발 간 등 록 번 호
11- 162000- 000885- 01

2021년 일반과제 실태조사
연구용역보고서

인공지능(AI) 개발과 활용에서의 인권 가이드라인 연구
(최종보고서)

2021. 11.

인공지능(AI) 개발과 활용에서의 인권 가이드라인 연구 (최종보고서)

2021년 국가인권위원회 일반과제 실태조사
연구용역 최종보고서를 제출합니다.

2021. 11. .

연구수행기관 : 사단법인 정보인권연구소
연구책임자 : 김기중 (사단법인 정보인권연구소 이사)
공동연구원 :
오정미 (법무법인 이공 변호사)
오철우 (대구경북과학기술원 기초학부 겸임교수)
장여경 (사단법인 정보인권연구소 이사)
전치형 (KAIST 과학기술정책대학원 교수)
보조연구원 : 김민 (진보네트워크센터 활동가)

이 보고서는 연구용역수행기관의 결과물로서,
국가인권위원회의 입장과 다를 수 있습니다.

목 차

표 목차

그림 목차

【 요 약 】

1. 서론

최근 몇 년간 공공기관과 민간 기업에서 인공지능을 기반으로 하는 신기술의 도입과 사용이 크게 증가하였다. 이미 인공지능은 인터넷과 모바일 서비스의 추천 알고리즘 등으로 일반시민의 일상생활은 물론 소상공인의 생계에 큰 영향을 미치고 있으며, 채용, 노동, 금융, 행정, 복지, 치안 등 사회 전반에서 완전히 또는 부분적으로 인공지능에 의해 내려진 조치나 결정이 인간의 기본적 권리와 삶에 직접적인 영향을 미치고 있다. 특히 완전히 자동화된 행정처분과 전자정부서비스를 규정한 제정 「행정기본법」과 개정 「전자정부법」의 시행으로 공공부문 의사결정에서 인공지능의 관여와 역할이 커질 것으로 예상되고 있다.

인공지능의 발전과 확산은 생산성·편의성을 높여 국가경쟁력과 삶의 질을 높일 것으로 기대되지만, 개인정보 및 사생활 침해와 차별 등 기본적 인권을 침해하는 문제들도 제기되고 있다.

특히 인공지능이 막대한 양의 데이터를 분석하고 이들 간의 연결 관계를 식별함으로써 사생활 침해가 발생할 수 있으며, 데이터셋에 내재된 편향성에 영향을 받아 여성과 남성, 백인과 흑인, 혹은 내국인과 외국인 집단 간에 차별적인 결과를 생성하는 문제가 발생하기도 한다. 업계 일부에서는 인간에게도 편견이 있다며 인공지능이 인간사회의 차별을 학습했을 뿐이라고 주장하지만, 유럽연합 <인공지능 백서>는 인공지능 의사결정에서 작용하는 편견의 경우 사회적 통제 메커니즘이 없으면 훨씬 더 많은 사람들에게 장기간 영향을 준다고 지적하였다.

그러나 인공지능으로 영향을 받는 당사자들은 인공지능의 도입, 운영, 결정에 대하여 의견 제시와 참여의 기회를 보장받고 있지 못하다. 또한 인공지능 분야에 대한 법률적·제도적 규제 장치가 부족한 상황에서 부분적이고 개별적으로 이루어지는 소관부처별 접근 방식으로는 인공지능으로 인한 인권침해에 대하여 효과적인 권리구제를 보장하지 못한다.

최근 유엔 인권이사회는 신기술이 갖는 영향력, 기회, 도전 과제에 대응하는 인권의 증진 및 보호에 있어 전체적, 포용적, 포괄적 접근방식(holistic, inclusive and comprehensive approach)이 필요하다고 강조하고 있다. 세계 각국이 인공지능을 규율하는 법제도를 적용 중이거나 도입을 준비 중이라는 사실 또한 매우 시사적이다.

본 연구는 인공지능 관련 인권적 이슈를 조사 정리 및 분석하고, 정책적으로 적용가능한 인권기준 및 제도적 방안을 모색하고자 한다.

2. 인공지능 도입 현황

가. 공공영역의 인공지능 활용

1) 디지털 뉴딜과 스마트 시티

2020년 7월 정부가 발표한 디지털 뉴딜의 대표적 과제로 '데이터 댐'과 '공공데이터 개방·활용 활성화' 사업을 들 수 있다. 해당 사업을 위해 과학기술정보통신부와 지능정보사회진흥원은 2021년 6월 인공지능 허브 홈페이지를 통해 헬스케어, 비전, 안전, 자율주행 등 8대 분야 170종의 데이터 4억 8천만 건을 개방했다. 또한 '스마트 의료 및 돌봄 인프라 구축' 과제는 '건강취약계층 스마트 건강관리'를 포함하고 있다. 이는 2025년까지 노인, 장애인 등 건강취약계층 12만 명을 대상으로 사물인터넷(IoT), 인공지능을 활용한 디지털 돌봄 시범사업을 추진하며 사물인터넷 센서, 인공지능 스피커를 통해 노인과 장애인 등 대상자의 맥박·혈당·활동 등을 감지하고 말벗·인지기능을 지원하는 한편 돌봄로봇 4종을 개발하겠다는 계획이 포함되어 있다.

그러나 이렇게 수집·공유된 데이터의 상당수는 개인의 동의 하에 연출된 일부 데이터를 제외하고는 대부분 실제 환경에서 수집된 데이터로서, 대규모 학습용 데이터에 국민의 개인정보나 민감정보가 포함되어 유출될 수 있다는 우려가 있다. 특히 헬스케어 분야 데이터나 건강취약계층을 대상으로 한 디지털 돌봄 사업에서 수집되는 데이터의 경우 질병정보 등 실제 건강에 관한 정보를 포함하고 있어 사생활 침해 위험성이 큰 민감정보이다. 기업이 국가를 통해 국민의 민감정보를 무료로 제공받아 다시 이익을 창출하는

게 맞지 않다는 지적도 제기된다. 정부는 배포 데이터에 대하여 개인정보가 포함되지 않도록 비식별화 처리를 거쳐 가공된 데이터라고 설명하지만, 가명 데이터는 다른 정보와의 결합 등을 통해 정보주체를 식별할 수 있는 가능성이 완전히 사라지지 않는 이상 정보인권 침해에 대한 우려가 있을 수밖에 없다.

자율주행, 안전 분야 데이터 또한 지하철 역사, 도로, 버스 내부와 같은 공공장소 CCTV로부터 수집된 영상을 가공하여 배포하고 있는데 이 역시 정보주체의 동의나 고지 없이 가공된 것이다. 이러한 영상 정보의 경우 사람 얼굴과 차량 번호판 등의 개인정보를 모자이크 또는 블러링 처리하는 방식으로 비식별화하였다고 하지만, 딥러닝 기술의 발전으로 추후 얼굴이 검출되거나 인식되는 문제가 발생할 수 있다. 무엇보다 샘플데이터 등을 불특정 다수에 공개하거나 과학적 연구 목적의 범위를 벗어난 민간 사업에 가명정보를 공개하는 것은 개인정보보호법에 위배된다.

도시 공간의 디지털 혁신 과제는 비용절감, 생산성 향상 등 도시의 새로운 발전양상에 대한 기대와 동시에 보안 위협과 사생활 침해, 신기술 위주 정책으로 인한 불평등한 정보접근성 등의 문제점을 안고 있다. 특히 한국에서 진행 중인 스마트 시티 논의는 특정한 도시 문제 해결보다는 기술을 보유한 민간이 주체가 되어 다양한 인프라 간 데이터 공유를 추구하려는 성격이 강해 그 우려가 커질 수밖에 없다. 스마트 시티의 계획과 추진은 일반 감시 카메라보다 고도화된 스마트 폴(Pole), 스마트 보도 등 지능화 기기의 설치를 통해 방대한 시민들의 개인정보를 수집하는 방향으로 전개되고 있지만 개인정보보호 권리 침해에 대한 검토가 제대로 이뤄지지 않고 있다.

2) 법무부 얼굴인식 시스템

법무부는 특정인의 얼굴과 지문을 입국 과정에서 수집된 외국인의 생체정보와 비교·분석하는 생체인식 정보 시스템을 운영하고 있으며, 최근에는 출입국 심사 과정 및 공항에서 위험인물을 실시간으로 식별하기 위한 얼굴인식 시스템을 구축하는 과정에서 수억 명의 생체인식 정보를 활용하고 있다.

출입국관리법을 통하여 입국하는 모든 외국인으로부터 수집한 얼굴과 지문 등 생체정보가 별도의 법적 근거 없이 얼굴인식정보 등 생체인식정보로 가공되어 수사지원 등 출

입국관리 외 목적으로 활용되어 왔다. 수집된 생체인식 정보를 2019년부터 정보주체 당사자 동의 없이 공항 내 실시간 얼굴인식 시스템 구축을 위한 학습용 데이터로 활용하여 왔다. 또한 법무부가 이렇게 보유한 생체인식 정보는 사업 참여 민간기업 다수에 학습용 데이터로 제공하는 등 목적 외로 사용되고 있는 실태가 심각하다. 법무부가 보유한 생체인식 정보, 즉 출입국 관리를 통해 수집한 외국인의 얼굴 사진은 약 2억에서 4억장에 달하는 것으로 확인되는데 2020년 사업을 진행하며 이중 최소 1.2억장이 이미 참가기업에게 제공되어 얼굴인식 알고리즘의 학습과 검증에 이용되었다. 또한 얼굴 사진을 포함한 내국인의 출입국 심사정보도 인공지능 실증랩 데이터베이스에 이관되어 이용된 것으로 확인되며, 더불어 공항 출입국 관리 구역에 일련의 카메라 시스템을 구축해 인공지능 학습을 위한 얼굴인식 및 행동인식 용도의 실제 데이터도 수집하고 있다.

3) 과학기술정보통신부와 공공분야 지능정보화

공공서비스에 대한 신기술을 지원하는 과학기술정보통신부의 ‘디지털 공공서비스 혁신 프로젝트’는 재난 시뮬레이션 모의훈련 등 일반적인 자동화와 행정처리의 고도화 사업이 주를 이루지만, 취약계층을 대상으로 부정수급을 감시하는 서비스, 생체인식정보 등 특별한 보호가 필요한 민감정보를 광범위하게 이용하는 서비스, 여러 나라에서 고위험 또는 금지되는 인공지능으로 규제되는 공공장소 원격 생체인식 서비스 등이 포함되어 있다.

이를테면 인공지능을 활용해 특정한 부정수급의 양상을 예측하고 확인하여 적발하는 사회보장정보원 사업은 불투명하고 위법한 개인정보 처리를 수반할 수 있다. 특히 결과 예측이나 오작동을 수정하기 쉬운 단순 행정자동화에 비해, 행정처리가 어떤 과정을 거쳐서 특정한 결론에 도달했는지 설명 여부가 쉽지 않은 불투명한 인공지능 시스템의 경우 처분의 당사자인 시민이 설명을 들을 수 있는 권리나 구제를 받을 권리를 보장받지 못하는 등 헌법상 적법절차의 원칙에 어긋날 수 있다. 이러한 인공지능 적발시스템이 실제 복지를 위해 유용하게 기여하는가에 대한 문제제기 또한 있을 수 있다.

경기도 부천시의 지능형 역학시스템 사업은 얼굴인식을 기반으로 한 추적 인공지능 시스템이 연계된 CCTV를 통해 확진자의 동선을 추적하고 마스크 미착용자나 확진자와

2M 이내 접촉한 자를 자동으로 파악하고 그 휴대전화 기지국 접속정보를 통해 신원을 확인하려는 목적을 가지고 있다. 나아가 교통과 방범 등의 목적으로 설치 및 운영하는 CCTV를 통해 수집되는 부천시민의 실제 영상 데이터를 민간 인공지능 학습용 데이터로 사용하겠다는 계획도 포함하고 있다. 그러나 현행법 상 민감정보로 정의된 생체인식 정보인 얼굴인식정보가 정보주체의 동의 없이 처리되는 것이 위법은 아닌지 의문이다.

과학기술정보통신부는 이처럼 공공개발사업을 통하여 방대한 국민 개인정보를 보유하고 있는 여러 공공기관에서 공공장소 얼굴인식 및 행동인식 인공지능 시스템을 개발 및 배치하도록 지원하고 있다. 이들 사업은 CCTV 등을 통해 확보한 국민의 실제 데이터를 생체인식정보로 가공한 후 사업 참여 민간기업 다수에 학습용 및 검증용으로 제공하였거나 제공할 계획이 있다는 공통점이 있다. 해당 사업들은 보안 및 데이터 외부 유출을 막기 위해 '실증랩' 공간에서만 데이터에 접근하도록 조치를 취했지만, 대구 수성구 사업에 참여하는 민간기업이 주민 얼굴 영상 10만여 건을 실증랩 밖으로 무단 반출하였다는 사실이 밝혀져 논란이 되었다.

4) 경찰청 얼굴인식과 범죄예측 시스템

경찰청은 얼굴인식 기술을 통해 CCTV, 블랙박스, 채증 장비 등의 다양한 영상정보 처리기기에서 수집된 영상을 적극 활용하고자 얼굴인식 시스템을 개발하고, 기존 구속피의자의 사진 자료를 다양한 각도로 인식할 수 있도록 3D로 변환시켰다. 2017년에는 전국 경찰서에 3D 영상 촬영 시스템을 보급해 구속피의자에 대한 3D 촬영을 시작하였다. 이 구속피의자 3D 데이터베이스의 경우 2018년 18만 명, 2019년 19만 명 이상의 3D 얼굴 사진이 포함되어있는 것으로 밝혀졌다. 경찰은 이후 매년 시스템을 고도화여 왔으며, 2024년에는 얼굴인식 시스템을 실시간으로 CCTV와 연계하여 검색하려고 계획하고 있다. 현재 경찰 등 수사기관이 3D 등 얼굴인식 데이터베이스를 구축하고 이를 자동으로 비교·대조할 수 있는 시스템을 운영하도록 허용하는 법적 근거는 없다. 공개된 장소에 설치된 공공기관 CCTV만 최소 100만 대에 이르는 등 영상을 통한 감시망이 촘촘한 환경을 고려해 보았을 때, 아무런 통제 장치 없이 얼굴인식 기술이 도입된다면 다수의 정보주체 시민이 공공장소에서 은밀한 대량 감시의 대상이 되는 등 정보인권이 침해될 수

있다.

한편 경찰청은 치안·공공데이터를 통합한 빅데이터를 인공지능으로 분석하여 지역별 범죄위험도와 범죄발생 건수를 예측하는 '범죄위험도 예측분석 시스템(Pre-CAS)'을 개발하여 전국에서 운영하고 있다. 그러나 데이터 기반 기술을 활용한 범죄 예측에는 상당한 위험성과 차별의 가능성이 뒤따른다. 경찰청은 미국의 범죄 예측 프로그램 프레드폴(PredPol)을 참조하여 한국의 치안 환경 특성에 맞춰 구축하는 것을 목표로 해당 시스템을 개발하였는데, 프레드폴은 유색인종과 빈곤한 거주 지역에 대한 편향성 문제와 사생활의 권리 침해로 이미 많은 비판을 받아온 시스템이다. 한국의 범죄예측 시스템 또한 이러한 위험성을 피해갈 수 없을 것으로 보인다.

한편 2018년 8월, 서울시 민생사법경찰단은 불법대부, 다단계 판매와 같은 민생범죄 수사에 '인공지능 수사관'을 도입하였다가 개인정보보호위원회의 위법성 지적 이후 이를 철회한 바 있다. 개인정보보호위원회는 민생사법경찰단이 해당 시스템을 이용해 인터넷 SNS 등 온라인에 공개된 게시물을 광범위하게 수집하고, 범죄 관련성이 높다고 판단되는 게시물을 분석해 해당 게시물에 포함된 성명, 아이디, 전화번호, 주소, 업체명 등을 이용한 것은 사실상 구체적 법령에 근거하지 않고 이뤄지는 온라인 불심검문과 유사한 것으로 판단하였다.

5) 군사안보 분야의 인공지능 활용과 개발

국방부는 무기체계의 지능화, 훈련체계의 고도화, 스마트한 병영환경 조성 등 크게 세 가지 분야에서 스마트 국방혁신을 추진하고 있다. 그러나 일부 사업과 정책은 장병의 개인정보 침해, 자율살상무기로의 발전 등 우려되는 지점이 보인다.

특히 2020년부터 진행 중인 육군의 '스마트 부대 구축사업' 및 '지능형 출입통제 체계 구축사업'과 해군의 '스마트 전투함 사업'은 효율적인 부대 지휘통제와 감시를 목적으로 장병의 심박수 등 건강정보 및 위치정보를 실시간으로 수집하고 처리하는 내용으로 논란이 되었다. 육군은 부대 생활관 복도에 실시간 얼굴인식 등이 가능한 지능형 CCTV 설치 계획을 밝히기도 했다.

한편, 한국은 이미 준자율 수준의 무인 무기는 물론이고, 목표물을 감지하면 자동으로

조준해서 사격할 수 있는 자동발사기능을 탑재하여 자율살상무기로 분류되는 센트리건(Sentry Gun)을 비무장지대에 배치하는 등 인공지능의 무기화를 빠르게 추구하는 것으로 보인다. 동시에 한국 정부는 인공지능 기술이 적용된 무기체계인 '자율살상무기체계(LAWS)'에 관한 국제사회의 논의에서 사실상 규제에 반대하는 태도를 보이고 있어 비판을 받고 있다.

나. 민간영역의 인공지능 활용

1) 인공지능 채용 도구

2019년 조사 자료에 의하면 인공지능을 활용해 신규 채용을 진행하고 있는 기업이 11.4%, 활용 계획이 있는 기업이 10.7%였으며, 공공기관의 경우 350곳의 공공기관 중 최소 37곳이 인공지능을 활용해온 것으로 확인된다.

그러나 인공지능과 데이터 기반 기술이 과거의 누적된 차별로 대표성이 충분치 않은 데이터에 의존할 경우 차별이 모방되고 확산될 위험이 있다. 고용률과 임금 격차에 대한 노동 인구 데이터에는 성별, 성적 지향과 성 정체성, 장애, 나이, 인종과 민족, 학력, 지역 등 다양한 요소에 기반하여 역사적으로 누적되어 온 오랜 기간의 차별이 반영되어 있는데 이러한 편향성은 인공지능 기술이 불평등과 차별이 반영된 데이터를 학습하는 결과로 이어질 수 있다. 대표적인 사례로 2018년 인공지능을 활용한 채용 시스템을 폐기한 전자상거래 기업 아마존의 경우를 들 수 있다. 남성 비율이 높은 기술 업계의 데이터를 기반으로 만들어진 아마존의 채용 시스템이 여성과 관련된 키워드를 자동으로 감점 요소로 분류하는 등 차별적 결정을 내렸다는 사실이 밝혀진 것이다. 채용과 노동은 사람들의 미래와 현재의 경제 상황에 직접적인 영향을 미칠 수 있기에 불투명한 인공지능 기술을 도입하는 것이 특히 위험한 영역이라고 할 수 있다.

채용 과정에서 성별, 연령, 신체조건, 용모, 출신지역 등으로 차별하지 말 것을 규정한 채용 공정화 관련 법제도를 인공지능을 활용한 채용 도구가 잘 준수하고 있는지에 대해 점검이 필요해 보인다. 국회 국정감사에서 인천공항공사는 인공지능 면접을 도입할 때 지원자 평가방법과 알고리즘에 대한 기술적 검토 또는 외부 자문을 거치지 않았다는 사

실이 드러났으며 한국방송통신전파진흥원은 자기관이 탈락시킨 지원자에 대하여 인공지능 면접이 어떤 기준을 적용하였는지 파악하고 있지 못한 것으로 나타났다.

2) 금융 서비스

ICT서비스를 통해 소비자로부터 다양한 데이터를 수집해온 기술기업부터 전통적인 대형 금융사까지, 금융 영역 전반적으로 비금융정보를 이용하는 새로운 신용평가 모델이 도입되고 있다. 이러한 '대안신용평가'는 기존의 신용평가사가 개인의 대출, 카드, 연체 등 이력을 이용해 책정한 신용등급에서 한 발 나아가 스마트폰 등을 통해 수집되는 수많은 디지털 기록 및 비금융정보를 알고리즘으로 분석하는 등 개인을 프로파일링하여 신용등급을 내리는 것으로 정의된다.

대안신용평가는 제도권 금융에서 소외된 취약계층의 신용도를 평가할 수 있다는 장점이 있으나 비금융정보의 해석 과정에서 오히려 취약계층을 배제하는 방향으로 악용될 수 있다. 주로 활용되는 비금융정보로는 통신정보, 결제정보, 휴대폰 사용기록이 있다. 그러나 이들 비금융정보가 실제 사용자의 연체율과 유의미한 상관관계가 있는지, 합리적인 평가가 이뤄질 수 있는지 의문이 제기된다. 개인이 민감하게 생각하는 기록까지 축적해 분류하는 만큼 사생활 침해 위험성이 커지고, 디지털 기록이 남는 모든 활동을 평가 대상으로 하여 개인의 표현의 자유가 위축되는 등 일상생활에 지장이 생길 위험도 있다.

또한 부정확한 데이터의 사용으로 잘못된 평가로 이어질 수 있으며, 정확하게 기록된 데이터를 사용할지라도 평가 대상자는 물론이고 사용자에게도 불투명하게 작동하는 인공지능 프로파일링이 예측할 수 없던 결과나 의도치 않은 차별을 야기할 수도 있다. 민감한 데이터의 직접적인 이용을 배제하더라도 다른 대리지표 데이터를 통하여 결과적으로 차별적 속성을 기준으로 한 평가를 낳을 수 있다. 미국의 연방거래위원회는 대출 상환 기록이 저조한 사람들과 같은 상점을 이용한 방문자들의 경우 동반하여 신용한도가 낮아진다는 증거를 찾아낸 바 있다.

3) 플랫폼 노동

플랫폼 노동 기업은 노동 통제를 위해 인공지능 기술을 적극 이용하고 있다. 주요 배달 플랫폼들은 인공지능 배차의 알고리즘이 빠르고 효율적인 배차를 이뤄낸다고 주장하지만, 당사자인 배달 노동자들은 알고리즘이 노동통제 및 사고 발생과 불공정한 배달료로 인한 임금삭감의 문제를 유발한다고 반박하고 있다. 인공지능 시스템을 통해 픽업 위치와 배달지역을 일방적으로 정해주는 ‘AI추천배차’ 방식의 경우, 교통상황과 실제 환경을 무시한 인공적 배달 경로를 제시해 배달노동자의 사고 발생 가능성을 높이고 있다는 점과, AI추천배차를 거부하면 경고, 계정 정지, 평가 하락, 배달료 차감 등 불이익을 주는 점 또한 논란이 되고 있다.

한편 플랫폼 기업의 인공지능 활용은 광범위한 이용자 데이터 축적과 불투명하고 독점적인 데이터 사용으로 이어져 노사 관계를 더욱 불균형하게 만든다는 지적을 받고 있다. 배달노동자와 소비자, 음식점주는 자신의 활동 결과인 데이터에 대한 접근은 커녕 어떠한 종류의 개인정보가 수집되는지 파악하기조차 힘든 것에 비해, 플랫폼 기업은 모든 데이터를 독점하고 이를 통해 주문비용, 중개비용, 배달비용에 대한 기준을 정하면서도 인공지능 시스템의 뒤에 숨어 사용자성을 감추고 있다는 것이다.

