 인권을 침해할 수 있다.

자동번역은 우리가 일상생활에서 가장 쉽게 사용하는 인공지능 중 하나이다. 방대한 양의 언어를 학습한 인공지능이 단어와 단어를 교환하는 방식이 아닌 문맥에 따른 번역 결과를 제공하며 활용성이 높아졌다. 하지만 인공지능은 현실 세계의 데이터를 통하여 편향된 학습을 진행하였고, 이로 인해 편향된 번역 결과를 제공하였다.

6) https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

2017년 사이언스지(Science)에 게재된 논문 “Semantics derived automatically from language corpora contain human-like biases“[7]에 따르면 번역 인공지능이 성차별적 결과물을 산출했다고 한다. 연구자들은 터키어의 성 중립적 3인칭 대명사인 “o”를 의사와 간호사를 번갈아가며 함께 입력한 결과, 의사와 함께 “o”를 입력하는 경우 “o”를 남성으로 표시했으며, 간호사와 함께 “o”를 입력하는 경우 “o”를 여성으로 표시했다. 즉, 인공지능에게 성별을 입력 값으로 주지 않았음에도 불구하고 인공지능이 기존에 학습한 데이터를 바탕으로 의사는 남성으로, 간호사는 여성으로 결과를 산출한 것이다.

진행된 실험에서 연구자들은 터키어 “o birdoktor” and “o birhemşire”를 구글 번역기에 입력하여 영어로 변환하였고 그 결과 “he is a doctor” and “she is a nurse.”라는 결과 값을 받았다. 인공지능이 과거 데이터를 통해 의사는 남성, 간호사는 여성인 경우가 많은 것을 학습하여 이를 바탕으로 번역 작업한 결과물을 산출한 것이다. 물론 현실 세계를 반영한 결과가 아니냐는 의견이 있을 수 있지만, 이는 성차별 및 고정관념을 재생산하는 결과를 만드는 것이다. 이러한 번역 결과가 널리 활용되는 경우 사회의 언어가 과거의 성차별적 수준에 그대로 머무르게 되고 사회 인식 변화에 큰 걸림돌로 작용할 수 있다.

2.3. 인사관리 분야

인공지능의 활용이 다양한 분야로 확대되면서 인사관리 분야에도 영향을 미치고 있다. 인사관리에서 인공지능을 활용하는 것

7) Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan, "Semantics derived automatically from language corpora contain human-like biases" Science Vol. 356, Issue 6334, pp. 183-186. (April 14, 2017), https://science.sciencemag.org/content/356/6334/183, Retrieved from Princeton University, “Biased bots: Artificial-intelligence systems echo human prejudices”

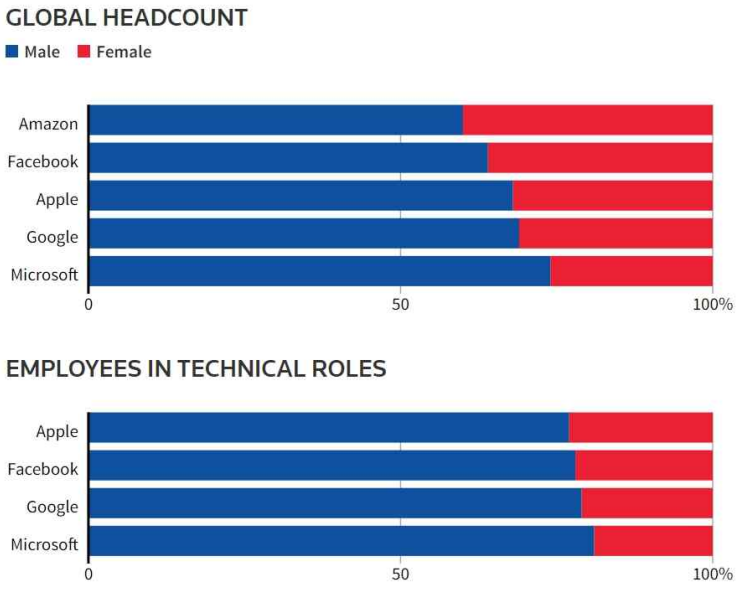
은 다른 분야보다 더 신중하게 접근해야 하는데 그 이유는 문제가 발생할 경우 다른 분야보다 더 큰 피해를 불러일으킬 수 있기 때문이다. 특히, 인공지능이 채용과 성과평가와 같이 인사관리 안에서도 파급력이 큰 분야에서 활용되므로 이러한 과정과 결과를 면밀하게 관찰할 필요가 있다.

널리 알려진 아마존의 채용 사례에서는 바로 위에서 검토해 본 의사, 간호사의 번역 사례로 살펴본 데이터의 성별 편향에 따른 문제를 다시 한 번 명확하게 보여준다. 로이터 통신에 따르면 아마존은 2014년부터 실험적으로 채용을 위한 인공지능 도구를 개발하여 운영했다.[8] 이 인공지능은 입사지원자의 지원서에 1점에서 5점 사이의 별점을 주고 점수가 높은 지원자를 사용자(인사담당자)에게 제시했다. 보통 상품과 서비스에 부여하는 별점을 사람에게 그대로 적용한 점도 비난받았지만 더 큰 문제는 해당 도구가 남성 편향된 데이터를 학습하여 여성 지원차를 차별했다는 점이었다.

이 인공지능은 아마존에 이미 고용되어 있는 직원의 이력서 데이터베이스로 머신러닝 훈련을 받았는데, 아마존의 소프트웨어 개발자와 기술관련 업무에 종사하는 직원은 여성보다 남성이 압도적으로 많았다. 해당 보도에서는 기술직에서 성별 비중이 얼마나 차이나고 있는지 다음과 같이 제시하였다.

<표 2-2> 인공지능 기업 임직원 성별 비중

8) Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters (October 10, 2018).



출처: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

위 그래프에서 파랑색은 남성, 빨강색은 여성 직원의 비율을 말하고 있으며 남성 비율이 적게는 약 60% 많게는 약 70%인 것을 보여주고 있다. 그래프에 나타난 페이스북(현 메타), 애플, 구글, 마이크로소프트 직원의 성별 비율은 인공지능 관련 주요 기업의 남성 편중을 그대로 보여주고 있다.

특히, 두 번째 그래프에서 보여주고 있는 기술직의 남성 편중 심화 현상은 인공지능이 성차별적 결과물을 도출하는 이유를 간접적으로 설명하고 있는 부분이기도 하다. 남성 개발자의 입장에서는 그동안 자신이 경험한 현실 세계를 바탕으로 '인간적으로' 인공지능을 개발하지만 이는 여성의 입장에서 현실세계를 이해한 부분이 포함되지 않은 한계를 가지고 있다.

2.4. 안면인식 분야

안면인식 인공지능은 방대한 양의 얼굴사진 데이터 학습을 통해 발전을 거듭했으며, 특히 보안 분야에서 신분 확인을 위한 도구로 사용되고 있다. 미 상무부(U.S. Department of Commerce)의 국립표준기술연구소(National Institute of Standards and Technology, NIST)는 2013년에서 2018년까지 안면인식 기술이 비약적으로 향상되었다고 평가했다. 본 연구소의 발표에 따르면 주요 안면인식 기술을 시험한 결과 오류발생 비율이 2010년 5%, 2014년 4%에서 2018년 0.2%로 급격하게 감소했다고 한다.[9)]

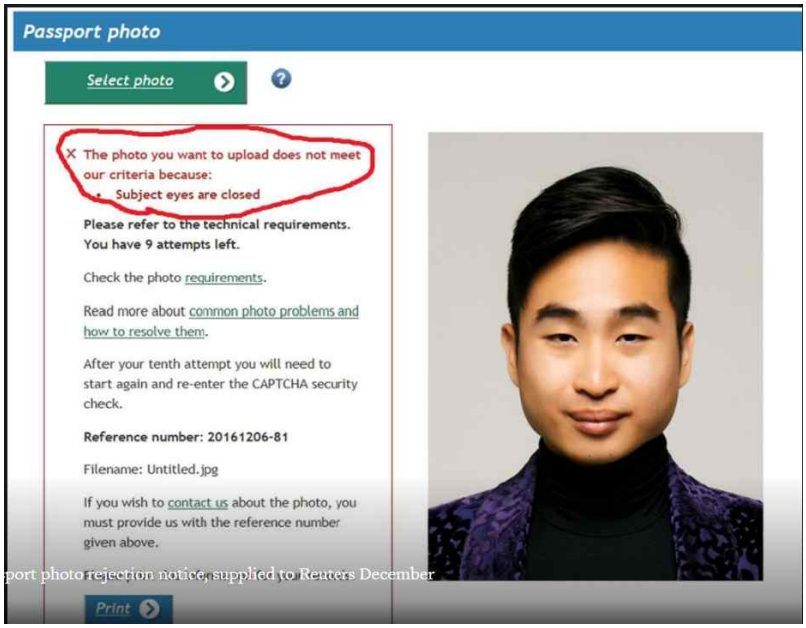
하지만 인공지능이 다양한 데이터를 학습하지 못하는 경우 학습 데이터에 반영되지 않은 인구집단은 신분 확인 시 오류가 발생하는 등의 불이익을 받을 수 있다. 안면인식 인공지능은 정부와 민간의 신분확인을 비롯한 다양한 분에에서 비교적 오랜 기간 동안 사용되었기 때문에 연관된 사례를 다수 찾을 수 있었다. 아래에서는 각 사례를 조금 더 깊게 살펴보면서 인공지능이 어떠한 피해를 줄 수 있는지 알아보도록 한다.

2016년 한 아시아계 뉴질랜드인은 자동 시스템을 통해 여권 발급을 신청하던 중 자신의 사진이 여권 발급에 부적합하다는 응답과 함께 여권 발급이 거절되었다.[10)] 그 이유는 AI가 이 여권발급 신청자가 눈을 감고 사진을 찍었다고 판단했기 때문이다. 만약 이 AI가 아시아인을 포함한 다양한 인종에 대해 학습을 충분히 했다면, 해당 사진에 대해 부적합 판정을 내리지 않았을 것으로 보인다.

<그림 2-2> 여권 사진 오류 사례

9) Patrick Grother, Mei Ngan, Kayee Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification," NISTIR 8238 (November 2018), https://doi.org/10.6028/NIST.IR.8238

10) Reuters Staff, "New Zealand passport robot tells applicant of Asian descent to open eyes," Reuters (December 7, 2016).

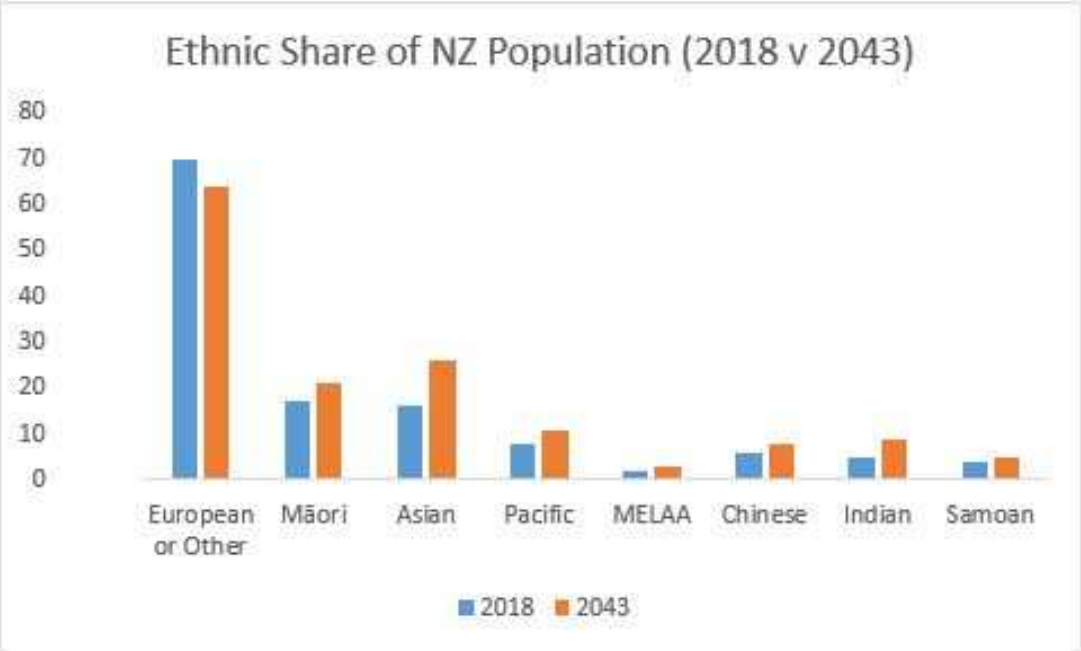


sport photo rejection notice, supplied to Reuters December

물론 해당 업무를 관장하는 뉴질랜드 담당부처의 대변인은 이 사례가 아시안에 대한 차별이라고 인정하지는 않았다. 대변인에 따르면 온라인으로 제출되는 여권사진의 약 20%가 거절되며, 가장 빈번한 사유는 눈을 감았다는 이유라는 것이다. 즉, 아시안 계 뉴질랜드인의 외모 때문이 아니라 기계의 통상적인 오류라고 주장하는 것처럼 들릴 수 있으나, 다른 시각으로 보면 이러한 대변인의 답변은 AI의 편향에 따른 오류인 것을 더 정확히 말해주고 있다.

해당 사안이 발생한 2018년 기준으로 뉴질랜드의 아시안 계 인구는 전체의 15.1%이다. 거절된 20%의 여권사진 중 눈을 감은 사유의 비율을 밝히지는 않고 있어 정확하지는 않지만, 아시안 계 비율 15%와 상당히 근접한 수치가 눈을 감은 이유로 거절당했을 확률이 높아 보인다. 만약 이 수치가 연관성이 부족하다고 해도 뉴질랜드 외교부에서 밝힌 20%의 거절된 여권 신청자들은 외모를 이유로 행정서비스 이용에 있어서 다르게 행정처분을 받은 것으로 보인다.

〈표 2-3〉 뉴질랜드 인구의 인종 비율



Source: Stats NZ, National Ethnic Population Projections 2018 base – 2043.xls

물론, 이 사안은 피해자에게 돌이킬 수 없는 불이익을 주거나 중대한 편익을 감소시켰다고 보기는 어렵다. 이 내용을 언론에 제보한 피해자 역시 자신의 여권사진에 대한 부적합 판정은 정부부처의 공무원이 내린 결정이 아닌 기계의 결정이므로 이에 대해 본인은 괜찮다(정부에 책임을 묻지 않겠다)고 인터뷰에서 답변했다. 하지만 이 피해자가 해당 내용을 기록하고 본인의 사진을 언론에 제공하면서까지 문제제기를 한 것으로 미루어 보면, 이에 대한 피해의식이 상당했을 가능성이 크다.

사법 분야에서의 인공지능을 통한 신분확인에서 문제가 발생할 경우에는 그 피해가 막대하다. 미국 콜로라도주 덴버에서는 은행 강도와 얼굴이 비슷하다는 이유로 스티브 탤리(Steve Talley)를 체포하고 수개월 간 구치소에 감금하고 조사했다. 스티브는 은행 강도 사건이 발생했을 당시 사무실에서 파이낸셜 애널리스트로서 업무를 위해 통화한 기록까지 있는 등 명확한 알리바이가 있었

지만 미 연방수사국(FBI)과 덴버 경찰에서는 스티브를 강력한 용의자로 지목하고 감금 조사했다. 스티브는 명확한 알리바이가 있어 풀려났지만 이후에도 당국에서는 오랜 기간동안 감시조치를 이어갔으며 수개월 뒤에는 새로운 증거를 찾았다며 또다시 체포하기도 했다.[11]

<그림 2-3> 스티브 탤리 사례

(Source: CBS4 Denver; Denver court documents)

최초 체포 시 덴버 경찰은 스티브의 집에 경찰특공대(SWAT)를 보내 섬광탄을 투척하고 다수의 요원이 스티브 위에 눌러 앉는 등 강력하게 제압했다. 이로 인해 스티브는 갈비뼈가 부러지고 치아가 손상되는 등의 큰 부상을 입었다. 신체적 부상 외에도 스티브는 실직으로 인한 파산 등 재정적인 어려움에도 처했다.

스티브는 매일 뮤추얼 펀드를 판매하고 다른 파이낸셜 어드바이저와 연락을 취해야 하는 등의 업무를 해야 했지만 장기간 조사로 인한 공백과 이후 이어진 당국의 감시, 재구속 등으로 인해 스티브는 일자리를 잃었으며 이후 복직하지 못했다. 이후 스티브는 덴버 경찰과 FBI를 상대로 미화 천만 달러 규모의 피해보상 소송을 제기했다.

11) 관련기사: https://denver.cbslocal.com/2016/09/15/former-financial-advisor-wrongly-accused-of-bank-robbery-fights-to-win-life-back/

이상 살펴본 안면인식 분야의 인공지능 차별 사례에서 몇 가지 시사점이 있다. 먼저 여권사진 오류 사례에서 주목해야 할 점은 정부부처에서 대국민 행정서비스 제공을 위해 사용한 인공지능에 문제가 발생했다는 점이다. 일반 기업의 경우 유사사례 발생 시 해당 기업에서 제공하고 있는 서비스 중 인공지능을 활용한 특정 서비스 이용자가 피해자가 되므로 그 범위가 제한적이다. 하지만 정부에서 대국민 민원서비스에 인공지능을 적용할 경우 문제 발생 시 피해자의 범위가 크게 확대된다.

책임소재 파악 역시 정부가 기업에 비해 더 복잡하고 시간이 오래 걸린다. 기업의 경우 대표이사와 기업 내에 있는 해당 인공지능 개발부서(또는 외주를 주었을 경우 계약 당사자)가 문제 발생 시 책임을 지고 신속하게 오류 개선 등 후속 조치를 취하면 되지만 정부부처의 경우 자체 인공지능 개발부서가 없는 경우가 많으므로 외주 업체, 담당 공무원과 관리자가 애매하게 책임을 나누는 구조이다. 책임 소재가 불명확하면 후속 조치에 있어서도 의사결정 참여자가 많으므로 속도를 내기 어렵다.

위에서 검토한 여권사진과 은행강도 수사 사례를 비교해보면 사법 분야의 인공지능 사용이 얼마나 큰 영향을 미치는지 알 수 있다. 사법 분야에서 인공지능이 잘못 사용된다면 민원서비스보다 피해내용이 훨씬 더 심각하며, 피해자는 돌이킬 수 없는 불이익을 받을 수 있다. 사법기관은 인공지능 사용 시 기계에 대한 과도한 신뢰를 지양할 필요가 있다.

수사 당국은 은행의 감시카메라에 기록된 은행 강도의 얼굴, 특히 귀의 형태가 스티브와 유사한 점, 그의 지인이 감시카메라 영상을 보고 스티브라고 말한 점 등을 근거로 스티브를 강력한 용의자로 지목하고 수사를 지속했다. 하지만 스티브가 명백한 알리바이가 있고 범행 동기가 부족한 점에 대해서는 간과했는데 이는 신기술에 대한 과도한 의존에서 비롯한 것으로 보인다.

신기술에 대한 의존은 앞으로 더 강화될 것으로 보인다. 정부는 업무의 객관성, 정확성, 신속성 등 여러 측면에서 인공지능을 통한 업무개선을 이룰 수 있는 기회가 많다. 하지만 인공지능이 위 사례와 같이 바람직하지 않은 결과를 도출하는 경우 큰 사회적 비용을 유발하며, 특히 사법 분야에서 인공지능이 오류를 발생시킬 경우 돌이킬 수 없는 피해를 만들 수도 있다.

3. 인공지능과 빅데이터의 운영원리와 차별

3.1. 인공지능의 주요 개념

인공지능 분야는 최근 기술발달에 힘입어 급격한 성장을 이루었다. 매년 인공지능의 기술범위가 폭발적으로 확장되고 있어 인공지능의 용어를 단적으로 정의하기 어려울 정도이다. 한 인공지능 관련 기업은 인공지능을 '기계가 인간의 지능을 모방하거나 넘어서는 것을 가능하게 하는 여러 기술을 통칭하는 것'이라고 말하고 있다.[12)]

인공지능에 포함되는 여러 기술에는 예를 들어 알고리즘, 신경망, 머신러닝, 딥러닝, 로보틱스 등을 포함시킬 수 있다. 이렇게 인공지능이라고 통칭되는 여러 기술들은 서로 연결되어 운영되고 있다. 예를 들어 머신러닝은 신경망 기술에 기반하여 운영되고 있고 딥러닝은 머신러닝의 학습방법 중 하나이다.

우리나라 국민에게 널리 알려진 인공지능은 구글 자회사인 딥마인드의 인공지능 '알파고(AlphaGo)'이다. 바둑 인공지능 프로그램인 알파고는 다수의 인공신경망을 조합한 프로그램[13)]으로 2016년 이세돌 9단을 이기며 인간을 넘어서는 능력을 보여주었다. 우리나라는 물론 전 세계 200만 명이 이를 관람하였고, 많은 사람들이 인공지능의 급속한 성장에 놀라는 계기가 되었다.

사람들은 알파고의 승리를 비롯하여 다양한 분야에서의 인공지능의 성과를 목격하였고, 컴퓨터 프로그램이 사람보다 더 우수하다는 인식을 가지기 시작했다. 사람의 사고력을 압도하는 컴퓨터

12) "What is AI," Qualcomm, https://www.qualcomm.com/products/artificial-intelligence/what-is-ai-faq.

13) https://www.deepmind.com/research/highlighted-research/alphago

프로그램에 대한 막연한 두려움은 인공지능과 관련된 영화나 드라마 등에 그대로 반영되었다. 인공지능은 사람을 모방하여 행동하고 반복적이고 지속적인 의사결정 권한을 가지고 있으며 복잡한 의사결정을 단시간 내에 내릴 수 있다. 이와 같은 특성들로 인해 사람들은 인공지능이 마치 사람보다 높은 위치에 있는 것처럼 느끼며 인공지능 프로그램의 결과를 과신하는 부작용을 발생시키기도 했다.