4) 챗봇 이루다 논란

‘이루다’는 2020년 12월 출시된 페이스북 메신저 기반의 인공지능 챗봇 서비스로, 일부 남성 이용자들이 이루다를 성적 대상으로 취급하고, 발화내용에 여성・성소수자・장애인・흑인을 혐오하는 내용이 포함되어 인공지능 윤리와 차별 논란을 빚었다. 무엇보다 개인정보보호법 위반 논란이 크게 일어 개인정보보호위원회는 이루다 개발사 (주)스캐터랩에 대한 조사를 실시하고 확인된 위법 행위에 대하여 총 1억 330만원의 과징금과 과태료 등을 부과하였다.

스캐터랩은 자사의 과거 서비스인 ‘텍스트앳’과 ‘연애의 과학’에서 수집한 카카오톡 대화를 자사 다른 서비스인 이루다의 인공지능 개발과 운영에 이용하였다. 그 과정에서 카카오톡 대화에 포함된 이름, 휴대전화번호, 주소 등의 개인정보를 삭제하거나 암

호화하는 등의 조치를 전혀 하지 않고 약 60만 명에 달하는 이용자의 카카오톡 대화문장 94억여 건을 이용하였고, 이루다 서비스 운영에서는 20대 여성의 카카오톡 대화문장 약 1억 건을 응답 DB로 구축하고 이용하였다.

보호위원회는 스캐터랩이 텍스트앳과 연애의 과학 개인정보 처리방침에 '신규 서비스 개발'을 포함시켜 이용자가 로그인함으로써 동의한 것으로 간주한 것으로는, 이용자가 이루다 개발과 운영 목적의 이용을 예상하거나 동의하였다고 보기 어렵다고 보았으며 결국 수집 목적을 벗어나 개인정보를 이용하였다고 판단하였다. 또한 스캐터랩이 깃허브(Github)에 카카오톡 대화문장 1,431건과 함께 인공지능 모델을 게시한 것에 대하여는, 가명정보를 불특정 다수에게 제공하면서 '특정 개인을 알아보기 위하여 사용될 수 있는 정보'를 포함한 것으로서 위법하다고 판단하였다.

3. 인공지능과 국제 규범

인공지능과 관련한 국제 규범은 비교적 자율적인 윤리 규범으로부터 논의가 시작되어 최근에는 의무적인 법규의 형식으로 발달하고 있다.

유엔 규범의 경우 인공지능에 대하여도 국제 인권법 체계에 따른 인권 보호, 존중, 구제의 실현을 요구하고 있다. 기업 또는 각국 정부가 주도해 온 인공지능 윤리에 대하여도 국제인권규범을 반영하고 법률에 기반하여 규제할 것을 요구하여 왔다.

유럽평의회는 2020년 회원국 권고 CM/Rec(2020)1에서 인권 침해를 예방, 탐지, 금지 및 구제하는 효과적이고 예측 가능한 입법을 요구하고, 공공 및 민간 부문 행위자가 그 법적 의무를 이행하지 않는 경우 책임을 져야 한다고 강조하였다. 유럽연합은 2021년 4월 21일 공공과 민간 부문 고위험 인공지능에 요구사항을 적용하는 인공지능법(안)을 발의하였다.

가. 유엔의 기준 및 제도

1) 신기술과 인권 문제 검토

유엔 인권최고대표실은 유엔 인권이사회 특별절차에서 '신기술(new technologies)'과

인권 문제를 공통 관심 주제로 제시하였다. 초국적기업 및 기타 사업체의 인권에 대한 실무그룹, 평화적 집회 및 결사의 자유 특별보고관을 비롯하여 분야별로 임무를 맡은 특별절차 수임자들은 로봇공학, 자동화, 인공지능, 드론, 치명적인 자율 무기 시스템을 비롯하여 신기술의 인권 영향을 다각도로 다루어 왔으며, 국가는 물론 민간 기업에 대하여 신기술 환경에서 인권 보장을 위한 여러 권고를 제시해 왔다.

2018년 의사 표현의 자유 특별보고관은 인공지능이 인권에 미칠 영향에 대한 보고서를 발표하면서 이러한 기술적 역능의 확장에 대응하여 인권을 보장해야 하는 국가의 의무와 기업의 책임을 강조하였다. 특별보고관은 공공과 민간 인공지능에 대하여 영향평가 및 독립적인 감사 실시, 의견 수렴, 책임성과 이의제기를 보장할 것 등을 권고하였으며, 특히 기업에 대하여는 인권 원칙 존중 및 의견 수렴, 고지와 투명성, 편향과 차별 방지, 인권영향평가와 당사자 참여, 감사 가능성, 구제 수단 등을 보장할 것을 권고하였다.

2020년 유엔 사무총장은 인공지능 등 신기술 환경에서 사회권 보장을 위한 국가와 민간 기업들의 조치로서, 인권 보장, 부작용 방지 등의 입법 조치, 정보 및 기술 격차 해소, 기술 변화로부터 사회권과 노동권의 보호, 공공부문에서 신기술 정보 전파, 신기술 의사결정에 대한 이해관계자 참여 보장 및 공공부문 인공지능 의사결의 설명가능성 보장, 인권 실사, 책임성 보장 법제도, 감독 체계 및 구제 수단의 마련, 신기술 접근에 대한 차별 및 편견 해소, 인권 보고 및 검토 등을 요구하였다.

2) 디지털 시대 프라이버시권 검토

유엔 기구들의 여러 노력을 배경으로 유엔 인권이사회는 2019년 제42차 회기에서 <디지털 시대 프라이버시권>을 결의하면서 인공지능을 비롯한 신기술에 대한 규율을 요구하였다. 인권이사회는 인공지능 등 신기술(new and emerging technologies)의 사용, 배치 및 이후 발전은 프라이버시권 및 기타 인권의 향유에 영향을 미칠 수 있다는 사실을 인정하고, 프라이버시권에 대한 위험을 최소화할 수 있고 하여야 한다는 사실을 강조하였다. 특히 국가는, 기업이 인공지능을 비롯한 기술의 설계, 개발, 배치 및 평가할 때 프라이버시권 및 기타 관련 인권을 완전히 보장하도록 보장하고, 그 권리가 침해되거나 남용당했을 수 있는 개인에게 배보상 및 반복 금지 보장 등 효과적인 구제 수단에 접근할

수 있도록 규정하는 법률, 규제 및 정책을 도입하여야 한다. 기업은, 유엔 <기업과 인권 이행지침>에 따라 프라이버시권을 비롯한 인권에 대한 존중이 자동화된 의사 결정 및 머신 러닝 기술의 설계, 운영, 평가 및 규율에도 반영되도록 보장하여야 하며, 그 원인이 되었거나 원인에 기여한 인권 침해에 대하여 보상하여야 한다. 이 유엔 인권이사회 결의의 주요 내용은 2020년 12월 유엔 총회 결의로 채택되었다.

인공지능과 인권 문제를 해결하기 위한 여러 유엔 권고들을 최근에 종합한 것으로는 2021년 9월 유엔 인권최고대표의 보고서가 있다. 최고대표는 인공지능 시스템이 프라이버시권을 침해하게 되는 기능적 요인으로서 개인정보를 포함한 대용량 데이터셋 의존성, 자동화된 추론과 예측의 확률적 특성과 불확실성을 들었다. 또한 편향된 데이터셋이 인공지능 시스템에 기반한 차별적 의사결정으로 이어지는 경우와 의사 결정 과정의 불투명성을 우려하였다. 이러한 문제를 해결하기 위한 최고대표의 권고는 '인권 기반 접근'으로 요약할 수 있다. 최고대표는 공공과 민간 인공지능 사용의 부정적인 인권 영향을 방지하며 피해자를 구제하는 입법 및 규제 체계의 도입을 요구하고, 이러한 조치들이 인공지능의 편익을 누리면서 유해한 결과물을 방지할 수 있다고 강조하였다. 특히 원격 실시간 얼굴 인식과 같은 잠재적인 고위험 기술의 경우, 그 사용의 인권 준수가 보장될 때까지 유예(모라토리엄)하여야 한다고 발표하여 주목을 받았다. 입법 조치는 인공지능에 대한 적정하고 독립적이고 공정한 감독 체계를 포함해야 하고, 감독 시스템은 개인정보 보호 감독기구, 소비자 보호 기관, 부문별 규제 기관, 차별 방지 기구 및 국가 인권 기구를 포함하여야 한다. 인권영향평가를 포함한 인권 실사 또한 강력히 요구하였다. 인공지능의 사용이 인권과 양립할 수 없는 것으로 드러나는 경우 그 사용을 중지하여야 하며, 특히 국가는 공공재 또는 서비스 전달에서 인권에 위험한 인공지능을 사용하지 말아야 한다. 마지막으로 최고대표는 인공지능 사용의 투명성 강화를 요구하였다. 특히 공공부문은 인권에 중대한 영향을 미칠 수 있는 모든 인공지능 의사결정에 설명가능성을 보장하여야 하며, 감사가 불가능한 인공지능 시스템을 사용하지 말아야 한다. 더불어 인공지능에 대한 민관 파트너십이 독립적인 인권 감독의 대상이 되어야 하며, 인권에 대한 정부 책무를 포기하는 결과로 이어지지 않도록 보장할 것을 요구하였다.

나. 유럽연합의 기준 및 제도

1) 유럽연합 개인정보보호 일반규정

2016년 제정된 유럽연합 개인정보보호 일반규정(GDPR)은 인간의 개입이 전혀 없는 완전 자동화 의사결정에 대하여 정보주체가 그 대상이 되지 않을 권리를 규정하였다. 다만 계약의 체결 또는 이행을 위해 필요한 경우, 법률이 허용하는 경우, 정보주체의 명시적인 동의에 근거한 경우는 예외적으로 완전 자동화 의사결정이 허용된다. 완전 자동화 의사결정이 민감정보에 근거하여 이루어질 수 있는 경우로는 정보주체의 명시적인 동의에 의하거나, 법률에 기반한 상당한 공익상의 이유로 처리가 필요하며 정보주체를 보호 조치가 존재하는 경우 뿐이다.

개인정보를 처리하는 자가 예외적으로 완전 자동화 의사결정을 실시하는 경우, 프로파일링 및 완전 자동화 의사결정 유무, 관련된 로직에 관한 구체적이고 유의미한 정보, 처리의 중대성과 이로 인해 발생할 수 있는 결과 등을 정보주체에게 사전적으로 설명하고, 그에 대하여 정보주체가 인간의 개입을 요구할 권리, 본인의 의견을 피력할 권리, 결정에 대한 설명을 들을 권리 및 결정에 이의를 제기할 권리 등을 보장하여야 한다. 더불어 개인정보를 처리하는 자는 처리한 데이터셋에서 편견이 있는지 확인하고, 이를 해결할 수 있는 방법을 개발해야 한다. 또 알고리즘을 검사하고 자동화 의사결정의 정확성과 관련성을 주기적으로 검토하여 그 개선에 반영하여야 한다.

이와 같은 규정에 근거하여 네덜란드 법원은 2021년 3월 차량공유 플랫폼의 완전 자동화 의사결정에서 플랫폼 노동자들의 권리를 인정하였다. 이탈리아 개인정보보호 감독기구는 2021년 6월 배달 플랫폼 푸디뉴가 라이더에게 차별적인 알고리즘을 사용한 데 대하여 2백6십만 유로의 과징금을 부과하였다.

2) 유럽연합 인공지능법(안)

가) 인공지능법(안)의 내용

2021년 4월 21일 유럽 집행위원회는 인공지능법(안)을 발의하였다. 이 법의 제정목적은

유럽연합 시장에 출시·사용되는 인공지능 시스템의 안전성을 보장하고 기본적 권리와 유럽연합의 가치를 존중하고, 인공지능에 대한 투자·혁신이 촉진되도록 법적 안전성을 보장하며, 인공지능시스템에 적용될 수 있는 기본권 및 안전기준 관련 법령의 효율적 집행을 강화하는 한편으로, 합법적이고 안전하며 신뢰할 수 있는 인공지능의 개발을 촉진하는 데 있다.

유럽연합은 인공지능시스템 시장에서 인공지능으로 인한 혜택과 위험이 유럽연합 수준에서 적절히 잘 다루어질 수 있도록 입법조치가 필요했다고 설명하였다. 인공지능법(안)은 최초로 인공지능에 관하여 포괄적 규제프레임워크를 제시하였으며 '위험 기반 접근법(risk-based approach)'을 취하고 있는 것이 큰 특징이다.

법안은 인공지능 시스템을 그 위험성에 따라 허용되지 않는 위험, 고위험, 제한적인 위험, 최소 위험으로 구분하고 각각의 법적 의무를 부여하였다. 첫째, 국민의 안전과 생계, 권리에 대한 명백한 위협으로 간주되는 인공지능 시스템은 원칙적으로 금지된다. 사용자의 자유의지를 우회하기 위해 인간의 행동을 조작하는 인공지능 시스템 또는 애플리케이션들(예: 미성년자의 위험한 행동을 조장하는 음성 지원을 사용하는 장난감)과 정부의 '사회적 점수체제'를 허용하는 시스템이 여기에 포함된다.

둘째, 고위험 인공지능의 경우 시장에 출시되기 전에 엄격한 의무가 부과된다. 이는 적절한 위험관리 및 완화 시스템, 학습·검증·테스트 데이터셋의 관련성·오류방지·완전성 등 고품질 보장, 활동 기록 보관 및 문서화, 사용자에 대한 투명성, 인간의 관리·감독, 높은 수준의 견고성·정확성 및 사이버 보안 등의 의무이며, 모든 고위험 인공지능 제공자는 적합성 평가, 유럽연합 데이터베이스 등록, 모니터링 보고 등의 의무 또한 가지고 있다. 고위험 인공지능에 해당하는 인공지능은 제품의 안전성 요소(장난감, 자동차, 의료기기, 기계류, 항공 등) 분야, 시민의 생명과 건강을 위험에 빠뜨릴 수 있는 주요 인프라 분야(예: 도로 교통, 물, 가스, 난방, 전기 등), 교육 또는 직업훈련 평가 분야, 고용·근로자 관리 및 자영업에 대한 접근 분야, 민간·공공 필수서비스 분야(신용평가, 사회복지급여 등), 법집행 분야, 출입국관리 분야, 사법 분야에서 사용되는 인공지능이다.

특히 실시간 및 사후적으로 이루어지는 모든 원격 생체인식 시스템은 고위험으로 간주되며 엄격한 요건을 따라야 한다. 다만 법 집행을 목적으로 이 기술들을 공개적으로 접근할 수 있는 공간에서 실시간으로 사용하는 것은 금지되며 법원의 허가 등에 의해

제한적으로만 허용된다.

셋째, 제한적인 위험의 인공지능은 특정한 투명성 의무가 있다. 챗봇과 같은 인공지능 시스템을 사용할 때는 사용자가 기계와 상호 작용하고 있다는 점을 인지하고 이 정보에 입각해 계속하거나 취소하는 결정을 내릴 수 있어야 한다.

넷째, 최소 위험의 인공지능은 자유롭게 사용할 수 있다. 인공지능 탑재 게임이나 스팸 필터 등의 애플리케이션을 비롯한 대다수의 인공지능 시스템이 이 범주에 속한다.

인공지능 규제 거버넌스에 있어 법안은 각국 시장 감독 소관 기관들이 감독하도록 규정하였다. 시장 감독기관은 인공지능 제공자가 사용하는 학습, 검증 및 테스트 데이터셋에 전면적으로 접근할 수 있어야 한다. 유럽 전체적으로는 이들 시장 감독기관들과 유럽연합 개인정보보호 감독관이 참여하는 유럽인공지능위원회를 설립하여 그 집행을 촉진하고 표준 개발을 주도할 것을 제안하였다.

금지되는 인공지능시스템을 출시하거나 서비스한 경우 최대 3천만 유로(약 400억 원) 또는 전세계 연매출의 6% 중 더 높은 금액의 과징금이 부과된다. 국가 관할 당국 조사 등 협력의무 위반시 최대 2,000만 유로 또는 전세계 연매출의 4% 중 더 높은 금액의 과징금이 부과된다. 인증기관과 관할 당국에 잘못된 정보를 제공하는 경우 최대 1,000만 유로 또는 전세계 연매출의 2% 중 더 높은 금액의 과징금이 부과된다.

나) 인공지능법(안)에 대한 의견

유럽 개인정보보호 기관의 의견

유럽연합 주요 개인정보보호 감독기구인 EDPS와 EDPB는 2021년 6월 18일 인공지능법(안)에 대한 공동의견을 발표하였다. EDPB와 EDPS는 인공지능법(안)이 인공지능 위험성에 대한 규제를 추진하고 있다는 점을 긍정적으로 평가하면서도, 몇 가지 쟁점에서 법안을 수정할 것을 제안하였다. 특히 EDPB와 EDPS는 인공지능법(안)과 개인정보보호 일반규정(GDPR) 등 기존 개인정보보호 법률들의 명시적 준수 및 일관성을 주문하였다.

이들 기관은 인공지능 시스템에 의해 야기되는 사회적/집단별 위험성(예: 집단적 차별이나 공공장소 의사 표현 등)을 동등하게 평가하고 완화시킬 것을 권고하였다. 또한 개인정보보호와 관련된 측면이 적용되는 만큼, '기본권에 대한 위험'의 개념이 GDPR 등

과 일치되어야 한다고 지적하였다. 특히 법안 본문에 인공지능 시스템의 영향을 받는 개인에 대한 어떠한 언급도 없다는 사실을 사각지대로 보았다. 사람들에게 영향을 미친 행위자에게 부과되는 의무는 더 구체적으로 도출되어야 하기 때문에, 입법자들은 인공지능 시스템의 대상자가 이용할 수 있는 권리와 구제 수단을 법안에 명시적으로 언급하여야 한다.

금지되는 인공지능과 관련하여, 법집행 목적의 침입적 형태의 인공지능은 고위험으로 분류되는 것이 아니라 금지된 인공지능시스템으로 분류되어야 하며, 사회신용점수는 공공기관 뿐 아니라 민간기업에서도 금지되어야 한다. 얼굴 뿐 아니라 걸음걸이, 지문, DNA, 음성, 키보드 입력 및 기타 생체인식 신호 또는 행동 신호와 같은 인적 특성을 자동으로 인식하는 인공지능을 공개적으로 접근할 수 있는 공간에서 사용하는 것에 대하여, 실시간과 법집행 목적에 한정하지 않고 온라인을 비롯한 어떤 상황에서든 전면 금지되어야 한다. 생체인식으로 개인을 인종, 성별, 정치적 또는 성적 지향, 기타 차별을 금지하고 있는 기준에 따라 집단으로 분류하는 인공지능 시스템도 마찬가지로 금지되어야 한다. 과학적 타당성이 입증되지 않았거나 본질적으로 침해적인 거짓말탐지기도 금지되어야 한다. 법집행기관이 자연인의 개인적 위험 평가를 하는 인공지능 시스템도 금지되어야 한다. 감정인식 또한 원칙적으로 금지되어야 한다.

고위험 인공지능과 관련하여, 현재 보험료 결정이나 의료적 처치 또는 건강 연구 목적의 사용과 같이 상당한 위험을 수반하는 일부 유형의 사용 사례가 누락되어 있다고 지적하고, 해당 목록이 정기적으로 갱신되어야 한다고 강조하였다. 또한 사전 적합성평가에서 ‘고위험’으로 평가하지 않았더라도, 시스템의 사용자인 개인정보처리자가 개인정보보호법에 따른 개인정보보호 영향평가를 독립적으로 실시할 필요가 있다. 더불어 고위험 인공지능의 사전 적합성평가는 항상 제3자가 수행하여야 한다.

투명성과 관련하여, EDPB와 EDPS는 고위험 인공지능 시스템을 공공데이터베이스에 등록하도록 한 것을 환영하였다. 다만 범죄를 탐지・예방・수사・기소하는 데 사용되는 인공지능 시스템 모두에 투명성 의무가 적용되지 않는다는 사실은 너무 광범위한 예외라고 보았다. 비밀유지 대상이 되는 인공지능의 경우에는 일반 대중이 아니라도 소관 감독기관에 등록할 필요가 있다.

거버넌스 문제와 관련하여서는, 감독기관의 독립성을 명확히 확립해야 한다. 시장 감

독기관 및 개인정보보호 감독기구 모두 기존 법률에서 독립성을 보장받는데, 이 법안은 감독기관이 독립적일 것을 요구하지 않으며, 이들 기관에 집행위원회 보고를 요구하고 있다. EDPB와 EDPS는 국가 개인정보보호 감독기구들이 국가 인공지능 감독 기관으로 지정되는 것이 바람직하다는 의견이다. 유럽 인권기구인 유럽연합 기본권청 또한 인공지능위원회 참관 기관으로 고려되어야 한다.

시민사회의 의견

시민사회는 유럽연합의 인공지능 규제 구상을 전반적으로 환영하면서도, 인공지능법이 인권을 보다 강력하게 보호할 것을 요구하고 여러 예외로 인한 규제 공백을 보완할 것을 요구하였다.

독일 기반 알고리즘왓치는 생체인식 대량감시 등 인권 침해가 큰 인공지능들을 금지 대상에서 제외하거나 예외를 인정한 부분을 개선할 것을 요구하였다. 예를 들어 법집행기관 외 공공기관 및 민간회사의 원격 생체인식은 금지 대상에서 제외되었으며, 금지에서도 "자연인의 생명 또는 신체적 안전에 대한 구체적이고 실질적이며 임박한 위협 또는 테러 공격의 방지"의 경우에는 예외로 하였고, 예외 사유의 경우 법원 등의 사전 허가를 받도록 하였음에도 '긴급 상황'에서는 또다시 예외로 하였다. 사회신용점수의 금지 또한 공공기관에 한정되어 있다. 한편, '예측 치안'용 인공지능 시스템은 인권 침해 우려가 큼에도 불구하고 금지 대상이 아니라 고위험으로 분류되어 있다. 더불어 고위험 인공지능에 대한 사전 적합성평가를 이해관계가 있는 회사 및 기관 자체적으로 수행하도록 한 점이 부적절하다. '제한된 위험'에 속하는 감정 인식, 생체인식 분류, 딥페이크 또한 개인 뿐 아니라 사회적으로 중대한 위험이 될 수 있고 특히 감정 인식의 과학적 근거는 논쟁 중이다. 이 분야 인공지능에 요구되는 최소한의 투명성 요구가 범죄 탐지, 방지, 수사에서 또다시 예외가 된 것 역시 문제이다. 모든 고위험 인공지능에 대한 등록과 공개 제도는 매우 고무적이지만, 모든 공공기관 인공지능은 위험성 여부와 무관하게 일반에 공개되어야 한다. 무엇보다 사람 중심의 인공지능 체계를 제안했던 유럽 집행위원회 법안에서 인공지능 시스템으로 영향을 받는 사람들에 대한 책무성 체계를 누락하고 있다는 점은 매우 큰 결함이다. 사람들의 삶에 영향을 미치는 결과를 낳는 인공지능에는 투명성을 보장하는 것에서 더 나아가 그 영향을 받은 사람이나 집단이 결과에

이의를 제기하고 번복하거나 재고를 요청할 수 있는 구제 수단에 수월하게 접근하고 법적인 권리를 행사할 수 있도록 보장하여야 한다.