하지만 어디까지나 인공지능은 사람을 모방하여 사람과 비슷한 의사결정을 내리기 위해 개발된 컴퓨터 프로그램이다. 물론 비교할 수 없는 강력한 연산능력, 언제 어디서나 운영할 수 있는 존재감 등 사람보다 강점이 많은 것은 사실이다. 그러나 인공지능은 인간을 이롭게하기 위하여 개발된 것이다. 따라서 인공지능의 우월성에 사로잡혀 인간의 능력을 평가절하 하는 등 사람의 가치가 왜곡되는 일은 없어야 한다.

〈알고리즘〉

알고리즘은 최근 인공지능 규제와 관련하여 많이 언급되고 있는 단어 중 하나이다. 많은 법률안이나 관련 보고서들이 알고리즘을 규제 대상으로 삼는 경우가 많다. 알고리즘은 컴퓨터가 일련의 정보를 처리하는 과정을 통칭하는 것이다. 인공지능과 관련된 개념들이 다양하고 생소하다보니 알고리즘을 특정 컴퓨터 프로그램 운영 방식으로 여기는 경우도 있다.

알고리즘 역시 인공지능과 마찬가지로 모두가 동의하는 정확한 정의가 있는 것은 아니다. 하지만 다수의 보고서를 검토한 결과 알고리즘은 정보를 처리하는 과정으로 이해하고 있는 것을 알 수 있었다. EU에서 발간한 보고서에 따르면 알고리즘이란 컴퓨터 내에서 다수의 명령어에 의해 입력값이 출력값으로 산출되는 과

정[14]으로 보고 있다. 또한, 미국 백악관에서 발간한 보고서에서는 알고리즘을 데이터에 연속적으로 적용할 수 있는 지침과 과정으로 보고 있다.[15]

단순화시킨 예시를 통해 알고리즘을 설명하면 다음과 같다. 우리가 가나다순에 따라 명단을 작성한다고 할 때, 우리는 순서가 정해지지 않은 명단을 컴퓨터에 입력한 후 컴퓨터에게 가나다순으로 정리하게 명령한다. 그러면 컴퓨터는 명단에 있는 이름을 서로 비교해가며 순서를 재배열한 후 가나다순으로 정리된 결과를 산출한다. 여기서 입력 값은 순서가 정해지지 않은 명단이며 출력 값은 가나다순으로 정리된 명단이다. 그리고 컴퓨터가 명단에 있는 이름을 가나다순으로 정리하기 위하여 각각의 이름을 비교하고 순서를 재배열한 것이 입력 값과 출력 값 사이에서 진행되는 다수의 명령어인 것이다.[16]

<신경망>

알고리즘과 관련해 조금 더 실제적인 논의를 위해 알고리즘에 사용되는 신경망(Neural Networks)의 내용을 알아보는 것이 필요하다. 신경망은 인간의 뇌를 흉내내어 만든 정보처리 기술이다. 인간의 뇌가 수많은 신경세포의 연결에 의하여 작동하는 것을 모방하여 컴퓨터에도 수많은 정보를 입력시켜 그것들을 서로 연결시키고 영향을 주고받도록 설계한 것을 신경망이라고 말한다.

---

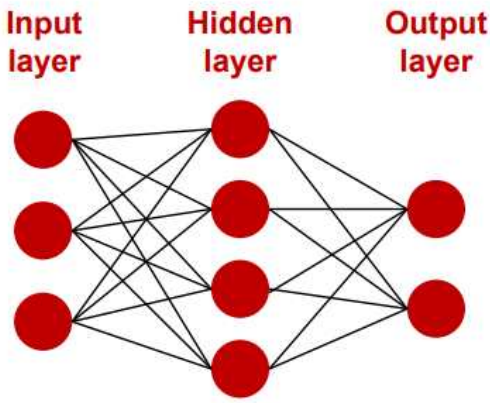
14) European Union Agency for Fundamental Rights, “BigData: Discrimination in data-supported decision making,” FRA Focus, (2018) p. 4, https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making

15) White House, Executive Office of the President, “Big data: seizing opportunities, preserving values,” (Washington : 2014), p. 46.

16) European Union Agency for Fundamental Rights, “BigData: Discrimination in data-supported decision making,” FRA Focus, (2018) p. 4, https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making

앞서 알고리즘 설명에서 등장했던 입력 값과 출력 값, 중간 처리과정과 연결하여 설명해보자면 다음과 같다. 컴퓨터에 자료를 입력(Input layer)하면 이 자료들이 서로 정보를 주고받으며 입력 값을 변경시키는 과정(hidden layer)을 거친다. 이러한 과정을 거친 후 출력 값(output layer)을 얻는데 여기서의 출력 값은 중간 변경 과정에 의해 입력 값과는 다른 결과를 도출한다.

<그림 3-1> 신경망



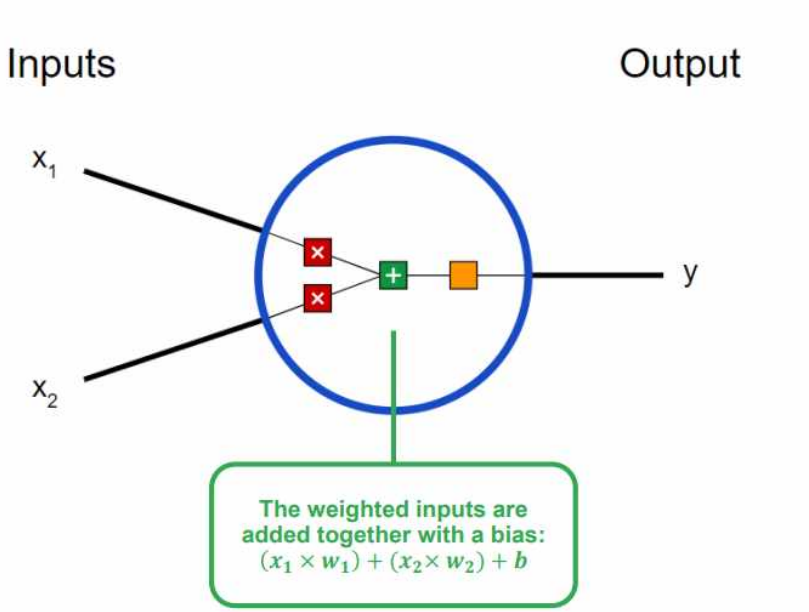
출처: Gonzalo A. Bello, ExploreCSR workshop (April 23, 2022) presentation material.

이러한 신경망은 내부에서 정보를 주고받을 때 가중치(weights)와 편향(bias)을 통해 출력 값의 정확성을 높인다. 즉 데이터와 데이터 사이에서 정보를 주고받을 때 어떠한 정보에 더 무게를 두고 데이터를 처리할지 정하고 그에 따라 출력 값을 조정하는 것이다.

머신러닝이란 실제 값과 컴퓨터가 산출한 값을 비교하여 컴퓨터가 산출한 결과가 실제에 더 가까워질 수 있도록 가중치와 편향을 조정하는 과정이다. 이러한 과정을 반복하여 컴퓨터가 실제와 가까운 값을 낼 수 있도록 하는 것이 머신러닝이다.

아래 그림은 한 개의 뉴런이 두 개의 입력 값을 받아서 각각 가중치를 곱하고 편향을 조정하며 새로운 출력 값을 산출하는 과정을 도식화한 것이다. x 부분에서는 가중치가 곱해지며 + 부분에서는 편향이 더해진다. 이러한 뉴런이 수없이 많이 연결되어 하나의 레이어(layer)를 구성하므로 컴퓨터는 입력 값을 반복적인 연산을 통해 실제 값과 유사한 출력 값으로 만든다.

<그림 3-2 뉴런>



출처: Gonzalo A. Bello, ExploreCSR workshop (April 23, 2022) presentation material.

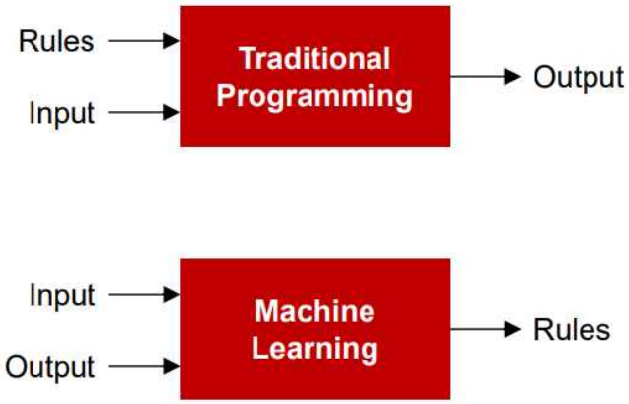
<머신러닝>

머신러닝은 외부의 프로그램 없이 기계가 스스로 학습하는 것을 말한다. 머신러닝이 기존의 컴퓨터 프로그램과 다른 점은 이러한 신경망에 의해 스스로 훈련이 가능하다는 것이다. 기존의 컴퓨터 프로그램은 사전에 입력되어 있는 계산방식에 따라 입력 값을 처리한 후 결과를 도출하고 더 이상의 처리과정이 진행되지 않

았다. 하지만 머신러닝은 정해놓은 규칙에 따라 처리한 결과 값을 다시 입력 값으로 받아들이면서 이를 반복 연산한다.

기존의 컴퓨터 프로그램은 반복 처리 시 사람의 명령에 의해 진행이 되었지만 머신러닝은 스스로 결과 값을 다시 입력 값으로 변환시키므로 사람의 명령 없이 학습을 진행할 수 있다. 이렇게 반복적인 처리과정을 통해 프로그램이 스스로 실제 값에 가까워지도록 출력 값을 조정하고 발전해나가는 것이 머신러닝의 강점이다. 이를 가능하게 한 것은 앞서 살펴본 신경망의 작동원리이다.

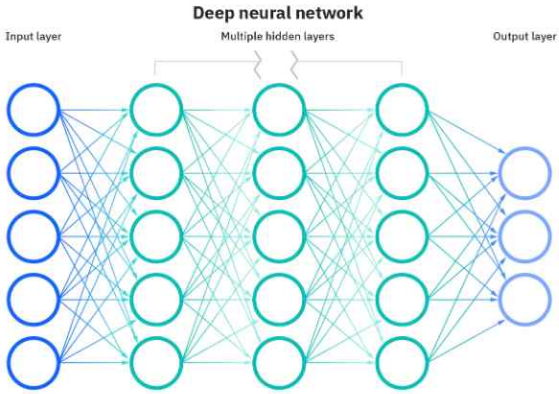
<그림 3-3> 머신러닝과 기존 컴퓨터 프로그램



출처: Gonzalo A. Bello, ExploreCSR workshop (April 23, 2022) presentation material.

<딥러닝>

<그림 3-4> 딥러닝의 구조



출처: https://www.ibm.com/cloud/learn/neural-networks

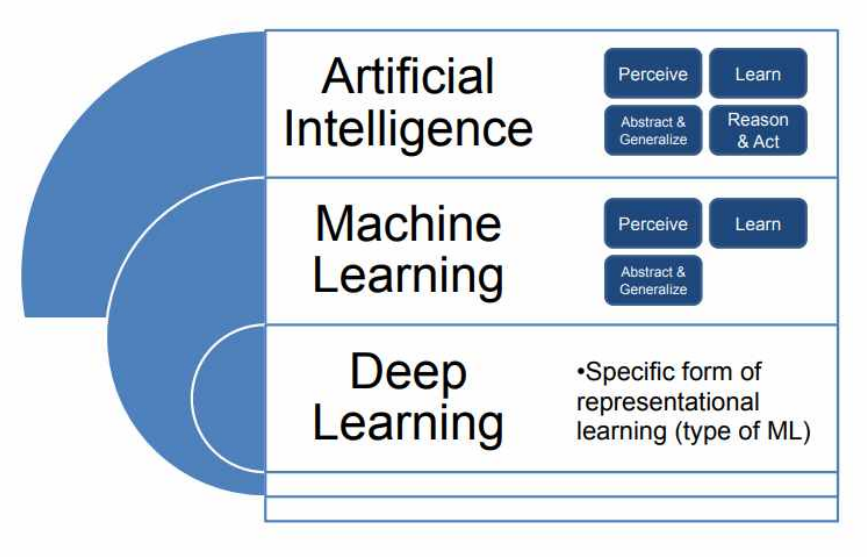
딥러닝은 이름은 머신러닝과 비슷하지만 그 내용은 신경망과도 연결되어 있다. 머신러닝이 기계가 스스로 반복 학습하는 과정을 말하는 것이라면, 딥러닝은 신경망의 층(layer)의 형태에 대해 말하는 것이다. 앞서 살펴본 <그림 3-1> 신경망 설명 그림을 다시 확인하면서 살펴보면, 신경망은 입력층(input layers), 은닉층(hidden layers), 출력층(output layers)으로 이루어져있다.

<그림 3-1>에서는 중간의 은닉층(hidden layer)이 한 개만 있었지만 위의 그림 3-4와 같이 다수의 은닉층이 있는 경우 훨씬 더 많은 연산 작업을 반복하므로 컴퓨터가 학습하는 양이 증대한다. 이에 따라 컴퓨터는 결과값을 실제값과 더 유사하게 조정할 수 있는 기회를 많이 가지게 되고, 결국 정확도가 높아지는 효과를 얻을 수 있다.

여기서 은닉층의 갯수가 적으면 얕다고 표현하고, 비교적 갯수가 많으면 깊다고 표현하면서 딥(deep)러닝(learning)이라는 용어를 사용하고 있는 것이다. 위 그림에서 다수의 은닉층(Multiple hidden layers)라고 표현한 중간 부분은 3개를 말하는 것이 아니라 수십개의 은닉층을 말하고 있는 것이다. 1960년대에 신경망 연구가 처음 시작될 때에는 은닉층이 한 개에 불과했지만 1980년대에는 2~3개의 은닉층을 사용했고, 지금은 50개가 넘는 은닉층을 사용하고 있다.[17] 머신러닝 시 학습의 난이도가 높은 경우 많은(또는 깊

은) 은닉층을 사용하며 비교적 난이도가 낮은 경우 소수의(또는 얕은) 은닉층을 사용한다.

이상 살펴본 인공지능, 머신러닝, 딥러닝에 대한 이해를 돕기 위해 이들 간의 관계를 도식화하면 아래와 같다.



출처: FTC 공청회 자료[18]

3.2. 인공지능 운영원리와 차별

위 그림과 같이 복잡한 은닉층을 활용하여 자료를 처리하는 경우 입력 값이 어떠한 과정을 통해 출력 값에 이르렀는지 설명하기가 거의 불가능하다. 입력된 자료가 수많은 은닉층에서 자료를 주고받으며 서로 영향을 미치고 값을 변경시키기 때문에 어느 시점에서 어떤 원인으로 결과가 도출되었는지에 대한 추적이 어렵다.[19]

17) Larry Hardesty, "Explained: Neural networks," MIT News, April 14, 2017, https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

18) FTC, Hearing #7 on Competition and Consumer Protection in the 21st Century (November 13, 2018), p. 24. https://www.ftc.gov/system/files/documents/public_events/1418693/cpc-hearings-howard_11-13-18.pdf

19) European Union Agency for Fundamental Rights, "BigData: Discrimination in data-supported decision making," FRA Focus, (2018) p. 6,

인공지능은 이렇게 복잡하고 연속적인 의사결정 과정 때문에 결과에 대한 설명이 어렵고, 이로 인하여 투명성이 부족하다는 지적을 받는다. 머신러닝을 사용하지 않는 컴퓨터 프로그램은 사람이 입력한 프로그램 명령어에 의하여 작동하므로 어떤 명령어와 코드에 따라 컴퓨터가 결과 값을 냈는지 추적이 가능하다. 하지만 머신러닝이 작업한 결과 값은 컴퓨터 프로그램이 스스로 수많은 명령을 내리므로 결과 값에 영향을 미친 입력값과 관련된 연산과정을 찾기 어렵다.

마찬가지로 인공지능의 차별적 의사결정의 원인을 찾아내는 것 역시 굉장히 힘든 작업이다. 차별을 찾기 위해서는 인공지능이 내린 수많은 결정 가운데에서 어떠한 연산 작업에서 문제를 발생시켰는지 알아내야 하지만 이와 같이 복잡한 과정 속에서 원인을 특정하는 것은 굉장히 어려운 일이기 때문이다.

대안으로 제시할 수 있는 것은 인공지능에게 같은 조건에서 다른 정보를 입력한 후 산출된 결과 값을 비교해보는 것인데, 이 또한 중간 연산 과정을 설명할 수 없으므로 결과론적인 논의에서 머무는 한계를 가지고 있다. 이러한 문제를 해결하기 위하여 인공지능 관련 주요 기업들이 인공지능을 설명해주는 서비스를 제공하기도 한다. 최근 들어 구글은 Explainable AI, IBM은 Explainable artificial intelligence(XAI)를 통해 인공지능의 결정에 대한 설명을 제공하는 소프트웨어 서비스를 제공하기 시작했다.

---

https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making

3.3. 빅데이터의 편향성

<빅데이터(Big Data)>

2014년 당시 미국 오바마 행정부는 새롭게 부상하고 있는 빅데이터에 대한 이해를 높이기 위하여 백악관 차원에서 워킹 그룹을 설치하여 관련 이해당사자들로부터 광범위하게 정보를 수집했다. 워킹 그룹의 활동 기한은 90일이었지만 해당 기간 안에 기업, 학계, 시민사회, 연방정부 등 수백 명을 만나며 빅데이터의 순기능을 강화시키는 한편 역기능을 최소화하기 위한 방안을 고민했다.

해당 보고서는 당시 미국의 기업 등 민간분야는 물론 정부를 포함한 공공분야까지 미국 사회 전반에 걸쳐 많은 영향을 미쳤으며, 지금까지도 미국 내에서 빅데이터와 관련한 주요 문서로 여겨지고 있다. 미국 사회가 빅데이터에 대한 관심을 높이고, 다양한 분야에서 빅데이터를 수집하고 활용하는 계기를 만들었다는 평가를 받고 있다. 따라서 비교적 많은 시간이 지난 지금까지도 빅데이터에 관한 주요 문서로 논의되고 있다. 이 보고서는 작성 당시 Massachusetts Institute of Technology(MIT), New York University(NYU), University
of California, Berkeley(UC Berkeley)와 같이 빅데이터 관련 기술을 선도하고 있는 학교와 공동으로 컨퍼런스를 개최하는 등 빅데이터에 대한 심도 있는 이해를 하고자 노력했다.

본 보고서는 빅데이터가 무엇인지를 설명하면서 The “3 Vs: Volume, Variety and Velocity” (p. 4.)를 제시했다. 빅데이터는 크기(Volume)가 무척 크며, 다양하고(Variety), 빠른 속도로 변화한다(Velocity)고 설명하고 있다. 빅데이터가 일반 데이터보다 크기가 훨씬 큰 점에 대해서는 쉽게 수긍할 수 있지만 다양성과 변화성에 대해서는 개념이 모호할 수 있다. 따라서 아래 설명을 통해 빅데이

터가 왜 다양하고 빠른 속도로 변하고 있는지 알아보도록 한다.

빅데이터는 Internet of Things 등 디지털 기기 증가에 따라 급속하게 성장하고 있다. 동 보고서는 현재는 디지털과 아날로그 자료가 빅데이터로 만들어지고 있지만, 앞으로 디지털의 비중이 급격히 높아질 것으로 내다보았다. 여기서 디지털 자료는 컴퓨터의 자료처리에 의해 생성되는 데이터를 말하는 것으로 이메일, 웹 검색, GPS 정보 등을 예로 들 수 있다. 아날로그 자료는 물리적인 실제 세계로부터 생성되었으나 아날로그로 변환된 것이다. 전화와 카메라를 통한 음성과 영상 정보, 심장 박동수와 같은 신체활동 데이터가 아날로그에서 디지털로 변환된 빅데이터로 볼 수 있다. 기존의 아날로그 기기들이 디지털로 전환되는 것에 더하여 이를 발전된 통신기술을 통하여 실시간으로 수집하여 빅데이터를 구축하고 있다. 이러한 데이터의 디지털화는 향후 급속히 발전할 것으로 보인다.

빅데이터가 기존의 데이터와 큰 차별성을 가지는 점은 바로 조합(fusion)을 통하여 또 다른 빅데이터를 구축하고, 더 다양한 정보를 알아낼 수 있다는 점이다. 단적인 예로 핸드폰에 있는 위치추적 기능과 건강정보 수집 기능을 조합하면 거주 지역별 건강상태 데이터를 생성할 수 있다. 또한, 위치 추적 기능과 결제 정보를 조합하면 거주 지역별 소비 패턴을 파악할 수 있다. 여기서 한발 더 나아가 위치 추적 기능, 건강정보 수집 기능, 결제 정보를 모두 조합한다면 거주 지역별 건강상태와 소비 패턴을 함께 비교해볼 수 있다.

이러한 방법으로 데이터 조합을 늘려간다면 굉장히 다양한 경우의 수를 만들어낼 수 있다. 위치 추적 기능을 사용했으므로 저녁과 아침에 집에서의 건강상태와 낮에 다른 곳으로 이동한 후의 건강상태를 비교할 수 있으며, 지역에 따른 개인의 소비 패턴 변화 양상도 파악할 수 있다. 이와 같이 데이터 조합은 개별 빅데이터가

의도하지 않았던 정보 생산을 가능하게 함으로써 기업과 소비자들에게 더 큰 편익을 제공할 수 있는 한편, 오용될 경우 개인정보 침해로 이어질 수 있다.

이상 살펴본 빅데이터의 특성과 생성 경로를 바탕으로 빅데이터를 간략하게 정의해보자면, 빅데이터는 반복적으로 축적된 자료를 뜻한다. 물론, 빅데이터 역시 통용되는 정의가 존재하는 것은 아니다. 앞서 언급한 백악관의 빅데이터 보고서에서도 빅데이터에 대한 정의가 다수 존재한다고 설명하고 있다. 빅데이터에 대한 정의는 컴퓨터 과학자, 금융 분석가, 기업가 등 분야별로 조금씩 다르기 때문이다. 빅데이터에 대한 여러 정의 중 하나로 미국 국립과학재단(National Science Foundation)의 정의를 빌려 말하자면, 빅데이터는 "디지털 소스(도구, 센서, 인터넷, 이메일, 비디오 등)를 통해 생성되는 크고, 다양하고, 복잡하고, 시계열적(longitudinal)인" 특성을 가진 자료라고 설명할 수 있다.