네덜란드 기반 ECNL은 인공지능 위험 수준과 관계 없이 모든 인공지능 시스템에 최소한의 투명성 및 인권 준수를 촉구하고, 제공자는 물론 사용자 역시 의무를 준수할 것을 요구하였다. 특히 인공지능 시스템에 대한 인권영향평가의 실시가 필요하다. 현재는 제공자에게만 평가 의무가 부여되어 있는데, 인공지능 시스템의 사용자가 그 사용 전에 인권영향평가를 비롯한 인권 실사를 실시하여야 한다는 것이다. 더불어, 인공지능 시스템으로부터 영향을 받는 집단이나 사람의 피해에 대한 배상권 및 구제 수단이 법안에 반드시 추가되어야 한다. 무엇보다 영향을 받는 사람들을 비롯한 이해관계자들에 대한 통지와 의미있는 참여가 보장되어야 한다. 시정조치, 통지 등의 집행 절차 뿐 아니라 영향평가에 시민단체를 포함한 이해관계자 참여를 보장하여야 하고, 규제기구의 결정사항에 대한 제3자 이의제기도 가능하여야 한다.

다. 인공지능 (인권)영향 평가의 기준 및 제도

1) 인권 기반 접근법

인공지능에 대한 위험 기반 접근법과 또 다르게, 인공지능에 대하여 인권 기반 접근법을 취하고 있는 국제 규범들이 존재한다. 인권 기반 접근법의 주요 요소 중 하나는 인권영향평가이다. 인권영향평가 일반에 대한 기존 국제 규범으로는 유엔 <기업과 인권 이행지침>, 최종 사용 관련 인권 위험 식별 및 평가, 다국적 기업을 위한 OECD 지침이 있다. 인공지능과 관련된 인권영향평가와 관련한 국제 규범으로는 유럽평의회 인권위원장, 유럽평의회 권고, 유럽 집행위원회 인공지능 고위전문가그룹을 비롯하여 학계 및 시민사회에서 발표한 사항들을 참고할 만 하다.

2) 유엔 인권최고대표의 <최종 사용에서 인권 위험 식별 및 평가>

유엔 인권최고대표는 2020년 <최종 사용에서 인권 위험 식별 및 평가>를 발표하였다.

이 문서는 제품과 서비스에서 인권 위험을 식별하고 평가할 때 <기업과 인권 이행 지침>의 기본적인 요구를 이해하고자 하는 기술 기업 경영진을 대상으로 한다. 관련하여 <기업과 인권 이행 지침>은 기업에 대하여 1) 발생할 수 있는 영향에 대하여 광범위한 관점을 갖고, 2) 가장 심각한 해악에 초점을 맞추고, 3) 이해관계자들과 의미 있게 관계를 맺고 소통할 것을 요구하였다.

3) 유럽평의회 인권위원장의 <인공지능 블랙박스 해제: 인권 보호를 위한 10단계>

유럽평의회 인권위원장은 2019년 <인공지능 블랙박스 해제: 인권 보호를 위한 10단계>에 대한 권고를 발표하였다. 이 권고는 인공지능이 인권에 미치는 부정적인 영향을 방지하고 완화하기 위한 방안을 제시하였다. 권고가 초점을 맞춘 10가지 조치 영역은 1) 인권영향평가, 2) 공개적인 의견 수렴, 3) 민간부문의 인권기준 이행을 촉진해야 하는 회원국 의무, 4) 정보와 투명성, 5) 독립적 감독, 6) 차별금지와 평등, 7) 개인정보보호 및 프라이버시, 8) 표현의 자유, 집회 및 결사의 자유, 노동권, 9) 구제 수단, 10) 인공지능 리터러시 증진 등이다.

4) 유럽평의회 <알고리즘 시스템의 인권 영향에 대한 대응 지침>

유럽평의회는 2020년 4월 8일 알고리즘 시스템의 인권 영향에 대한 각료위원회 권고 CM/Rec (2020)1를 채택하였다. 이 권고와 부록 <알고리즘 시스템의 인권 영향에 대한 대응 지침>은 알고리즘 시스템의 설계 및 개발과 관련하여 국가 및 민간의 행위자들에게 지침을 제시하여 기술 개발로부터 유럽인권협약에 규정된 인권과 개인의 자유를 보호하는 것을 목표로 한다. 특히 권고는 1) 국가는 법적 체제를 기술적 상황에 적용하기 위한 검토를 실시하여야 하며, 2) 민간 행위자는 <기업과 인권 이행 지침>에 따라 법을 준수하고 인권을 존중해야 한다고 명시하였다.

5) 캐나다 정부 <알고리즘 영향 평가 도구>

캐나다 정부는 2019년부터 <캐나다 알고리즘 영향평가 도구(Canadian Algorithmic Impact Assessment Tool)>를 사용하고 있다. 이 평가 도구는 캐나다 재정위원회의 <자동화된 의사결정 훈령(Directive on Automated Decision-Making)>을 지원하기 위하여 개발되었으며 공공기관이 의무적으로 실시하여야 하는 위험 평가 도구이다. 평가 도구는 48개의 위험성과 33개의 완화 질문들로 구성되어 있으며, 답변 결과에 따라 공공기관 자동화된 의사결정 시스템의 영향 수준이 결정된다. 해당 질문들은 캐나다 재정위원회 사무국이 학계, 시민 사회 및 기타 공공 기관과의 협의를 통해 수립하였으며, 정부 정책, 윤리 및 행정법적 고려 사항에 따라 자동화된 의사결정 시스템의 위험성 영역별로 구성되어 있다. 전반적으로 이 평가 도구는 각 부처 및 공공기관들이 자동화된 의사결정 시스템과 관련된 위험을 더 잘 이해하고 관리할 수 있도록 설계되었다. 알고리즘 영향평가는 프로젝트 설계 단계의 초기 시점에 실시되어야 하며, 그 결과는 캐나다 공식 언어로 접근 가능한 형식으로 발표되어야 한다.

라. 공공기관 모범 기준 및 제도

2017년 폴란드 법원은 정부의 실업자 점수 알고리즘이 국회에 입법한 법률에 근거를 두고 있지 않은데 대하여 위헌이라고 결정하였다. 미국 텍사스 휴스턴의 연방지방법원은 민간 기업에서 조달한 교육청의 교사 평가 알고리즘에 대하여 투명성과 적법절차 부족을 이유로 운영을 중단시켰다. 특히 법원은 민간 기업의 영업비밀과 국민의 헌법상 권리인 적법절차를 모두 충족하기 위해서는 공공기관의 중요한 의사결정에 비밀 알고리즘을 사용해서는 안된다고 설시하였다. 2020년에는 네덜란드 헤이그 지방법원이 사회복지급여 부정수급 탐지 시스템에 대하여 투명성 부족과 개인정보보호법 위반을 이유로 운영을 중단하라는 내용의 판결을 내렸다.

일부 국가는 조달 지침으로 공공부문에 도입되는 인공지능에 대하여 더욱 엄격한 기준과 절차를 요구하고 있다.

해외 일부 지방자치단체는 자치단체가 도입한 인공지능에 대한 사항을 시민들에게 공

개하는 정책을 시행 중이다. 네덜란드 암스테르담과 핀란드 헬싱키 시는 2020년 시민들에 알고리즘 등록부를 시범적으로 공개하였다. 프랑스 앙티브시는 「디지털 공화국법」에 따라 2021년 2월부터 알고리즘 주민공개 제도를 실시하고 있다.

마. 시사점

인공지능 등 신기술과 관련하여 인권 보장을 요구하는 국제적인 기준 및 제도에서 비교적 공통적이고 인권적 함의가 높은 기준을 추려보면 다음과 같다. 첫째, 인공지능 개발과 활용에서 인권과 책임성을 보장하는 법률과 감독 체계 수립을 요구한다. 둘째, 인공지능 개발과 활용에서 투명성과 자기결정권 보장을 요구한다. 셋째, 여러 인권 영역 중에 특히 인공지능 개발과 활용에서 문제가 되는 개인정보 권리를 보호하기 위한 규범 준수가 요구된다. 넷째, 인공지능의 개발과 활용에서 편향과 차별을 금지하기 위한 조치로서, 데이터셋의 사전적인 품질 관리와 사후적인 모니터링을 요구하고 있다. 다섯째, 인공지능이 인권과 안전에 미치는 위험을 식별하고 완화시키기 위하여 다양한 영향평가가 제안되고 있으며, 특히 인권영향평가에 대한 요구가 두드러진다.

4. 인공지능 관련 국내 기준 및 제도

가. 국내 기준

1) 인공지능 국가전략

2019년 12월 17일 정부는 대통령 주재로 열린 국무회의에서 과학기술정보통신부를 비롯한 전 부처가 참여하여 마련한 <인공지능(AI) 국가전략>을 발표하였다.

이 국가전략은 'IT 강국을 넘어 AI 강국으로' 라는 제목 하에 3대 분야의 9대 전략과 100대 실행과제를 배치하였다. 특히 "사람 중심의 인공지능 구현"을 위한 포용적 일자리 안전망 구축, 역기능 방지와 윤리체계 마련 분야에서 13개 과제가 선정됐다.

그러나 <인공지능 국가전략>에 대한 비판적 지적 중 하나는 혁신을 위해 국가적 역량

을 총결집할 것을 선언하면서도 인공지능으로부터 복합적인 영향을 받게 되는 다양한 시민사회 이해당사자의 목소리를 반영하고 참여를 보장하기 위한 계획을 포함하고 있지 않다는 것이다. 무엇보다 인권 보장의 의무를 이행하여야 할 국가로서 국가전략을 수립함에 있어 인권에 대한 고려와 언급을 전혀 하지 않은 것은 중대한 결함이라 할 것이다.

2) 과학기술정보통신부

국내 인공지능 기준과 관련하여 대표적으로 언급되는 것은 과학기술정보통신부의 <인공지능(AI) 윤리기준>이다. 그러나 이 윤리기준은 구속력 있는 '법'이나 '지침'이 아닌 도덕적 규범이자 자율규범으로서 '자율적'인 준수를 목표로 하며, 기업 자율성을 존중하고 인공지능 기술발전을 장려하며 기술과 사회변화에 유연하게 대처하고자 하는 지향점을 명확히 밝혔다. 이는 인권 보호에 대한 국가의 의무와 인권 존중에 대한 기업의 책임, 피해자 구제의 실현을 요구하는 국제 인권 규범과 차이가 있다.

과학기술정보통신부는 2020년 12월 23일 발표한 <인공지능 법·제도·규제 정비 로드맵> 역시 '민간자율 우선'을 그 추진 방향의 하나로 설정하였다. 2021년 5월 14일 <신뢰할 수 있는 인공지능 실현전략>에서는 「지능정보화 기본법」에 기반하여 인공지능 영향평가와 고위험 인공지능에 대한 기준 도입 방침을 밝혔다.

3) 방송통신위원회

방송통신위원회는 2019년 11월 11일 <이용자 중심의 지능정보사회를 실현하기 위한 원칙>과 2021년 6월 30일 <인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙>을 발표하였다. 그러나 이들 원칙은 전적으로 '자율적인' 규범으로 서비스 제공자에게 제시되었다는 점에서, 그 규범적 효력에 의문이 남는다.

4) 공정거래위원회

공정거래위원회는 온라인플랫폼 중개거래 투명성·공정성 제고를 취지로 「온라인 플

랫폼 중개거래의 공정화에 관한 법률」 제정법률안을 2021년 1월 28일 국회에 발의하면서 온라인 플랫폼 거래 계약서 필수 기재 사항에 플랫폼 알고리즘에 의하여 상품이 노출되는 순서, 형태 및 기준을 포함하였다. 계약 내용 변경 시에도 해당 내용 및 사유를 이용자에게 미리 통보하도록 하였다.

또한 2021년 3월 5일 소관「전자상거래 등에서의 소비자보호에 관한 법률」에 대한 전부개정법률안을 입법예고하면서 조회수, 판매량, 상품 가격, 광고비 지급 여부 등 검색·노출 순위를 결정하는 주요 기준을 표시하도록 하는 등 알고리즘 검색 결과 및 순위 등에서 소비자의 합리적 선택을 위한 정보 제공을 강화하였다.

5) 개인정보보호위원회

개인정보보호위원회는 이루다 사건 이후 2021년 5월 31일 <인공지능(AI) 자율점검표>를 발표하고 인공지능의 개발·운영에 참여하는 자의 개인정보보호에 대한 인식을 제고하고 개인정보보호법 준수를 요구하였다.

더불어 2021년 9월 28일 국회에 발의한 개인정보보호법 개정법률안은 ‘자동화된 결정에 대한 정보주체의 권리 등’에 대한 조항을 신설하고 정보주체가 완전 자동화 의사결정을 거부하거나 설명 등을 요구할 수 있도록 하였다.

6) 금융위원회

금융위원회는 2021년 7월 8일 <금융분야 인공지능(AI) 가이드라인>의 시행을 발표하였다. 가이드라인은 신용정보법상 ‘자동화평가 결과에 대한 설명 및 이의제기 등(제36조의2)’ 조항을 금융 실무에 적용할 수 있도록 “고객에 대한 설명의무가 있는 금융서비스 등에 AI 시스템을 활용하는 경우 또는 고위험 서비스에 AI 시스템을 활용하는 경우 설명가능 인공지능 기술 등 적절한 인공지능 기술을 투명하게 적용하여 맥락에 맞는 설명이 도출되는지 여부를 확인하고, AI 시스템의 안정성·신뢰성 등을 훼손하지 않는 범위 내에서 설명가능성을 합리적인 수준으로 개선하기 위해 노력”할 것 등을 규정하였다. 금융위원회는 준비기간을 거쳐 연내 가이드라인을 시행하고, 금융업권 및 기능·서

비스별 특성을 고려하여 가이드라인을 구체화한 세부 실무지침도 마련할 계획이라고 밝혔다.

그러나 가이드라인이 전반적으로 금융회사의 자율적 조치에 의존하고 있다는 점에서 운영원칙과 점검기준이 모호하다는 문제점을 지적받고 있다.

7) 서울특별시 교육청

2021년 9월 서울특별시교육청은 <인공지능(AI) 공공성 확보를 위한 현장 가이드라인>을 발표하였다. 가이드라인은 학교에서 인공지능을 도입할 때 '인공지능(AI) 등급 평가 매트릭스'와 '인공지능(AI) 영향평가 체크리스트'를 사용해 인공지능에 기반한 결정의 영향을 평가하도록 하였다.

나. 국내 제도

1) 인공지능 관련 법령

'인공지능'이라는 용어는 다양한 법령에서 사용하고 있다. 예를 들면, 「행정기본법」은 "행정청은 법률로 정하는 바에 따라 완전히 자동화된 시스템(인공지능 기술을 적용한 시스템을 포함한다)으로 처분을 할 수 있다."(제20조)는 규정을 두고 있다. 하지만, '인공지능'의 개념을 정의하거나 인공지능에 대한 대응방안을 규정한 법령 규정은 없다.

다만, 「지능정보화 기본법」은 "전자적 방법으로 학습·추론·판단 등을 구현하는 기술"을 '지능정보기술'의 하나로 정의하고 '지능정보기술'에 대한 진흥과 규제 정책을 규정하고 있다. 「지능정보화 기본법」은 '지능정보사회 기본원칙'으로 국가와 지방자치단체에게 인간의 존엄·가치를 보장하고 차별을 방지할 책무를 부여하면서 동시에 인간의 존엄·가치 보장, 지능정보기술의 역기능 방지, 개인정보의 보호/사생활 비밀·자유 보장의 책무는 "사회의 구성원 전체"에 부여하는 안전장치를 마련하였다. 그러나 「지능정보화 기본법」의 대부분은 지능정보기술의 개발, 보급, 표준화, 인력양성

등의 진흥정책에 할애하고 있고, 인공지능 개발 주체에게 안전성, 신뢰성, 공정성을 확보하도록 하였으나 '노력할 의무'에 그치므로 선언적 성격의 의미를 벗어나기는 어렵다.

2021년 3월 23일 개정 「전자정부법」은 행정기관이 인공지능을 활용하여 전자정부서비스를 제공할 수 있도록 하였으나(제18조의2), 이 규정 또한 행정부문에서 인공지능의 활용에 대한 근거만을 두었을 뿐이다.

2) 인공지능을 직접 규율 대상으로 하는 법률안

국회에는 인공지능을 직접 규율 대상으로 하는 제정법률안들이 다수 제안되어 있다. 그러나 제안된 법률안 대다수는 인공지능 산업의 기반이나 기술 개발을 진흥하는 것을 주요 목표로 하고 있다. 규정들 역시 산업 육성에 치우쳐 인공지능이 추구해야 할 원칙들을 담보할 내용에 대한 고민이 부족해 보이며, 이에 따라 인공지능과 상호작용하는 국민에 대한 구체적인 내용은 찾아보기 어렵다.

다만 정필모 의원이 대표발의한 인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안(의안번호: 2111261)의 경우, 의료, 전기·가스·수도·핵시설 등 에너지·기간서비스, 범죄수사 생체인식, 개인에 대한 평가·의사결정, 공공기관 사용 부문에서 '사람의 생명·신체에 위험을 줄 수 있거나 부당한 차별 및 편견의 확산 등 인간의 존엄성을 해칠 위험이 있는 인공지능'을 정의(제2조 제2호)하고 규율하고자 했다는 점에서 유럽연합 등에서 추진하고 있는 위험 기반 접근법과 유사한 지향을 가지고 있다.

한편, 이상민 의원이 대표발의한 평등에 관한 법률안(의안번호: 2110822)의 경우 '인공지능 디지털 기술 등에 대한 동일 적용'에 대한 조항을 두고 이 법이 인공지능, 빅데이터 등 디지털 기술을 기반으로 한 모든 영역에도 동일하게 적용함을 규정하였다(안 제8조).

「지능정보화 기본법」에도 불구하고 인공지능에 대한 독자적인 제정 법률을 추진하기 위하여는 「지능정보화 기본법」이 이미 그 목적으로 밝히고 있는 인공지능 관련 정책의 수립·추진 및 산업 경쟁력 여건 조성에 대한 내용을 반복적으로 규정하기 보다는, 최근 국제 규범에서 강조되고 있는 인공지능 규율을 포괄하는 것이 바람직할 것이다. 즉, 인공지능의 개발과 활용으로부터 국민의 안전과 인권을 보호하기 위한 원칙을 밝히고,

그 이행을 준수하는 감독 체계를 수립하며, 영향을 받는 사람들의 의견을 수렴하고 참여를 보장하는 거버넌스를 갖추고, 책임성 체계를 갖추어 피해 국민을 구제할 수 있는 규정이 마련되어야 할 것이다.

5. 인공지능과 국가인권기구

가. 인공지능과 국가인권기구 관련 해외 논의

인공지능에 대한 인권적 감독의 중요성이 커질수록 국가인권기구의 이 분야 역할과 개입에 대한 요구도 커져 왔다.

국가인권기구는 독립성, 전문성 및 진정사건 처리 경험을 보유하고 있다는 점에서 인공지능의 인권 준수에 대한 독립적인 감독은 물론 인권영향평가 등 관련 지침과 지원 역시 전문적이며 효과적으로 수행할 수 있을 것으로 기대되고 있다. 인공지능 기술의 발전은 국가인권기구의 권리구제 활동에도 변화를 가져올 수 있기 때문에 이 분야 국가인권기구의 대응 역시 필연적으로 이루어질 수밖에 없다. 또한 차별 시정 업무를 수행하는 국가인권기구는 관련 권리구제 뿐 아니라 인공지능 시스템으로부터 차별적인 영향을 받는 사람들과의 협의 역할을 수행할 수 있다.

다만 인공지능에 대한 전문지식 측면에서 국가인권기구의 역량과 자원이 현재까지 부족한 측면이 있기 때문에 이 관련된 인적, 재정적 자원을 보다 더 지원할 필요가 있다는 점 또한 지적되고 있다.

1) 유럽연합과 유럽평의회 인권기구의 검토

유럽연합의 인권기구인 유럽연합 기본권청은 2020년 국가인권기구의 강화 및 효과적인 활동에 대한 보고서에서 인공지능 기술의 발전으로 사생활권, 개인정보보호권, 차별금지 관련 조항이 동반하여 문제가 되고 있다고 지적하면서, 이 문제에서 독립성, 전문성 및 진정사건 처리 경험을 보유하고 있는 국가인권기구들이 알맞은 역할을 수행할 수 있다고 보았다.

더불어 기본권청은 2020년 12월 <인공지능과 기본권> 보고서에서 인공지능 시스템이 기본권에 미치는 부정적인 영향을 감독하고 이를 해결할 수 있는 효과적인 책임 체계의 구축이 필요하다고 지적하면서, 국가인권기구를 비롯한 기존의 감독 전문 조직을 더 잘 활용할 것을 유럽연합과 회원국들에 권고하였다.

이 보고서에서 기본권청은 인공지능 관련 기술을 사용할 때 기본권 영향평가에 포함될 수 있는 주요 요소를 제안하였다. 우선 여러 인권 영역 중에서도 주요하게 평가할 대상으로는 개인정보보호, 차별 금지, 권리 구제를 꼽았다. 첫째, 개인정보에 대한 처리는 개인정보보호법을 준수하여 합법적이어야 한다. 둘째, 인공지능 프로세스는 차별 우려가 있는 집단에 대한 부당한 대우나 차별을 금지하여야 한다. 이때 불이익의 정도는 그 특성(위해 유형), 심각성(위해 정도), 중대성(다른 집단에 비해 불이익을 받을 위험이 높은지)에 따라 다르다. 셋째, 인공지능의 대상이 되는 사람들은 이의를 제기하고 효과적인 구제 수단에 접근할 수 있어야 한다.

현행 개인정보보호 감독기구, 평등기구, 옴부즈만기구, 국가인권기구 등은 인공지능이 각자의 전문 영역에서 기본권에 미치는 잠재적 영향에 대하여 협의하고 감독하는 역할을 수행할 수 있다.

한편, 유럽평의회 인권위원장도 2019년 보고서에서 인공지능의 인권보장 체계에서 국가인권기구의 역할이 중요하다는 점을 지적하였다.

2) 국가인권기구들의 활동 사례

이미 몇 년 전부터 인공지능에 대한 효과적이고 독립적인 감독 체계 수립에 있어 국가인권기구들의 역할과 관여가 요구되어 왔다. 국제적으로 권위 있는 개인정보보호 감독기구 국제협회는 2018년 발표한 <인공지능 윤리 및 개인정보보호에 대한 선언>에서 개인정보보호 감독기구들이 인권기구들과 협력할 필요성이 있다고 강조하였다. 특히 인공지능의 차별 문제와 관련한 국가인권기구들의 역할이 요구되어 왔다. 영국에서는 2020년 공직생활윤리위원회가 모든 공공기관 인공지능으로 하여금 현행 법률을 준수하고 그 내용을 공표하도록 권고하고, 영국 평등인권위원회에 공공부문 인공지능의 평등법 준수지침 개발과 역할을 요청하였다.

최근 각국 국가인권기구들의 인공지능 정책 관련 개입도 증가하였다. 뉴질랜드 국가인권기구는 신기술 환경에서 프라이버시, 데이터, 기술 문제를 아우르는 새로운 법적 규제를 제안하면서 △국제인권기준 준수 △뉴질랜드 법제도 준수 △프라이버시 침해의 예외 △감독 체제, 투명성, 목적 제한 등 안전조치를 요구하였다. 네덜란드 국가인권기구는 디지털 채용 절차에서 나타날 수 있는 노동 시장 차별 가능성에 대해서 검토하면서 동등한 대우를 받을 권리를 강조하였다. 스웨덴 국가인권기구와 평등 옴부즈만은 인공지능의 차별금지법 준수를 감독하고 있다. 독일의 연방차별금지국은 2019년 <알고리즘 사용에 관련된 차별 위험>에 대한 연구에서 알고리즘 차별 방지를 위하여 정부에 대한 정책권고를 요구하였다. 호주 국가인권위원회는 2018년부터 일련의 '인권과 기술' 프로젝트 사업을 수행하고 2021년 6월 <인권과 기술> 최종보고서를 발간하였다.

나. 인공지능과 국가인권위원회의 역할

국가인권위원회는 국가인권위원회법과 국제인권법에 의해 설립된 국가인권기구로서 대한민국 헌법과 국제인권규범에서 보호하는 모든 사람의 인권과 자유를 인공지능을 비롯한 지능정보사회에서 보호하고 향상시켜야 할 책무가 있다. 국가인권위원회는 국가인권위원회법에 따라 인공지능에 관한 법령·제도·정책 등 인권 개선에 관한 권고 또는 의견을 표명할 수 있고, 공공기관 인공지능의 인권침해 및 차별, 법인·단체·사인의 인공지능에 의한 차별에 대해 조사 및 구제 활동을 할 수 있으며, 인권침해의 유형, 판단기준 및 그 예방 조치 등에 관한 지침의 제시 및 권고를 할 수 있으므로, 공공기관이 직접 또는 조달을 통하여 개발하거나 활용하려는 인권지능에 대한 인권 기준을 제시할 수 있다. 유엔 <기업과 인권 이행 지침>에 따른 기업의 인권경영이 인공지능에 대하여도 이행될 수 있도록 관련 인권 기준을 제시할 수 있다.

특히 국가인권위원회는 인공지능에 대한 인권영향평가가 시행되도록 지침을 마련하고 권고하는 것이 바람직하다. 이미 국가인권위원회는 「공공기관 인권경영 매뉴얼」 적용 권고를 통해 공공기관에 대한 인권영향평가가 시행되도록 한 바 있으므로, 적어도 공공기관에 대한 인권영향평가에서 인공지능을 도입할 경우의 영향평가 기준을 마련하고 시행을 권고할 필요가 있다.

또한 국가인권위원회는 인공지능과 인권에 관한 정보와 이슈를 상시 관측하면서 우리 사회의 인공지능 거버넌스에도 적극 참여해야 한다. 우리 사회에서 잘 드러나지 않은 채 일어나고 있거나 앞으로 일어날 것으로 우려되는 인권 침해 문제를 폭넓게 관측하는 역할은 인공지능 거버넌스에서 기존의 하향 접근이나 상향 접근과는 구분되는 중간적인 접근으로서 이해될 수 있다.

6. 인공지능 개발과 활용에서의 인권 가이드라인(안)

제1장 의의

제1절 제정 배경

1. 사회 전반에 걸쳐 인공지능(AI: Artificial Intelligence)의 활용이 증가함에 따라 고용, 금융, 행정, 복지 등 거의 모든 분야에서 인간의 기본적인 삶과 인권에 영향을 미치는 사례가 증가하고 있습니다.

2. 앞으로도 막대한 양의 빅데이터를 분석하고 학습하는 과정을 통해 다양한 영역에서 사람의 판단을 대신할 수 있는 인공지능은 점차 적용영역을 넓혀 사회 전반과 개인의 삶에 강력한 영향력과 파급력을 행사할 것입니다.

3. 인공지능의 발전과 확산은 국가경쟁력과 개인 삶의 질을 높일 것으로 기대되지만 개인정보 및 사생활 침해, 차별 등과 같은 인권을 침해하는 문제들도 대두되고 있습니다.

4. 반면 인공지능으로 영향을 받는 당사자들은 인공지능의 도입, 운영, 결정에 대하여 참여의 기회를 보장받고 있지 못하며, 인공지능으로 인한 인권침해가 발생한 경우에도 적절하고 효과적인 권리구제를 받을 수 있는 절차와 방법이 미흡한 상황입니다.

5. 따라서 인공지능을 개인의 삶과 사회적 공익에 기여할 수 있도록 설계하며, 인간의 존엄성과 차별금지, 자기결정권 보장 등 기본적 인권에 기반을 두도록 하는 것이 매우 중요합니다.

6. 본 가이드라인은 인권적 관점에서 인공지능의 개발과 활용의 전 과정에서 인권침해를 예방하기 위하여 준수해야 할 기본 원칙과 주요 내용을 제시하고자 마련되었습니다.

제2절 목적 및 의미

7. 국가인권위원회는 「국가인권위원회법」 제19조 제6호에 따라 인권침해의 유형, 판단기준 및 그 예방조치 등에 관한 지침을 제시할 권한을 가지고 있습니다.

8. 이에 국가인권위원회는 인공지능으로 인한 인권침해와 차별을 판단하고, 개선 등의 권고나 구제절차를 마련하는 데 필요한 기준을 제시하고자 합니다.

9. 본 가이드라인은 인공지능의 개발과 활용에 있어서 우리 사회의 인권적 가치가 훼손되지 않고 인간의 존엄성을 보장하며 기본적 인권을 실현하는 방향으로 나아가도록 하는 데 목적이 있습니다.

10. 또한 인공지능의 개발과 활용 과정에 적용할 인권원칙을 제시하고, 인공지능 서비스 이용자, 영향을 받는 당사자 등에게 주어진 권리 및 피해구제수단을 제공하며, 정부에게 적절한 법령과 정책을 수립할 수 있는 가이드라인을 마련하고자 합니다.

제2장 적용범위 및 정의

제1절 적용 범위

11. 인공지능의 개발과 활용을 위한 가이드라인에서 의미하는 인권은 「대한민국 헌법」 및 법률에서 보장하거나 대한민국이 가입·비준한 국제인권조약 및 국제관습법에서 인정하는 인간으로서의 존엄과 가치 및 자유와 권리를 의미합니다.

12. 본 가이드라인은 인공지능의 개발부터 활용에 이르는 모든 과정에 참여하는 사회 구성원들을 대상으로 하며, 여기에는 인공지능 개발자, 이용자, 영향을 받는 당사자, 정부 및 공공기관, 기업 등이 포함됩니다.

13. 인공지능 개발과 활용에 관한 법률을 제정하거나 개정할 때에는 본 가이드라인의 목적과 기본 원칙 및 주요 내용을 참고할 수 있습니다.

제2절 정의

14. 본 가이드라인에서 말하는 인공지능은 일차적으로는 학습과 추론, 판단을 전자적으로 구현하는 알고리즘과 해당 프로세스를 지칭하나, 이차적으로는 빅데이터 등 인공지능을 기능하게 하는 일련의 기술들을 포함하고 있습니다.

15. 본 가이드라인에서 기본 원칙은 인공지능의 개발과 활용에 있어서 기본이 되는 인권적 가치를 도출한 것으로 본 가이드라인이 추구하는 방향을 제시한 원칙입니다.

16. 본 가이드라인에서 주요 내용은 기본 원칙을 바탕으로 구체적으로 적용해야 할 과제들을 제시하였으며, 향후 입법 및 제도화 과정에서 주요 내용을 보다 구체화할 필요가 있습니다.

제3장 기본 원칙

제1절 인간의 존엄성

17. 「헌법」 제10조에서 보장하고 있는 인간의 존엄과 가치는 누구나 누려야 할 불가침의 기본적 인권으로, 모든 권리의 출발점인 동시에 종국적으로 보장되어야 할 인권적 가치입니다.

18. 어떠한 활동도 인간의 존엄에서 유래하는 다양한 권리들의 희생을 강요해서는 안 되며, 궁극적으로 모든 활동은 인간의 존엄과 가치를 증대시키는 방향으로 수행되어야 합니다.

19. 따라서 인공지능은 인간으로서의 존엄과 가치 및 행복을 추구할 권리에 부합하는 방식으로 개발 및 활용되어야 하며, 개인의 선택과 판단 및 행동을 강요하거나 자율성을 침해해서는 안 됩니다.

제2절 알 권리

20. 알 권리란 개인의 의사형성에 필요한 정보를 수집하고, 수집된 정보를 취사·선택

할 수 있는 자유를 의미합니다.

21. 개인은 알 권리를 통해 충분한 정보를 획득할 수 있고, 지식과 이해의 폭을 넓힐 수 있으며, 이를 바탕으로 합리적인 판단과 자신의 인격을 발현시킬 수 있습니다.

22. 개인의 의사 판단과 인격 발현의 중요한 요소인 알 권리의 보장을 위해 인공지능의 판단 과정 및 결과에 대한 합리적인 설명 등을 보장하도록 하고, 이를 뒷받침하는 기술적·제도적 장치를 마련해야 합니다.

제3절 자기결정권

23. 「헌법」 제10조가 규정하고 있는 인간의 존엄과 행복추구권에서 도출되는 자기결정권은 결정의 주체인 개인이 중대한 사안에 대해 외부의 강요 없이 누구나 스스로의 판단에 따라 타인의 간섭 없이 결정하고 행동할 수 있으며, 동시에 선택하지 않을 자유를 의미합니다.

24. 인공지능은 인간을 보조하여 서비스를 제공하는 것뿐만 아니라 주어진 데이터를 바탕으로 스스로 학습하여 의사결정을 내릴 수 있는 수준에 이르렀습니다.

25. 따라서 자기결정권은 인공지능의 개발과 활용의 전 과정에서 우선적으로 보장되어야 합니다.

제4절 평등과 차별 금지

26. 「헌법」 제11조는 모든 국민에게 평등권을 보장하고 있으며, 「국가인권위원회법」 제2조는 합리적인 이유 없이 성별, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향 등 개인 특성에 따라 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위를 평등권 침해의 차별행위로 정의하고 있습니다.

27. 인공지능의 개발과 활용은 개인의 행복과 사회적 공공성의 증진에 위배되어서는 안 되고, 인공지능을 통한 경제·사회·문화적 권리의 향유에 있어서 다양한 계층을 포용하고 참여의 기회를 동등하게 보장해야 합니다.

28. 또한 인공지능의 결정이 특정 집단이나 일부 계층에게 차별적이거나 부정적 영향

을 초래하지 않기 위해 개발 단계부터 다양한 계층의 의견을 수렴하고, 차별적 결과가 발생하지 않도록 필요한 조치를 취해야 합니다.

제4장 주요 내용

제1절 투명성과 설명 의무

29. 인공지능 및 이를 이용한 의사결정이 개인에게 미치는 영향력과 중요성이 갈수록 증가하고 있으므로, 인공지능의 판단과정과 그 결과에 대한 적절하고 합리적인 설명이 보장되어야 합니다. 학습 및 추론, 판단의 과정과 결과에 이른 이유를 설명하기 어려운 인공지능은 이에 대한 대응의 불확실성과 영향을 받는 당사자의 불안감을 유발하고, 인권 및 안전에 관한 법령과 정책의 집행효과를 불분명하게 할 수 있습니다.

30. 인공지능과 상호작용하는 사람에게는 언제나 그 상대방이 인공지능이라는 사실을 알려야 합니다.

31. 공공기관은 인공지능의 개발과 활용 계획 등을 사전 공개하여야 하고, 관련 당사자들의 의견을 공청회 등으로 수렴하여야 합니다. 공공기관이 인공지능을 통한 의사결정을 할 때 설명할 수 없는 인공지능을 활용해서는 안 되며, 특히 조달의 경우 입찰 단계에서부터 설명가능성이 보장되어야 합니다.

32. 공공기관이 개발하고 활용하는 모든 인공지능과 민간이 개발하고 활용하는 인공지능 중 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능(이하 ‘고위험 인공지능’이라 한다.)은 사용된 데이터와 인공지능 알고리즘의 주요 요인을 일반에게 공개하고 설명하여야 합니다.

33. 또한, 인공지능에 의한 자동화된 의사결정이 예정되어 있는 경우, 영향을 받는 당사자들은 사전에 그 사실을 알아야 합니다. 자동화된 의사결정에 의하여 영향을 받는 당사자는 그 결정의 이유에 대하여 설명을 듣고, 당사자 진술을 할 수 있으며, 이의를 제기할 수 있어야 합니다.

34. 특히, 완전히 자동화된 의사결정으로만 개인에게 법적 효력 또는 생명·신체·정신·재산에 중대한 영향을 미치는 일은 제한되어야 하고, 이러한 의사결정이 이루어진

경우에는 당사자가 해당 방식을 거부하거나 인적 개입을 요구할 수 있는 권리를 보장받아야 합니다.

제2절 사생활의 비밀 및 개인정보자기결정권 보장

35. 인공지능과 관련하여 정보주체의 권리는 처리된 개인정보에 관하여 고지를 받을 권리, 개인정보 접근 및 열람권, 개인정보처리 동의권 및 정정·삭제권, 처리정지권 등을 포함하며, 정보주체는 자신의 데이터가 사용되는 방법을 이해하고 그에 대한 통제권을 가지는 것이 중요합니다. 정보주체는 인공지능 서비스가 언제, 어디서 자신의 데이터를 수집하고, 어떻게 데이터를 처리하여 사용, 보관, 삭제되는지에 대해 알고 참여할 권리가 있습니다.

36. 인공지능의 개발과 활용에서 개인정보는 목적에 필요한 범위에서 최소한의 개인정보만을 처리하여야 하며, 처리목적 달성에 필요한 기간 동안만 보관되어야 합니다. 또한 이러한 개인정보 처리 원칙은 정보주체가 확인할 수 있도록 공개되어야 합니다.

37. 개인정보자기결정권은 정보주체의 자기 정보에 대한 통제력을 보장하기 위해 인정되는 것인데, 그 통제력 보장의 핵심은 정보주체의 동의권입니다. 따라서 개인정보 처리에 대한 정보주체의 동의는 단순한 외형적 의사 표시만이 아니라 정보주체가 개인정보 처리에 대한 제반 상황을 설명·제공받고 스스로의 자유의사에 기하여 결정할 수 있어야 합니다.

38. 인공지능의 개발과 활용에서 민감정보를 처리할 때에는 특별한 주의를 기울여 보호하여야 합니다. 더불어 의사결정의 내용과 관련성이 없거나, 부정확한 데이터에 기반한 의사결정이 이루어지지 않도록 데이터의 정확성, 완전성, 최신성을 보장해야 합니다.

제3절 차별 금지

39. 인공지능을 개발하고 활용할 때는 인공지능으로 인해 영향받는 사람의 다양성과 대표성을 반영하기 위해 노력해야 하고, 성별, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향 등 개인과 집단의 특성에 따라 편향적이고 차별적인 결과가 나오지 않도록

해야 합니다.

40. 데이터의 수집·선정 및 시스템 설계, 활용 등 인공지능 개발 전반에 걸쳐 편향이나 차별을 배제해야 하고, 이는 데이터 요소를 검사하고 차별적인 데이터를 조정하는 등의 조치를 포함합니다.

41. 특히 학습용 데이터가 인공지능의 판단에 직접적인 영향을 미치는 상황을 고려할 때, 학습용 데이터의 수집 단계부터 차별적 요소를 통제하고 데이터 편향성을 최소화하여 인공지능을 통한 의사결정이 특정 집단에 부정적 영향을 미치지 않도록 해야 합니다.

42. 개발한 인공지능에 대해 주기적인 모니터링을 거쳐 데이터 품질과 위험을 관리하고, 차별적 결과나 의도치 않은 결과에 대해 개선의 조치를 주기적으로 수행해야 합니다.

43. 인공지능 기술 및 서비스에 대한 접근성과 인공지능이 주는 혜택은 사회적 약자와 취약계층을 포함하여 모든 사회구성원에게 평등하게 제공되어야 합니다.

제4절 인공지능 인권영향평가 시행

44. 국가는 인공지능의 개발과 활용에 있어 인권적 가치가 우선시 되도록 하여야 하며, 인공지능으로 인해 발생할 수 있는 인권침해와 차별에 대하여 사전적 또는 사후적으로 관리 감독을 할 의무가 있습니다.

45. 국가는 인공지능의 개발과 활용에 있어서 인권 침해와 차별의 가능성 및 정도, 영향을 받은 당사자의 수, 사용된 데이터의 양 등을 고려하여 인권영향평가를 실시해야 합니다. 특히 인공지능 기술이 적용되어 기존 제도로 관리되거나 감독될 수 없는 새로운 분야는 인권영향평가 제도를 도입해야 합니다.

46. 인권영향평가 내용에는 인공지능의 특성, 상황, 범위 및 목적을 감안하여 본 인권가이드라인이 제시한 기본 원칙 및 주요 내용, 국제 인권 기준, 관련 법률에서 정한 의무 등이 포함되어야 하며, 인권 침해 위험요인의 분석, 개선 사항 등을 도출해야 합니다.

47. 인권영향평가는 개발 및 출시 전에 실시하고 인공지능의 기능 또는 범위 변경 시 평가를 갱신하여야 합니다.

48. 인권영향평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하기 위한 조치사항을 수립하여 적용하여야 하며, 원칙

적으로 그 내용이 공개되어야 합니다. 또한, 이를 방지하거나 완화하는 조치를 취하기 전에는 그 개발과 활용을 중단해야 합니다.

49. 국가는 인권영향평가를 인권전문성과 독립성을 확보한 기관이 담당하도록 하고, 해당 기관은 인권영향평가의 활성화를 위하여 관계 전문가의 육성, 영향평가 기준의 개발 및 보급 등 필요한 조치를 마련해야 합니다.

제5절 위험도 등급 및 관련 법제도 마련

50. 국가는 공공기관과 민간이 개발하고 활용하는 인공지능에서 인권과 책임성을 보장하도록 관련 법률과 감독 체계를 보완해야 하며, 특히 인권을 보장하기 위한 구체적인 지침과 정책을 마련하여야 합니다.

51. 국가는 개인의 인권과 안전에 미치는 위험성이 매우 높아 인공지능이 금지되는 영역, 상당한 제한이 필요한 인공지능 고위험 영역, 위험성이 거의 없는 영역 등 적절하게 위험성 단계를 구분하고, 그에 맞는 규제 수준과 인적 개입이 이루어지도록 법과 제도를 마련하여야 합니다.

52. 특히, 당사자의 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능은 투명성과 설명가능성, 개인 정보 보호, 차별 금지 등 그 규제에 있어서 공공과 민간의 구분 없이 엄격하게 적용되어야 합니다.

53. 감독 기관은 공공기관과 민간의 위법한 인공지능 개발과 활용 여부를 조사하고 피해 구제 및 조치를 취하기 위하여 상세 정보에 접근할 수 있어야 합니다. 이를 위하여 공공기관 인공지능 및 민간 고위험 인공지능 개발자 및 운영자는 사용된 데이터와 알고리즘의 주요 요소 등을 기록하고 문서화하여 일정 기간 보관하여야 합니다.

54. 국가는 인공지능을 독립적이고 효과적으로 감독할 수 있는 체계를 수립하여 개인의 인권과 안전을 보장하고 피해를 구제하여야 합니다. 인공지능 국가 감독 체계는 독립적이고 효과적이어야 하며, 진정 또는 인지로 접수한 사건을 조사하기에 충분한 자원, 권한 및 전문지식을 구비해야 합니다.

55. 국가는 인공지능으로 인하여 인권을 침해당하거나 차별을 받은 사람이 진정을 접수하여 권리를 구제받을 수 있는 기회를 보장하는 등 국가기관의 구제수단에 대한 접근

을 보장해야 합니다. 인공지능을 개발하고 활용하는 공공기관과 민간은 언제든지 구제가 가능하도록 그 책임자에 대한 정보는 물론, 이의를 제기할 수 있는 기관과 방법에 대한 정보를 일반에 공개하여야 합니다.

56. 특히 국가는 대량 감시와 차별로 이어질 위험이 높은 얼굴인식 등 원격 생체인식 기술의 사용을 공공장소에서 금지하고, 특별한 경우에 한하여 사용을 허용하되, 인권 침해나 차별의 위험성이 드러난 경우 이를 방지하거나 완화하는 조치를 취하기 전에는 사용을 중단해야 합니다. 또한 국가는 생명의 존엄성 및 윤리를 훼손할 가능성이 높은 자율살상무기에 대하여 인도주의적으로 접근하고 그 연구, 개발, 생산 및 활용을 금지하는 국제 규범을 준수하고, 이에 대한 논의에 적극적으로 참여해야 합니다.

제1장 서론

제1절 연구 목적 및 필요성

인공지능(Artificial Intelligence, AI)은 통상 “전자적 방법으로 학습·추론·판단 등을 구현하는”(「지능정보화 기본법」 제2조 제4호 가목) 알고리즘과 해당 프로세스를 지칭하지만, 빅데이터 등 “데이터를 전자적 방법으로 수집·분석·가공 등 처리하는 기술”(동법 동조 동호 나목)을 포함하는 의미로도 사용되고 있다. 완전히 또는 부분적으로 의사결정을 자동화하는 측면을 강조하는 경우 2018년 미국에서 처음으로 인공지능 책임성 관련 법률로 시행된 뉴욕시 <알고리즘 책임법>[1]이나 2019년 캐나다 정부 훈령으로 시행된 <자동화된 의사결정 지침>[2]처럼 ‘자동화된 의사결정 시스템’으로 지칭되기도 한다. 2021년 4월 발의된 유럽연합(EU)의 인공지능법(안)에서는 인공지능 시스템에 대하여 기계 학습, 논리·지식기반 또는 통계적 접근방식으로 개발되고, 인간이 정의한 목표를 위해 그것이 상호 작용하는 환경에 영향을 미치는 콘텐츠, 예측, 추천, 결정 등의 아웃풋을 생성할 수 있는 소프트웨어로 정의하였다.[3]

최근 몇 년간 공공기관과 민간 기업에서 인공지능을 기반으로 하는 신기술의 도입과 사용이 크게 증가하였다. 이미 인공지능은 인터넷과 모바일 서비스의 추천 알고리즘 등으로 일반시민의 일상생활은 물론 소상공인의 생계에 큰 영향을 미치고 있으며, 채용, 노동, 금융, 행정, 복지, 치안 등 사회 전반에서 완전히 또는 부분적으로 인공지능에 의해 내려진 조치나 결정이 인간의 기본적 권리와 삶에 직접적인 영향을 미치고 있다.

특히 행정처분 등 법적 또는 그에 준하는 효력을 갖는 공공부문 의사결정에서 인공지능의 관여와 역할이 커질 것으로 예상되고 있다. 2021년 들어 국회는 「행정기본법」을 제정하면서 행정청이 인공지능 기술을 적용한 시스템을 포함하여 완전히 자동화된 시스템으로 행정처분을 할 수 있도록 하였고(제20조), 「전자정부법」을 개정하여 행정기관이

1) A Local Law in relation to automated decision systems used by agencies. <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0 (검색일: 2021. 11. 1.)>.
2) Directive on Automated Decision-Making.
3) Artificial Intelligence Act. 부속서 I.

인공지능을 활용하여 전자정부서비스를 제공할 수 있도록 하였다(제18조의2).