<데이터 편향>

앞서 검토한 바와 같이 컴퓨터가 머신러닝을 통해 더욱 정확한 결과를 도출하기 위해서는 많은 양의 데이터가 필요하다. 인공지능의 운영목표는 인간을 모방하고 인간과 비슷한 의사결정을 내리는 것이기 때문에 가상의 데이터가 아닌 실제로 인간이 실생활을 통해 축적한 데이터를 학습에 활용하는 것이 효과적이다.

하지만 인간에게 내재되어 있는 편향으로 인해 실제 데이터는 항상 편향을 내포하고 있다. 훈련용 데이터를 별도 구축하여 인공지능의 훈련에 사용한다 해도 이는 실제 데이터에 기반한 것으로 이미 편향된 구조 안에 머무는 한계가 있다. 즉, 머신러닝에 사용되는 데이터는 모두 직간접적으로 편향되어 있다.

이러한 편향이 잘못되었다는 것은 아니다. 인간과 사회가 각자의 특성을 가지는 것은 너무나 당연한 일이다. 다만 여기서 말하고자 하는 내용은 인공지능이 사용하는 데이터에는 인간의 특성이 이미 반영되어 있고 이것이 편향을 불러일으킨다는 것이다. 역사적으로 인간이 쌓아온 차별과 모순이 우리 사회에 자연스럽게 자리잡고 있으므로, 사회를 반영한 데이터 안에도 편향이 뿌리 깊게 내재되어 있다.

인간의 편향과 불평등은 비단 현대사회만의 문제가 아니라 역사적으로 이어져 내려온 현상이다. 하지만 사람들이 이것을 부정적으로 인식하는 이유는 편향과 불평등이 사회의 발전을 저해한다고 인식했기 때문이다. 이러한 문제 인식과 개선 노력은 사회를 발전시키는 원동력으로 작용한다. 이러한 맥락에서 인공지능이 편향된 의사결정을 하지 않도록 개발하려는 노력 역시 우리 사회가 한 단계 더 발전을 이룰 수 있는 방법이다.

EU기본권청(European Union Agency for Fundamental. Rights: FRA)은 이러한 데이터 품질 문제에 대해 여러 보고서를 통해 문제를 제기했다. 편향되어 있거나 결점이 있는 데이터를 활용하여 인공지능을 학습시키는 경우 차별을 포함한 인권침해의 결과가 발생할 것을 우려했다. 특히, 인공지능이 고용과 복지자원 배분 등 사회경제적 권리와 관련된 분야에서 적극적으로 활용되고 있는 점에 주목하며 인공지능이 공정하고 효과적으로 작동해야 하는 점을 강조했다.[20]

2장에서 제시한 인공지능의 역기능을 상기해본다면 EU기본권청의 우려를 충분히 이해할 수 있다. 아마존이 테스트했던 채용 인공지능 사례는 편향된 데이터가 어떻게 차별을 재확산할 수 있

20) European Union Agency for Fundamental Rights, “Data quality and artificial intelligence,” FRA Focus, (2019), p.9. https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect

는지 보여주었다. 만약, 다수의 기업이 활용하는 인공지능에서 이러한 차별이 지속적이고 반복적으로 발생한다면 사회의 불평등이 이전보다 급속하게 확산될 수 있다. 그동안 우리 사회가 수많은 노력을 통해 조금씩 이룩해낸 발전이 잘못된 인공지능 하나로 인해 물거품이 될 수도 있는 것이다. 인공지능의 역기능을 제대로 관리하지 못한다면 인간을 돕기 위해 만들어진 인공지능이 오히려 사회의 발전을 저해할 수 있다.

3.4. 데이터의 질적 및 양적 편향

데이터의 편향은 질적 측면과 양적 측면으로 구분하여 접근할 수 있다. 먼저, 질적 측면은 개인의 시각에 따른 편향이다. 사람은 자신의 시각으로 정보를 습득하고 해석한다. 이렇게 개인의 주관이 반영되어 해석된 정보들이 스마트기기를 통해 데이터로 축적된다. 예를 들어 어린 아기가 있는 사진을 보고 그 사진에 이름을 붙일 때 백인의 입장에서는 백인 아기를 '아기'로 이름을 붙이지만 흑인 아기는 그냥 '아기'가 아닌 '흑인 아기'로 분류한다.[21]

다음으로 양적 측면으로는 성, 연령, 지역, 인종 등 다양한 인구 분류에 있어서 실제 인구와 유사하게 데이터가 구성되어 있지 않아 발생하는 편향이다. 경제적 발전을 통해 디지털 기기를 다수 활용하는 지역의 사람들은 빅데이터에 자신의 필요(Needs)를 반영시킬 수 있는 반편, 비교적 발전되지 않은 지역의 사람들은 디지털 기기와 인터넷 연결 부족으로 자신의 정보를 빅데이터에 반영하기 어렵다.

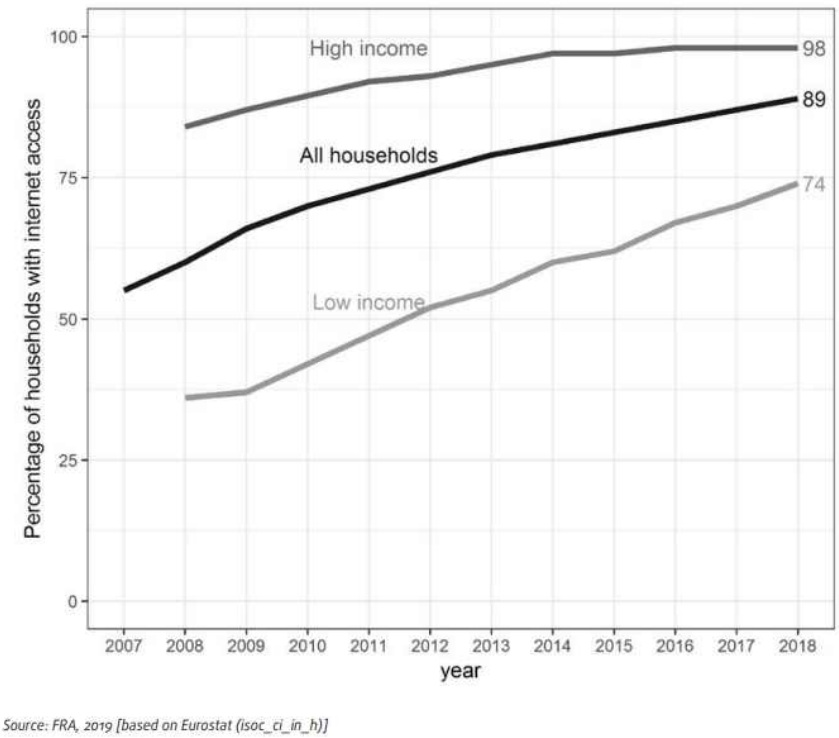
만약 정부에서 국토 균형 발전을 위하여 도로나 버스 노선

21) European Union Agency for Fundamental Rights, "BigData: Discrimination in data-supported decision making," FRA Focus, (2018) p. 5, https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making

신설을 위하여 수요를 조사할 때 핸드폰 위치기록 데이터를 사용한다고 가정하자. 핸드폰을 잘 들고 다니지 않는 노인이나 아직 핸드폰이 없는 어린이 그룹은 자신의 동선을 정부가 수집하는 데이터에 반영할 수 없으므로 사회자본의 혜택에서 멀어질 수밖에 없다.

아래 그래프는 EU에서 진행한 인터넷 사용가구의 소득을 분석한 자료이다. 2018년을 기준으로 EU 전체 가구의 89%가 인터넷에 접근할 수 있었지만 이를 소득별로 구분해보면 저소득 가구의 경우 74%만 인터넷에 접속이 가능했다. 한편 소득에 따른 인터넷 접근 차이는 점점 줄어들고 있었다. 2008년에는 고소득 가구의 약 80%가 인터넷을 사용했지만 저소득 가구의 경우 약 40%에 머물고 있었다.

<그림 3-5> 유럽의 가구소득별 인터넷 접근성



출처: European Union Agency for Fundamental Rights, “Data quality and artificial intelligence,” FRA Focus, (2019), p.7.

미국의 경우 의료 연구 자료 수집 시 정부가 아메리카 원주민과 알래스카 원주민의 자료를 충분히 수집하지 못해 고작 몇 명의 표본이 인구집단 전체를 반영하는 사례도 있었다.[22] 이렇게 데이터가 부족하거나 왜곡될 경우 정부의 보건의료 정책 수립 시 이들이 필요한 의료 혜택을 받지 못하는 부작용이 발생할 수 있다. 데이터의 양적 편향 문제를 해결하기 위하여 버클리 데이터과학연구소(Berkeley Institute for Data Science)의 Steph Eaneff 교수는 기존 데이터 수집이 부족한 인구집단에 대해 추가로 데이터를 수집하여 전체 빅데이터에 반영하는 방법을 대안으로 제시했다.[23]

또한 국제적인 시각으로 이 문제에 접근해본다면, 저개발국가에 거주하는 주민들은 디지털 기기의 부족에 따른 빅데이터 부족으로 다른 국가에서는 사용할 수 있는 서비스를 누리지 못하는 결과를 가져올 수 있다. 선진국에서는 데이터에 기반하여 의료나 교육 등 다양한 서비스를 개발하여 활용하고, 서비스 수출입을 통하여 다른 나라에서도 이러한 서비스를 활용할 수 있다. 하지만 데이터를 보유하고 있지 않는 국가의 경우 빅데이터에 기반한 다양한 서비스를 활용할 수 없고, 이는 국가 간 발전 격차를 심화시키는 요인이 된다.

지금까지 검토한 바와 같이 데이터는 질적 및 양적 측면에서 구조적인 편향을 가지고 있다. 따라서 인공지능이 더욱 공정한 결과를 도출하기 위해서는 알고리즘이 편향을 줄이는 역할을 해야 한다. 편향되어 있는 인간의 데이터를 입력하더라도 공정한 결과를 도출해야 하는 것이다. 인권기반의 인공지능 개발은 바로 이러한 점을 고려하여 이루어져야 한다. 인공지능 개발 시 데이터에 내재한 편향을 고려하고 어떻게 하면 이러한 사회의 불평등을 개선할 수 있을지 고민하는 것이 필요하다.

22) https://newsroom.ucla.edu/releases/data-american-indian-alaska-native-health

23) Eaneff, Stephanie, Obermeyer, Ziad and Butte, Atul J, “The Case for Algorithmic Stewardship for Artificial Intelligence and Machine Learning Technologies,” JAMA, 2020;324(14): p. 1397. doi:10.1001/jama.2020.9371.

또한, 인공지능 개발 후 테스트 과정에서 편향된 사회의 모습이 데이터와 인공지능에 어떻게 반영되어 있는지 찾아내는 과정 역시 필요하다. 데이터에 내재된 편향을 완전하게 제거할 수는 없으므로 기 편향된 데이터가 어느 부분에서 문제를 발생시킬지에 대한 위험 탐지 활동이 필요하다. 이러한 노력이 반복된다면 인공지능의 역기능을 예방할 수 있을 뿐만 아니라 인공지능이 데이터 편향을 극복하고 사회의 발전에 기여할 수 있다.

4. 미국의 인공지능 규제 동향

4.1. 미국 알고리즘책임법안

미국의 대표적인 인공지능 규제 동향을 살펴볼 수 있는 것은 2022년에 발의된 알고리즘책임법안(The Algorithmic Accountability Act of 2022)이다. 주요 내용을 살펴보면 미국 연방거래위원회(Federal Trade Commission: FTC)에 알고리즘을 모니터링하는 담당부서 신설, 일정 규모 이상의 인공지능을 운영하는 회사의 인공지능 영향평가 의무화, 연방거래위원회 인공지능 담당부서의 동향 보고서 작성 및 공개 등이다.

본 법안은 미국 연방정부 차원에서 인공지능을 어떻게 관리할 것인지에 대한 기본적인 접근방법을 보여주고 있어 향후 인공지능과 관련된 규제 논의에 있어 중요한 역할을 할 것으로 보인다. 본 법안에서는 먼저, 미국 연방거래위원회 내에 알고리즘 담당 부서를 설치하여 연방정부가 인공지능과 관련한 전문성을 가지도록 정하고 있다. 이후, 해당 부서가 인공지능 운영 기업들로부터 알고리즘 관련 자료를 제출받아 위원회가 알고리즘과 관련된 내용을 이해하고 관련 사안을 결정할 수 있도록 지원하도록 하였다. 또한 인공지능 운영회사가 알고리즘을 활용한 제품의 이용 및 판매 시 이것이 이용자에게 어떠한 영향을 미칠 것인지에 대해 자체 평가하고 이를 연방정부에 제출하도록 규정했다.

법안의 주요 내용을 조금 더 자세히 살펴보면 다음과 같다. 인공지능 운영회사 즉, 알고리즘이 자동적으로 중요한 의사결정을 내리는 시스템을 사용하는 회사는 알고리즘의 의사결정에 의한 영향을 사전에 평가해야 한다. 이 영향 평가는 법 이후에 개발되어 활용되는 알고리즘뿐만 아니라 이전에 활용되고 있는 알고리즘까지 모두 포함한다.

물론, 이미 개발되어 있고 앞으로도 개발할 모든 알고리즘을 평가하는 것은 아니다. 해당 법안에서는 비교적 큰 영향을 미치는 알고리즘에 대해 법을 적용하는데, 그 범위는 연간 매출이 5천만 달러 이상이거나 자본 가치가 2억 5천만 달러 이상인 회사가 백만 이 넘는 사람(또는 가구)이나 기계의 데이터를 수집하는 경우이다. 우리가 일상생활에서 활용하고 있는 알고리즘의 경우 대기업의 제품이 대부분이므로 해당 법에서 규정하고 있는 평가 대상에 속한다고 볼 수 있다.

각 회사에서 알고리즘의 영향을 자체적으로 평가할 때 측정 내용이 불명확하거나 회사마다 기준이 다르다면 평가의 실효성이 떨어질 수 있다. 따라서 이 법은 알고리즘에 대한 회사 자체의 평가가 표준화될 수 있도록 미국 연방거래위원회(FTC)가 이러한 평가의 기준을 안내하는 가이드라인을 마련하도록 정하고 있다. 해당 가이드라인은 영향평가 진행 및 보고 방법에 대한 구체적인 기준을 제시한다.

이 가이드라인은 본 법안에서 중요한 위치를 차지하고 있는데 그 이유는 영향평가가 향후 알고리즘 규제에 있어 시작점과 같은 역할을 하기 때문이다. 먼저 알고리즘을 활용하는 회사가 영향평가 결과에 대한 책임을 지도록 명시하고 있다. 만약 영향평가 시 누락되거나 오류가 있는 경우 해당 회사는 법에서 요구하는 사항을 충분히 따르지 않았으므로 법을 위반하게 되는 것이다.

만약 외부 회사에서 알고리즘을 개발하여 해당 알고리즘을 활용하는 회사에 납품하는 경우 외부 개발사도 이러한 영향평가 책임에서 벗어날 수 없다. 영향평가는 알고리즘에 문제가 있을 경우 책임소재를 파악하기 위한 첫 단계이다. 만약 운영사와 개발사가 서로에게 책임을 떠넘기는 경우 책임소재가 불명확하여 다음 단계로 넘어갈 수 없다. 따라서 이를 예방하기 위하여 알고리즘을 개발하여 납품한 업체 역시 영향평가에 따른 책임을 부여하는 것이다.

이렇게 작성된 영향평가 자료는 연방거래위원회에 제출되며 연방거래위원회는 아카이브를 구축하여 이를 보관한다. 연방거래위원회는 향후 알고리즘에 문제가 발생하면 알고리즘 운영사에서 절차에 따라 영향평가를 진행하고 문제 예방을 위한 노력을 기울였는지 확인할 수 있다.

이러한 접근 방식은 자율통제를 통하여 연방정부의 업무 부담을 줄이는 한편, 기업의 자발적인 올바른 인공지능 사용문화 확산을 의도한 것으로 보인다. 민간의 인공지능 개발인력 규모와 기술발전 속도를 정부에서 따라잡기란 거의 불가능하다. 따라서 월등히 앞서가고 있는 민간을 규제하기 위해서는 이를 효율적으로 관리하기 위한 권한과 자원이 필요하다.

이 법안에서는 연방거래위원회에 알고리즘 규제를 위하여 국(Bureau) 단위의 조직을 설치하고 50명의 인력을 배치할 것을 정하고 있다. 물론 제한된 예산으로 인하여 처음부터 대규모의 인적, 물적 자원을 배정받기는 힘들 것으로 보인다. 하지만 감시해야 할 기업들의 규모를 고려한다면 이 정도 규모의 조직으로는 실효성이 부족할 수도 있다.

우리가 실생활에서 쉽게 활용할 수 있는 알고리즘을 운영하는 기업인 구글과 애플의 경우 직원 수가 각각 약 13만 5천명, 15만 4천명으로 알려져 있다(2021년 기준). 물론, 연방정부에서는 법으로부터 부여받은 감독권한을 가지고 업무를 수행하므로 직접적으로 인원을 비교하는 것에는 무리가 있지만, 보다 실질적인 감독을 위해서는 법안이 정하고 있는 규모로는 한계를 가질 것으로 보인다.

알고리즘책임법안은 각 기업의 알고리즘에 대한 관리와 함께 전체 알고리즘 관련 산업의 동향을 연방거래위원회가 파악하고

이를 대중에 알리도록 정하고 있다. 연방거래위원회가 각 기업으로부터 제출받는 알고리즘 영향평가 보고서는 인공지능 관련 대기업들의 최신 알고리즘을 다루게 될 것이기 때문에 이 동향 보고서는 인공지능의 발전동향 파악, 잠재적 위험 가능성 예측 등에 있어서 중요한 자료가 될 것으로 보인다.

또한 위원회는 각 기업으로부터 제출받은 알고리즘 영향평가 자료의 내용을 대중에게 직간접 적으로 공개한다. 직접적인 공개는 연방거래위원회의 영향평가 자료 보관소를 통해 이루어진다. 해당 법안은 연방거래위원회가 별도의 영향평가 자료보관소를 설치하여 소비자와 관련 옹호자들이 이를 검토할 수 있도록 정하고 있다. 이를 통해 소비자들은 자신이 이용하는 서비스에서 어떠한 부분이 자동화되었는지, 또한 어떠한 데이터가 사용되었는지에 대해서도 알 수 있다.

간접적인 공개는 연방거래위원회의 알고리즘 동향분석보고서를 통하여 이루어진다. 법안은 연방거래위원회가 각 회사가 제출한 알고리즘 영향평가 자료를 바탕으로 전체적인 알고리즘의 동향에 대해 검토하고, 전체적인 기술의 움직임에 대해 분석하는 보고서를 작성하여 대중에 공개하도록 정하고 있다. 이 보고서는 각 회사의 이름을 밝히지는 않지만 알고리즘 이용자들이 알고리즘 기술이 어떻게 발전하고 있는지, 이것이 이용자들에게 어떠한 영향을 미칠 수 있는지에 대해 안내한다.

따라서 알고리즘에 대한 정보 공개는 이용자의 권리 증진에 큰 역할을 할 것으로 보인다. 알고리즘 동향 보고서를 통해 이용자는 최신 알고리즘 제품이 어떠한 방식으로 정보를 수집하여 의사결정을 진행하는지에 대해 알 수 있다. 따라서 이용자가 향후 자신의 권리를 보호하고 불이익을 받지 않을 수 있는 방법을 고민할 수 있다.

한편, 정부의 규제가 산업 발전과 기업 경쟁력에 미치는 영향 역시 고려해야 한다. 본 법안의 내용을 살펴본 결과, 주로 큰 규모의 알고리즘을 운영하는 회사를 대상으로 하고 있는데 그 이유는 알고리즘의 영향력이 크고 자체 통제 역량을 보유하고 있는 기업을 선도적으로 관리하는 한편, 규모가 작은 회사의 경우 신기술 개발에 있어 비효율성이 발생하지 않도록 배려한 것으로 보인다. 영향력이 작은 알고리즘까지 규제할 경우 영향평가에 소요되는 비용보다 평가에 따른 편익이 적을 것으로 보이기 때문이다.

하지만 여기서 주의해야 할 점은 규모가 작고 백만 개 이하의 데이터를 활용하는 기업이라도 실제 알고리즘 적용을 통한 영향력이 큰 경우가 있을 수 있으므로 이에 대한 고민이 필요할 것으로 보인다. 만약 의료, 금융, 언론과 같이 개인의 건강, 경제적 안정, 정치성향 형성에 직접적인 영향을 미치는 서비스의 경우 역기능에 의한 피해가 심각할 수 있다. 본 법안이 목표로 하고있는 것은 알고리즘이 사회에 미칠 수 있는 위험을 예방하는 것이므로, 이와 같이 규모가 작음에도 불구하고 영향력이 큰 사례도 함께 고려해야 할 것으로 보인다.

알고리즘책임법안은 오레곤주의 Ron Wyden 의원이 주축이 되어 미국 국회에 상정되었지만, 아직 국회 내에서 폭넓은 지지를 받고 있다고 보기에는 어렵다. 2019년 최초로 발의된 후 통과되지 못했으며, 2022년에 다시 발의했지만 역시 큰 진행상황을 보이고 있지는 않다. 하지만 이 법안은 그동안 미국의 주의회나 시의회 차원에서 논의되었던 알고리즘 관련 법안 내용을 아우를 수 있는 통합적인 구상을 담고 있다. 따라서 우리나라도 이와 관련된 논의가 진행되고 있는 만큼, 본 법안과 관련한 논의 동향을 지속적으로 살펴볼 필요가 있다.