2016년 한국 사회에 알파고가 던진 충격 이후 인공지능의 놀라운 발전은 언론, 산업계, 학계, 정부와 국회에서 큰 관심을 받았지만, 이러한 관심 대부분은 대체로 산업적 가치에 초점을 맞추어 왔다. 인공지능의 발전과 확산은 생산성·편의성을 높여 국가경쟁력과 삶의 질을 높일 것으로 기대되지만, 개인정보 및 사생활 침해와 차별 등 기본적 인권을 침해하는 문제들도 제기되고 있다.[4)]

인공지능은 막대한 양의 데이터를 분석하고 이들 간의 연결 관계를 식별함으로써 사생활 침해가 발생할 수 있으며, 데이터셋에 내재된 편향성에 영향을 받아 여성과 남성, 백인과 흑인, 혹은 내국인과 외국인 집단 간에 차별적인 결과를 생성하는 문제가 발생하기도 한다. 온라인 챗봇이 공개 후 불과 몇시간 만에 인종주의자가 되었고, 많은 기관에서 사용되는 얼굴인식기술이 흑인 여성을 잘 식별하지 못하였다. 인공지능은 채용지원자 중 여성을 선호하지 않았고, 흑인들의 재범 위험성을 백인보다 2배 높다고 판단하였다. 영국에서는 코로나로 인해 취소된 대학입학시험 대신 인공지능으로 성적을 부여하였다가 부유한 지역 학생이 높은 점수를 받고 가난한 지역 학생이 낮은 점수를 받은 것으로 드러나 큰 사회적 논란을 겪기도 하였다.[5)] 이에 대하여 업계 일부에서는 인간에게도 편견이 있다며 인공지능이 인간사회의 차별을 학습했을 뿐이라고 주장하지만, 유럽연합 <인공지능 백서>는 인공지능 의사결정에서 작용하는 편견의 경우 사회적 통제 메커니즘이 없으면 훨씬 더 많은 사람들에게 장기간 영향을 준다고 지적하였다.[6)] 페이스북의 캠브리지 애널리티카 사건 이후로는 플랫폼 기업의 공론장 알고리즘이 선거와 민주주의에도 영향을 미쳤다는 경각심도 일고 있다.[7)] 인공지능은 고도화된 얼굴인식 지능형 CCTV 시스템이나 자발적 정보제공이 이루어지는 스마트폰 앱 등을 통해 개인의 일상적인 습관을 추적하고 분석할 수 있어, 인공지능에 의한 감시와 추적, 고용주에 의한 노동자 통제 및 차별 등의 잠재적인 위험이 존재한다.

국내에서도 공공기관의 '인공지능 채용'에서 불합격된 사람들은 어떻게 평가되고

4) 이하 주요 사례는 오요한·홍성욱 (2018); European Commission (2020a); European Union Agency for Fundamental Rights (2020b).
5) 사는 곳으로 성적을 결정했다 : 가난한 지역 공립학교 학생들에게 낮은 점수 준 알고리즘, 실패한 실험이 남긴 과제. 한겨레21 제1329호 보도 (2020. 9. 7).
6) European Commission (2020a), p11.
7) 무심코 누른 '좋아요'가 당신의 투표심리를 조종한다. 서울경제 보도 (2020. 4. 10).

채용 탈락으로 결정되었는지 알지 못한 실정이 드러났으며, 공공기관들은 인공지능을 적용한 면접이나 서류평가가 공정한지, 면접대상의 외모나 사투리를 차별하지 않는지 전혀 검증하지 않고 사용해 왔다.[8] 최근에는 플랫폼 기업의 불투명한 알고리즘은 노동자와 자영업자들에게 고통을 주고 있다는 문제제기가 커지고 있다.[9] 포털 뉴스의 비밀 알고리즘이 공론장의 편향성에 영향을 미치고 있다는 비판이 계속되고 있다.[10]

그러나 인공지능으로 영향을 받는 당사자[11]들은 인공지능의 도입, 운영, 결정에 대하여 의견 제시와 참여의 기회를 보장받고 있지 못하며, 인공지능 분야에 대한 법률적·제도적 규제 장치가 부족하여 인공지능으로 인한 인권침해가 발생한 경우에도 효과적인 권리구제를 받을 수 있는 절차와 방법이 부족한 실정이다.

우리나라의 법률 및 제도는 선진국에 비해 인공지능 기술 및 산업의 성장에 중점을 두어 인공지능으로 인한 인권 문제를 다루는 데는 미흡한 측면이 있고,[12] 각 정부 부처와 기관은 각자의 소관업무와 관련하여 개별적 이슈로 접근하고 있다. 그러나 개별적 접근은 권리주체를 위한 보호나 지원이 부분적으로만 이루어질 수 있고, 이해관계자의 입장에 따라 자의적 해석과 충돌이 발생할 수 있으며, 개발자나 사업자가 편의에 따라 권리주체를 위한 보호 장치를 취사선택할 수 있는 문제 등이 생길 수 있다. 유엔 인권이사회는 신기술이 갖는 영향력, 기회, 도전 과제에 대응하는 인권의 증진 및 보호에 있어 전체적, 포용적, 포괄적 접근방식(holistic, inclusive and comprehensive approach)이 필요

8) 오정미 (2021). “공정성, 투명성, 책임성 제고를 위한 인공지능 법제 방향”. 토론회 <인공지능의 공정성, 투명성, 책임성 보장을 위한 법제 정비 방안>, 정필모 국회의원 등 주최 (2021. 2. 17.); 김민 (2021). “인공지능 채용 도구의 공정성과 투명성”. 같은 토론회 자료 참조.

9) 화장실 갔다고 일감 '뚝'…쉴 틈 없앤 요기요. MBC뉴스 보도 (2021. 4. 23.); 박정훈 (2021). “우리는 데이터가 아니다”. 경향신문(2021. 6. 15).

10) 집중분석 네이버 '뉴스홈'. MBC 스트레이트 114회 (2020. 12. 13.); 네이버 모바일 뉴스홈 분석. MBC 스트레이트 114회 (2021. 3. 7).

11) 개인정보보호 관련한 권리주체의 경우 ‘정보주체’로 표현되어 왔으며, 인터넷 서비스 제공과 관련한 주체의 경우 ‘이용자’ 또는 ‘이해관계자(stakeholeder)’라는 표현이 사용되어 왔다. 그러나 인공지능 개발과 활용에서 ‘영향을 받는’ 사람들은 개발자나 이용자와 또 다른 이해당사자로서, 인공지능 관련 인권 규범에서는 이들의 권리 보장을 중요한 문제로 다루고 있다. 특히 유엔인권규범은 <기업과 인권 이행 지침> 이래로 인권에 부정적인 ‘영향을 받는 개인(affected individuals)’ 또는 ‘영향을 받는 집단(affected groups)’이 인권실사 과정에 참여하여 인권의 부정적인 영향을 완화하거나 방지하여야 한다고 강조하여 왔다. 본 보고서는 해당 분야별로 주로 사용되는 표현을 사용하되, 인공지능 환경, 특히 그 결정의 대상에 대해서는 ‘영향을 받는’ 개인·집단·당사자라는 표현을 사용하였다.

12) 이순기 (2020). 인공지능의 윤리적 사용을 위한 개선과제. <이슈와 논점> 제1759호, 국회입법조사처.

하다고 강조하고 있다.[13)]

특히 챗봇 이루다 사건은 우리 사회의 편견과 혐오를 재생산하는 모습으로 우리에게 많은 충격과 분노를 야기하였다. 이에 인공지능 제품과 서비스 역시 고용과 서비스 등에서 성별, 장애, 연령, 인종, 지역 등을 이유로 차별을 금지하는 현행 법률을 준수하여야 하고 국가인권기구 등 소관 기관이 그 준수를 감독하여야 할 필요성이 대두된다.

최근 세계 각국이 인공지능을 규율하는 법제도를 적용 중이거나 도입을 준비 중이라는 사실은 매우 시사적이다. 캐나다, 뉴질랜드에서는 공공기관에서 도입한 인공지능 시스템에 대하여 영향평가를 실시하고 위험한 인공지능에 대하여 데이터셋 관리, 투명성 보장, 인간의 감독 등에 대한 의무를 부여하였다. 유럽 집행위원회는 여기서 더 나아가 공공부문과 민간부문의 모든 인공지능에 대하여 규제하는 법안을 유럽의회에 발의하였다. 특히 유엔 등 국제인권기구는 물론 지역별, 국가별 인권기구들이 발표하는 인공지능 인권 기준과 정책 역시 주목을 받고 있다.

인공지능을 이용한 자동화는 제품과 시스템의 효율성과 예측성을 크게 높일 수 있고 인간이 기존에 수행할 수 없었던 업무 영역을 확장시켰다. 그러나 인공지능이 사회적으로 적절히 통제되지 않는다면 그 편향성과 위험성이 인권과 안전에 중대한 영향을 미칠 것이다. 이러한 문제를 해결하기 위해서는 인공지능의 개발과 활용에 있어 기본권 보호를 위한 기술적·규제적 장치가 필요하고, 궁극적으로 인공지능 알고리즘의 설명가능성과 판단 책임성의 확보가 중요하다. 따라서 원칙적이고 자율규범의 성격이 강한 기존의 ‘인공지능 윤리기준(과학기술정보통신부)’ 및 관련 법률에 적용가능한 구체적인 인권기준의 정립이 시급히 요구되는 바이다.

본 연구는 인공지능 관련 인권적 이슈를 조사 정리 및 분석하고 정책적으로 적용가능한 인권기준 및 제도적 방안을 모색하고자 한다. 이를 위하여 인공지능 규범과 관련한 여러 국제문헌들과 국내외 관련 법제도 검토를 통하여 인공지능 기술에 대한 감독 체계, 투명성과 참여 보장, 인권영향평가, 권리 구제 등 관련 기준 및 법제도를 살펴보고자 한다. 이로써 향후 구체적인 법제화를 위해 필요한 인권적 핵심 기준과, 적용가능한 인권적 대응 방안(제도)을 도출하고자 한다.

---

13) Resolution adopted by the Human Rights Council on 13 July 2021, 47. 23. New and emerging digital technologies and human rights. 유엔문서 A/HRC/RES/47/23 (2021. 7. 16).

제2절 연구 내용 및 범위

본 연구는 다음과 같은 연구 범위를 가지고 있다. 첫째, 인공지능 개발과 활용에 있어 인권적 기준을 정립한다. 둘째, 개인정보보호, 알 권리, 절차적 참여권 등 인공지능 관련 기본권을 보장하기 위한 방안을 제시한다. 셋째, 인권적 기준에 따른 구체적 법률 및 제도 정비 방안 마련한다. 넷째, 인공지능의 인권적 규율 규범을 마련하였거나 추진 중인 해외사례와 비교 분석을 수행한다.

이에 우선 2장에서 국내 인공지능 도입 현황을 공공영역과 민간영역으로 나누어 특히 위험성이 높다고 보여지는 분야에서 인공지능 활용 문제를 살펴보았다.

3장에서는 인공지능 규율과 관련된 국제 규범을 집중적으로 살펴보았다. 유엔 인권이사회와 인권최고대표를 비롯한 인권 기구들 뿐 아니라 유엔 사무총장과 조약 기구 등이 인공지능 등 신기술과 관련하여 권고하는 기준 및 제도를 검토하였고, 유럽연합・유럽평의회 등 유럽에서 도입하였거나 도입을 준비중인 기준 및 제도를 검토하였다. 이어서 각국에서 도입하였거나 도입을 준비 중인 인공지능 영향평가 제도에 대하여 주의깊게 살펴보았다. 특히 국제인권기구들이 주목하고 있는 인권 기반 접근법과 인권 영향 평가에 대하여 소개하였다.

이어 4장은 국내 기준과 제도 현황을 살펴보았다. 우선 인공지능 국가전략, 과학기술정보통신부, 방송통신위원회, 공정거래위원회, 개인정보보호위원회, 금융위원회 등 각 부처 및 행정기관이 소관하는 법률에 기반하여 발표한 인공지능 관련 기준을 살펴보았다. 이어서 「지능정보화 기본법」 등 인공지능 관련 기존 기준 및 제도 현황을 검토하고 국회에 발의되어 있는 관련 법률안에서 제시하고 있는 기준 및 제도 또한 살펴보았다.

더불어 5장에서는 인공지능과 관련한 국가인권기구의 역할과 국가인권위원회의 관련 기능을 검토하였다.

6장에서는 가이드라인(초안)을 개발하여 이에 대한 심층면접조사를 실시하고 그 결과를 요약하였다.

7장에서는 이상에서 검토한 내용을 토대로 인공지능 개발과 활용에서의 인권 가이드라인(안)을 개발하여 제시하였다. 특히 인권 기준 및 제도가 자율적이고 기술적인 기존의 인공지능 윤리 가이드라인과 공통적이거나 차이가 있는 지점을 규명하고, 인공지능 개발

과 활용에 있어 인권적 관점에서 준수해야 할 기준을 수립하고 적용가능한 실행 기준으로 인권 제도 및 법제 정비 방향을 제시하고자 하였다.

마지막으로 8장에서는 연구진의 결론으로 정리된 가이드라인(안)을 제시하였다.

제3절 연구 방법

본 연구는 우선 국내·외 문헌에 대한 연구를 통하여 인공지능과 인권 기준 및 제도에 대한 선행연구를 검토하여 유엔 등 관련 국제 문헌들과 국내 공공 문헌들을 살펴보았다. 해외 각국 관련 법률 및 법안 등 주요 해외 사례에 대한 상세한 검토도 수행하였다.

본 연구는 이러한 검토 결과를 토대로 인공지능 개발과 활용에서의 인권 가이드라인(안)을 개발하여 제시하였다. 우선 가이드라인(초안)을 개발하여 관련 전문가 및 당사자에 대한 심층면접조사를 비대면으로 실시하였다. 일대일 개별서면조사와 집단심층면접조사를 병행하여 개발자, 법률가, 학계 및 기관 연구자 등 전문가 뿐 아니라 인공지능 개발과 활용에서 영향을 받는 당사자들로서 노동자, 수급대상자, 여성, 장애인 등 당사자 및 당사자 권리를 옹호하는 단체의 의견을 수렴하고자 하였다. 또한 면접조사 결과를 검토하여 결론의 가이드라인(안)에 반영하였다.

제2장 인공지능 도입 현황

제1절 공공영역의 인공지능 활용

1. 디지털 뉴딜과 스마트 시티

2020년 7월, 정부는 경기 회복 및 산업 구조의 전환 대응을 목적으로 데이터 댐, 지능형 정부 등 10대 대표과제를 기반으로 한 한국판 뉴딜 계획을 발표했다. 또한 ‘DNA(데이터·네트워크·인공지능)생태계 강화’, ‘교육인프라 디지털 전환’, ‘비대면 산업육성’, ‘SOC 디지털화’ 등 4대 분야와 12대 과제 그리고 세부적인 31개의 대표사업으로 구성된 한국판 뉴딜 내 디지털 뉴딜 정책을 발표했으며, 정부는 2025년까지 한국판 뉴딜 정책에 투입될 160조원의 예산 중 총 58.2조원의 예산을 디지털 뉴딜에 투자할 예정이다.

디지털 뉴딜의 대표적 과제로 ‘데이터 댐’과 ‘공공데이터 개방·활용 활성화’ 사업이 있다. 주요 목표는 데이터와 인공지능 등 다양한 분야의 신기술 산업을 위해 대규모 공공데이터를 개방해 데이터 경제를 가속화하고 촉진시키는 것으로, 해당 사업을 위해 과학기술정보통신부와 지능정보사회진흥원은 2021년 6월 인공지능 허브 홈페이지를 통해 헬스케어, 비전, 안전, 자율주행 등 8대 분야 170종의 데이터 4억 8천만 건을 개방했다. 한편 디지털 뉴딜의 ‘스마트 의료 및 돌봄 인프라 구축’ 과제는 ‘건강취약계층 스마트 건강관리’를 포함하고 있다. 이는 2025년까지 노인, 장애인 등 건강취약계층 12만 명을 대상으로 사물인터넷(IoT), 인공지능을 활용한 디지털 돌봄 시범사업을 추진하며 사물인터넷 센서, 인공지능 스피커를 통해 노인과 장애인 등 대상자의 맥박·혈당·활동 등을 감지하고 말벗·인지기능을 지원하겠다는 계획이다. 또한 2021년까지 노인·장애인의 신체활동과 간호·간병인의 업무보조 지원을 위해 욕창예방, 배설보조, 식사보조, 이동보조기구 탑승을 보조하는 돌봄로봇 4종을 개발하겠다는 계획도 포함되어 있다.[14]

14) 한국판 뉴딜 종합계획. 기획재정부 보도자료 (2020. 7. 14).

<표 1> 디지털 뉴딜 31개 대표사업

4대 분야	12대 과제	대표사업
DNA 생태계 강화	1. 데이터 구축・개방・활용	1)데이터 댐 2)공공데이터 개방・활용 활성화
		3)디지털 집현전
	2. 5G・AI 융합확산	4)문화유산 실감체험 및 XR 플래그십 프로젝트
		5)자율주행차량/자율운항선박
		6)스마트 공장 7)스마트 건설
		8)비대면디지털 기업육성 및 스마트 대한민국 펀드
	3. 5G・AI기반 지능형 정부	9)모바일 신분증 10)인공지능 국민비서
	4. K-사이버 방역체계 구축	11)양자 암호통신 인프라
		12)ICT 중소기업 보안강화 및 시스템안전진단
교육 인프라 디지털 전환	5. 디지털 교육인프라조성	13)K-에듀통합플랫폼 구축
	6. 온라인 교육강화	14)K-MOOC 활성화 15)공공스마트직업훈련 플랫폼
비대면 산업육성	7. 스마트의료 및 돌봄인프라 구축	16)안전하고 편리한 스마트 병원
		17)닥터 앤서2.0 18)모바일 건강지킴이
	8. 중소기업 원격근무 확산	19)중소기업 비대면 전환 20)공동화상 회의실
	9. 소상공인 온라인 비즈니스 지원	21)소상공인 맞춤형 온라인 판로지원
		22)스마트 상점・공방
SOC 디지털화	10. 인프라 디지털 관리체계 구축	23)국민안전 스마트인프라
		24)디지털 트윈 25)스마트 재해위험 알리미
	11. 도시・산단의 공간 디지털 혁신	26)스마트 시티 27)스마트 산단
	12. 스마트 물류체계 구축	28)스마트 육상물류 29)스마트 해운물류
		30)농축산물 유통 플랫폼
디지털 격차해소		31)디지털 배움터

그러나 이렇게 수집·공유된 데이터의 상당수는 개인의 동의 하에 연출된 일부 데이터를 제외하고는 대부분 실제 환경에서 수집된 데이터로서, 대규모 학습용 데이터에 국민의 개인정보나 민감정보가 포함되어 유출될 수 있다는 우려가 있다. 특히 헬스케어 분야 데이터나 건강취약계층을 대상으로 한 디지털 돌봄 사업에서 수집되는 데이터의 경우 질병정보 등 실제 건강에 관한 정보를 포함하고 있어 사생활 침해 위험성이 큰 민감정보이다. 공공데이터 소유권에 대한 문제로서, 기업이 국가를 통해 국민의 민감정보를 무료로 제공받아 다시 이익을 창출하는 게 맞지 않다는 지적도 제기된다. 정부는 배포 데이터에 대하여 개인정보가 포함되지 않도록 비식별화 처리를 거쳐 가공된 데이터라고 설명하지만, 가명 데이터는 다른 정보와의 결합 등을 통해 정보주체를 식별할 수 있는 가능성이 완전히 사라지지 않는 이상 정보인권 침해에 대한 우려가 있을 수밖에 없다.[15)]

자율주행, 안전 분야 데이터 또한 지하철 역사, 도로, 버스 내부와 같은 공공장소 CCTV로부터 수집된 영상을 가공하여 배포하고 있는데 이 역시 정보주체의 동의나 고지 없이 가공된 것이다. 이러한 영상 정보의 경우 사람 얼굴과 차량 번호판 등의 개인정보를 모자이크 또는 블러링 처리하는 방식으로 비식별화한다[16)]. 그러나 딥러닝 기술의 발전으로 이를 복원할 수 있는 디블러링(Deblurring)이나 슈퍼 레졸루션(Super Resolution)을 이용해 추후 얼굴이 검출되거나 인식되는 문제가 발생할 수 있다. 무엇보다 샘플데이터 등을 불특정 다수에 공개하거나 과학적 연구 목적의 범위를 벗어난 민간 사업에 가명정보를 공개하는 것은 개인정보보호법에 위배된다. 개인정보보호위원회는 2021년 인공지능 챗봇 이루다 사건에 대한 2021. 4. 28. 결정 제2021-007-072호에서 “개인정보처리자는 가명정보를 제공받는 자를 고려하여 ‘특정 개인을 알아보기 위하여 사용될 수 있는 정보’를 배제하여야 하는데, 불특정 다수에게 가명정보를 공개하는 행위는 불특정 다수 중 누군가는 공개하는 정보에 포함된 ‘특정 개인을 알아보기 위하여 사용될 수 있는 정보’와 결합하여 개인을 알아볼 수 있다는 점에서 허용된다고 보기 어렵다.“라고 지적한 바 있다.

도시 공간의 디지털 혁신 과제는 교통·방범 등 CCTV 연계 통합플랫폼의 구축, 스마

15) 국가인권위원회는 2016년「신용정보의 이용 및 보호에 관한 법률 일부 개정 법률안」, 2019년「개인정보보호법」등 ‘데이터 3법’ 개정에 대한 의견을 밝히며 비식별 개인정보 활용 요건을 구체화하고 안전성을 확보하는 등 정보주체의 개인정보 권리를 보호해야 한다고 지적했다.

16) 예를 들어 대전도시철도공사 (2021). 인공지능 데이터 구축·활용 가이드라인 : CCTV영상 AI 데이터, 2.4.1.4.절 “영상 데이터 비식별화 처리” 참조.

트 시티 솔루션 확산 및 스마트 시티 시범도시 조성 등의 내용을 포함하고 있다. 교통·방범(112)·방재(119) 등 과거 목적별로 운영되던 CCTV를 통합·연계하는 지자체 CCTV 통합플랫폼을 구축하고 스마트 횡단보도·수요응답형 대중교통·드론 배송 등 현장효과가 검증된 스마트 솔루션을 2025년까지 75개 이상 지자체로 확산시키겠다는 계획이다. 스마트 시티로의 발전은 비용절감, 생산성 향상 등 도시의 새로운 발전양상에 대한 기대와 동시에 보안 위협과 사생활 침해, 신기술 위주 정책으로 인한 불평등한 정보접근성 등의 문제점을 안고 있다. 특히 한국에서 진행 중인 스마트 시티 논의는 주로 4차 산업혁명시대의 신산업에 대응하는 차원에서 진행되는 바, 특정한 도시 문제 해결보다는 기술을 보유한 민간이 주체가 되어 다양한 인프라 간 데이터 공유를 위해 하나의 플랫폼으로서 도시 모델을 추구하려는 성격이 강해[17] 그 우려가 커질 수밖에 없다.

이러한 스마트 시티의 계획과 추진은 일반 감시 카메라보다 고도화된 스마트 폴(Pole), 스마트 보도 등 지능화 기기의 설치를 통해 방대한 시민들의 개인정보를 수집하는 방향으로 전개되고 있지만 개인정보보호 권리 침해에 대한 검토는 제대로 이뤄지지 않고 있다. 스마트 폴은 가로등 또는 신호등과 같은 도시 인프라에 감시 카메라, 공공 와이파이, 비상벨, 유동인구센서 등 사물인터넷 기술을 기반으로 주변기기를 설치하는 일련의 지능형 CCTV 시스템을 말하는데, 서울, 부천, 대전 등의 도시와 기초 자치구 단위에서 진행 중인 스마트 시티 조성 사업을 통해 지속적으로 구축되고 있다. 이는 특히 대다수의 지방자치단체에서 구체적 법적 근거 없이 운영 중인 CCTV 통합관제센터[18] 및 지능형 CCTV와 연계되어 사생활 침해의 가능성을 더 높이는 것으로 보인다. 유엔 프라이버시 특별보고관은 2021년 6월 25일 발표한 한국보고서[19]에서 제주지역 스마트 시티 사업을 검토한 바 있다. 특별보고관은 프라이버시권에 중대한 영향을 미치는 이들 사업에서 개인정보 영향평가조차 시행되지 않은 것에 우려를 표하고 가능한 한 빨리, 실행 단계 전에 그 실시를 요구하였다. 더불어 이들 사업은 ‘프라이버시 기본설계’ 및 ‘프라이버시 기본설정’을 고려하고 존중하며 반영해야 한다고 지적하였다(33문).

---

17) 김상민, 임태경 (2020). 지방자체단체의 스마트 시티 혁신 정책 추진 방향 - 스마트 시티와 사회혁신의 융합적 접근 모색 -. 한국지방행정연구원 연구보고서 2020-12.

18) 국가인권위원회는 법률적 근거 없이 운영 중인 CCTV 통합관제센터가 인권침해라고 지적하고 개선을 권고한 바 있다. 국가인권위원회 결정 2018. 6. 29. 폐쇄회로 텔레비전 통합관제센터 설치 및 운영에 대한 개선 권고.

19) Visit to the Republic of Korea - Report of the Special Rapporteur on the right to privacy, Joseph Cannataci. 유엔문서 A/HRC/46/37/Add.6 (2021. 6. 25).

2. 법무부 얼굴인식 시스템

법무부는 특정인의 얼굴과 지문을 입국 과정에서 수집된 외국인의 생체정보와 비교·분석하는 생체인식 정보 시스템을 운영하고 있으며, 최근에는 출입국 심사 과정 및 공항에서 위험인물을 실시간으로 식별하기 위한 얼굴인식 시스템을 구축하는 과정에서 수억 명의 생체인식 정보를 활용하고 있다.

법무부 출입국외국인정책본부는 2010년 G20 정상회의를 앞두고 전국 22개 공항과 항만에 지문인식기와 얼굴인식기를 설치해 외국인의 지문과 얼굴 정보를 확인하는 등 입국 심사 절차를 강화하였고 이어 출입국관리법 개정을 통해 입국하는 모든 외국인에 대해 생체정보 제공을 의무화하고 수집된 생체정보를 데이터베이스화하여 이용하여 왔다. 해당 데이터베이스는 얼굴인식 시스템인 바이오정보전문분석시스템(Biometrics Analysis System for Experts, BASE) 프로그램을 통해 외국인 용의자의 신원을 특정하거나 입국 심사 단계에서 활용되고 있는데, 보도에 의하면 2019년 기준 1억 5천만 장의 사진이 데이터베이스에 저장되어 있으며 2015년부터 2019년 6월까지 총 8,435건의 분석 의뢰를 받아 2,089건의 신원을 식별해냈다고 한다. 이는 특히 출입국심사 목적을 넘어 외국인 용의자의 신원 식별 등 수사 지원의 목적으로도 사용되고 있다.[20]

나아가 이러한 과정으로 수집된 생체인식 정보는 정보주체 당사자 동의 없이 공항 내 실시간 얼굴인식 시스템 구축을 위한 학습용 데이터로 활용되어 왔다. 2019년 법무부와 과학기술정보통신부는 업무 협약을 맺어 인공지능 기술을 기반으로 출입국 심사를 단순화하고 공항 내 위험인물을 실시간으로 식별·추적하는 시스템을 구축하고 고도화하기 위한 목적으로 ‘19년 인공지능 식별추적 시스템 실증 및 검증 사업’을 시작했다.

해당 사업은 크게 ①법무부가 보유한 생체인식 정보를 ‘실증랩’ 공간에서 사업 참여 민간기업 다수에 학습용 데이터로 제공 ②공항 출입국 관리 구역에 카메라를 설치하여 공항 내 이상행동에 대한 실제 환경 데이터를 수집해 학습용 데이터로 제공하는 것으로 나뉘며, 이후 감시카메라를 통해 공항 이용자의 신원 및 위험 상황을 실시간으로 식별하고 탐지하는 시스템을 갖추는 것을 목표로 하고 있다. 법무부가 보유한 생체인식

20) 바이오정보(얼굴사진, 지문) 분석으로 위험인물 입국 막는다. 법무부 보도자료 (2018. 1. 23), 콧수염-뿔테안경으로 가려도… 1초만에 “얼굴 일치도 78%”. 동아일보 보도 (2019. 12. 14).

정보, 즉 출입국 관리를 통해 수집한 외국인의 얼굴 사진은 약 2억에서 4억장에 달하는데 2020년 사업을 진행하며 이중 최소 1.2억장이 이미 참가 기업에게 제공되어 얼굴인식 알고리즘의 학습과 검증에 이용되었다. 또한 얼굴 사진을 포함한 내국인의 출입국 심사 정보도 인공지능 실증랩 데이터베이스에 이관되어 이용된 것으로 확인되며, 더불어 공항 출입국 관리 구역에 일련의 카메라 시스템을 구축해 인공지능 학습을 위한 얼굴인식 및 행동인식 용도의 실제 데이터도 수집하고 있다.[21]

3. 과학기술정보통신부와 공공분야 지능정보화

과학기술정보통신부는 공공부문에 인공지능, 빅데이터 등 신기술을 선도적으로 적용하는 ‘디지털 공공서비스 혁신 프로젝트’을 통해 매년 100~200억 원의 예산을 투입하여 여러 공공기관 및 지자체의 공공서비스에 신기술을 시범 적용하는 사업을 주도하고 있다. 여기에는 재난 시뮬레이션 모의훈련, 고전문헌 자동번역, 전화 금융사기 예방 등 일반적인 자동화와 행정처리의 고도화 사업이 주를 이루지만 취약계층을 대상으로 부정수급을 감시하는 서비스, 생체인식정보 등 특별한 보호가 필요한 민감정보를 광범위하게 이용하는 서비스, 여러 나라에서 고위험 또는 금지되는 인공지능으로 규제되는 공공장소 원격 생체인식 서비스 등이 포함되어있다.

먼저 2020년 사회보장정보원 주관의 ‘머신러닝·RPA기반의 사회서비스 바우처 부정수급 탐지시스템 구축’ 사업은 빅데이터 분석 및 인공지능 기반 머신러닝 예측 모델을 활용해 복지급여를 지급하거나 부정수급을 탐지하는 것을 목표로 하고 있다.[22] 그러나 인공지능을 활용해 특정한 부정수급의 양상을 예측하고 확인하여 적발하는 인공지능 기반의 행정처리는 불투명하고 위법한 개인정보 처리를 수반할 수 있다.[23] 특히 결과 예측

---

21) 정보통신산업진흥원 (2020). AI식별추적시스템구축 사업 의의와 성과. 이슈리포트 2020-제20호(2020. 12. 31.); 정부, 출입국 얼굴사진 1억7천만건 AI업체에 넘겼다. 한겨레 보도(2021. 10. 21).

22) 과학기술정보통신부, 한국정보화진흥원 (2020). 2020년도 ICT기반 공공서비스 촉진사업 온라인 설명회.

23) 2020년 네덜란드 헤이그 지방법원은 사회복지급여 부정수급 탐지 시스템에 대하여 투명성 부족과 개인정보보호법 위반을 이유로 운영을 중단하라는 취지의 판결을 하였다. Welfare surveillance system violates human rights, Dutch court rules. The Guardian 보도(2020. 2. 5).

이나 오작동을 수정하기 쉬운 단순 행정자동화에 비해, 행정처리가 어떤 과정을 거쳐서 특정한 결론에 도달했는지 설명 여부가 쉽지 않은 불투명한 인공지능 시스템의 경우 처분의 당사자인 시민이 설명을 들을 수 있는 권리나 구제를 받을 권리를 보장받지 못하는 등 헌법상 적법절차의 원칙에 어긋날 수 있다. 이러한 인공지능 적발시스템이 실제 복지를 위해 유용하게 기여하는가에 대한 문제제기 또한 있을 수 있다. 유엔 극빈 및 인권에 관한 특별보고관은 2019년 보고서를 통해 사회복지 분야에서 인공지능 등 신기술의 활용이 사생활에 대한 권리, 차별 금지 등의 전통적인 시민적, 정치적 권리에 위험을 가져온다고 경고하며 투명성 보장과 정책 결정 과정에 대한 광범위한 개입을 통해 인권을 존중해야 한다고 밝혔다.[24)]

2021년 경기도 부천시와 함께 진행한 CCTV 기반 지능형 역학시스템 사업은 특히 문제적이다. 해당 사업은 코로나19 대응을 위한 확진자 동선 추적 및 접촉자 파악과 같은 역학조사 과정을 보완하기 위해 ‘CCTV 영상의 탐색을 자동화’ 시키고 ‘기존 역학조사 지원시스템과 연계’ 함과 동시에, ‘양질의 인공지능 데이터셋 구축’ 을 통해 인공지능 시스템의 지속적인 고도화 기반을 조성하는 것을 목표로 한다.[25)] 해당 시스템은 보건소로부터 확진자의 기초 신상과 사진 정보를 취득한 후 얼굴인식을 기반으로 한 추적 인공지능 시스템이 연계된 CCTV를 통해 확진자의 동선을 추적하고 마스크 미착용자나 확진자와 2M 이내 접촉한 자를 자동으로 파악한다. 이에 더해 휴대전화 기지국 접속정보를 교차 검색하여 접촉자의 방문 장소와 이동 경로 및 신원 정보까지 확인하겠다는 계획을 내세우고 있다. 해당 사업은 또한 실제 현장 CCTV 영상 및 데이터셋을 수집하고 이를 어노테이션 및 라벨링할 수 있는 환경까지 구축하려는 목표를 가지고 있다. 부천시가 교통과 방범 등의 목적으로 설치 및 운영하는 CCTV를 통해 수집되는 부천시민의 실제 영상 데이터를 민간 인공지능 학습용 데이터로 사용하겠다는 계획이다.

---

24) Report of the Special Rapporteur on extreme poverty and human rights. 유엔문서 A/74/493 (2019. 10. 11).

25) 부천시 (2021a). “인공지능과 CCTV 영상을 이용한 지능형 역학시스템 구축”, 2021년도 ICT기반 공공서비스 촉진사업 제안요청서 (2021. 2. 2.); 부천시 (2021b). “인공지능과 CCTV 영상을 이용한 지능형 역학시스템 구축”, 2021 디지털 공공서비스 혁신 프로젝트 온라인 사업설명회 (2021. 2. 25.) 자료집(과학기술정보통신부·한국지능정보사회진흥원).

<그림 1> 경기도 부천시 지능형 역학시스템 추진목표

*자료: '인공지능과 CCTV 영상을 이용한 지능형 역학시스템 구축' 제안요청서 (2021. 2. 2.)

<그림 2> 경기도 부천시 지능형 역학시스템 흐름도

*자료: '인공지능과 CCTV 영상을 이용한 지능형 역학시스템 구축' 사업설명회 (2021. 2. 25.)

부천시는 CCTV 영상 데이터 사용을 위해 「개인정보보호법」 제28조의2(가명정보의 처리 등) 저촉 여부를 검토하였다고 밝혔지만 현행법 상 민감정보로 정의된 생체인식 정보인 얼굴인식정보가 정보주체의 동의 없이 처리되는 것이 위법은 아닌지 의문이다. 또한 2020년 개인정보보호위원회 등이 발행한 <보건의료 데이터 활용 가이드라인>에서는 지문 등 생체인식 정보에 대하여 가명처리 적용 가능성에 대한 판단을 유보하였으며 본인의 동의에 기반해 사용해야 함을 명시하고 있다. 부천시 등은 모자이크, 흐림, 삭제 등의 얼굴 마스킹 기술을 통해 개인의 얼굴을 익명·비식별화한다고 밝혔으나, 원본정보의 비식별 여부와 관계없이 변환된 얼굴인식 데이터에 일련번호가 부여되는 등 개인을 추적하고 식별할 수 있는 템플릿으로 처리되는 것은 여전히 개인정보보호법 상의 민감정보의 처리에 해당할 수 있다.

과학기술정보통신부는 이처럼 공공개발사업을 통하여 방대한 국민 개인정보를 보유하고 있는 여러 공공기관에서 공공장소 얼굴인식 및 행동인식 인공지능 시스템을 개발 및 배치하도록 지원하고 있다. 과학기술정보통신부는 앞서 살펴본 법무부, 부천시 사업 외에도 대구 수성구와 함께 '인공지능(AI)융합 국민안전 확보 및 신속대응 지원 사업' 또한 진행하였다. 그런데 이들 사업은 CCTV 등을 통해 확보한 국민의 실제 데이터를 생체인식정보로 가공한 후 사업 참여 민간기업 다수에 학습용 및 검증용으로 제공하였거나 제공할 계획이 있다는 공통점이 있다.[26] 해당 사업들은 보안 및 데이터 외부 유출을 막기 위해 '실증랩' 공간에서만 데이터에 접근하도록 조치를 취했지만, 수성구 사업에 참여하는 민간기업이 주민 얼굴 영상 10만여 건을 실증랩 밖으로 무단 반출하였다는 사실이 밝혀져 논란이 되었다.[27]

---

26) 과학기술정보통신부는 이들 업체가 개인정보 처리를 위탁받은 수탁자에 해당한다고 해석해 왔다. [보도설명] 출입국 얼굴사진 민간업체 이관 관련(한겨레). 과학기술정보통신부 보도자료 (2021. 10. 21.) 참조. 그러나 해당 사업의 개발자로 특정 업체를 선정하지 않은 채 다수 기업을 동시에 참여시키고 참여 기업 대다수가 자사 솔루션의 기능을 독자적으로 향상시키고 지적재산권을 취득해 왔다는 점에서 불법적인 제3자 제공에 해당할 가능성이 있다.

27) 정부가 '연구용' 줬더니, 얼굴 영상 10만건 빼돌렸다. 한겨레 보도 (2021. 11. 17).

4. 경찰청 얼굴인식과 범죄예측 시스템

경찰청은 9대 수법범죄자를 대상으로 3D 얼굴인식 데이터베이스 및 검색 시스템과 3D 영상 촬영 시스템을 운영하고 있다. 경찰청은 얼굴인식 기술을 통해 CCTV, 블랙박스, 채증 장비 등의 다양한 영상정보 처리기기에서 수집된 영상을 적극 활용하고자 2014년 '3D 얼굴인식 및 3D 얼굴영상 변환 시스템 개발 사업'으로 얼굴인식 시스템을 개발하고, 기존 구속피의자의 사진 자료를 다양한 각도로 인식할 수 있도록 3D로 변환시켰다.[28] 2017년에는 전국 경찰서에 3D 영상 촬영 시스템을 보급해 구속피의자에 대한 3D 촬영을 시작하였다. 이 구속피의자 3D 데이터베이스의 경우 2018년 18만명,[29] 2019년 19만명 이상의 3D 얼굴 사진이 포함되어있는 것으로 밝혀졌다.

경찰은 이후 매년 이를 고도화시키기 위해 2019년까지 총 19억 3천만원의 예산을 투입했으며, 2024년에는 얼굴인식 시스템을 실시간으로 CCTV와 연계하여 검색하려는 사업 계획을 갖고 있다.

<표 2> '범죄예방 3D얼굴인식시스템' 연차별('20년~'24년) 고도화 사업 계획

년도	요구 예산	얼굴인식 성능 고도화 사업 내용(예정)
'20	2억	○ 어두운(저조도 40Lux이상) 범죄현장에서 찍힌 용의자 얼굴인식 개발(CCTV 각도 : 상단 30도 & 좌우 45도 이내) ※ 일반조도 : 300 ~ 500Lux
'21	2억	○ 어두운(저조도 40Lux이상) 범죄현장에서 찍힌 용의자 얼굴인식 개발(CCTV 각도 : 하단 15도 & 좌우 45도 이내)
'22	2억	○ 마스크 착용 등 얼굴 사진이 일부 가릴 경우 얼굴인식(정면) 개발 등
'23	2억	○ 범죄현장 동영상 속 얼굴 사진 인식 등
'24	2억	○ 실시간 CCTV 연계 얼굴인식 등

* 자료: 2019년 정기국회 국정감사 자료(정인화 의원)

28) 경찰청 (2014). 3D 얼굴인식 및 3D 얼굴영상 변환 시스템 개발 사업 제안요청서 (2014. 3).
29) CCTV 찍힌 범인이 흐리다면?…'3D 얼굴 인식'의 진화. KBS 보도 (2018. 12. 17).

<표 3> 범죄별 DB 구축 현황(9대 수법범죄)

구분	계	강도	절도	사기	위・변조	약취유인
DB(건)	198,330	18,777	77,068	49,439	8,911	963
구분	공갈	방화	강간	장물	기타	
DB(건)	6,368	3,121	29,421	1,707	2,555	

*자료: 2019년 정기국회 국정감사 자료(정인화 의원)

현재 경찰 등 수사기관이 3D 등 얼굴인식 데이터베이스를 구축하고 이를 자동으로 비교・대조할 수 있는 시스템을 운영하도록 허용하는 법적 근거는 없다. 공개된 장소에 설치된 공공기관 CCTV만 최소 100만 대에 이르는 등 영상을 통한 감시망이 촘촘한 환경을 고려해 보았을 때, 아무런 통제 장치 없이 얼굴인식 기술이 도입된다면 다수의 정보주체 시민이 공공장소에서 은밀한 대량 감시의 대상이 되는 등 정보인권이 침해될 수 있다. 농성에 참가한 노동조합원이나 철거민은 물론 집회시위 참가자들까지 광범위하게 그 채취와 보관 대상으로 규정한 우리나라 디엔에이신원확인정보 데이터베이스의 사례를 상기하여 보았을 때,[30] 광범위한 대상이 얼굴인식 등 민감한 생체인식정보의 수집과 저장, 검색 및 활용 대상이 될 우려가 있다.

한편 경찰청은 치안・공공데이터를 통합한 빅데이터를 인공지능으로 분석하여 지역별 범죄위험도와 범죄발생 건수를 예측하는 '범죄위험도 예측분석 시스템(Pre-CAS)'을 개발하여 전국에서 운영하고 있다. 2019년 9월 경찰청은 '빅데이터 통합 플랫폼'을 통해 각종 경찰 치안 빅데이터를 통합・관리하고 한국형 인공지능 기반 범죄예측 시스템을 구축하겠다는 목표를 갖고 '스마트 치안 구현단'과 '빅데이터 전담 부서'를 신설했다.[31] 이후 2019년 11월 경찰청은 행정안전부 국가정보자원관리원과 함께 인천 지역의 빅데이터를 분석한 바탕으로 '범죄위험도 예측분석 시스템'의 개발을 시작했으며 이를 자체연구 등을 통해 고도화하여 2021년 5월 전국 운영을 시작했다.[32]

그러나 데이터 기반 기술을 활용한 범죄 예측에는 상당한 위험성과 차별의 가능성이 뒤따른다. 경찰청은 미국의 범죄 예측 프로그램 프레드폴(PredPol)을 참조하여 한국의 치

30) 용산참사 11년 지났는데…철거민 DNA '막무가내 채취'. CBS노컷뉴스 보도 (2020. 1. 30).
31) 인공지능·빅데이터 활용, 치안 서비스 수준을 높인다. 경찰청 보도자료 (2019. 9. 6).
32) 빅데이터를 통한 범죄예측, 첫 발을 내딛다. 경찰청 보도자료 (2019. 11. 29.); 경찰, 빅데이터·인공지능(AI) 활용한 범죄예방활동 전국 확대. 경찰청 보도자료 (2021. 4. 30).

안 환경 특성에 맞춰 구축하는 것을 목표로 해당 시스템을 개발하였는데, 프레드폴은 데이터의 질과 유형으로 인한 편향성 문제와 사생활의 권리 침해로 이미 많은 비판을 받아온 시스템이다. 먼저 경찰이 보유한 기존의 데이터에 대한 문제가 제기될 수 있다. 미국 경찰이 유색인종 위주 거주 지역을 우범지대로 분류하여 순찰과 검문을 강화한 결과, 이들 지역의 범죄율이 백인 위주 거주 지역에 비해 높게 형성되었으며, 이 데이터를 학습한 범죄 예측 시스템은 또다시 유색인종 거주 지역의 범죄율을 높게 예측하고 이는 반복적으로 경찰이 그 지역을 집중하여 순찰하게 만드는 피드백 순환 구조를 갖는다. 또한 분석 데이터에 경범죄가 포함될 경우, 빈곤한 거주 지역일수록 거리에서 경범죄에 해당하는 행위가 이루어지기 때문에 경찰에 보다 쉽게 적발되고, 이는 결과적으로 경찰이 편향적으로 빈곤 지역을 집중 순찰하는 불균형을 강화시킬 수 있다. 이러한 악순환은 공통적으로 낙후된 지역에 대한 낙인 효과를 강화시킴과 동시에 데이터 기반 범죄예측을 스스로 정당화시킨다는 특징이 있다.[33]

한국의 범죄예측 시스템 또한 이러한 위험성을 피해갈 수 없을 것으로 보인다. 경찰청은 '범죄위험도 예측분석 시스템'을 위해 기존의 치안 데이터뿐 아니라 경제활동인구, 실업률과 고용률, 건물노후도, 공시지가 등 경제적 지표를 함께 활용한다. 또한 범죄 데이터에는 강도, 절도와 같은 주요 범죄 뿐 아니라, 경범죄에 해당하는 '무질서 행위'를 포함하고 있는데[34] 여기에는 시비, 행패, 소란, 청소년비행, 무전취식, 무임승차, 주취, 보호조치 대상자, 위험행위, 소음, 노점상 등이 해당할 수 있다.

<표 4> '범죄위험도 예측분석 시스템' 위험도 예측을 위한 분석 데이터 종류

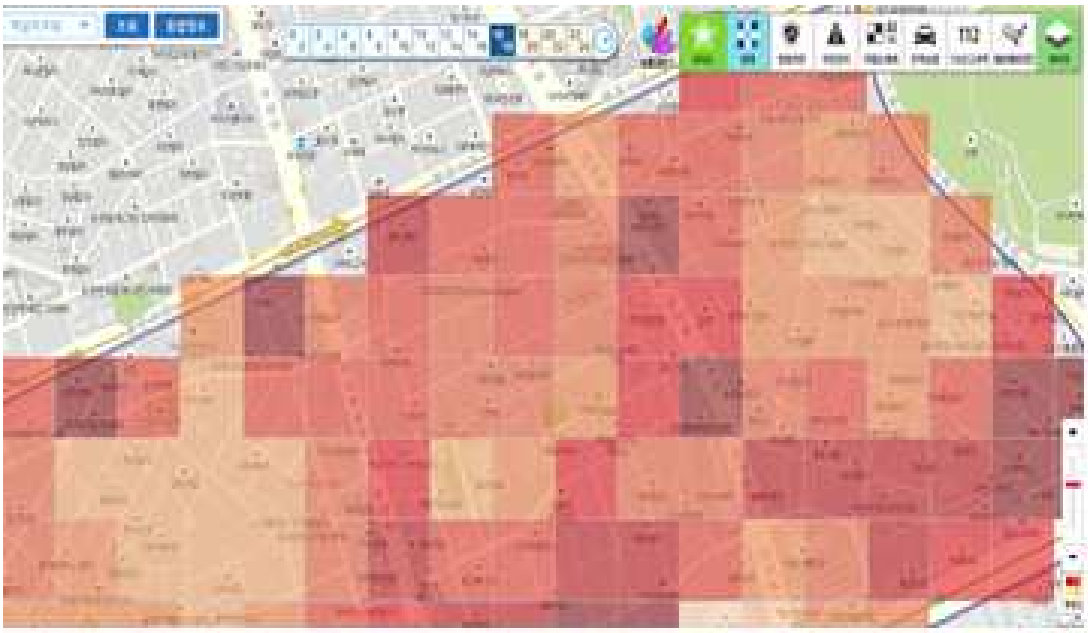
구분	종류
치안데이터	범죄(KICS), 112신고, 유흥시설 수, 교통사고 수, 경찰관 수 등
공공데이터	인구(전입・전출・거주), 기상, 요일, 면적, 경제활동참가율, 실업률, 고용률, 건물유형・노후도, 공시지가, 토지용도, 학교, 공원, 관광지, 소상공인 업소 등

* 자료 : 인공지능(AI)으로 범죄예방의 첫걸음 내디딘다. 경찰청 보도자료 (2021. 3. 2.)