4.2. 뉴욕시의 인공지능 규제 사례

뉴욕시는 학생들의 학교배정, 교사의 업무평가, 의료사기 대응, 건축물 (안전) 평가, 공공임대, 푸드 스템프 등 다양한 시정 운영에 인공지능을 활용했다. 뉴욕시가 주요한 시정에서 인공지능을 활용하자 시 의회에서는 이에 대한 규제방안을 마련했다. 인공지능이 뉴욕 시민에게 있어 중요한 의사결정을 내린다면 이러한 인공지능이 보다 투명해야 한다고 여겼기 때문이다.

2017. 8. 24. James Vacca 뉴욕시의원(지역구: East Bronx)은 알고리즘을 비롯한 모든 '자동화된 의사결정 시스템'의 소스 코드(source code) 공개를 주요 내용으로 하는 법안(Int. 1696)을 뉴욕시의회(New York City Council)에 발의했다. 여기서 '자동화된 의사결정 시스템(automated decision systems: ADS)'은 머신러닝을 통하여 생성된 알고리즘과 여러 형태의 데이터 분석 형식을 포함하는 넓은 개념으로 뉴욕시가 의사결정 시 직접 또는 간접적으로 활용하는 알고리즘을 총칭한다.[24]

Vacca 의원이 이 법안을 발의하게 된 배경은 자신의 지역구인 브롱스(Bronx)에 경찰력이 예상보다 적게 배치되는 것을 이상하게 여기면서 본격적으로 시작되었다.[25] 이 문제에 대해 관심을 가진 Vacca의원은 뉴욕 경찰에 경찰력 배치 기준에 대한 설명을 요구했다. 뉴욕 경찰은 Vacca 의원에게 경찰력 배치는 과거 범죄발생 자료를 인공지능에게 학습시키고 그 결과에 따라 범죄가 발생할 가능성이 높은 지역에 우선적으로 경찰력을 배치한다고 설명했다.

---

24) Rashida Richardson, ed., "Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force," AI Now Institute, December 4, 2019, https://ainowinstitute.org/ads-shadowreport-2019.html.

25) Vacca 의원의 법안 발의 배경에는 다양한 요소가 있었다. Vacca 의원은 한 언론사와의 인터뷰에서 법안 발의에 프로퍼블리카의 'Machine Bias' 기사가 부분적으로 영향을 미쳤다고 언급한 바 있다. https://statescoop.com/in-first-meeting-nyc-algorithm-task-force-talks-racial-biases-dirty-data/

당시 뉴욕시의회 기술위원회 위원장이기도 했던 Vacca의원은 뉴욕 경찰의 이러한 답변이 충분하지 않다고 여겼다. 왜냐하면 가장 중요한 부분인 '어떠한 기준'에 대한 설명이 없었기 때문이다. 경찰은 인공지능이 사용한 데이터에 대한 정확한 정보(생성 시기, 항목 등), 해당 인공지능의 의사결정 기준 등 인공지능이 결론에 도달하기까지의 과정에 대한 답변을 하지 않았다. Vacca 의원은 비단 경찰력 배치 문제뿐만 아니라 인공지능을 활용한 뉴욕시의 업무결정 활동 전반에 대한 설명이 굉장히 부족하다고 느끼고 이를 개선할 법안을 준비했다.[26]

Vacca 의원실에서는 2017년 5월부터 법안을 준비하여 8월경에 초안을 완성했다. 그 주요 내용은 뉴욕시가 치안활동, 벌금부과 또는 서비스 제공 등을 결정하기 위하여 인공지능을 사용한다면 여기에 사용되는 인공지능의 소스 코드를 대중에 공개하도록 정한 것이었다. 또한 시정에 인공지능을 활용하는 경우 실제적으로 인공지능을 업무에 적용하기 전에 뉴욕시민의 데이터를 가지고 시뮬레이션을 진행하여 사전에 위험성을 판단해보도록 정했다.

이후 2017년 10월 뉴욕시의회 기술위원회 공청회에서 해당 법안이 논의되었는데, 이해관계자들은 제안된 법안에서 정하고 있는 규제를 완화할 것을 강력하게 요구했다. 이해관계자들로부터 가장 집중적으로 공격을 받은 부분은 인공지능의 소스 코드를 대중에 공개하는 내용이었다. 소스 코드 공개 부분은 인공지능 관련 기업은 물론, 정책 전문가들까지도 반대했다. 기업의 입장에서는 영업상 비밀을 이유로 공개를 거부했고, 정책 전문가들은 인공지능의 결정과정을 대중에 공개하게 되면 사람들이 의사결정 시스템을 파악하여 이를 악용할 수 있다는 이유였다.

하지만 이러한 주장은 시민보다 기업의 이익에 더 중점을

26) Julia Powles, "New York City's Bold, Flawed Attempt to Make Algorithms Accountable," The New Yorker, December 20, 2017.

둔 주장이라고 볼 수 있다. 이미 인공지능을 활용한 의사결정으로 인해 시민들이 부당함을 느끼고 있었고 이로 인한 다양한 형태의 피해를 입고 있는 상황이었다. 뉴욕시에서 시민들의 이익을 최우선으로 고려했다면, 시에서 할 수 있는 모든 방법을 동원하여 알고리즘의 부작용을 예방하고자 했을 것이다. 그런데 뉴욕시는 기업의 지적재산권 보호를 이보다 우선시하고 소스 코드 공개를 거부했다.

법안 초안의 소스 코드 공개 부분이 강력한 저항에 부딪히자 법안을 옹호하는 측에서는 소스 코드 전체를 공개하는 대신 주요 결정사항과 관련된 일부 내용만 공개하는 것으로 대안을 제시했으나 뉴욕시 측에서는 이마저도 받아들이지 않았다. 회사가 독점하고 있는 정보, 즉 고유 기술을 시에서 공개할 수 없다는 이유에서였다.

뉴욕시는 인공지능 제공 회사와 서비스 사용계약 체결 시 회사의 정보를 외부에 공개하지 않겠다는 조건을 명시했으므로 이를 이행하기 위해 법안의 내용이 변경되어야 한다고 주장했다. 또한 뉴욕시는 소스 코드를 공개한다면 뉴욕시와 계약을 체결할 인공지능 서비스 회사는 없어지게 되고, 결과적으로 이러한 서비스를 활용할 수 없을 것을 우려했다.

인공지능을 활용한 행정 서비스 제공의 목적이 시민의 이익 증진이었다면 뉴욕시가 소스 코드 공개에 적극적인 입장이었을 것으로 추정한다. 하지만 뉴욕시는 현업에서 큰 역할을 하고 있는 인공지능을 지속적으로 활용할 수 있는 방법을 찾고 있었고 이로 인해 시민의 이익 보다 기업의 이익에 더 중점을 둔 입장을 보였다.

이러한 뉴욕시의 입장 설명 내용 중 주목해야 할 부분은 뉴욕시가 이미 인공지능 도구 없이는 현업을 유지하기 어려운 단계에 이르렀다는 것이다. 뉴욕시 대변인은 법안 통과 시 뉴욕시가 인공지능 회사와의 계약을 위반하게 된다는 내용 등을 언급하면서

마지막 부분에 뉴욕시가 기술을 활용한 혁신적인 솔루션을 통해 현업에서 많은 업무를 해결하고 있고, (소스 코드 공개 시) 이를 활용하지 못하게 된다고 언급했다.[27)]

즉, 뉴욕시는 인공지능 서비스를 활용할 수 없을 경우 현업에 지장이 발생할 것을 우려한 것으로 보인다. 인공지능 서비스 회사와의 계약과 앞으로의 업무 관계를 유지를 위해 시민들의 알 권리 보호, 행정서비스의 투명성 제고 등 정부가 지켜야 할 중요한 가치를 후순위에 두는 듯한 모습을 보였다. 인공지능 기술을 활용한 행정서비스가 2010년대에 등장한 것을 고려하면, 인공지능이 불과 몇 년 만에 대체 불가능한 정도의 편익을 제공하고 있는 것으로 보인다.

결론적으로 법안의 초안 내용 중 소스 코드를 공개하는 부분은 제외되었다. 대신 사실조사 태스크포스(fact-finding task force)를 구성하여 인공지능을 감시하는 역할을 담당하도록 정했다. 수정된 법안은 태스크포스가 뉴욕시에서 활용하는 인공지능에 대해 다음과 같은 내용을 조사하고 권고할 것을 정하고 있었다.

이러한 과정을 거치며 Vacca 의원의 법안은 당초에 의도했던 목표를 달성하기 힘든 수준까지 변경되었다. 법안 초안은 시민들이 인공지능의 의사결정에 대한 의미 있는 설명을 들을 수 있는 장치를 마련하였으나 변경안과 같이 소스 코드 공개가 이루어지지 않는 이상 태스크 포스가 실질적인 역할을 하기는 어렵기 때문이다. 태스크 포스의 핵심 활동은 인공지능의 의사결정에서 어떠한 부분이 잘못되었는지 파악하고 이에 대한 수정을 권고해야 하지만 소스 코드에 접근할 수 없다면 테스크 포스의 문제 진단과 해결방안 제안이 추상적인 단계에 머물 수밖에 없다.

---

27) New York City Council Legislative Research Center, “Legislation,” https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0. 2017년 10월 16일 뉴욕시의회 기술위원회 공청회에서의 뉴욕시 정보기술통신국 직원의 발언내용.

「Int. 1696」 수정안

사실조사 태스크 포스(fact-finding task force) 역할

- 뉴욕시에서 사용하는 인공지능 중 어떠한 인공지능을 규제할지에 대한 권고
- 일반 시민들이 인공지능에 대해 의미 있는 정보를 얻을 수 있는 방안에 대한 권고
- 시민들이 자신에게 영향을 미치는 결정이 어떻게 이루어졌는지에 대해 설명을 들을 수 있는 방법에 대한 권고
- 시 정부가 사용한 인공지능에 의해 피해가 발생할 경우 이에 대해 어떻게 처리할 것인지에 대한 권고

Vacca 의원은 뉴욕시의회 기술위원회 회의에서 “만약 우리가 앞으로 기계와 알고리즘, 데이터에 의해 통치된다면, 그것들은 좀 더 투명해야 한다.” 고 언급했다. 수년이 지난 지금 Vacca 의원의 주장이 시사하는 바는 더욱 크다. 그동안 인공지능에 의한 사회 양극화 심화, 사회적 약자에 대한 불이익 심화 등 인공지능의 역기능에 따른 문제들이 지속적으로 발생하고 있기 때문이다.

2017년 12월 11일, 소스 코드 공개를 제외시키고 테스크 포스를 추가한 수정된 법안이 의회를 통과했다. 법안을 발의한 9명의 시의원 모두 민주당에 속해 있었으므로 이를 다른 당에서 견제하더라도 이상한 상황이 아니었으나 해당 법안 초당적인 지지를 얻어 반대표 없이 통과되었다. 하지만 이러한 초당적 지지와 반대로, 통과된 법은 하지만 뉴욕 시장의 서명을 받지는 못했다. 뉴욕시는 소스 코드 공개를 제외한 수정안 이후에도 해당 법에 대해 지속적으로 부정적인 입장을 가지고 있었기 때문이다. 결국 2018년 1월 17일 뉴욕 시장의 서명 없이 해당 법안은 법(Local Law 49 of 2018)으로서의 효력을 발휘하였다.

Local Law 49(LL49)는 뉴욕시의 지지를 얻지 못한 상태로 시행된 만큼 이후 진행과정에서 많은 난관을 겪었다. 법에서 정한 자동화된 결정 시스템 테스크 포스(Automated Decision Systems (ADS) Task Force)를 발족시켜야 했지만, 뉴욕시는 이에 미온적인 태도를 견지했다. 이를 예상했던 Vacca의원과 시민사회는 적극적으로 테스크 포스 설치를 위해 노력했다. 법이 의회에서 통과된 후 4일 만에 Vacca의원은 뉴욕 시장에게 테스크 포스에 포함될 인사들을 추천하는 편지를 발송했다. 또한 법 효력 발생 이후 5일 만에 관련 시민사회가 연대하여 테스크 포스 설치를 권고하는 편지를 발송하기도 했다.

수개월이 지난 2018년 5월, 뉴욕 시장은 ADS 테스크 포스의 의장과 함께 테스크 포스에 포함될 인사들을 발표했다. 이후 테스크 포스가 한동안 뚜렷한 활동을 보이지는 않았던 것으로 보인다. 6개월 후인 2018년 11월에 테스크 포스를 소개하는 웹페이지가 뉴욕시 홈페이지 안에 개설되었으며, 이듬해 2월이 되어서야 테스크 포스 맴버 전체의 약력을 담은 소개 웹페이지가 개설되었다. 이러한 경과를 지켜본 시민사회는 ADS 테스크 포스에 더딘 진행상황을 염려하는 서한을 발송했다.

2019년 3월 26일, 뉴욕시 감사원이 테스크 포스 의장에게 뉴욕시가 사용한 8가지 유형의 자동결정시스템에 대한 정보를 요청하는 서한을 보냈다. 테스크 포스는 감사원의 요청에 대해 이전과는 다른 신속한 반응을 보였다. 테스크 포스는 감사원의 요청을 받은 바로 다음 날 2번의 공청회 개최 계획과 함께 그해 여름에 걸쳐 지역별로 미팅을 개최할 것을 발표했다.

위 사례는 인공지능 관련 규제가 이해관계자의 지지를 얻지 못하는 경우 그 효용성이 현저하게 낮아지는 것을 보여주고 있다. 물론 이것은 비단 인공지능과 관련된 규제에만 해당하는 내용은 아니다. 다른 분야 역시 이해관계자가 참여하지 않아 유명무실한

규제로 머무르는 경우가 있다. 하지만 인공지능 기술이 보편화되고 표준화된 기술이 아닌 점, 정부가 인공지능을 통제할 전문성을 보유하지 못한 점을 고려해볼 때 국회, 시민사회, 정부가 인공지능 기업에 미칠 수 있는 영향이 제한적임을 염두해야 할 필요가 있다.

비록 Vacca 의원이 의도했던 모습은 아니었지만, 뉴욕시는 알고리즘을 규제하는 법안을 시행하였으며 이러한 사례는 우리에게 여러 참고점을 제시하고 있다. 특히 Vacca 의원의 초안인 정부에서 활용하는 인공지능의 소스 코드(source code) 공개, 인공지능 활용 전 실제 데이터를 활용한 시뮬레이션 수행, 기업의 지적재산권 보호 등의 내용은 향후 인공지능에 대한 규제 검토 시 눈여겨 보아야 할 부분이다.

4.3. 미국 FDA의 Pre-Cert Program

의료분야는 인공지능이 광범위하게 활용되고 있는 대표적인 분야 중 하나이다. 미국 식품의약국(US Food and Drug Administration: FDA)은 향후 의료분야에서 사용되는 인공지능 규제 모델을 개발하기 위해 Digital Health Software Precertification (Pre-Cert) Program을 마련하고 시범적으로 인공지능 소프트웨어를 규제하고 있다.

FDA에서 공식적으로 밝히고 있는 이 프로그램의 이름은 소프트웨어 기반의 의료장비 (software-based medical devices)의 사전인증 프로그램이다. 여기에 인공지능이라는 단어가 들어가지는 않지만 FDA의 프로그램 운영 보고서를 확인해보면 임상의사결정(clinical decision making)을 지원하고 환자의 치료를 관리하는 등 질병의 진단 및 치료활동에서 소프트웨어의 사용이 증가했고 이러한 새로운 변화에 대응하기 위하여 본 프로그램을 도입했다고 밝히고 있다.

Pre-Cert 프로그램은 2017년부터 논의가 시작되었는데, 당시 디지털 의료 혁신이 이루어지고 있는 상황에서 환자의 안전을 보호하기 위한 방법을 고민하면서 프로그램의 비전을 수립하기 시작했다. 2018년에는 이해관계자의 의견을 모으고, 파일럿 프로그램 운영을 위하여 9개의 소프트웨어 회사를 모집했다. FDA는 2019년부터 해당 소프트웨어 회사와의 협력 하에 Pre-Cert 프로그램의 1단계 버전을 실험하기 시작했다. 이후 FDA는 파일럿 프로그램의 결과를 바탕으로 이와 같은 프로그램 운영을 확대할 예정이다.

<그림 5-1> Pre-Cert 프로그램 로드맵

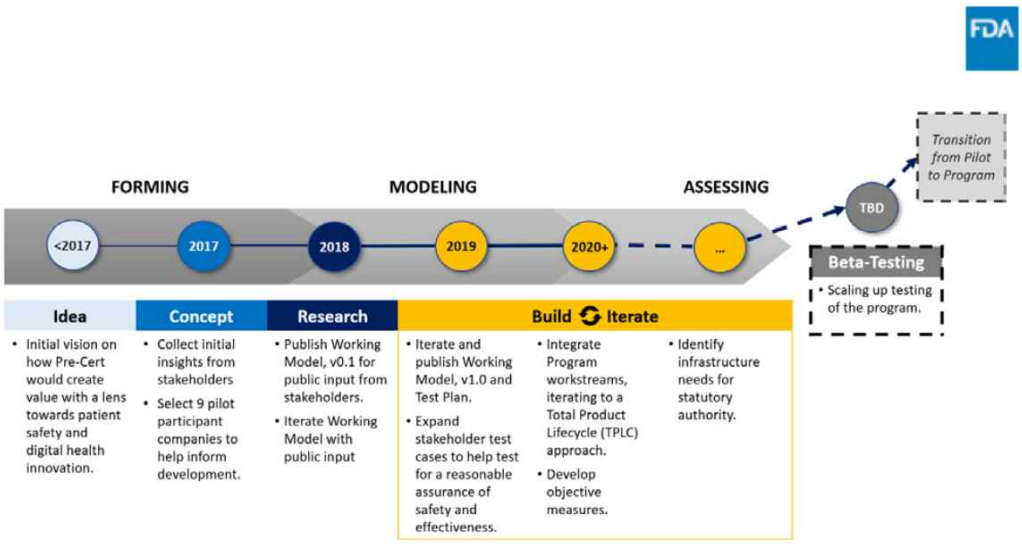


Figure 1. Software Pre-Cert Pilot Program Development Roadmap

출처: FDA, Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities, September 2020, p.2.

Pre-Cert 프로그램은 이미 개발되어 출시된 소프트웨어를 점검하는 것이 아니라 개발 단계에 있는 소프트웨어를 테스트해보고 문제가 있다면 개발 단계에서 수정하는 것이 특징이다. FDA는 전통적으로 의약품 사용에 따른 사고 발생 시 돌이킬 수 없는 피

해를 초래할 수 있으므로 위험을 예방하는 것을 우선적 업무 목표로 정하고 있다.

FDA의 Pre-Cert 프로그램에 대한 접근 역시 이러한 기존의 접근 방식을 사용한 것으로 보인다. 기존 의약품 인허가 방식은 새로운 약의 위험성을 진단해보고 일반 국민들이 사용해도 안전하다고 판단되면 시장에 출시할 수 있도록 허가를 내주는 형식인데, 여기서 의약품을 소프트웨어로 바꾼 것이 Pre-Cert 프로그램이라고 볼 수 있다.

이러한 사전 예방적 접근방법은 기업의 개발비용을 절감할 수 있는 장점이 있는 한편, 창의성을 저해하는 단점도 가지고 있다. 먼저 개발비용의 측면에서, 소프트웨어에 결함이 있는 경우 시장에 출시되어 피해가 생기기 전에 진단 및 조치를 취하는 것은 기업은 물론 정부 입장에서도 비용을 줄일 수 있는 효율적인 방법이다. 소프트웨어 개발사 입장에서는 제품 개발 시 수정으로 인한 비용은 일부 상승할 수 있지만 제품 출시 이후 정부 규제에 따른 추가비용 발생 위험을 감소시킬 수 있다. 따라서 기업은 위험 비용을 절감하는 효과를 누릴 수 있고, 소비자 입장에서는 잘못된 소프트웨어의 사용에 의한 피해 비용을 예방할 수 있다.

한편, 소프트웨어 개발 단계에서 정부가 관여하는 경우 이것이 기업의 창의성과 혁신성을 저해할 가능성이 있다. 소프트웨어 개발 시 기업이 정부의 간섭을 염두하여 새로운 시도를 꺼릴 수도 있으며, 기존 정부의 관여 사례가 향후 소프트웨어 개발 영역을 제한할 수 있다. 더군다나 정부와 기업의 전문성을 비교해볼 때 정부가 소프트웨어 개발에 있어 항상 최적의 결정을 내린다고 확신하기도 어렵다. 개발 단계의 정부 간섭은 기업의 신규 개발 시도를 소극적으로 변화시킬 위험이 있다. 개발 단계에서 정부의 관여로 인해 시장에 출시하지 못할 바에야 비용을 들여가며 개발시도를 할 이유가 없기 때문이다.

컴퓨터 소프트웨어는 기존에 개발되어 있는 소프트웨어를 바탕으로 새로운 소프트웨어를 개발한다. 따라서 최근에는 그동안 축적된 소프트웨어를 기반으로 신규 소프트웨어가 양적, 질적으로 급격히 늘어나고 있다. 다시 말하면, 기존 소프트웨어의 부분(module)을 조합하고 응용하여 새로운 소프트웨어를 만들기 때문에 향후 소프트웨어 개발속도는 더욱 빨라질 것으로 보인다. 이러한 상황에서 정부가 민간에서 개발되는 소프트웨어를 선제적으로 모니터링하고 관리하는 것은 큰 비용을 수반할 것으로 보인다. 또한, 제품 개발 단계에서 정부의 규제가 이루어지는 경우, 개발속도가 저하되고 경쟁력이 약화되는 결과를 발생시킬 가능성이 있다.

4.4. 캘리포니아주의 안면인식 및 기타 생체 감시 규제 법

앞서 살펴본 미국의 인공지능 관련 규제 동향은 법안 발의 단계이거나 국회에서 통과된 법이 이해관계자의 동의 부족으로 실효성이 부족하게 운영되는 등 인공지능을 관리하는 데에 있어 큰 역할을 하지는 못했다. 하지만 미국 캘리포니아주의 경우 경찰이 수집한 안면정보를 인공지능과 연결시키지 못하도록 금지하는 법안을 통과시켰고, 이에 따라 캘리포니아주 경찰은 2019년부터 2023년까지 안면인식 인공지능 활용에 있어 일부 제한사항을 적용받고 있다.

캘리포니아주의회는 경찰이 수많은 카메라를 활용하여 무차별적으로 주민의 얼굴 정보를 수집하고, 안면인식 기술을 사용해 주민을 감시하는 것에 대해 문제점을 제기하며 이를 규제하는 법안을 통과시켰다. 캘리포니아는 이 법을 통해 정부의 안면인식 및 기타 생체감시 기술사용이 헌법에 따른 프라이버시 보장 권리를 침해한 것을 재확인하였으며 이를 명백하게 금지하고 있다(제1조(a)).

캘리포니아주의회는 경찰의 제복에 부착되어 있는 카메라를 비롯한 다양한 영상정보 수집 장치가 주민들의 동의 없이 주민의 개인정보인 안면 정보를 수집하므로 이것이 무차별적 신원검사와 다름없다고 보았다. 의회는 이러한 경찰의 감시가 주민들의 발언권을 비롯한 자유권을 침해하는 것에 해당한다고 보았다(제1조 (b), (c)).

「의회 법안 제1215호(AB 1215)」

제 1 조.
이 법률은 이하 전문을 확인 및 선언한다:
(a) 캘리아포니아인들은 프라이버시를 개인의 자유를 구성하는 핵심 요소로서 가치 있게 여기며 캘리포니아 헌법 제I장 제1조에 따른 프라이버시 권리를 보장받는다.
(b) 안면 인식 및 기타 생체 감시 기술은 거주자 및 방문객들의 시민권과 시민적 자유에 독특한 방식으로 상당한 위협을 가한다.
(c) 안면 인식 및 기타 생체 감시 기술의 사용은 명시된 헌법적 권리를 위반하여 모든 이들에게 수시로 개인 사진이 부착된 신원확인카드를 요구하는 것에 상응하는 기능을 한다. 또한 이 기술은 동의 없이 사람들을 추적할 수 있다. 나아가 이 기술은 법을 준수하는 캘리포니아인들에 대한 대량의 데이터베이스를 생성할 수 있으며 공공장소에서의 자유발언권 행사를 어렵게 만들 수 있다.

이 법에서는 안면인식 인공지능 활용의 또 다른 문제점으로 인공지능의 오류 가능성을 지적했다. 안면인식 및 기타 생체감시기술이 여성, 청년 및 유색인종을 오인하는 비율이 더 높다고 지적하고 있다(제1조 (d)). 인종에 따라 안면인식의 오류비율이 달라지는 문제는 그동안 다른 사례에서도 반복적으로 확인된 바 있다. 사회

적 소수자에 대한 안면인식 기술의 높은 오류율에 대한 정확한 원인을 법에서 밝히고 있지는 않지만, 안면인식 인공지능의 학습데이터 부족, 피부색으로 인한 인공지능의 이미지 판별능력 저하 등이 원인일 것으로 전문가들은 추정하고 있다.

경찰 유니폼에 카메라를 부착하게 된 계기는 경찰의 과잉진압을 예방하고자 함이었다. 하지만 유니폼에 부착된 카메라가 상시 녹화상태로 운영되며 경찰과 마주치는 모든 주민들의 얼굴정보를 수집하게되었다. 결국, 캘리포니아 의회는 경찰의 카메라 장비 도입 및 운영이 본 취지인 공무집행의 투명성 제고와는 다른 결과를 낳는다고 여겼다.

「의회 법안 제1215호(AB 1215)」

제 1 조.

(d) 안면 인식 및 기타 생체 감시 기술은 여성, 청년 및 유색 인종을 오인하는 경우가 있으며 개인에게 치명적일 수 있는 “잘못된 일치(false positive)” 신원확인 결과의 위험을 높인다는 주장이 반복적으로 제기되고 있다.

(e) 안면 인식 및 기타 생체 감시는 집행관이 신체에 착용한 카메라를 투명성 및 책임성을 위한 도구가 아닌 무작위 감시 시스템으로 오용하여 해당 카메라 사용의 주요 목적을 퇴색시킬 수 있다.

인공지능을 활용한 치안 활동 시 많이 지적되는 문제가 잘못된 일치(false positive)와 지역 차별이다. 잘못된 일치(false positive)는 인공지능이 참이라는 결론을 내렸지만, 이것이 참이 아닌 거짓인 상황을 말하는 것이다. 인공지능은 데이터를 근거로 결론을 내렸지만 그러한 결론이 잘못 내려진 것이다. 앞서 스티브 탤

리 사례에서 확인된 바와 같이 얼굴이 비슷하게 생겼다는 이유로 범죄자로 잘못 지목될 수 있으며, 이것이 개인의 생활에 돌이킬 수 없을 정도의 큰 영향을 미칠 수 있다. 해당 법은 인공지능의 잘못된 일치에 따른 문제점 역시 지적하고 있다.

지역 차별과 관련해 이 법에서는 치안이 양호한 지역에 거주하는 사람과 그렇지 않은 지역에 거주하는 사람들이 시민적 자유를 불균형하게 누리는 문제를 지적했다(제1조 (f)).

「의회 법안 제1215호(AB 1215)」

제 1 조.

(f) 안면 인식 및 기타 생체 감시 기술의 사용은 치안이 매우 양호한 지역에 거주하는 사람들의 시민권 및 시민적 자유에 불균형적으로 영향을 미칠 수 있다. 또한 이 기술의 사용은 이러한 지역의 사람들(범죄 피해자, 밀입국자, 벌금미납자, 전과기록자 포함)이 경찰의 도움을 받거나 경찰을 지원하도록 할 수 있는 사기를 꺾어 효과적인 치안 유지 활동 및 공중의 안전을 저해할 수 있다.

이와 관련해 좀 더 구체적으로 설명하자면, 치안이 양호한 지역의 경우 카메라가 달린 제복을 입은 경찰이 소극적으로 안면 정보를 수집하는 한편, 치안이 좋지 않은 지역은 경찰이 유니폼에 부착된 카메라를 통해 더욱 적극적으로 안면 정보를 수집하므로 이러한 불균형을 문제점으로 제기한 것이다. 이러한 지역 차별 문제를 앞서 논의한 '잘못된 일치'와 함께 검토해본다면, 치안이 나쁜 지역의 주민의 안면정보는 더 많이 수집되고 경찰이 찾고자 하는 사람의 얼굴과 더 자주 비교되므로 '잘못된 일치'로 인한 피해 위험이 높아진다.

따라서 캘리포니아의 안면인식 및 기타 생체 감시 규제 법에서는 경찰이 유니폼에 부착되어 있는 카메라를 통해 영상을 수집할 시에 이 영상 정보를 경찰의 안면인식 인공지능과 연결시키지 않도록 금지하고 있다.

「의회 법안 제1215호(AB 1215)」

제 2 조.

(9) (B) (b) 사법기관 또는 법집행관은 법집행관의 카메라 또는 법집행관의 카메라로 수집한 데이터와 관련하여 어떠한 생체감시 시스템도 설치하거나, 작동시키거나, 사용하지 아니한다.

이 법은 인권침해를 유발할 위험이 있는 인공지능의 다양한 기술 중 안면인식 기술에만 제한되어 있고 인공지능의 차별보다는 침해와 관련된 성격이 짙다. 물론, 이 법은 일몰제 법안으로 일정 기간동안 효력을 발생하고 자동 폐지된다. 2023년 이후에는 추가 입법이 이루어지지 않는 이상 경찰이 다시 안면인식 기술을 사용할 수도 있다는 한계점을 가지고 있다.

하지만, 이 사례는 실제 입법을 통해 인공지능의 사용을 일부 제한했다는 점에서 본 연구에서 주요한 사례로 다루었다. 인공지능이 인권에 기반하여 개발되고 운영되지 않는다면 사람들은 자신의 권리를 침해한 인공지능을 막기 위해 문제를 제기하고 제도적 해결방안을 만들어나갈 것이다. 또한 이러한 해결 노력이 처음에는 부분적이고 단편적이지만 점점 축적되어 통합적인 방향으로 나아갈 것이다. 앞서 살펴본 미국 알고리즘책임법안은 기존에 미국에서 제기된 여러 가지 인공지능 관련 이슈를 통합적으로 대응하

기 위하여 발의되었다. 이러한 통합적인 법안은 산업계의 반발 등으로 쉽게 입법에 이르지 못하는 경우가 많다. 하지만 법안 내용에 담긴 문제의식과 문제 해결을 위한 제도 구상은 사회적 논의를 더욱 풍성하게 만들어주는 역할을 한다.

5. 인권기반 인공지능 개발 추진 방향

5.1. 인권기반 인공지능 개발 가이드라인

인권에 기반한 인공지능을 개발하기 위해서는 개발자를 포함한 조직 전체, 더 나아가 전 사회적 문화 조성이 필요하다. 근본적인 변화를 이루기 위해서는 사회 전체가 인공지능의 개발 및 활용에 있어 인권적 측면의 중요성을 숙지하고, 이를 우선적 가치로 여기는 문화를 가져야 한다. 일각에서는 마치 개발자가 인공지능 프로그램을 부족하게 만들어서 역기능이 나타나는 것처럼 인식하는 경향이 있다. 그러나 개발자는 조직의 지침에 따라, 조직의 목표를 달성하기 위하여 프로그램을 개발한 것이다. 게다가 조직의 목표는 소비자와 투자자의 수요에 기반한 것이다. 그러므로 사회 전체적인 변화가 수반되어야 인권에 기반한 인공지능 개발이 가능하다.

사회 전체적인 변화는 대국민 인식제고 사업을 통하여 이루어질 수 있으므로 본 연구에서는 사회의 변화보다 한 단계 낮은 조직의 변화를 목표로 가이드라인을 작성하였다. 따라서 이 가이드라인은 조직 전체 차원에서 인권에 기반한 인공지능 개발을 위하여 점검할 항목을 제시하고자 한다. 이를 위해 관련된 선례를 먼저 검토한 후 가이드라인을 제시하고자 한다.

<인공지능의 윤리 준수 프레임워크>

전기전자기술자협회(IEEE)는 인공지능 및 자율시스템의 윤리적 고려사항에 관한 IEEE 국제 지침을 펴내며 인공지능이 윤리적 기준에 따라 개발되고 운영될 수 있도록 적극적으로 노력하고 있다. 동 협회는 인공지능 사용기업에 대한 요구사항(“A Call to Action for Businesses Using AI”) 문서[28]를 발간하며 조직의 인공

지능 윤리 준수 단계를 측정할 수 있는 점검표를 제시했다.

먼저 가장 낮은 단계인 ‘도태 단계’에 있는 조직(기업)은 내부 직원이 AI 윤리 관련 자료를 스스로 찾고, 윤리적 가치보다는 조직 내부의 통제기준 준수에 더 집중한다. 경영진은 인공지능의 윤리 이슈가 중요하다는 인식을 가지고는 있으나 이를 우선순위에 두지는 않는다. 또한 조직의 핵심성과로 인공지능의 윤리 이슈를 원칙적인 수준에서만 논의하고 이를 측정하는 도구를 가지고 있지 않다.

인공지능의 윤리 준수 프레임워크(AI Ethics Readiness Framework)

	도태 단계
내부 트레이닝, 지원 및 인력	• 직원들이 적합한 AI 윤리 관련 자료를 스스로 찾아야 함 • 격려는 이루어지더라도 공식적인 지원은 없음 • 내부통제기준 준수에 더 집중되어 있음
경영진 수준에서의 지원	• 경영진의 AI 윤리에 관한 인지는 있으나 우선순위를 두지 않음
측정도구 및 핵심성과지표(KPIs)	• 일반적인 AI 윤리 원칙 수준에서만 머물며 명확하게 정성화된 측정도구의 부재
조직 내의 영향력	• 전반적인 조직의 변화 없음. 부서간 이기주의 극복 못함.

출처: IEEE SA, “A Call to Action for Businesses Using AI,” (https://ethicsinaction.ieee.org/#series).

다음 단계인 ‘기본 단계’는 내부 직원이 인공지능의 윤리 이슈에 대해 업무를 할 때 조직 차원의 공식적인 지원을 받을 수 있으며, 조직이 직원에게 일정 수준 이상의 이해를 요구한다. 경영진은 인공지능의 윤리 이슈에 대한 초급 수준의 교육을 받았으나 가지고 있는 지식은 아직 내부통제기준에 집중되어 있다. 다만 기본 단계의 조직은 핵심성과지표로 인공지능의 윤리 준수에 관한 정량지표(예를 들어 인권영향평가, 사회적 웰빙 지표 등)를 가지고 있다. 이는 조직이 인공지능의 윤리적 문제를 예방하는 것을 조직

28) IEEE SA, “A Call to Action for Businesses Using AI,” (출처: https://ethicsinaction.ieee.org/#series).

성과로 인정하고 있으며 이를 핵심적인 조직 목표로 인정하고 있는 것이기도 하다.

인공지능의 윤리 준수 프레임워크(AI Ethics Readiness Framework)

	기본 단계
내부 트레이닝, 지원 및 인력	• 팀원들에게 워크샵 및 수료증이 요구됨 • 관련 지원인력에 접근 및 지원 요청 가능 • 전문가 검토 위원회 설치
경영진 수준에서의 지원	• 초급 수준의 트레이닝만 이루어짐 • 내부통제기준에 집중된 지식
측정도구 및 핵심성과지표(KPIs)	• 정량화된 기본 측정도구 확립(인권영향평가, 사회적 웰빙 지표) • 사용자 리서치 프로세스에서 일부 측정도구 활용
조직 내의 영향력	• 전체 조직의 원칙 및 책임의식이 각 팀의 업무실행에도 반영됨

'발달 단계'는 조직 전반에 걸쳐 인공지능의 윤리에 대한 문화가 확산되어 있는 단계이다. 조직이 인공지능 윤리에 관한 자문위원회를 가지고 있으며 각 상품별로 인공지능 윤리를 담당하는 직원을 배치해야 한다. 또한 내부 직원들이 기존 업무에 인공지능의 윤리 준수와 관련한 사례를 반영할 수 있는 유연성을 가지고 있어야 한다. 경영진은 신규 프로젝트 검토 시 인공지능이 윤리기준을 준수하고 있는지 확인해야 하며, 경영진이 직접 각 부서가 윤리기준을 준수하기 위해 노력하는 상황을 보고받는다. 또한, 조직 전체 차원에서 인공지능이 윤리적 기준을 충족하도록 부서 간의 협력이 원활하게 이루어지며 조직 전체 구성원이 이에 대해 책임감을 가진다. 그리고 이러한 이슈와 관련하여 건설적 비판을 위한 기업 문화가 확립되어 있다.

인공지능의 윤리 준수 프레임워크(AI Ethics Readiness Framework)

	발달 단계
내부 트레이닝, 지원 및 인력	• 자문위원회 설치 • 각각의 상품/솔루션 별 핵심인력 및

	지원인력 배치 • 직원들이 기존 프로세스에 각자의 활용 사례를 반영할 수 있음
경영진 수준에서의 지원	• 신규 프로젝트에 AI 윤리기준 적용 • 단체협약에 AI 윤리 조항 포함 • 경영진에서 각 팀의 노력 사항들을 보고받음
측정도구 및 핵심성과지표(KPIs)	• 사용자의 신뢰 및 이해에 기반한 측정도구의 개발 및 유지 • 상품 평가에 가치 반영 • 문제해결을 우선순위로 두는 위험 분류표 마련 • 차등적인 개인정보보호(differential privacy)
조직 내의 영향력	• 프로세스 개선을 위한 각 부서들 간의 협력 • 조직의 경영진 및 직원 전원의 이해 및 책임의식 • 대화, 논의, 건설적 비판을 위한 기업문화의 확립

마지막으로 가장 높은 단계인 '선두 단계'의 조직은 내부적으로 인공지능의 윤리 이슈가 직원의 모든 역할과 의사결정에서 주요한 요소로 작용한다. 경영적 측면에서는 상품 개발 단계부터 윤리적 관점이 반영되며 윤리기준 준수에 따른 보상체계가 확립되어 있다. 또한, 경영진에서는 인공지능의 윤리 준수에 대한 업적을 치하한다. 다음으로 기업의 핵심성과지표를 사용자의 피드백에 따라 수정한다. 가장 근본적으로는 조직 전체가 사고방식의 변화를 보여야 한다.

인공지능의 윤리 준수 프레임워크(AI Ethics Readiness Framework)

	선두 단계
내부 트레이닝, 지원 및 인력	• AI 윤리가 업무의 일부로만 취급되지 않고 의사결정에 반영됨 • 모든 역할 및 직원교육과정에 영향을 미침
경영진 수준에서의 지원	• 상품개발전략에 윤리적 관점 및 실행지침이 반영됨

	• 윤리적 행위에 대한 보상체계 및 비윤리적 행위에 대한 조치 마련 • 경영진에서 AI 윤리 관련 업적 치하
측정도구 및 핵심성과지표(KPIs)	• 리서치에 기반한 건강한 신뢰 수준에 도달하기 위한 기획실행(sprints) 및 목표 설정 • 수정을 위한 지속적인 사용자 피드백 수용 • 수익 연계 • 공통취약성 및 노출(CVE)
조직 내의 영향력	• 상품의 방향성 및 사용자들과의 관계 변화 • 사고방식(mindset)의 변화

본 프레임워크는 전기전자기술자협회의 윤리기반설계 기업 위원회(Ethically Aligned Design(EAD) for Business Committee)에서 작성한 것으로 실무경험이 풍부한 개발자들이 모여 논의한 결과이다. 다양한 기업과 기관의 기술자가 참여했고 인공지능 관련 주요 기업의 경험 많은 개발자가 해당 위원회를 이끌었지만 이러한 단계 구분이 모든 기업에 적용되기에는 한계가 있다.

이상 검토한 프레임워크에서 중요한 점은 조직(기업)의 인공지능 윤리 준수는 경영진의 결정으로 이루어지지 않는다는 것이다. 윤리를 중요시하는 조직 문화는 조직의 결정이나 행동이 아닌 꾸준한 노력이 축적되어 변화가 이루어져야 달성할 수 있는 과제이다. 따라서 조직(기업)이 인공지능 개발 시 인권에 기반한 접근을 하기 위해서는 모든 조직 구성원의 끊임없는 노력과 단계적 발전이 필요하다.

이러한 관점에서 아래의 프레임워크를 바라본다면, 조직은 왼쪽의 도태 단계부터 오른쪽의 선두 단계로 나아가는 과정 내의 어딘가에 위치해 있는 것이다. 어떠한 조직(기업)은 내부 트레이닝 측면에서는 도태 단계에 머물러 있지만 조직 내의 영향력은 발달 단계에 해당할 수도 있다. 다만 중요한 것은 조직의 모든 부분이 선두 단계에 해당할 수 있도록 지속적인 변화 노력을 기울이는 것이다.

인공지능의 윤리 준수 프레임워크(AI Ethics Readiness Framework)

	도태 단계	기본 단계	발달 단계	선두 단계
내부 트레이닝, 지원 및 인력	• 직원들이 적합한 AI 윤리 관련 자료를 스스로 찾아야 함 • 격려는 이루어지더라도 공식적인 지원은 없음 • 내부통제기준 준수에 더 집중되어 있음	• 팀원들에게 워크샵 및 수료증이 요구됨 • 관련 지원인력에 접근 및 지원 요청 가능 • 전문가 검토 위원회 설치	• 자문위원회 설치 • 각각의 상품/솔루션별 핵심인력 및 지원인력 배치 • 직원들이 기존 프로세스에 각자의 활용 사례를 반영할 수 있음	• AI 윤리가 업무의 일부로만 취급되지 않고 의사결정에 반영됨 • 모든 역할 및 직원교육과정에 영향을 미침
경영진 수준에서의 지원	• 경영진의 AI 윤리에 관한 인지는 있으나 우선순위를 두지 않음	• 초급 수준의 트레이닝만 이루어짐 • 내부통제기준에 집중된 지식	• 신규 프로젝트에 AI 윤리기준 적용 • 단체협약에 AI 윤리 조항 포함 • 경영진에서 각 팀의 노력 사항들을 보고받음	• 상품개발전략에 윤리적 관점 및 실행지침이 반영됨 • 윤리적 행위에 대한 보상체계 및 비윤리적 행위에 대한 조치 마련 • 경영진에서 AI 윤리 관련 업적 치하
측정도구 및 핵심성과지표(KPIs)	• 일반적인 AI 윤리 원칙 수준에서만 머물며 명확하게 정성화된 측정도구의 부재	• 정량화된 기본 측정도구 확립(인권영향평가, 사회적 웰빙 지표) • 사용자 리서치 프로세스에	• 사용자의 신뢰 및 이해에 기반한 측정도구의 개발 및 유지 • 상품 평가에 가치 반영 • 문제해결을	• 리서치에 기반한 건강한 신뢰 수준에 도달하기 위한 기획실행(sprints) 및 목표 설정 • 수정을 위한

		서 일부 측정도구 활용	우선순위로 두는 위험 분류표 마련 • 차등적인 개인정보보호(differential privacy)	지속적인 사용자 피드백 수용 • 수익 연계 • 공통취약성 및 노출(CVE)
조직 내의 영향력	• 전반적인 조직의 변화 없음. 부서간 이기주의 극복 못함.	• 전체 조직의 원칙 및 책임의식이 각 팀의 업무실행에도 반영됨	• 프로세스 개선을 위한 각 부서들 간의 협력 • 조직의 경영진 및 직원 전원의 이해 및 책임의식 • 대화, 논의, 건설적 비판을 위한 기업문화의 확립	• 상품의 방향성 및 사용자들과의 관계 변화 • 사고방식(mindset)의 변화

<인공지능 및 알고리즘 활용 가이드>

미국 연방거래위원회(Federal Trade Commission, 이하 “FTC”)는 미국 알고리즘책임법안에 등장했던 기관으로, 미국 연방정부 차원에서 인공지능과 관련한 정책을 활발하게 펼치고 있다. 2020년 4월, FTC의 소비자보호국 국장 Andrew Smith는 인공지능 및 알고리즘의 활용에 관하여(Using Artificial Intelligence and Algorithms)라는 글[29]을 통해 기업들이 어떻게 인공지능 및 알고리즘 관련 소비자 보호 위험을 관리할 수 있을지에 대해 언급했다. Smith 국장의 글은 그동안 FTC가 다루었던 인공지능의 역기능 사례를 기반으로 작성되었으므로 기업에서 실질적으로 활용할 수 있는 유용한 내용으로 구성되어 있다.