33) 캐시 오닐 (2017). "5장 무고한 희생자들 : 가난이 범죄가 되는 미래". 「대량살상수학무기」. 흐름출판.; AI NOW Institute (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice.

34) 인공지능(AI)으로 범죄예방의 첫걸음 내디딘다. 경찰청 보도자료 (2021. 3. 2).

<그림3> ‘범죄위험도 예측분석 시스템’ 구역별 위험도 예측 화면



* 자료 : 인공지능(AI)으로 범죄예방의 첫걸음 내디딘다. 경찰청 보도자료 (2021. 3. 2.)

한편 2018년 8월, 서울시 민생사법경찰단은 불법대부, 다단계 판매와 같은 민생범죄 수사에 ‘인공지능 수사관’을 도입하였다가 개인정보보호위원회의 위법성 지적 이후 이를 철회한 바 있다. ‘인공지능 수사관’은 SNS, 블로그 등 온라인 플랫폼에 올라오는 게시글이나 이미지를 크롤링, 즉 실시간으로 수집 저장하고 불법성 컨텐츠의 패턴을 학습시킨 인공지능 시스템을 활용해 자동으로 불법성 컨텐츠를 찾아냄과 동시에 관련한 개인정보를 탐지하는 시스템으로 수사관이 직접 검색하는 등 육안으로 확인했던 업무를 자동화시킨 시스템이다. 그러나 2019년 5월, 개인정보보호위원회는 민생사법경찰단이 해당 시스템을 이용해 인터넷 SNS 등 온라인에 공개된 게시물을 광범위하게 수집하고, 범죄 관련성이 높다고 판단되는 게시물을 분석해 해당 게시물에 포함된 성명, 아이디, 전화번호, 주소, 업체명 등을 이용한 것은 사실상 구체적 법령에 근거하지 않고 이뤄지는 온라인 불심검문과 유사한 것으로 판단하여 개인정보보호법 제15조 제1항 제1호, 제2호 및 제3호에 위반된다는 결정을 내렸다.[35)]

5. 군사안보 분야의 인공지능 활용과 개발

국방부는 4차 산업혁명에 대응하여 ‘스마트 국방혁신 추진계획’을 수립하는 것을 시작으로 국방 분야에서 인공지능 등 신기술을 활용하기 위한 종합계획, 정책추진을 지속적으로 발표하고 진행하고 있다. 2019년 국방부는 스마트 국방운영, 국방혁신 촉진을 위한 기반 인프라 조성, 전략체계 관련 군사력 고도화 촉진 등을 위해 국방부 차관을 단장으로 하는 ‘4차 산업혁명 스마트 국방혁신 추진단’을 출범시켰다. 이러한 스마트 국방은 무기체계의 지능화, 훈련체계의 고도화, 스마트한 병영환경 조성 등 크게 세 가지로 나눠서 추진되고 있으며 일부 사업과 정책은 장병의 개인정보 침해, 자율살상무기로의 발전 등 우려되는 지점이 보인다.

특히 2020년부터 진행 중인 육군의 ‘스마트 부대 구축사업’ 및 ‘지능형 출입통제체계 구축사업’과 해군의 ‘스마트 전투함 사업’은 효율적인 부대 지휘통제와 감시를 목적으로 장병의 개인정보를 실시간으로 수집하고 처리하는 내용을 담고 있어 논란이 되었다.[36] 해당 사업은 개개인의 장병에게 지급한 웨어러블 기기를 통해 실시간으로 심박수, 혈압 등의 건강정보와 함께 위치정보를 수집하고 이를 통해 병력현황과 건강상태를 실시간으로 체크하는 등의 내용을 담고 있다. 또한 육군은 부대 생활관 복도에 실시간 얼굴인식 및 상황분석이 가능한 지능형 CCTV를 설치하여 싸움·구타·실신 등 안전사고를 파악하는 감시체계를 구축하겠다는 계획을 밝히기도 했다.[37]

한편, 한국은 이미 이스라엘 IAI사의 하피(Harpy)와 같은 자폭형 무인기, 패트리어트나 사드(THAAD) 등의 미사일 요격 시스템, 미래전 수행을 위해 창설된 육군의 드론봇전투단과 같은 준자율 수준의 무인 무기는 물론이고, 목표물을 감지하면 자동으로 조준해서 사격할 수 있는 자동발사기능을 탑재하여 자율살상무기로 분류되는 센트리건(Sentry Gun)을 비무장지대에 배치하는 등 인공지능의 무기화를 빠르게 추구하는 것으로 보인다.[38] 동시에 한국 정부는 인공지능 기술이 적용된 무기체계인 ‘자율살상무기체계(LAWS)’[39]에 관한 국제사회의 논의에서 사실상 규제에 반대하는 태도를 보이고 있어

35) 개인정보보호위원회 2019. 5. 27. 결정 제2019-09-130호.
36) 생활관 CCTV 설치, 실시간 장병 위치추적, 건강정보 수집… 기막힌 軍 스마트 사업, 군인권센터 보도자료 (2021. 1. 11).
37) 육군 병사 웨어러블기기 지급…생활관 CCTV 설치 추진, 뉴시스 보도 (2020. 12. 31).
38) 피스모모 (2021). 평화는 모두의 권리- 첨단기술과 평화권 애드보커시.

비판을 받고 있다. 학계의 보이콧 선언과 시민사회의 반대 캠페인 등 2010년 이후 자율살상무기체계에 대한 경각심과 우려가 급증하였고, 각국 정부들 또한 유엔 '특정재래식무기금지협약(CCW)' 체약국 회의를 통해 2014년부터 국제적 규제를 논의하고 있다. 그러나 한국은 미국과 러시아, 중국, 이스라엘 등의 국가와 함께 자율살상무기체계에 대한 국제적 규제가 불필요하다는 입장과 함께 그 도입과 사용에 관해 긍정적인 의견을 낸 것으로 보고되고 있다.[40]

39) 자율살상무기체계(LAWS)에 대해 국제적으로 표준화되고 합의된 정의는 없으나, 일반적으로 무기체계의 운용 과정에서 인간의 개입이나 통제 없이 스스로 목표를 탐색하고 정하여 공격을 실행할 수 있는 무기체계를 말한다.

40) 김민혁, 김재오 (2020). 자율살상무기체계에 대한 국제적 쟁점과 선제적 대응방향. 국방연구, 63(1), 171-204.

제2절 민간영역의 인공지능 활용

1. 인공지능 채용 도구

2018년 이후 인공지능을 활용한 채용 도구가 널리 도입되었다. 2019년 조사 자료에 의하면 인공지능을 활용해 신규 채용을 진행하고 있는 기업이 11.4%, 활용 계획이 있는 기업이 10.7%였으며,[41] 공공기관의 경우 350곳의 공공기관 중 최소 37곳이 인공지능을 활용해온 것으로 확인된다.[42] 인공지능을 활용한 채용 도구는 인공지능 면접과 인공지능 서류 평가 등이 있으며 자세한 내용은 다음과 같다.

인공지능 면접 또는 역량검사라고 불리우는 과정은 기본적으로 온라인 영상 면접으로 진행되며 일부 과정은 게임화된 인적성 검사를 추가로 포함하고 있다. 온라인 영상 면접의 경우 지원자의 표정・감정・안구 움직임 등의 얼굴 정보와 목소리・톤・크기・속도・음색 등의 음성정보를 추출하여 매력도・의사 표현・감정 전달력・호감도 등 외형적인 특성을 분석하고, 게임화된 인적성 검사 수행을 통해 정답과 오답・응답 속도・의사결정・학습 속도 등을 평가하고 분석한다고 알려져 있다.[43] 한 인공지능 면접 도구 개발 기업에 의하면 이를 위해 실제 재직 중인 고성과자와 저성과자의 인공지능 면접 응시 데이터・실제 성과 데이터, 우수 면접관・주요 대기업 인사총괄・산업심리학 교수 등이 면접 및 대화 영상을 분석하고 평가한 데이터 등이 사용된 것을 확인할 수 있다.[44]

인공지능 서류 평가는 지원자의 이력서, 자기소개서 등의 채용 서류를 분석해주는 도구로, 문장 및 맞춤법 오류 분석・표절 검사・주요 내용 요약과 같은 기본적 검수 기능과 더불어 채용 서류에 대한 점수와 직무 적합도를 인공지능이 평가하는 것으로 알려져 있다.[45] 채용 서류에 대한 인공지능 평가의 경우, 지원자 자기소개서 속 단어와 문장에서 능력・경험・신념・가치관・포부・지원동기 등으로 분류된 특성을 찾아내 고성과자

41) 대기업 채용“줄인다”34%“늘린다”18%. 한국경제연구원 보도자료 (2019. 9. 16).
42) <ALIO : 공공기관 경영정보 공개시스템> 및 각 기관 채용 공고를 통한 자체 조사에 의했다.
43) 투명성·공정성·신뢰성…AI면접 믿을 만할까?. 한겨레21 제1335호 보도 (2020. 10. 23).
44) 기업에 가장 적합한 인재를 과학적, 객관적으로 판단하는 분석 기술. 마이다스인 홍보자료 <https://www.midashri.com/intro/process/ai-interview/anal (검색일: 2021. 11. 1.)>
45) 카피킬러HR에 적용된 AI기술. 주) 무하유 홍보자료 <https://www.hr.copykiller.com/technology (검색일: 2021. 11. 1.)>

및 저성과자의 특성과 유사한지 여부를 판단한다고 주장하고 있으며, 해당 도구는 50만 건 이상의 자기소개서를 학습한 인공지능 모델을 활용하여 인공지능 서류 평가 도입 기업의 합격자 자기소개서 데이터를 기반으로 '우수 인재'의 패턴을 학습한 것으로 알려져 있다.

그러나 인공지능과 데이터 기반 기술이 과거의 누적된 차별로 대표성이 충분치 않은 데이터에 의존할 경우 차별이 모방되고 확산될 위험이 있다. 고용률과 임금 격차에 대한 노동 인구 데이터에는 성별, 성적 지향과 성 정체성, 장애, 나이, 인종과 민족, 학력, 지역 등 다양한 요소에 기반하여 역사적으로 누적되어 온 오랜 기간의 차별이 반영되어 있는데 이러한 편향성은 인공지능 기술이 불평등과 차별이 반영된 데이터를 학습하는 결과로 이어질 수 있다. 대표적인 사례로 2018년 인공지능을 활용한 채용 시스템을 폐기한 전자상거래 기업 아마존의 경우를 들 수 있다. 남성 비율이 높은 기술 업계의 데이터를 기반으로 만들어진 아마존의 채용 시스템이 여성과 관련된 키워드를 자동으로 감점 요소로 분류하는 등 차별적 결정을 내렸다는 사실이 밝혀진 것이다. 채용과 노동은 사람들의 미래와 현재의 경제 상황에 직접적인 영향을 미칠 수 있기에 불투명한 인공지능 기술을 도입하는 것이 특히 위험한 영역이라고 할 수 있다.

또한 채용 과정에서 성별, 연령, 신체조건, 용모, 출신지역 등으로 차별하지 말 것을 규정한 채용 공정화 관련 법제도를 인공지능을 활용한 채용 도구가 잘 준수하고 있는지에 대해 점검이 필요해 보인다. 보다 공정하고 투명한 인사절차를 거쳐야 할 공공기관마저 책임성 없이 인공지능 채용 도구를 무분별하게 도입하는 경향이 있다는 점은 특히 우려스럽다. 국회 국정감사에서 인천공항공사는 인공지능 면접을 도입할 때 지원자 평가 방법과 알고리즘에 대한 기술적 검토 또는 외부 자문을 거치지 않았다는 사실이 드러났으며 한국방송통신전파진흥원은 자기관이 탈락시킨 지원자에 대하여 인공지능 면접이 어떤 기준을 적용하였는지 파악하고 있지 못한 것으로 나타났다.[46]

46) 인천공항·한국공항공사, 인공지능의 차별 학습 및 편향성 대비 없이 무책임 AI면접 도입. 심상정 국회의원 보도자료 (2020. 10. 22.); 청년 앞길 막는 AI면접… AI윤리기준 투명성 공정성 필요. 정필모 국회의원 과기정통부 예산심사 질의서 (2020. 11. 5).

2. 금융 서비스

ICT서비스를 통해 소비자로부터 다양한 데이터를 수집해온 기술기업부터 전통적인 대형 금융사까지, 금융 영역 전반적으로 비금융정보를 이용하는 새로운 신용평가 모델이 도입되고 있다.[47] 이러한 '대안신용평가'는 기존의 신용평가사가 개인의 대출, 카드, 연체 등 이력을 이용해 책정한 신용등급에서 한 발 나아가 스마트폰 등을 통해 수집되는 수많은 디지털 기록 및 비금융정보를 알고리즘으로 분석하는 등 개인을 프로파일링하여 신용등급을 내리는 것으로 정의된다.

대안신용평가는 제도권 금융에서 소외된 취약계층의 신용도를 평가할 수 있다는 장점이 있으나 평가를 위한 비금융정보의 해석 과정에서 오히려 취약계층을 배제하는 방향으로 악용될 수도 있다. 주로 활용되는 비금융정보로는 통신정보, 결제정보, 휴대폰 사용기록이 있다. 구체적으로는 통신요금 납부 이력, 통화 사용량 및 시간대, 문자메시지 사용 패턴, 스마트폰의 배터리 충전 주기, 운영체제, 주로 사용하는 앱 카테고리, 타이핑 속도를 수집하고 분석해 사용자의 생활습관 및 의사소통의 성격을 파악하여 신용등급에 반영하는 방식이다. 그러나 위와 같은 비금융정보가 실제 사용자의 연체율과 유의미한 상관관계가 있는지, 합리적인 평가가 이뤄질 수 있는지 의문이 제기된다. 더불어 개인이 민감하게 생각하는 기록까지 축적해 분류하는 만큼 특정인의 일거수일투족이 노출되는 등 사생활 침해 위험성이 커지고, 나아가 디지털 기록이 남는 모든 활동이 평가의 대상이 될 수 있기에 개인의 표현의 자유가 위축되는 등 일상생활에 지장이 생길 위험도 있다.

또한 부정확한 데이터의 사용으로 인해 잘못된 평가로 이어질 수 있으며, 정확하게 기록된 데이터를 사용할지라도 평가 대상자는 물론이고 사용자에게도 불투명하게 작동하는 인공지능 프로파일링이 예측할 수 없던 결과나 의도치 않은 차별을 야기할 수도 있다. 특히 성별, 인종, 지역과 같은 민감한 데이터의 경우 그 직접적인 이용을 배제하더라도 설정된 판단기준과 데이터가 특정한 계층을 식별하는 대리지표 역할을 하여 결과적으로 차별적 속성을 기준으로 한 평가가 이루어질 수 있다. 프로파일링으로 인한 차별과 관련하여, 미국의 연방거래위원회는 대출 상환 기록이 저조한 사람들과 같은 상점을 이용한 방문자들의 경우 동반하여 신용한도가 낮아진다는 증거를 찾아낸 바 있다.[48] 이 경우 신

47) 휴대폰 충전 주기, 인터넷 검색시간…당신의 모든 게 신용이 된다. 한겨레 보도 (2021. 8. 26).

용한도가 낮아진 사람들은 특정한 사회경제적 상황에 속하기 때문이 아니라 해당 상황에 속한 사람들과 공통된 요소를 가지고 있다는 이유만으로 차별을 받게 된다.

3. 플랫폼 노동

플랫폼 노동 기업은 노동 통제를 위해 인공지능 기술을 적극 이용하고 있다. 특히 플랫폼 노동의 대표적 사례인 배달과 대리운전 등의 서비스를 운영하는 기업은 인공지능 시스템을 통해 소비자와 노동자 간 연결을 자동으로 주선하거나 기본적인 근태를 관리하고 감독하는 것으로 보인다.

주요 배달 플랫폼인 '배달의 민족', '쿠팡이츠', '요기요' 등은 인공지능 배차의 알고리즘이 빠르고 효율적인 배차를 이뤄낸다고 주장하지만, 당사자인 배달 노동자들은 알고리즘이 노동통제 및 사고 발생과 불공정한 배달료로 인한 임금삭감의 문제를 유발한다고 반박하고 있다. 기존에 플랫폼을 통한 배달노동은 픽업 위치와 배달지역 등 동선을 고려해 여러 건의 주문을 고려하여 배달하는 방식이 보편화되어 있었으나, 최근 플랫폼 측에서 인공지능 시스템을 통해 픽업 위치와 배달지역을 일방적으로 정해주는 'AI 추천배차' 방식을 도입하면서 당사자가 납득할 수 없는 경로의 배차가 강제되었다고 한다. AI추천배차가 교통상황과 실제 환경을 무시한 인공적 배달 경로를 제시해 배달노동자의 사고 발생 가능성을 높이고 있다는 점과, AI추천배차를 거부하면 경고, 계정 정지, 평가 하락, 배달료 차감 등 불이익을 주는 점 또한 논란이 되고 있다. AI추천배차와 관련하여 배달노동자 노동조합인 '라이더 유니온'은 주요 배달 플랫폼의 인공지능 알고리즘을 직접 검증하는 실험을 거쳐 다음과 같은 결과를 발표했다.[49]

48) Federal Trade Commission (2016). Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues (FTC Report).

49) 배달플랫폼의 AI노동통제, 라이더가 위험하다. 라이더유니온 AI검증결과 발표 기자간담회 자료 (2021. 6. 29).

AI 100% 수락시, 주행거리 늘고 노동 강도는 높아지는데 수익은 감소.

배민 AI가 설정한 직선거리 4.3km 배달을 15분 안에 가라는 배달 수락했더니 실 거리는 8.4km 나오고 배달시간은 24분 걸려.

요기요 AI는 먼저 들어온 주문을 나중에 배차하라고 지시.

검증에 참여한 쿠팡이츠 라이더, 둘째 날 자율적으로 거절하면서 배달했다가 계정 정지 당해.

요기요 라이더는 95% 수락율 유지하지 않으면 배달 한 건당 1,000원 마이너스, 등급하락 등으로 피해가 커, 둘째 날 실험에 참가하지도 못해.

배민은 과도한 거절시 다음에 배차가 지연될 수 있음을 경고.

AI알고리즘 시스템 효율적이지도 안정적이지도 않는데, AI 알고리즘 거절하는 것에 페널티를 주는 것은 부당.

라이더들 자율적으로 AI 배차를 거절하거나 선택할 경우 AI 100% 수락에 비해 주행거리 짧아지고, 수익은 늘어나는 경향.

6월 9일 교통법규 완벽히 지켜 배달할 경우 배달 한 건 당 약 30분 소요. 한 시간에 두 건이 최선. 안전배달료 등 건당요금체계 해결 없으면 단속만으로는 한계.

쿠팡이츠는 최소배달단가 2500원, 부산 배민은 2600원. 낮은 배달단가 1초마다 바뀌는 실시간 배달료에 라이더들 불안과 과속 유발.

한편 플랫폼 기업의 인공지능 활용은 광범위한 이용자 데이터 축적과 불투명하고 독점적인 데이터 사용으로 이어져 노사 관계를 더욱 불균형하게 만든다는 지적을 받고 있다. 배달노동자와 소비자, 음식점주는 자신의 활동 결과인 데이터에 대한 접근은 커녕 어떠한 종류의 개인정보가 수집되는지 파악하기조차 힘든 것에 비해, 플랫폼 기업은 모든 데이터를 독점하고 이를 통해 주문비용, 중개비용, 배달비용에 대한 기준을 정하면서도 인공지능 시스템의 뒤에 숨어 사용자성을 감추고 있다는 것이다.[50]

50) 박정훈 (2020).「배달의민족은 배달하지 않는다: 라이더가 말하는 한국형 플랫폼 노동」. 빨간소금.

4. 챗봇 이루다 논란

‘이루다’는 2020년 12월 출시된 페이스북 메신저 기반의 인공지능 챗봇 서비스로, 고위험 영역의 인공지능 시스템은 아니지만 국내에서 인공지능의 위법성 논란이 크게 불거진 서비스 중 하나다. 실제 스무살 여대생과 대화를 나누는 것 같은 자연스러운 대화 능력을 선보여, 출시 2주 만에 약 75만명에 달하는 이용자를 모으며 10~20대 사이에서 크게 인기를 끌었다. 그러나 일부 남성 이용자들이 이루다를 성적 대상으로 취급하고, 이루다 발화 내용에서 여성・성소수자・장애인・흑인 등을 혐오하는 내용이 발견되어 인공지능 윤리와 차별 논란을 빚었다. 무엇보다 이 서비스에서 개인정보보호법에 대한 중대한 위반 문제가 불거져 개인정보보호위원회가 개발 회사에 대한 행정조사를 실시하고 행정처분을 내린 바 있다.[51)]

개인정보보호위원회는 2021년 1월 12일부터 3월 25일까지 이루다 개발사 (주)스캐터랩의 개인정보 취급・운영 실태 및 개인정보보호 법규 위반 여부를 조사하였고, 4월 28일 확인된 위법 행위에 대하여 총 1억 330만원의 과징금과 과태료 등을 부과했다. 개인정보위원회의 조사 결과, 스캐터랩은 자사의 앱 서비스인 ‘텍스트앳’과 ‘연애의 과학’에서 수집한 카카오톡 대화를 2020년 2월부터 2021년 1월까지 페이스북 이용자 대상의 챗봇 서비스인 이루다의 인공지능 개발과 운영에 이용한 것으로 확인되었다. 스캐터랩은 이루다 인공지능 모델의 개발을 위한 알고리즘 학습 과정에서, 카카오톡 대화에 포함된 이름, 휴대전화번호, 주소 등의 개인정보를 삭제하거나 암호화하는 등의 조치를 전혀 하지 않고, 약 60만 명에 달하는 이용자의 카카오톡 대화문장 94억여 건을 이용하였고, 이루다 서비스 운영 과정에서는, 20대 여성의 카카오톡 대화문장 약 1억 건을 응답 DB로 구축하고, 이루다가 이 중 한 문장을 선택하여 발화할 수 있도록 운영하였다. 보호위원회는 텍스트앳과 연애의 과학 개인정보 처리방침에 ‘신규 서비스 개발’을 포함시켜 이용자가 로그인함으로써 동의한 것으로 간주하는 것만으로는 이용자가 이루다와 같은 ‘신규 서비스 개발’ 목적의 이용에 동의하였다고 보기 어렵고, ‘신규 서비스 개발’이라는 기재만으로 이용자가 이루다 개발과 운영에 카카오톡 대화가 이용될 것에 대해

51) 개인정보위, '이루다' 개발사 (주)스캐터랩에 과징금·과태료 등 제재 처분. 개인정보보호위원회 보도자료(2021. 4. 29).

예상하기도 어려우며, 이용자의 개인정보 자기결정권이 제한되는 등 이용자가 예측할 수 없는 손해를 입을 우려가 있다는 이유로 스캐터랩이 개인정보를 수집 목적을 벗어나 이용한 것이라고 판단하였다. 또한 스캐터랩이 개발자들의 코드 공유 및 협업 사이트로 알려진 Github에 2019년 10월부터 2021년 1월까지 이름 22건(성은 미포함)과 지명정보(구·동 단위) 34건, 성별, 대화 상대방과의 관계(친구 또는 연인) 등이 포함된 카카오톡 대화 문장 1,431건과 함께 인공지능 모델을 게시한 것에 대하여는, 가명정보를 불특정 다수에게 제공하면서 '특정 개인을 알아보기 위하여 사용될 수 있는 정보'를 포함하였다는 이유로 개인정보보호법 제28조의2제2항을 위반한 것이라고 판단하였다.