29) Andrew Smith, “Using Artificial Intelligence and Algorithms,” (2020. 4. 8.), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms

Smith 국장이 제시한 가이드는 총 5개 분야로 이루어져 있는데 투명성, 정보제공, 공정성, 안전성, 책임성으로 볼 수 있다. 이 가이드는 앞서 검토했던 인공지능의 윤리 준수 프레임워크와 같이 조직 전체 차원에서 활용할 수 있도록 작성되었다. 따라서 아래에서 제시하는 5개 분야 15개 과제가 기업의 경영진은 물론, 현장에서 프로그램을 개발하는 기술자에게까지 폭넓게 사용될 수 있을 것으로 보인다.

❶ 투명해야 한다. (Be transparent.)

○ 회사의 자동화 도구의 활용 방식에 관하여 소비자들을 기만하지 않아야 한다.
(Don't deceive consumers about how you use automated tools.)

○ 민감한 정보를 취합할 때에 투명하게 처리해야 한다.
(Be transparent when collecting sensitive data.)

○ 제3자로부터 취득한 정보에 기반하여 자동화된 의사결정을 내리는 경우 회사는 소비자에게 "불리한 조치" 통지를 제공하도록 요구될 수 있다.
(If you make automated decisions based on information from a third-party vendor, you may be required to provide the consumer with an "adverse action" notice.)

Smith 국장이 가장 처음 언급한 것은 투명성(Be transparent)이다. 기업이 자동화 도구의 활용 방식에 관하여 소비자들을 기만하지 않아야 함을 강조했다. 많은 경우에 인공지능이 백그라운드 차원에서 운영되며 이용자에게 드러나지 않는다. 하지만 이것은 소비자를 기만하는 행위로 보고 있다.

다음으로 강조한 것은 민감한 정보수집 시 투명성을 확보하는 것이다. 데이터가 커질수록 알고리즘이 개선되고 소비자를 위한 제품을 개선할 수 있다. 하지만 음성이나 영상 정보와 같은 민감한 자료를 비밀로 수집하고 이를 알고리즘 훈련에 사용하는 경우 FTC의 조사를 받을 수 있음을 알리고 있다.

❷ 설명해야 한다. (Explain your decision to the consumer.)

○ 알고리즘 의사결정에 기반하여 소비자에게 가치있는 무언가의 제공을 거부할 경우, 그 이유를 자세하게 설명해야 한다.
(If you deny consumers something of value based on algorithmic decision-making, explain why.)

○ 알고리즘을 활용하여 소비자에게 리스크 점수를 부여하고자 하는 경우, 해당 점수에 영향을 미친 주요 요인들과 고려사항들의 중요도를 공개해야 한다.
(If you use algorithms to assign risk scores to consumers, also disclose the key factors that affected the score, rank ordered for importance.)

○ 자동화 도구에 기반하여 거래 조건을 변경하고자 하는 경우, 반드시 소비자에게 알려야 한다.
(If you might change the terms of a deal based on automated tools, make sure to tell consumers.)

다음으로 제시되는 개념은 소비자에게 인공지능의 결정에 대한 설명을 제공하는 것이다. 먼저, 알고리즘의 결정을 통해 소비자의 요청을 거부하는 경우 그 이유를 자세하게 설명해야 한다. 이는 주로 금융과 관련된 회사가 신용정보를 기반으로 소비자의 요구를 받아들이지 않는 사례를 두고 말하는 것이다.

❸ 공정해야 한다. (Ensure that your decisions are fair.)

○ 취약 계층에 대한 차별이 있어서는 안 된다.
(Don't discriminate based on protected classes.)

○ 입력되는 자료(inputs)뿐만 아니라 결과물(outcomes)에도 주의해야 한다.
(Focus on inputs, but also on outcomes.)

○ 소비자에 관한 의사결정을 내릴 시 사용된 정보를 정정할 수 있는 기회와 접근 권한을 소비자에게 제공해야 한다.
(Give consumers access and an opportunity to correct information used to make decisions about them.)

공정성 부분에서는 입력자료와 함께 결과물도 함께 확인해야 한다는 점이 흥미롭다. 보통 빅데이터의 편향과 그에 따른 문제점을 언급하는 경우가 많지만 여기서는 이와 함께 결과물에도 확인이 필요하다고 주문한다.

FTC에서 알고리즘에 차별이 있는지 평가할 때 입력자료에 인종이 편향되지 않았는지 등을 확인함은 물론, 같은 입력값에 따른 산출 결과를 비교하여 만약 다른 결과가 나온다면 해당 알고리즘이 사용하고 있는 정의부터 다시 확인한다고 한다.

❹ 회사의 데이터와 모델은 안정적(robust)이며 경험적으로 건전해야 한다. (Ensure that your data and models are robust and empirically sound.)

○ 신용, 고용, 보험, 주택, 정부지원, 개인수표-현금 혹은 이와 유사한 거래들에 대한 소비자 접근에 관하여 결정을 내리고자 타인/타사에 소비자 관련 정보를 제공하는 경우, 해당 회사는 신용평가공정법(FCRA)을 반드시 준수해야 하는 소비자보고기관(consumer reporting agency)에 해당되어야 할 수 있다.

(If you provide data about consumers to others to make decisions about consumer access to credit, employment, insurance, housing, government benefits, check-cashing or similar transactions, you may be a consumer reporting agency that must comply with the FCRA, including ensuring that the data is accurate and up to date.)

○ 자동화된 의사결정에 사용하고자 회사의 소비자 관련 정보를 타인/타사에 제공하는 경우, 해당 회사는 소비자 보고기관이 아니더라도 관련 정보의 정확성을 확인해야 할 의무를 가질 수 있다.

(If you provide data about your customers to others for use in automated decision-making, you may have obligations to ensure that the data is accurate, even if you are not a consumer reporting agency.)

○ 회사의 AI 모델이 의도된 바에 따라 실행되고 불법적으로 차별하지 않도록 하고자 해당 AI 모델의 유효성을 반드시 확인하고 재확인해야 한다.

(Make sure that your AI models are validated and revalidated to ensure that they work as intended, and do not illegally discriminate.)

다음으로 안전성 및 건전성과 관련하여 Smith 국장은 FTC에서 진행되었던 유용한 정보를 공유했다. 이용자에게 금전적 이익을 결정하는 것과 관련하여 타 기관에 이용자의 정보를 제공하려면 소비자보호기관 자격이 있어야 하는데 RealPage라는 회사는 이를 지키지 않고 주택 임차 지원 희망자의 범죄 기록을 제공하다가 3백만 달러의 벌금을 부과받기도 하였다.

위 지침에서는 소비자에 관한 정보를 타 기관에 이전할 때 금전적 이익을 결정하는 정보가 아니더라도 정보의 정확성을 확인할 것을 요구하고 있다. 그 이유는 인공지능에 사용되는 데이터가 최대한 현실을 반영해야 하기 때문이다.

❺ 준법감시, 윤리, 공정성 및 비차별에 대해 책임을 져야 한다. (Hold yourself accountable for compliance, ethics, fairness, and nondiscrimination.)

○ 알고리즘을 사용하기에 앞서 질문을 해야 한다.
(Ask questions before you use the algorithm.)

- 회사의 데이터(data set)는 대표성을 갖는가?
(How representative is your data set?)

- 회사의 데이터 모델은 편향적인가?
(Does your data model account for biases?)

- 빅데이터에 기반한 회사의 예측이 얼마나 정확한가?
(How accurate are your predictions based on big data?)

- 회사의 빅데이터 의존으로 인해 윤리 혹은 공정성 문제가 야기되지 않는가?
(Does your reliance on big data raise ethical or fairness concerns?)

○ 미인가 사용으로부터 회사의 알고리즘을 보호해야 한다.
(Protect your algorithm from unauthorized use.)

○ 회사의 책무성(accountability) 메커니즘을 고려해야 한다.
(Consider your accountability mechanism.)

마지막으로 책임성 분야에서는 알고리즘 활용 전 4가지 질문사항을 눈여겨 볼만하다. 데이터의 대표성 확보 여부, 편향성 점검, 정확성 점검, 윤리적 또는 공정성과 관련한 문제 예측은 빅데이터를 사용하는 인공지능의 역기능을 예방하기 위하여 꼭 필요한 질문이다.

이상과 같이 FTC의 Smith 국장이 제시한 가이드라인을 검토해본 결과, 미국은 다양한 산업에서 인공지능을 활용하고 있으며 FTC가 인공지능에 대한 전문성을 바탕으로 인공지능의 역기능을 줄이기 위해 다수의 규제를 진행하고 있었다.

지금까지 검토한 내용은 기존에 다루어졌던 데이터의 중립성 확보 문제는 물론 이용자의 개인정보 보호, 입력값 점검뿐만 아니라 결과값 점검, 인공지능에 사용되는 데이터의 정확성 제고 등 인공지능의 역기능을 예방하기 위한 다양한 접근방법을 소개하고 있다.

이번 장에서 다루었던 IEEE의 인공지능의 윤리 준수 프레임워크, FTC의 인공지능 및 알고리즘 활용 (가이드)를 종합하여 인권기반 인공지능 개발 가이드라인을 구상해본다면 다음과 같다.

인권기반 인공지능 개발 가이드라인

본 가이드라인은 인공지능을 개발하는 조직 전체(관리자, 개발자, 행정지원인력 등)에 필요한 내용입니다. 따라서 모든 조직 구성원이 본 가이드라인에 제시된 질문을 통해 업무 방향을 수립 및 수정하는 것이 필요합니다. 이를 통해 인권에 기반한 인공지능 개발 문화가 조직에 자리 잡기를 바랍니다.

❶ 사람에게 온전히 도움을 주는가?

◇ 우리 조직이 개발한 인공지능이 인간에게 어떠한 도움을 제공하는가?

◇ 인공지능이 사람에게 온전하게 도움만 제공하는가? 아니면 도움과 함께 불편을 줄 수도 있는가?

◇ 인공지능의 역기능을 발견할 경우 수정할 수 있는가?

❷ 이용자의 권리를 알고 있는가?

◇ 인공지능 이용자는 어떠한 권리를 가지고 있는가?

◇ 인공지능이 이용자의 인권 증진을 위하여 사용되는가?

❸ 우리 조직의 목표에 부합하는가?

◇ 우리 조직의 핵심성과지표에 인권 증진이 포함되어 있는가?

◇ 경영진이 인권에 기반한 인공지능 개발을 조직의 최우선 목표로 하고 있는가?

◇ 내부고발자를 철저히 보호할 수 있는가?

❹ 데이터에 문제가 없는가?

◇ 데이터가 대표성을 가지고 있는가?

◇ 취약계층이 데이터에 반영되어 있는가?

◇ 이용자의 요구에 따라 데이터를 수정할 수 있는가?

❺ 제공자의 의무를 다하고 있는가?

◇ 이용자에게 알려야 하는 의무를 다하고 있는가?

◇ 법과 제도에서 정한 의무를 다하고 있는가?

◇ 사회적 의무를 다하고 있는가?

이상의 질문을 자유롭고 평등한 분위기에서 논의할 수 있어야 합니다. 또한 위 질문에 자신있게 답변할 수 있어야 합니다. 이러한 질문에 대한 충분한 고민과 조치가 이루어졌다면 각 질문을 자세하게 답변할 수 있을 것입니다.

5.2. 인공지능 기본원칙

최근 수년간 인공지능의 윤리적 문제에 관한 사회적 논의가 확산되면서 정부에서는 「인공지능 국가전략」('19. 8월), 「인공지능 윤리기준」('20. 12월) 등을 발표하며 인공지능에 관한 기본 원칙을 선언하기 시작했다. 사람이 중심이 되는 「인공지능 윤리기준」은 대통령직속 4차산업혁명위원회 회의에 심의안건으로 상정되어 의결된 내용으로, 우리나라 전체의 인공지능 윤리기준을 정부에서 선언한 것으로 볼 수 있다.

「인공지능 윤리기준」은 자율 규범적 성격으로 법이나 지침이 아님을 명시하고 있다. 이러한 포괄적 원칙은 강제력이 없는 도덕 규범으로서 기업의 자율성을 존중하기 위함임을 명시하고 있다. 따라서 본 기준은 우리 사회가 앞으로 인공지능을 개발하고 활용할 때 어떠한 기준으로 접근할지에 대한 선언의 의미가 있다고 볼 수 있다.

또한, 「인공지능 윤리기준」은 확장성을 강조하고 있다. 여기에 선언된 기준은 범용성을 가진 일반원칙으로서 각 기준이 더 세분화되고 발전되는 것을 허용(또는 독려)하고 있다. 「인공지능 윤리기준」에 따르면 발표된 기준이 각 인공지능 영역에서 "새롭게 제기되는 윤리적 이슈를 논의하고 구체적으로 발전시킬 수 있는 플랫폼으로 기능" 하도록 정하고 있다.

윤리기준은 인공지능이 윤리기준을 준수하기 위한 10대 핵심요건을 제시하고 있는데 인공지능의 1. 인권보장, 2. 프라이버시 보호, 3. 다양성 존중, 4. 침해금지, 5. 공공성, 6. 연대성, 7. 데이터 관리, 8. 책임성, 9. 안전성, 10. 투명성이다. 이 중 본 고의 주제인 '인권기반 인공지능 개발 모델'과 관련하여 더욱 비중 있게 참고할 부분은 1. 인권보장과 6. 연대성이다.

「인공지능 윤리기준」은 먼저 ‘인권보장’을 10대 핵심요건 중 가장 먼저 제시하며 「인공지능 윤리기준」에서 달성하고자 하는 우선 목표가 인권 보호 및 증진임을 보여주고 있다. 아래 명시된 요건은 인공지능의 개발과 활용에 있어 기본적 인권기준을 선언하는 의미를 가지고 있다.

「인공지능 윤리기준」 10대 핵심요건

① 인권보장

- 인공지능의 개발과 활용은 모든 인간에게 동등하게 부여된 권리를 존중하고, 다양한 민주적 가치와 국제 인권법 등에 명시된 권리를 보장하여야 한다.
- 인공지능의 개발과 활용은 인간의 권리와 자유를 침해해서는 안 된다.

인권보장과 함께 10대 핵심요건에서 주목해보아야 할 부분은 ‘연대성’이다. 인공지능 개발과 활용에 대한 민주적인 논의 진행과 이에 대한 참여 기회 보장은 인공지능의 부작용을 예방할 수 있는 효과적 방법이다. 인공지능 개발 및 운영 시 민주적인 의견 교환이 자유롭게 이루어진다면, 한 개발자나 기업이 발견하지 못한 인공지능의 역기능을 다른 개발자나 기업이 발견할 가능성을 높일 수 있다.

따라서 인권에 기반하여 인공지능을 개발할 수 있는 가장 효과적인 방법은 민주적 토의 절차를 거치는 것이다. 인공지능 개발자나 기업은 물론, 다양한 이해관계자가 연대하여 인공지능의 인권침해 위험성을 고민한다면 더욱 효과적으로 문제에 접근할 수 있다. 이러한 이유로 ‘연대성’은 향후 인공지능 개발 및 운영에 있어 더욱 중요한 요소로 여겨져야 할 부분이다.

「인공지능 윤리기준」 10대 핵심요건

⑥ 연대성

- 다양한 집단 간의 관계 연대성을 유지하고, 미래세대를 충분히 배려하여 인공지능을 활용해야 한다.
- 인공지능 전 주기에 걸쳐 다양한 주체들의 공정한 참여 기회를 보장하여야 한다.
- 윤리적 인공지능의 개발 및 활용에 국제사회가 협력하도록 노력해야 한다.

연대성과 관련하여 강조하고 싶은 부분은 내부신고자 보호 제도이다. 이러한 민주적인 토의 절차가 의미있는 결론을 도출하기 위해서는 실질적인 문제를 논의할 수 있어야 한다. 인공지능의 실질적이고 핵심적인 문제점을 토론의 장에 드러낼 수 있어야 하는 것이다. 하지만 많은 기업이 업무상의 비밀을 이유로 인공지능에 문제가 있더라도 이를 공개하지 않는 경향이 많다. 더욱이 문제가 심각할 경우 기업의 이익을 보호하기 위하여 이를 공개하지 않으려고 한다. 따라서 조직 구성원이 인공지능의 문제점을 외부에 알리고 이를 통해 문제를 해결할 수 있도록 내부신고자 보호 제도를 운영하는 것이 필요하다.

5.3. 인공지능 관련 법률(안)

<지능정보화 기본법>

4차 산업을 향한 국가 간 경쟁이 심화되자 우리나라는 데이터 · 인공지능 등 관련 핵심기술의 기반을 다지고 산업 변화로 인한 부작용에 대응하기 위하여 「지능정보화 기본법」을 제정했다.

이는 기존의 「국가정보화 기본법」을 전면 개정한 것으로, 4차 산업혁명에 따른 사회·경제적 변화를 선도하기 위한 범국가적 추진체계를 마련하고자 제정하였다.

여기서 “지능정보기술” 이란 인공지능, 빅데이터, 클라우드컴퓨팅 등 소위 4차 산업 주요 기술이라고 일컫는 개념들을 통칭하는 것이다. “지능정보화”란 앞서 언급한 지능정보기술 및 이와 관련된 기술을 활용하는 것을 의미한다. 마지막으로 “지능정보사회”란 “지능정보화를 통하여 산업·경제, 사회·문화, 행정 등 모든 분야에서 가치를 창출하고 발전을 이끌어가는 사회”를 말한다(「지능정보화 기본법」 제2조)

「지능정보화 기본법」 제44조, 제56조, 제62조에서 지능정보사회의 윤리에 관한 내용을 정하고 있는데 이는 본 연구주제인 인권기반 인공지능 개발과 밀접하게 관련이 있다. 먼저 제44조에서는 산업 변화에 따른 편익을 우리 사회가 보편적으로 향유하고 윤리를 확립하여 인간의 존엄과 가치를 존중해야 함을 명시하고 있다. 또한, 해당 조항에서는 4차 산업혁명의 역기능인 사생활 침해, 정보격차, 신기술에 대한 과의존, 이용자 권익 침해 등의 문제를 직시하고 이를 해결하도록 노력할 것을 주문하고 있다.

「지능정보화 기본법」

제44조(정보문화의 창달과 확산) ① 국가기관등은 인간의 존엄·가치가 존중되는 자유롭고 개방적인 정보문화 창달 및 확산을 위하여 다음 각 호의 사항이 이루어지도록 노력하여야 한다.

1. 지능정보사회 구현에 따른 편익의 보편적 향유
2. 지능정보사회윤리의 확립
3. 사생활의 비밀·자유와 개인정보의 보호
4. 지능정보화에 따른 정보격차의 해소
5. 지능정보서비스 과의존의 예방과 해소
6. 지능정보기술 및 지능정보서비스 이용자의 권익 보호

다음으로 제56조에서는 지능정보서비스가 사회에 어떠한 영향을 미치는지에 대한 ‘사회적 영향평가’에 대한 근거를 마련하고 있다. 앞서 제44조에서 지적한 지능정보기술의 역기능이 사회에 어떠한 영향을 미치는지에 대해 국가 및 지방자치단체가 이를 조사 및 평가할 수 있도록 정하고 있다.

또한, 과학기술정보통신부장관이 이렇게 진행된 사회적 영향평가의 결과를 공개하고, 지능정보서비스에 문제가 있는 경우 개선을 위해 필요한 조치를 국가기관이나 사업자에게 권고할 수 있다고 정하고 있다. 이는 4차 산업과 관련한 신기술의 안전성 및 신뢰성을 확보에 긍정적인 영향을 미칠 것으로 보이나, 과학기술정보통신부가 기업과 함께 관련 산업 발전을 선도하는 역할을 수행하는 점을 고려한다면 권고의 대상과 주체에 대한 더욱 구체적인 논의가 필요한 것을 알 수 있다.

「지능정보화 기본법」

제56조(지능정보서비스 등의 사회적 영향평가) ① 국가 및 지방자치단체는 국민의 생활에 파급력이 큰 지능정보서비스 등의 활용과 확산이 사회·경제·문화 및 국민의 일상생활 등에 미치는 영향에 대하여 다음 각 호의 사항을 조사·평가(이하 “사회적 영향평가”라 한다)할 수 있다. 다만, 지능정보기술의 경우에는 「과학기술기본법」 제14조제1항의 기술영향평가로 대신한다.

1. 지능정보서비스 등의 안전성 및 신뢰성
2. 정보격차 해소, 사생활 보호, 지능정보사회윤리 등 정보문화에 미치는 영향
3. 고용·노동, 공정거래, 산업 구조, 이용자 권익 등 사회·경제에 미치는 영향
4. 정보보호에 미치는 영향
5. 그 밖에 지능정보서비스 등이 사회·경제·문화 및 국민의 일상생활에 미치는 영향

② 과학기술정보통신부장관은 사회적 영향평가의 결과를 공개하고, 해당 지능정보서비스 등의 안전성·신뢰성 향상 등 필요한 조치를 국가기관등 및 사업자 등에 권고할 수 있다.

「지능정보화 기본법」 제62조는 지능정보사회윤리를 직접적으로 다루며 신기술의 인권적 측면을 직접적으로 다루고 있다. 조문에 언급되는 공공성, 책무성, 투명성 등의 개념은 위에서 알아보았던 「인공지능 윤리기준」의 주요 개념과 일맥상통한다.

제62조에서 주목할 것은 지능정보서비스가 인간의 존엄과 가치를 훼손하지 않도록 국가기관과 지방자치단체가 이를 교육하고 홍보함은 물론, 제도개선까지 적극적으로 추진해야 함을 명시한 점이다. 이는 공공분야의 모든 기관이 인공지능으로부터 인권을 보호하도록 적극적으로 노력해야 한다는 것을 명시한 것으로 볼 수 있다. 따라서 정부는 인공지능에 의한 인권침해 예방, 보다 투명한 인공지능 기술 구현, 이와 관련한 우리사회의 인식 증진 등을 위하여 앞으로 많은 활동을 펼쳐나가야 할 것이다.

「지능정보화 기본법」

제62조(지능정보사회윤리) ① 국가기관과 지방자치단체는 지능정보기술을 개발·활용하거나 지능정보서비스를 제공·이용할 때 인간의 존엄과 가치를 존중하고 공공성·책무성·통제성·투명성 등의 윤리원칙을 담은 지능정보사회윤리를 확립하기 위하여 다음 각 호의 사항을 포함한 시책을 마련하여야 한다.

1. 지능정보사회윤리 확립을 위한 교육, 전문인력 양성 및 홍보
2. 지능정보사회윤리 교육 콘텐츠 개발·보급
3. 지능정보사회윤리 관련 연구 및 개발
4. 지능정보사회윤리 관련 단체에 대한 지원

② 정부는 지능정보기술을 개발 또는 활용하는 자, 지능정보서비스를 제공 또는 이용하는 자가 인간의 존엄과 가치를 훼손하거나 기본적인 지능정보사회윤리를 침해하지 않도록 윤리교육·홍보 및 제도개선을 적극적으로 추진하여야 한다.

③ 정부는 제2항에 따른 윤리교육·홍보를 위하여 학교교육, 평생교육과 언론·인터넷 등을 통한 홍보 등에 사용하는 프로그램 및 콘텐츠를 개발하여 보급하여야 한다.

④ 정부는 지능정보기술 또는 지능정보서비스 개발자·공급자·이용자가 준수하여야 하는 사항을 정한 지능정보사회윤리준칙을 제정하여 보급할 수 있다.

<알고리즘 및 인공지능에 관한 법률안>

국회에서는 2020년 이후 인공지능과 관련된 법안 발의가 급격히 증가했다. 발의된 법안은 주로 인공지능 기술 발전을 촉진하고 인공지능으로 인한 부작용을 예방하는 내용을 담고 있다.

<표 6-1> 인공지능 관련 법률안 목록 (2020 ~ 2021, 최신순)

번호	의안번호	의안명	제안일자	심사진행
1	2113509	알고리즘 및 인공지능에 관한 법률안(윤영찬의원 등 12인)	2021-11-24	소관위심사
2	2111573	인공지능에 관한 법률안(이용빈의원등31인)	2021-07-19	소관위심사
3	2111261	인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안(정필모의원 등 23인)	2021-07-01	소관위심사
4	2111106	정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(류호정의원 등 12인)	2021-06-25	소관위심사
5	2110148	인공지능교육진흥법안(안민석의원 등 10인)	2021-05-17	소관위심사
6	2104772	인공지능 기술 기본법안(민형배의원 등 10인)	2020-10-29	소관위심사

7	2104564	인공지능 집적단지의 육성에 관한 특별법안(송갑석의원 등 11인)	2020-10-19	소관위심사
8	2103515	인공지능산업 육성에 관한 법률안(양향자의원 등 23인)	2020-09-03	소관위심사
9	2101823	인공지능 연구개발 및 산업 진흥, 윤리적 책임 등에 관한 법률안(이상민의원 등 11인)	2020-07-13	소관위심사

위 법률안 중 인공지능에 의한 부작용을 다루고 있는 법안은 ‘알고리즘 및 인공지능에 관한 법률안,’ ‘인공지능에 관한 법률안,’ ‘인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안,’ ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안’으로 볼 수 있다. 이 중에서 인공지능의 규제내용을 폭넓게 담고 있는 ‘알고리즘 및 인공지능에 관한 법률안’을 중심으로 인권기반 인공지능 개발을 어떻게 추진할 수 있을지 알아보도록 한다.

윤영찬 의원 등이 발의한 ‘알고리즘 및 인공지능에 관한 법률안’에서는 제15조에서 ‘고위험인공지능 심의위원회’를 국무총리 산하에 설치할 것을 정하고 있다. 심의위원회는 인공지능 관련 정책수립, 공공부분의 인공지능 활용 검토, 윤리원칙 마련 등 인공지능과 관련된 전반적인 정책검토에서부터 국제협력에 이르기까지 폭넓은 업무 범위를 가지고 있다.

「알고리즘 및 인공지능에 관한 법률안」

고위험인공지능 심의위원회의 심의·조정 사항

1. 고위험인공지능과 그 알고리즘 규율에 관한 기본원칙의 수립
2. 고위험인공지능과 그 알고리즘의 규율에 관한 관련 정책의

수립

3. 고위험인공지능과 관련된 사회적 변화 양상과 정책적 대응에 관한 사항

4. 공공 부분에서의 고위험인공지능의 도입에 관한 사항

5. 윤리원칙 및 관련 법령의 제정 및 개정에 관한 사항

6. 국제협력에 관한 사항

7. 정책 추진에 필요한 재원 조달 및 운용에 관한 사항

8. 제6조제4항에 따른 알고리즘 및 인공지능 윤리위원회에 관한 사항

9. 제19조제3항에 따라 고위험인공지능 이용자가 요구한 자료 제출에 대한 심사

10. 그 밖에 사회적으로 중대한 영향을 미칠 수 있다고 판단하여 위원장이 회의에 부치는 사항

해당 위원회의 심의위원장 2인은 국무총리와 대통령이 위촉하는 사람이며 간사는 과학기술정보통신부장관이다. 위원은 30인 이내로 비교적 많은 숫자를 가지고 있고 임기는 2년이다. 위원장의 직급이나 위원회의 규모를 고려했을 때, 해당 위원회는 국가 전체의 인공지능을 관할할 수 있을 것으로 보인다. 위원회가 30인을 위원으로 확보할 경우 인공지능 업무 분야별로 소위원회를 구성하여 폭넓은 주제를 다룰 수 있기 때문이다.

또한, 이 법안에서는 제17조에서 고위험인공지능개발사업자의 책무를 아래와 같이 정하고 있다. 고위험인공지능의 개발과 관련하여 인공지능 개발 주체에 대해 아래와 같은 책무를 부여하는 것은 인공지능 개발과 활용에서의 책임성을 강화하는 데에 큰 역할을 할 것으로 보인다.

「알고리즘 및 인공지능에 관한 법률안」

제18조(고위험인공지능개발사업자의 책무) ① 고위험인공지능의 개발과 관련된 경제활동을 영위하는 자(이하 “고위험인공지능개발

사업자”라 한다)는 다음 각 호의 사항을 준수하여야 한다.

1. 고위험인공지능 개발 관련 위험관리시스템의 구축
2. 고위험인공지능 개발 단계별 문서의 전자화
3. 고위험인공지능의 개발 결과의 추적을 위한 기록
4. 고위험인공지능 이용자에 대한 정보제공
5. 사람에 의한 고위험인공지능의 관리·감독
6. 고위험인공지능 개발 과정에서의 사이버 보안 강화
7. 국민의 생명이나 신체적 안전에 중대한 위험성이 있는지에 대한 위험 평가

② 고위험인공지능개발사업자는 이용자 및 이해관계자에게 고위험인공지능 알고리즘 등의 동작원리를 알려야 한다. 이 경우 영업비밀로서 대통령령으로 정하는 경우에는 그러하지 아니하다.

③ 고위험인공지능이용사업자는 이용자로 하여금 인공지능을 통하여 업무가 처리됨을 알 수 있도록 하여야 한다.

제18조(고위험인공지능이용사업자의 책무) ① 고위험인공지능을 이용하여, 이용자에게 서비스를 제공하는 자(이하 “고위험인공지능이용사업자”라 한다)는 고위험인공지능이용사업이 이용자의 권리를 침해하지 않도록 인공지능모니터링시스템을 구축·운영하여야 한다.

② 고위험인공지능이용사업자는 이용자에게 사전에 인공지능을 이용한 서비스가 제공되고 있음을 고지하여야 한다.

③ 고위험인공지능이용사업자는 고위험인공지능을 이용함으로써 이용자의 생명 및 신체의 안전에 중대한 위험을 야기할 가능성이 있는 경우에는 그 위험성에 대하여 이해할 수 있도록 설명하여야 한다.

④ 과학기술정보통신부장관은 제1항에 따른 인공지능모니터링시스템의 구축 및 운영을 행정적·재정적으로 지원할 수 있다.

또한, 제2항에서 고위험인공지능 이용자에게 알고리즘 등의 동작원리를 알려야 한다고 정한 부분 역시 큰 의미를 가지고 있다. 인공지능 서비스를 사용하는 소비자의 알 권리를 법률에서 재확인한다면, 이후 소비자의 권리를 강화할 때 소비자의 알 권리 보장을

발전시킬 근거가 될 수 있기 때문이다. 다만, 그동안 많은 사례에서 기업이 영업비밀 유지를 위해 알고리즘 작동 원리와 관련된 내용을 공개하지 않은 선례를 고려하여, 실효성을 제고할 수 있는 방안을 마련하는 것이 필요하다.

다음으로 고위험인공지능 개발 과정과 결과를 기록하도록 정한 부분은 향후 인공지능 문제 발생 시 원인에 대한 접근을 가능하게 하므로 중요한 부분으로 보인다. 이와 유사하게 미국의 알고리즘책임법안 역시 인공지능 개발 과정을 문서화할 것을 요구하고 있다. 하지만 알고리즘책임법안의 경우 문서화한 자료를 정부에게 제출하도록 정하고 있으므로 이는 양 법안의 다른 점으로 볼 수 있다.

인공지능 개발문서를 관리하는 문제는 기업의 영업비밀과 직결되어 있어 조심스러운 접근이 필요하다. 고위험인공지능 개발문서에 기업의 영업비밀을 포함시키고 기업에서 관리하도록 해야 할지, 개발 문서에 기업의 핵심적 영업비밀을 제외하고 정부에서 문서를 관리할지, 영업비밀 전반을 제외하고 제3의 전문기관에서 문서를 관리할지 등 합리적인 방법을 찾기 위해 다양한 대안을 고려해야 한다.

마지막으로, 이용자에게 인공지능의 의사결정이 이루어지고 있는 부분을 고지하도록 명시한 것은 이용자의 알 권리 보호에 있어 실질적인 역할을 할 것으로 보인다. 그동안 다수의 인공지능 서비스가 의사결정의 주체가 인공지능인 것을 명시적으로 밝히지 않아 소비자들이 이를 미처 인식하지 못하는 경우가 있었다. 소비자는 자신에게 영향을 미치는 결정이 누구에 의해 이루어졌는지 알 권리가 있으므로 이를 적극적으로 보호하는 것이 필요하다.

알고리즘 및 인공지능에 관한 법률안은 고위험인공지능 이용자의 권리를 아래와 같이 규정하였는데, 이는 소비자 권리보호

측면에서 큰 역할을 할 것으로 보인다.

「알고리즘 및 인공지능에 관한 법률안」

고위험인공지능 이용자의 보호(제19조)

1. 고위험인공지능을 이용한 기술 또는 서비스에 대한 설명요구권

2. 고위험인공지능을 이용한 기술 또는 서비스에 대한 이의제기권 또는 거부권

② 고위험인공지능 이용자는 알고리즘에 따른 부당한 처우가 있었는지를 확인하기 위하여 사업자에게 자료를 요청할 수 있다.

③ 제2항에 따른 요청을 받은 사업자는 다른 법률에 특별한 규정 또는 정당한 사유가 없는 한 제6조제4항에 따른 알고리즘 및 윤리위원회의 심의를 거쳐 이를 처리하여야 한다. 이 경우 제6조제4항에 따른 알고리즘 및 인공지능 윤리위원회가 자료의 제출을 거부하는 경우에는 심의위원회에 자료의 제출을 위한 심사를 요청할 수 있다.

④ 이용자는 제공받는 서비스가 알고리즘에 따라 처리된다는 사실을 제공받아야 한다. 이 경우 이용자는 제공받는 서비스가 알고리즘에 따라 처리되는 것을 거부할 권한이 있다.

여기서 인공지능 이용자에게 서비스에 대한 이의제기권 또는 거부권을 부여한 것은 인공지능을 선택하지 않을 권리를 보장하는 것으로 볼 수 있다. 즉, 기계(인공지능)가 아닌 사람에 의해 결정받을 권리를 부여하는 것이다. 특히 본 법률안에서 규정하고 있는 고위험인공지능이 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미치는 인공지능인 것을 고려할 때 이러한 권리 부여는 이용자에게 필수적이다.

본 법안은 고위험 인공지능에 대한 정의를 아래와 같이 내리고 있는데 이는 현재 실생활에서 활용되고 있는 인공지능의 주요 분야를 폭넓게 포괄하고 있는 것으로 보인다.

「알고리즘 및 인공지능에 관한 법률안」

고위험인공지능의 정의(제2조 제3항)

“고위험인공지능”이란 국민의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미치는 인공지능으로 다음 각 목의 어느 하나에 해당하는 인공지능을 말한다.

가. 인간의 생명과 관련된 인공지능

나. 생체인식과 관련된 인공지능

다. 교통, 수도, 가스, 난방, 전기 등 주요 사회기반시설의 관리·운용과 관련된 인공지능

라. 채용 등 인사 평가 또는 직무 배치의 결정에 이용되는 인공지능

마. 응급서비스, 대출 신용평가 등 필수 공공·민간 서비스 관련 인공지능

바. 수사 및 기소 등 기본권을 침해할 수 있는 국가기관의 권한 행사에 이용되는 인공지능

사. 문서의 진위 확인, 위험평가 등 이민, 망명 및 출입국관리와 관련된 인공지능

다만, 위 정의에서 인간의 생명에 영향을 미치는 요소가 다양하기 때문에 여기에 해당하는 인공지능의 범위를 좀 더 구체화할 필요가 있다. 또한, 여기서 생명과 직결된 부분만 다룬다면 신체 및 정신적 건강에 영향을 미칠 수 있는 다른 인공지능의 경우 적용 대상에서 제외된다. 의료분야의 인공지능 활용이 비약적으로 성장하고 있는 점, 의료행위가 이용자에게 돌이킬 수 없는 영향을 미치는 경우가 많은 점 등을 고려하여 해당 정의를 구체화할 필요가 있다.

마지막으로 위 목록에 해당되지는 않았지만 국민의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미치는 인공지능이 앞으로 개발될 수 있으므로 이를 포괄할 수 있는 내용이 법안에

포함되는 것이 필요하다. 최근 몇 년간 인공지능의 활용 분야가 급속히 확장되고 있으므로 이를 고려하여 이용자의 권리를 폭넓게 보호해야 하기 때문이다.

5.4. 인권기반 인공지능 개발 모델

이상 국내외의 논의를 검토해본 결과, 인공지능이 인권을 침해하지 않도록 예방하고, 인권침해 발생 시 이에 대응할 수 있는 제도를 마련하는 것이 필요하다는 것은 누구나 동의하고 있다는 것을 알 수 있었다. 하지만 이용자의 알 권리와 인공지능 개발사의 영업상 비밀을 보장받을 권리가 충돌하므로 이를 조정하기 위한 제도가 필요하다.

인공지능의 인권침해를 예방하기 위한 가장 효율적인 방법은 인공지능 설계 단계에서부터 인권침해를 예방하기 위한 노력을 기울이는 것이다. 이를 위하여 인권에 기반한 인공지능 개발 가이드라인이 필요하며 이를 5장 초반부에서 제시하였다. 인공지능 개발자는 인공지능이 편향된 데이터로 훈련을 진행한다는 것을 명심하고 이러한 편향을 줄일 수 있는 방안을 고민하면서 인공지능을 개발할 필요가 있다. 이러한 개발자의 문제 인식은 인공지능에 의한 인권침해 예방에 큰 역할을 할 수 있을 것이다.

이와 함께 인공지능 기업의 경영진과 투자자, 그리고 이용자 역시 인공지능의 인권 존중을 최우선 가치로 여기고 행동해야 한다. 개발자는 회사의 개발 의지를 실현하기 위하여 노력한다. 회사의 인공지능 개발 원칙과 방향 설정, 조직성과평가 체계 등 다양한 조직적 요소가 인권에 기반한 인공지능의 구현을 목표로 할 때 개발자도 이에 따르는 것이다.

편향이 줄어든 데이터를 통해 인공지능을 훈련하는 방법 역

시 인권침해 예방에 일부 도움이 될 것으로 보인다. 인구통계학적인 고려를 통해 사회 내의 특정 그룹이 소외되지 않도록 데이터를 관리하는 것은 우리 사회가 공정하게 인공지능을 활용하는 데에 있어 큰 역할을 할 것이다.

앞으로 인공지능의 인권침해 예방과 관련하여 더 논의해야 할 부분은 인공지능 이용자의 선택권 보장이다. 인공지능의 제안이 아닌 이용자 본인의 선택으로 인공지능 서비스를 이용할 수 있어야 한다. 즉, 서비스 제공 시 인공지능에 의한 서비스와 사람에 의한 서비스를 선택할 수 있어야 한다.

예를 들어, 의료분야에서 건강 진단 시 환자가 스스로 의사와 인공지능 중 어떠한 대상을 통해 진단받을지에 대한 결정을 내릴 수 있도록 의료사업자는 이용자에게 선택권을 보장해야 한다. 개인에게 미치는 영향이 큰 의료분야뿐만 아니라 인공지능을 활용하는 모든 분야에서는 이용자의 선택권을 보호하기 위한 고민을 해야 한다.

인공지능에 의해 이용자의 선택권이 축소된 사례 중 우리가 쉽게 찾아볼 수 있는 것의 온라인 음원사이트이다. 음원사이트가 소비자의 선택할 권리에 비중을 두었다면 인공지능이 추천하는 음악 대신 소비자가 자신이 원하는 음악을 찾아 나갈 수 있도록 도와주는 방법으로 화면을 구성했을 것이다. 하지만 음원사이트는 사용자의 성, 연령, 이전 재생목록 등 사용자의 프로파일을 활용한 음악 추천 메뉴를 사용자가 접근이 편한 위치에 배치하여 소비자의 선택보다는 자사 인공지능이 추천하는 음악을 들어보도록 유도한다.

음원사이트 이용자가 인공지능이 제시하는 음악목록을 선택하거나 자신이 원하는 음악을 검색할 때 동등한 조건에서 불편함 없이 결정할 수 있어야 이용자의 선택권이 보장된 것이라고 볼 수

있다. 만약 소비자가 원하는 메뉴를 선택할 수 있더라도 그 선택 과정을 복잡하게 만드는 등 접근성을 저하시키는 경우, 서비스 이용에 불편함이 생기고 이는 온전히 선택할 권리를 침해할 위험이 있다.

이용자의 선택권 보호와 관련하여 공정거래위원회는 맞춤형 광고 수신 여부를 이용자가 선택하도록 하는 방안을 제시했다. 공정개래위원회의 「전자상거래 등에서의 소비자보호에 관한 법률」 개정안에서 온라인사업자가 소비자에게 맞춤형 광고 수신 여부를 선택할 수 있도록 선택권을 부여하는 방안을 제시했다.

「전자상거래 등에서의 소비자보호에 관한 법률
전부개정법률(안) 입법예고

마. 정보의 투명성 확보조치 신설(안 제16조)

1) 검색결과·순위, 사용자 후기는 소비자의 선택에 절대적 영향을 미치므로, 기만적 소비자 유인 행위로부터의 합리적 선택권 보호를 위해 정보제공의 투명성을 확보할 필요가 있음.

2) 허위·과장·기만적 소비자유인행위를 사전에 차단하기 위해 재화등의 거래와 관련된 검색결과를 제공할 때 광고를 구분하여 표시하도록 하고, 검색순위를 결정하는 주요결정 기준을 표시하도록 의무화함. 또한 사업자가 이용후기 게시판을 사용하는 경우 이용후기의 수집·처리에 관한 정보를 공개하도록 함.

바. 맞춤형 광고 등 정보이용 시 고지의무 강화(안 제18조)

타겟형 광고 등 맞춤형 광고가 크게 늘어나고 있으나, 소비자는 이를 일반광고와 구분할 수 없어 합리적 선택을 제약받게 되므로, 맞춤형 광고 제공시 그 사실을 고지하도록 하고, 소비자가 맞춤형

광고를 거부할 경우 일반광고를 선택할 수 있도록 규정함.

위 개정안은 또한 이용자의 알권리 보장에도 큰 도움을 줄 수 있을 것으로 보인다. 그동안 많은 인공지능 서비스 제공업체들이 인공지능의 결정과 사람의 결정을 구분하지 않고 함께 배치하거나 제안하는 일이 있었다. 하지만 위와 같이 고지의무를 강화한다면 앞으로 어떠한 서비스가 인공지능에 의한 것인지 구분할 수 있게 된다.

2장에서 검토한 바와 같이 온라인 광고는 소비자의 프로파일을 이용하여 소비자와 연관된 광고를 송출한다. 물론 추천 광고가 소비자가 필요하다고 느낄 수도 있는 내용을 제시하여 흥미를 유발시키는 역할을 할 수도 있겠지만 이것이 반복되는 경우 예상하지 못한 부분에서 차별로 인한 피해가 발생할 수 있다. 따라서 이러한 선택권 보장은 소비자가 다양한 제품에 대한 정보를 얻고 자신이 원하는 제품(서비스)을 구매할 수 있도록 환경을 조성하는 것이 필요하다.

이렇게 자발적인 의사결정의 기회를 열어놓을 때 사람들은 과거 자신의 선택에 귀속되어 제자리에서 맴돌지 않고 새로운 선택을 할 수 있는 기회를 가질 수 있다. 수많은 인터넷 미디어 사업자의 알고리즘이 과거 이용기록을 활용한 연관 자료를 제시하며 자사 서비스의 연속적인 이용을 유도하고 있다. 물론 이용자 입장에서는 자신의 기호에 맞는 선택을 쉽게 할 수 있으므로 편리하다고 느끼겠지만 이는 사회적 양극화를 낳는 주요 원인이 되었다.

인권기반 인공지능 개발을 위해 다음으로 검토해야 할 사안은 규제의 다양화이다. 앞서 검토한 인공지능 관련 법률안은 규제의 일원화를 통해 관련 자원과 정보를 축적시켜 인공지능에 대한 정부 대응력을 높일 수 있는 것을 목표 했다. 하지만 인공지능이 사회 전 분야에 걸쳐 확산되고 있는 상황에서 소수의 기관이 모든

인공지능을 감시하고 통제하기에는 무리가 따를 수 있다. FDA의 Pre-Cert 프로그램 사례에서 보여준 바와 같이 해당 분야에 대한 전문성을 가진 정부기관이 인력을 확보하여 인권 침해를 예방한다면 시간과 비용이 상당히 소요될 것이다.

마지막으로 인권에 기반한 인공지능이 만들어지게 하기 위해서는 인공지능의 인권침해 예방과 관련된 사회적 논의가 지속적으로 확산되는 것이 필요하다. 국회의 입법과 정부의 정책 등 실질적인 대응 조치는 모두 사회적 논의에서부터 비롯된다. 이러한 논의 확산을 위해 사회 내에 다양한 주체를 세워서 인공지능에 의한 권리침해를 예방하는 문화를 확산해야 한다.

앞서 검토한 법률안에서 등장한 '민간자율인공지능윤리위원회'는 사회 곳곳에서 인권기반 인공지능 문화 확산에 큰 역할을 할 수 있을 것으로 보인다. 이러한 민간자율위원회가 관심을 가지고 사회 곳곳에서 활동하며 인공지능의 인권침해 예방 문화를 확산시킨다면 인공지능의 감시 주체가 급격히 증가하게 되므로 문제 해결에 큰 기여를 할 수 있다.

'민간자율인공지능윤리위원회'는 각 학교와 연구기관에서 연구윤리 준수를 위해 설치하고 있는 연구윤리심의위원회(Institutional Review Board: IRB)를 모델로 한 것으로 보인다. IRB는 모든 생명을 대상으로 과학 연구를 하는 경우 이에 대한 윤리적 타당성을 검토하고 연구 대상을 보호하기 위한 제도이다. IRB는 비단 의학이나 생명공학에서만 활용되는 것이 아니라 심리학과 같은 사회과학까지 모두 적용되는 폭넓은 제도이다.

IRB는 연구자 윤리지침 마련, 연구종사자 교육, 연구대상자 보호 대책 수립 등의 활동을 한다. 또한 IRB는 위원회가 승인한 연구에 대해 조사하고 감독하는 권한이 있으므로 해당 연구가 종료될 때까지 관여할 수 있다. 이는 인공지능에 대한 조사 및 감독에

도 적용 가능할 것으로 보인다. 인공지능이 인권에 기반하여 개발되고 운영되도록 기본 지침을 마련하고, 인공지능 개발자에게 관련된 교육을 실시하며, 인공지능 사용자를 위한 보도 대책을 수립한다면 인권에 기반한 인공지능 개발에서 주요한 역할을 담당할 수 있는 것이다.

기업이나 정부 등 인공지능을 통한 서비스를 제공하는 측의 입장에서는 인공지능을 활용하여 비용절감 및 업무효율성 제고를 달성하고자 한다. 하지만 이것이 이용자의 인권을 침해한다면 오히려 더 큰 비용을 발생할 가능성이 높다. 따라서 인권에 기반한 인공지능 개발 및 활용은 윤리적 측면은 물론 비용적 측면에서도 타당한 주장이다. 우리나라는 이미 사회적으로 이에 대한 논의가 활발하게 진행되고 있는 만큼 관련 제도가 마련될 것으로 보인다. 인공지능의 전문성에 있어 민간과 정부의 차이가 크므로 정부의 감시와 통제에는 목표 달성에 한계가 있다. 따라서 향후 관련 제도 마련 시 사회 전체적인 변화에 초점을 맞추는 것이 필요하다.

참고문헌

Ananny, Mike. “Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness.“ Science, Technology, & Human Values 41, no.1 (2016).

Anderson, Monica. “Social Media Conversations About Race.“ Pew Research Center (August 15, 2016). https://www.pewresearch.org/internet/2016/08/15/social-media-conversations-about-race/

Andrew Smith. “Using Artificial Intelligence and Algorithms.” (2020. 4. 8 .) . https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms

Caliskan, Aylin., Bryson, Joanna J., Narayanan, Arvind. “Semantics derived automatically from language corpora contain human-like biases.“ Science Vol. 356, Issue 6334, (April 14, 2017). https://science.sciencemag.org/content/356/6334/183.

Commissioner for Human Rights of The Council of Europe. “Unboxing Artificial Intelligence: 10 steps to protect Human Rights.” https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights

Council of Europe. “Algorithms and Human Rights.” Council of Europe Study DGI (2017) 12. (March 2018). https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.

Dastin, Jeffrey. “Amazon scraps secret AI recruiting tool that showed bias against women.“ Reuters (October 10, 2018). https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Datta, Amit., Tschantz, Michael Carl., Datta, Anupam. “Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination,“ Proceedings on Privacy Enhancing Technologies Volume 2015: Issue 1 (April 18, 2015). https://content.sciendo.com/configurable/contentpage/journals$002fpopets$002f2015$002f1$002farticle-p92.xml

Devlin, Hannah. “AI programs exhibit racial and gender biases, research reveals.” The Guardian, (April 13, 2017). https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals

Eaneff, Stephanie., Obermeyer, Ziad. and Butte, Atul J. “The Case for Algorithmic Stewardship for Artificial Intelligence and Machine Learning Technologies.” Journal of American Medical Association 2020;324(14).

European Union Agency for Fundamental Rights. “BigData: Discrimination in data-supported decision making.” FRA Focus. (2018). https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making.

European Union Agency for Fundamental Rights. “Data quality

and artificial intelligence." FRA Focus. (2019). https://fra.europa.eu/en/publication/2019/data-quality-and-artificial-intelligence-mitigating-bias-and-error-protect

Gibbs, Samuel. "Women less likely to be shown ads for high-paid jobs on Google, study shows." The Guardian (July 8, 2015). https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study.

Hardesty, Larry. "Explained: Neural networks." MIT News. (April 14, 2017). https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

IEEE SA. "A Call to Action for Businesses Using AI." https://ethicsinaction.ieee.org/#series.

Irons, Meghan E. "Caught in a dragnet." Boston Globe (July 17, 2011). http://archive.boston.com/news/local/massachusetts/articles/2011/07/17/man_sues_registry_after_license_mistakenly_revoked/.

The New York City Council Legislative Research Center. "Legislation." https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0

Powles, Julia. "New York City's Bold, Flawed Attempt to Make Algorithms Accountable." The New Yorker. December 20, 2017. https://www.newyorker.com/tech/elements/new-york-citysbold-flawed-attempt-to-make-algorithms-accountable

Princeton University. “Biased bots: Artificial-intelligence systems echo human prejudices.” https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices.

Qualcomm, “What is AI.” https://www.qualcomm.com/products/artificial-intelligence/what-is-ai-faq

Rashida Richardson, ed., “Confronting Black Boxes: A Shadow Report of the New York
City Automated Decision System Task Force,” AI Now Institute, December 4, 2019, https://
ainowinstitute.org/ads-shadowreport-2019.html.

Sample, Ian. “AI watchdog needed to regulate automated decision-making, say experts.“ The Guardian (January 27, 2017). https://www.theguardian.com/technology/2017/jan/27/ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions

Staff and agencies, The Guardian. “Microsoft ’deeply sorry’ for racist and sexist tweets by AI chatbot.” The Guardian, (March 26, 2016). https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot.

Staff, Reuters. “New Zealand passport robot tells applicant of Asian descent to open eyes.“ Reuters (December 7, 2016). https://www.reuters.com/article/us-newzealand-passport-error/new-z

ealand-passport-robot-tells-applicant-of-asian-descent-to-open-eyes-idUSKBN13W0RL.

Turque, Bill. "Creative ... motivating' and fired." The Washington Post (March 6, 2012). https://www.washingtonpost.com/local/education/creative—motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html

United States Congress, Tom Lantos Human Rights Commission. "Artificial Intelligence: The Consequences for Human Rights," accessed on October 15, 2020. https://humanrightscommission.house.gov/events/hearings/artificial-intelligence-consequences-human-rights

White House, Executive Office of the President. "Big data: seizing opportunities, preserving values." (Washington : 2014). https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

부 록

미 캘리포니아주의 안면인식 및 기타 생체 감시 규제 법 (AB-1215)

출처:

https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB1215

Date Published: 10/09/2019 09:00 PM

BILL START

Assembly Bill No. 1215

CHAPTER 579

An act to add and repeal Section 832.19 of the Penal Code, relating to law enforcement.

[Approved by Governor October 08, 2019. Filed with Secretary of State October 08, 2019.]

LEGISLATIVE COUNSEL'S DIGEST

AB 1215, Ting. Law enforcement: facial recognition and other biometric surveillance.

Existing law states the intent of the Legislature to establish policies and procedures to address issues related to the downloading and storage of data recorded by a body-worn camera worn by a peace officer, and requires that those policies and procedures be based on best practices. Existing law requires law enforcement agencies, departments, or entities to consider certain best practices regarding the downloading and storage of body-worn camera data when establishing policies and procedures for the implementation and operation of a body-worn camera system, as specified.

This bill would prohibit a law enforcement agency or law enforcement officer from installing, activating, or using any biometric surveillance system in connection with an officer camera or data collected by an officer camera. The bill would authorize a person to bring an action for equitable or declaratory relief against a law enforcement agency or officer who violates that prohibition.

The bill would repeal these provisions on January 1, 2023.

DIGEST KEY

Vote: majority Appropriation: no Fiscal Committee: no Local Program: no

BILL TEXT

THE PEOPLE OF THE STATE OF CALIFORNIA DO ENACT AS FOLLOWS:

SECTION 1.

The Legislature finds and declares all of the following:

(a) Californians value privacy as an essential element of their individual freedom, and are guaranteed a right to privacy in Section 1 of Article I of the California Constitution.

(b) Facial recognition and other biometric surveillance technology pose unique and significant threats to the civil rights and civil liberties of residents and visitors.

(c) The use of facial recognition and other biometric surveillance is the functional equivalent of requiring every person to show a personal photo identification card at all times in violation of recognized constitutional rights. This technology also allows people to be tracked without consent. It would also generate massive databases about law-abiding Californians, and may chill the exercise of free speech in public places.

(d) Facial recognition and other biometric surveillance technology has been repeatedly demonstrated to misidentify women, young people, and people of color and to create an elevated risk of harmful "false positive" identifications.

(e) Facial and other biometric surveillance would corrupt the core purpose of officer-worn body-worn cameras by transforming those devices from transparency and accountability tools into roving surveillance systems.

(f) The use of facial recognition and other biometric surveillance would disproportionately impact the civil rights and civil liberties of persons who live in highly policed communities. Its use would also diminish effective policing and public safety by discouraging people in these communities, including victims of crime, undocumented persons, people with unpaid fines and fees, and those with prior criminal history from seeking police assistance or from assisting the police.

SEC. 2.

Section 832.19 is added to the Penal Code, immediately following Section 832.18, to read:

832.19.

(a) For the purposes of this section, the following terms have the following meanings:

(1) “Biometric data” means a physiological, biological, or behavioral characteristic that can be used, singly or in combination with each other or with other information, to establish individual identity.

(2) “Biometric surveillance system” means any computer software or application that performs facial recognition or other biometric surveillance.

(3) “Facial recognition or other biometric surveillance” means either of the following, alone or in combination:

(A) An automated or semiautomated process that captures or analyzes biometric data of an individual to identify or assist in identifying an individual.

(B) An automated or semiautomated process that generates, or assists in generating, surveillance information about an individual based on biometric data.

(4) “Facial recognition or other biometric surveillance” does not include the use of an automated or semiautomated process for the purpose of redacting a recording for release or disclosure outside the law enforcement agency to protect the privacy of a subject depicted in the recording, if the process does not generate or result

in the retention of any biometric data or surveillance information.

(5) “Law enforcement agency” means any police department, sheriff’s department, district attorney, county probation department, transit agency police department, school district police department, highway patrol, the police department of any campus of the University of California, the California State University, or a community college, the Department of the California Highway Patrol, and the Department of Justice.

(6) “Law enforcement officer” means an officer, deputy, employee, or agent of a law enforcement agency.

(7) “Officer camera” means a body-worn camera or similar device that records or transmits images or sound and is attached to the body or clothing of, or carried by, a law enforcement officer.

(8) “Surveillance information” means either of the following, alone or in combination:

(A) Any information about a known or unknown individual, including, but not limited to, a person’s name, date of birth, gender, or criminal background.

(B) Any information derived from biometric data, including, but not limited to, assessments about an individual’s sentiment, state of mind, or level of dangerousness.

(9) “Use” means either of the following, alone or in combination:

(A) The direct use of a biometric surveillance system by a law enforcement officer or law enforcement agency.

(B) A request or agreement by a law enforcement officer or law enforcement agency that another law enforcement agency or other third party use a biometric surveillance system on behalf of the requesting officer or agency.

(b) A law enforcement agency or law enforcement officer shall not install, activate, or use any biometric surveillance system in connection with an officer camera or data collected by an officer camera.

(c) In addition to any other sanctions, penalties, or remedies provided by law, a person may bring an action for equitable or declaratory relief in a court of competent jurisdiction against a law enforcement agency or law enforcement officer that violates this section.

(d) This section does not preclude a law enforcement agency or law enforcement officer from using a mobile fingerprint scanning device during a lawful detention to identify a person who does not have proof of identification if this use is lawful and does not generate or result in the retention of any biometric data or surveillance

information.

(e) This section shall remain in effect only until January 1, 2023, and as of that date is repealed.

[부록] 미 캘리포니아주의 안면인식 및 기타 생체 감시 규제 법(AB-1215)

[비공식 번역문]

발행일: 2019년 10월 9일 오후 9시

의회 법안 제1215호

제 579 장

법률집행에 관한 형법전(Penal Code) 제832.19조의 추가 및 폐지를 위한 법

[2019년 10월 8일 주지사 승인. 2019년 10월 8일 주무장관 앞 제출]

법률고문인 주요요약서

(LEGISLATIVE COUNSEL'S DIGEST)

의회 법안 제1215호(AB 1215), Ting 의원. 법률집행: 안면 인식 및 기타 생체 감시.

기존의 법은 보안관이 착용한 카메라로 기록되는 데이터의 다운로드 및

저장에 대한 사안들을 해결하기 위한 방침 및 절차를 수립하고자 하는 법률의 취지를 담고 있으며, 이러한 방침 및 절차는 최선의 관행(best practices)에 기반해야 함을 요구하고 있다.

기존의 법은 인체 착용 카메라 시스템의 실행 및 운영에 관한 방침 및 절차를 수립할 시 사법당국과 법 집행부 등 유관기관들이 인체 착용 카메라 데이터의 다운로드 및 저장에 관한 최선의 관행을 고려해야 할 것을 요구한다.

이 법안은 법 집행관의 카메라 또는 해당 카메라로 수집된 데이터와 관련하여 사법당국 또는 법 집행관이 여하한 생체 감시 시스템을 설치, 작동, 사용하는 것을 금지시키고자 한다. 이 법안은 동 금지 규정을 위반하는 사법당국이나 집행관을 상대로 개인이 공정한 구제(equitable relief) 또는 선언적 구제(declaratory relief)를 위한 소송을 제기할 수 있는 권한을 부여하고자 한다.

이 법안은 2023년 1월 1일자로 관련 조문을 폐지하고자 한다.

핵심요약

의결: 과반수　예산안: 없음　연간위원회: 없음　지역프로그램: 없음

법안 내용

캘리포니아 주민들은 다음과 같이 제정한다.

제 1 조.

이 법률은 이하 전문을 확인 및 선언한다:

(a) 캘리아포니아인들은 프라이버시를 개인의 자유를 구성하는 핵심 요소로서 가치 있게 여기며 캘리포니아 헌법 제I장 제1조에 따른 프라이버시 권리를 보장받는다.

(b) 안면 인식 및 기타 생체 감시 기술은 거주자 및 방문객들의 시민권과 시민적 자유에 독특한 방식으로 상당한 위협을 가한다.

(c) 안면 인식 및 기타 생체 감시 기술의 사용은 명시된 헌법적 권리를 위반하여 모든 이들에게 수시로 개인 사진이 부착된 신원확인카드를 요구하는 것에 상응하는 기능을 한다. 또한 이 기술은 동의 없이 사람들을 추적할 수 있다. 나아가 이 기술은 법을 준수하는 캘리포니아인들에 대한 대량의 데이터베이스를 생성할 수 있으며 공공장소에서의 자유발언권 행사를 어렵게 만들 수 있다.

(d) 안면 인식 및 기타 생체 감시 기술은 여성, 청년 및 유색인종을 오인하는 경우가 있으며 개인에게 치명적일 수 있는 “잘못된 일치(false positive)” 신원확인 결과의 위험을 높인다는 주장이 반복적으로 제기되

고 있다.

(e) 안면 인식 및 기타 생체 감시는 집행관이 신체에 착용한 카메라를 투명성 및 책임성을 위한 도구가 아닌 무작위 감시 시스템으로 오용하여 해당 카메라 사용의 주요 목적을 퇴색시킬 수 있다.

(f) 안면 인식 및 기타 생체 감시 기술의 사용은 치안이 매우 양호한 지역에 거주하는 사람들의 시민권 및 시민적 자유에 불균형적으로 영향을 미칠 수 있다. 또한 이 기술의 사용은 이러한 지역의 사람들(범죄 피해자, 밀입국자, 벌금미납자, 전과기록자 포함)이 경찰의 도움을 받거나 경찰을 지원하도록 할 수 있는 사기를 꺾어 효과적인 치안 유지 활동 및 공중의 안전을 저해할 수 있다.

제 2 조.

형법전 제832.18조의 후속 조항으로 제832.19조가 다음의 내용으로 추가된다:

832.19.

(a) 이 조항의 목적 상, 이하의 용어는 다음과 같은 의미를 갖는다:

(1) “생체 데이터(biometric data)”란 단독으로 혹은 다른 생체 데이터와 함께 또는 기타 정보와 함께 개인의 신원(identity)을 확인하는 데에 사용될 수 있는 생리적, 생물학적 또는 행동적 특징을 의미한다.

(2) “생체 감시 시스템(biometric surveillance system)”이란 안면 인식 또는 기타 생체 감시를 수행하는 컴퓨터 소프트웨어나 어플리케이션을

의미한다.

(3) “안면 인식 또는 기타 생체 감시(facial recognition or other biometric surveillance)”란 단독으로 혹은 혼합하여 다음 중 하나를 의미한다:

(A) 개인의 신원을 확인해주는 또는 신원확인에 도움을 주는 한 개인의 생체 데이터를 수집 또는 분석하는 자동화된 혹은 반자동화된 프로세스.

(B) 생체 데이터에 기반하여 개인에 관한 감시 정보를 생성하는 또는 생성에 도움을 주는 자동화된 혹은 반자동화된 프로세스

(4) “안면 인식 또는 기타 생체 감시”는 사법당국 이외의 배포 혹은 공시의 경우 기록물에 묘사된 대상의 프라이버시를 보호하고자 기록을 삭제하기 위한 목적으로 사용되는 자동화된 혹은 반자동화된 프로세스를 포함하여 의미하지 아니한다. 단, 해당 프로세스가 어떠한 생체 데이터나 감시 정보를 생성하지 아니하고 생성물을 보관하는 결과를 가져오지 않는 경우에 한한다.

(5) “사법기관(law enforcement agency)”이란 경찰서, 보완관 부서, 지방검사, 자치주 보호관찰부, 교통기관경찰국, 학군 경찰부서, 고속도로 순찰대, 캘리포니아대학(University of California), 캘리포니아주립대학(California State University) 또는 커뮤니티 대학의 캠퍼스 경찰부서, 캘리포니아 고속도로 순찰부서 및 사법부를 의미한다.

(6) “법 집행관(law enforcement officer)”이란 사법기관의 임원, 부대리인, 직원 또는 에이전트를 의미한다.

(7) “법 집행관의 카메라(officer camera)”란 법 집행관의 신체나 의복에 장착되어 있거나 법 집행관이 소지하고 다니는 것으로서 이미지 혹은 사운드를 기록 또는 전송해주는 신체 착용 카메라 또는 그와 유사한 장치를 의미한다.

(8) “감시 정보(surveillance information)”란 단독으로 혹은 혼합하여 다음 중 하나를 의미한다:

(A) 알려지거나 알려지지 않은 개인에 관한 여하한 정보를 의미하며 개인의 성명, 생일, 성별 또는 범죄기록을 포함하나 이에 한정하지 아니한다.

(B) 생체 데이터로 축출된 여하한 정보를 의미하며 개인의 정서, 정신상태 또는 위험도에 관한 평가/측정을 포함하나 이에 한정하지 아니한다.

(9) “사용(use)”이란 단독으로 혹은 혼합하여 다음 중 하나를 의미한다:

(A) 법 집행관이나 사법기관에 의한 생체 감시 시스템의 직접적인 사용.

(B) 또 다른 사법기관이나 제3자가 해당 요청을 한 집행관이나 에이전트를 대신하여 생체 감시 시스템을 사용할 수 있도록 하는 법 집행관이나 사법기관의 요청 또는 동의.

(b) 사법기관 또는 법 집행관은 법 집행관의 카메라 또는 법 집행관의 카메라로 수집한 데이터와 관련하여 어떠한 생체 감시 시스템도 설치하거나, 작동시키거나, 사용하지 아니한다.

(c) 법에서 명시한 기타 제재, 벌금 또는 구제에 더하여, 이 조항을 위반하는 사법기관이나 법 집행관을 상대로 개인이 공정한 구제(equitable

relief) 또는 선언적 구제(declaratory relief)를 위한 소송을 관할법원에 제기할 수 있는 권한을 부여하고자 한다.

(d) 이 조항은 사법기관이나 법 집행관이 합법적 구금을 집행하는 가운데 신원 증빙이 없는 사람의 신원 확인을 위하여 모바일 지문 인식기기를 사용하는 것을 금지하지는 아니한다. 단, 해당 사용은 합법적이어야 하며 어떠한 생체 데이터나 감시 정보를 생성하지 아니하고 생성물을 보관하는 결과를 가져오지 않는 경우에 한한다.

(e) 이 조항은 2023년 1월 1일까지 유효하며 해당 일자로 폐지된다.