제3장 인공지능과 국제 규범

인공지능과 관련한 국제 규범은 비교적 자율적인 윤리 규범으로부터 논의가 시작되어 최근에는 의무적인 법규의 형식으로 발달하고 있다.

유엔 규범의 경우 인공지능에 대하여도 국제 인권법 체계에 따른 인권 보호, 존중, 구제의 실현을 요구하고 있다. 유엔 인권규범에 따르면 국가는 적절한 정책·규제 및 심사를 통해 기업을 포함한 제3자에 의한 인권 침해로부터 사람들을 보호할 의무가 있고, 기업은 타인의 권리 침해를 방지하고 인권에 부정적 영향을 주는 문제들을 해결하는 등 인권 존중에 관한 책임이 있으며, 사법적·비사법적 메커니즘을 포함하여 피해자의 접근성이 보장되는 실효성 있는 구제수단 마련이 필히 요구된다. 유엔은 기업 또는 각국 정부가 주도해 온 인공지능 윤리에 대하여도 국제인권규범을 반영하고 법률에 기반하여 규제할 것을 요구하여 왔다.

유엔 의사 표현의 자유 특별보고관은 2018년 보고서[52]에서 "윤리 강령과 관련 제도적 구조의 개발은 인권 책무를 중요하게 보완할 수는 있지만 대체물은 아니다.(48문)"라고 지적하면서 "민간이 주력하고 공공이 추구하는 인공지능 윤리는 인권 기반 규제에 대한 반발을 내포한 경우가 많다. 인공지능 분야에서 윤리는 특정 과제를 해결하는 중요 체계를 제공하지만, 윤리가 모든 국가에서 법률로 묶여있는 인권을 대체하는 것은 아니다. 기업과 정부는 윤리 강령과 지침을 개발하는 중에도 인공지능 운영의 모든 측면에 인권 고려사항과 책임이 확실하게 통합시켜야 한다.(46문)"이라고 명확히 밝혔다. 더불어 "인공지능 기술의 윤리적 영향에 대한 가이드라인이나 규약을 만드는 모든 노력은 반드시 인권의 원칙에 기반을 두어야 한다.(65문)"고 권고하였다. 유엔 사무총장 또한 2020년 보고서에서[53]에서 인공지능 등 신기술 환경에서 사회권을 보장하기 위하여 "국가는 민간 부문 활동에 관한 조치를 포함하여 입법 조치를 취해야 할 의무를 재확인하고 준수(62(b)문)"할 것을 요구하고, 특히 인공지능 등 신기술이 사용되는 상황에 대해 책임성을 완전하게 보장하는 적절한 법률체계와 절차방법을 마련하여 감독 체제를 수립하고 구제 수단을 구비할 것을 요구하였다(62(h)문).

52) 유엔문서 A/73/348 (2018. 8. 29).
53) 유엔문서 A/HRC/43/29 (2020. 3. 4).

유럽평의회는 2020년 <알고리즘 시스템의 인권 영향에 대한 권고> 및 부록 지침[54]에서 회원국에게 인권 침해를 예방, 탐지, 금지 및 구제하는 효과적이고 예측 가능한 입법을 요구하고, 공공 및 민간 부문 행위자가 그 법적 의무를 이행하지 않는 경우 책임을 져야 한다고 강조하였다.

유럽연합은 2021년 4월 21일 공공과 민간 부문 고위험 인공지능에 요구사항을 적용하는 인공지능법(안)[55]을 발의하였다. 유럽 집행위원회는 많은 알고리즘의 불투명성이 불확실성을 유발하고 안전 및 기본권리에 대한 기존 법률의 효과적인 집행을 방해할 수 있다는 점을 지적하고, 이러한 문제를 해결하기 위하여 입법 조치에 이르렀다고 설명하였다. 특히 이 법안은 유럽 집행위원회가 2019년 4월 채택한 <신뢰가능 인공지능 윤리 가이드라인>의 원칙을 법규로 발전시킨 것이다.[56] 2019년 윤리 가이드라인은 인간의 기본권 존중을 인공지능 개발의 가장 기본적 원칙으로 삼고, 신뢰할 만한 인공지능의 속성으로 적법성(lawful), 윤리성(ethical), 견고성(robust)을 제시하였다. 나아가 신뢰할 만한 인공지능의 7대 요구사항을 인간행위자와 감독, 기술적 견고성과 안전, 프라이버시와 데이터 거버넌스, 투명성, 다양성・차별금지・공정성, 사회・환경적 복지, 책임성으로 정리했다.[57] 가이드라인을 준비한 인공지능 고위전문가그룹은 여기서 더 나아가 윤리 기준을 실현하기 위한 세부 평가 목록(assessment list)을 공개하였다. 이 평가 목록은 140여 가지 항목으로 각 요구사항이 인공지능 개발 수명주기에서 잘 충족되는지를 실무에서 점검할 수 있도록 구성되어 있으며, 인공지능 영향평가의 초기 형태로 널리 알려졌다. 평가 목록의 공개와 함께 유럽연합의 인공지능 윤리 가이드라인은 기존에 원칙 위주로 이루어져 온 윤리 선언과 다른 실천력을 보여주었다고 평가되었다. 이 윤리 가이드라인은 발표 시점부터 "새롭고 구체적인 규제 마련을 장려"[58]하려는 지향점을 가지고 있었으며, 신뢰할 만한 인공지능에 대한 유럽연합의 접근법은 일찌기 법적 규제를 예비하고 있었다.[59] 유럽 집행위원회는 윤리 가이드라인 채택 후 <인공지능 백서(2020. 2)>,[60] <인공지

54) Council of Europe (2020).
55) Artificial Intelligence Act.
56) High-Level Expert Group on Artificial Intelligence (2019).
57) 유재흥, 추형석, 강송희 (2021). 유럽(EU)의 인공지능 윤리 정책 현황과 시사점 : 원칙에서 실천으로. 소프트웨어정책연구소 ISSUE REPORT IS-114(2021. 3. 25).
58) High-Level Expert Group on Artificial Intelligence (2019). 11p.
59) European Commission Ethics Guidelines for Trustworthy AI. Simmons & Simmons LLP (2019. 4. 12).

능 공공조달 백서(2020. 5)>[61]를 연달아 발표하며 가이드라인을 실현하는 인공지능 규제 프레임워크(regulation framework)를 제시하여 왔다. 이러한 지향점과 노력들이 2021년 4월 유럽 집행위원회가 유럽 의회에 인공지능법(안)을 발의하면서 구체적인 법규의 형태를 띄게 된 것이다.

이하에서는 국제 규범이 인공지능의 개발과 활용의 행위자인 국가와 기업에 요구하는 법률적이고 정책적인 의무와 책임에 대하여 좀더 상세하게 살펴본다.

제1절 유엔의 기준 및 제도

1. 신기술과 인권 문제 검토

유엔 인권최고대표실은 유엔 인권이사회 특별절차에서 '신기술(new technologies)'과 인권 문제를 공통 관심 주제(thematic issues and crises of common interest)로 제시하였다.[62] 초국적기업 및 기타 사업체의 인권에 대한 실무그룹, 평화적 집회 및 결사의 자유 특별보고관을 비롯하여 분야별로 임무를 맡은 특별절차 수임자들은 로봇공학, 자동화, 인공지능, 드론, 치명적인 자율 무기 시스템을 비롯하여 신기술의 인권 영향을 다각도로 다루어 왔으며, 국가는 물론 민간 기업에 대하여 신기술 환경에서 인권 보장을 위한 여러 권고를 제시해 왔다(부록 VI 참조). 그 가운데 특히 두드러진 주목을 받은 문헌을 소개하면 다음과 같다.

2018년 8월 의사 표현의 자유 특별보고관은 인공지능이 인권에 미칠 영향에 대한 보고서를 발표하였다.[63] 보고서는 인공지능이 의사 표현의 자유에 미치는 영향에 대해 조사하면서, 인공지능이 이제 정보 환경의 중요한 부분이 되어 개인의 권리 향유에 혜택과

<https://www.simmons-simmons.com/en/publications/ck0bjufy9o7su0b33hylmt08f/120419-european-commission-ethics-guidelines-for-trustworthy-ai (검색일: 2021. 9. 1.)>.

60) European Commission (2020a).

61) European Commission (2020b).

62) 2021년 9월 현재 유엔 인권이사회 특별절차의 공통 관심 주제로는 신기술, 지속가능한 개발 목표, 이주, 기후 변화등이 있다. <https://www.ohchr.org/EN/HRBodies/SP/Pages/CrosscuttingThematicIssues.aspx (검색일: 2021. 9. 1.)> 참조.

63) 유엔문서 A/73/348 (2018. 8. 29).

위험을 동시에 주고 있다고 지적하였다. 특별보고관은 이러한 기술적 역능의 확장에 대응하여 인권을 보장해야 하는 국가의 의무와 기업의 책임을 강조하면서 국가와 기업이 각기 실행할 수 있는 구체적인 조치들을 제안하였다.

우선 국가에 대한 권고로는 △인공지능 시스템의 조달이나 사용 이전에 공공적으로 의견을 수렴하고 인권영향평가 또는 알고리즘 영향평가를 수행하는 한편, 독립적인 외부 전문가의 정기적인 감사를 실시할 것(62문), △민간 기업의 인공지능 설계, 배치 및 구현에 있어서도 인권이 중심에 올 수 있도록 개인정보보호 규제, 영향평가 및 감사 실시, 효과적인 외부 책임성 메카니즘을 보장해야 하고, 인공지능 관련 규제 수립에 있어 시민사회, 인권단체 및 소외되거나 과소대표된 최종 이용자의 대표자들로부터 폭넓은 의견을 수렴할 것(63문), △다양하고 다원적인 정보 환경 조성을 위한 정책 및 입법 환경을 조성하고 이의제기 보장 수단을 강구할 것(64문) 등이 권고되었다.

또한 기업에 대한 권고는 다음과 같다. △인공지능 윤리 지침이나 행동강령, 기업 규칙 및 기술 지침은 물론 플랫폼 이용약관 또한 인권 원칙에 기반하여야 하며, 인공지능의 개발과 배치에서 시민사회의 의견을 수렴할 것(65문), △플랫폼, 서비스 및 애플리케이션에서 인공지능 기술과 자동화 기술이 어디서 어떻게 사용되는지 명시하고, 인공지능 시스템의 사용이 인권 향유에 미치는 영향을 사용자가 이해하고 다루는 데 필요한 내용을 고지하여야 하며, 콘텐츠 표시 뿐 아니라 콘텐츠 삭제 및 그에 대한 이의제기 데이터도 공개할 것(66문), △인공지능 시스템의 입력 및 출력 수준에서 차별을 방지하고 설명하며, 인공지능 시스템을 설계하고 배치하는 인력이 다양하고 차별에 반대하는 태도를 갖추고 데이터셋의 선택 및 시스템 설계에서 편향과 차별 방지에 우선순위를 둘 것을 포함하며, 인공지능 시스템의 차별적 결과에 대한 적극적인 모니터링을 시행할 것(67문), △신규 인공지능 시스템의 설계와 배치에서 사전적인 인권영향평가와 공개적인 의견 수렴을 실시하고 그 결과를 공개하며, 의견 수렴에 시민사회, 인권 활동가, 소수자 및 과소대표된 최종 이용자의 대표자들을 참여시킬 것(68문), △모든 인공지능 코드가 완전히 감사가능하도록 조치하고 인공지능 시스템에 대한 외부 독립 감사를 추진하며, 감사 결과를 공개할 것(69문), △인공지능 시스템이 인권에 대하여 부정적인 영향을 미친 경우 해당 개인은 구제 수단에 접근할 수 있어야 하며, 이용자 불만과 이의제기에 시기적절한 인간의 검토와 구제 조치를 보장하고, 인공지능 시스템이 불만과 구제 요청의 대상이 되

는 빈도와 구제 수단의 요형과 효과성에 대한 데이터를 정기적으로 공개할 것(70문) 등이 권고되었다.

<표 5> 유엔 의사 표현의 자유 특별보고관 '인공지능에 대한 법적 체계'

원칙	내용
인권 원칙	인공지능은 모든 다른 기술과 마찬가지로 국제인권법에 따른 국가의 의무와 민간 기업의 책임을 준수하여 설계되어야 하고 개발되어야 하고 도입되어야 함. 기업은 그 표준, 규정, 시스템 설계를 보편 인권 원칙에 맞추어야 함
투명성	인공지능 시스템은 개인에게 적극적으로 공개되어야 하며, 이들이 인공지능 절차에 자신의 데이터를 적용하거나 투여한다는 사실을 이해할 수 있는 방식으로 공개되어야 함. 기업과 정부는 인공지능 가치 체계의 각 측면에 걸쳐 투명성을 수용해야 함. 기업은 개인 이용자에게 인공지능 시스템의 존재 여부, 그 목적, 구성 및 영향에 대해 교육해야 함. 기업은 얼마나 많은 내용이 삭제되고, 얼마나 자주 삭제를 요청받는지, 얼마나 자주 삭제에 대한 이의가 제기되는지를 공개해야 함
인권영향평가	정부와 기업은 인공지능 시스템을 면밀히 조사하고 개념에서 구현에 이르기까지 이의를 제기할 수 있는 조치를 취해야 함. 인권영향평가는 인공지능 시스템의 인권 영향 문제를 해결하기 위한 하나의 도구임
감사	인공지능 시스템의 외부적 검토를 촉진하는 것은 엄격하고 독립적으로 투명성을 보장하는 데 중요함
개인의 자율성	인공지능이 개인의 의견 형성 및 보유 역량과 정보 환경에서 접근하고 표현하는 역량을 비가시적으로 대체하거나 조작하거나 방해해서는 안 됨. 개인의 자율성을 존중하는 것은 최소한 이용자가 지식, 선택 및 통제권을 갖도록 보장하는 것을 의미함
고지 및 동의	기업은 플랫폼, 사이트 또는 서비스의 이용에 자사 인공지능이 어떻게 관여하고 있는지를 이용자에게 충분히 알려야 함
권리구제	인공지능 시스템이 인권에 악영향을 미친다면 관련 기업이 이를 구제하는 것이 가능해야 하고 구제되어야 함

특히 의사 표현의 자유 특별보고관은 인공지능에 대한 법적 체계(The Legal Framework for AI)에 반영되어야 할 원칙으로서 △인권 원칙 준수 △투명성 보장 △인권 영향평가 실시 △감사 실시 △개인의 자율성 보장 △고지 및 동의 원칙 △권리구제 보장 등의 7가지 원칙을 제시하였다.[64]

한편, 인권 기구 외에도 조약 기구를 비롯한 유엔의 주요 기구들 역시 신기술 환경에서 인권 보장을 위한 국가와 기업의 조치를 촉구해 왔다.

2020년 3월 유엔 사무총장은 <사회권의 실현에 있어 신기술의 역할에 대한 보고서>[65]를 발표하고 신기술 환경에서 사회권 보장을 위한 국가와 민간 기업들의 조치를 요구하였다(62문). 즉, (a)신기술의 개발, 사용 및 거버넌스에 있어 모든 인권의 보호 및 강화를 중심 목표로서 전적으로 수용하고, 모든 인권에 대하여 온라인과 오프라인에서 동등한 존중과 이행을 보장해야 하고, (b)국가는 민간 부문 활동에 관한 조치를 포함하여 입법 조치를 취해야 할 의무를 재확인하고 준수하여야 하며, 이로써 신기술은 경제・사회・문화적 권리를 포함한 모든 사람들의 인권에 대한 완전한 향유에 기여하고 인권에 미치는 부작용이 방지되어야 하고, (c) 국가 간 및 국가 내적으로 정보 격차 및 기술 격차를 해소하기 위한 노력을 가속해야 하며, 신기술의 접근성, 가용성, 경제성, 적응성 및 품질을 개선하기 위한 포괄적인 접근 방식을 촉진해야 하고, (d) 기술 변화 등에 의해 야기되는 변화와 불안정성으로부터 복원력을 수립할 수 있는 사회적 보호의 권리에 투자해야 하며, 모든 고용 형태의 노동권을 보호해야 하고, (e) 공공부문에서 신기술, 특히 인공지능의 이용에 관한 정보를 대중에게 전파하기 위한 노력을 대폭 증진해야 하고, (f) 신기술의 개발 및 도입에 관한 의사결정에 모든 관련 이해관계자의 참여를 보장하며, 특히 공공부문에서 인공지능이 지원하는 의사결정에 대하여 적절한 설명가능성이 보장될 필요가 있고, (g) 인권의 향유에 중대한 영향을 미칠 수 있는 신기술 시스템, 특히 인공지능 시스템의 전체 수명주기 동안 체계적으로 인권 실사를 실시해야 하고, (h) 신기술이 사용되는 상황에 대해 책임을 완전하게 보장하는 적절한 법률체계와 절차방법을 마련해야 하며, 이를 위해 국내 법제도의 공백을 검토 및 평가하고 필요한 경우 감독 체제를 수립하

64) United Nation's Human Rights Special Procedures (2018). Factsheet : Artificial Intelligence Technologies and Freedom of Expression. <https://www.ohchr.org/Documents/Issues/Expression/Factsheet_3.pdf (검색일: 2021. 9. 1.)>.

65) 유엔문서 A/HRC/43/29 (2020. 3. 4).

며, 신기술로 인한 피해에 대해 사람들이 이용할 수 있는 구제 수단이 구비되어야 하고, (i) 신기술의 개발 및 사용, 특히 경제・사회・문화적 권리의 향유에 필수적인 제품 및 서비스에 대한 접근에 있어서 차별과 편견을 해소해야 하고, (j) 정례인권검토(UPR)와 인권 조약 기구 하에서 이루어지는 보고 및 검토에 있어 신기술이 경제・사회・문화적 권리에 미치는 영향에 특히 주의를 기울여야 한다.

2020년 12월 유엔 조약 기구인 인종차별철폐위원회는 <법집행관의 인종적 프로파일링 방지와 대응을 위한 일반 권고>[66]를 발표하고 각국 정부에 (A) 입법적・정책적 조치 (B) 인권 교육 및 훈련 조치 (C) 채용 조치 (D) 공동체 정책 (E) 세분화된 데이터 관리 (F) 책무성 조치 (G) 국제인권법을 준수하는 인공지능 등의 조치를 권고하였다(Chapter. Ⅷ). 특히 국가는 법집행관에 의한 인종적인 프로파일링을 정의하고 금지하는 법과 정책을 개발하고 효과적으로 시행하여야 한다(38문). 이때 법집행기관 내부와 외부 모두에서 효과적이고 독립적인 감독 메카니즘을 수립하고 위반 행위에 대한 징계 조치를 포함해야 하며, 독립 전문가가 정기적인 감사를 실시하고 그 결과는 투명하게 공개하여야 한한다(39문). 또한 국가는 인종차별로부터 모든 사람을 보호하고 구제하며, 인종차별 피해에 대한 공정하고 적절한 배보상을 보장하여야 하고(40문), 피해자 중심적인 관점(victim-centred approaches)으로 사법 당국을 비롯하여 관련 기관, 공동체, 차별의 교차성을 경험하는 집단의 대표자를 비롯한 시민단체는 물론 국가인권기구 간의 협력을 증진하여야 한다(41문).

최근 특히 주목을 받은 것은 또다른 조약 기구인 유엔 아동권리위원회가 발표한 디지털 아동권리 일반 논평이다. 유엔 아동권리위원회는 2021년 3월 2일, 인공지능을 비롯한 디지털 환경에서 아동을 안전하게 보호하고 아동의 권리를 보장하기 위한 일반논평[67] 제25호(이하 ‘일반논평’)를 채택했다.[68] 디지털 환경에서도 모든 아동의 권리가 존중되고 보호받고 충족되어야 한다는 원칙 아래, 아동권리위원회는 디지털 기술의 혁신이 심지어 아동들이 스스로 인터넷에 접속하지 않는 곳에 이르기까지 아동의 삶과 권리에 광범위하고 상호의존적인 형태로 영향을 미친다고 보고, 디지털 포용(digital inclusion)이 달

66) General recommendation No. 36 (2020) on preventing and combating racial profiling by law enforcement officials. 유엔문서 CERD/C/GC/36 (2020. 12. 17).
67) 유엔의 일반논평(general comments)은 국제 조약 규정을 해석하는 기준으로 볼 수 있다.
68) General comment No. 25 (2021) on children's rights in relation to the digital environment. 유엔문서 CRC/C/GC/25 (2021. 3. 2).

성되지 않으면 이미 존재하는 불평등은 악화되고 새로운 불평등이 발생할 수 있다고 지적하였다(4문). 따라서 국가가 디지털 환경에서 아동권리협약 이행을 위해 취해야 하는 조치들로 (A) 입법 (B) 종합 정책 및 전략 (C) 조정 (D) 자원 배분 (E) 정보 수집 및 연구 (F) 독립적 모니터링 (G) 정보 제공, 인식 제고 및 교육 (H) 시민사회 협력 (I) 아동 권리와 기업 부문 (J) 상업 광고와 마케팅 (K) 사법과 구제 수단 접근 보장 등을 들 수 있다. 더불어 일반논평은 기업의 의무를 강조한다. 기업은 디지털 환경과 관련된 서비스와 제품의 제공으로 아동의 권리에 직간접적인 영향을 미친다. 따라서 기업은 아동의 권리를 존중하고 디지털 환경과 관련된 아동의 권리 침해를 예방, 구제해야 하고, 국가는 기업이 이러한 책임을 다할 수 있도록 보장할 의무가 있다(35문).

2. 디지털 시대 프라이버시권 검토

유엔 기구들의 여러 노력을 배경으로 유엔 인권이사회는 2019년 제42차 회기에서 <디지털 시대 프라이버시권>을 결의하면서 인공지능을 비롯한 신기술에 대한 규율을 요구하였다.[69] 인권이사회는 인공지능 등 신기술(new and emerging technologies)[70]의 사용, 배치 및 이후 발전은 프라이버시권 및 기타 인권의 향유에 영향을 미칠 수 있다는 사실을 인정하고, 프라이버시권에 대한 위험을 최소화할 수 있고 하여야 한다는 사실을 강조하였다. 이는 인공지능 등 신기술의 설계, 개발 및 배치에 있어 국제인권법을 고려하는 등 적정한 규제 및 기타 적절한 메커니즘을 도입하고, 안전하고 보안을 보장하며 고품질인 데이터 인프라를 확보하며, 인간의 감사 메커니즘 및 시정 메커니즘을 개발함으로써 이루어질 수 있다. 특히 국가는, 기업이 인공지능을 비롯한 기술의 설계, 개발, 배치 및 평가할 때 프라이버시권 및 기타 관련 인권을 완전히 보장하도록 보장하고, 그 권리가 침해되거나 남용당했을 수 있는 개인에게 배보상 및 반복 금지 보장 등 효과적인 구제 수

69) Resolution adopted by the Human Rights Council on 26 September 2019, 42/15. The right to privacy in the digital age. 유엔문서 A/HRC/RES/42/15 (2019. 10. 7).

70) 유엔 사무총장은 앞서 2020년 3월 보고서에서 '신기술(new technologies, frontier technologies, 또는 emerging technologies)'의 정의에 대한 보편적 합의는 없지만, 디지털 기술(인공지능, 빅데이터 분석, 사물인터넷, 로봇공학, 블록체인 등), 생명공학(줄기세포 기술, 건강 모니터링 기술 등), 첨단소재(나노소재 등), 에너지 및 환경(드론, 마이크로 위성, 전기차, 바이오 연료 등)의 4개 분야를 주로 아우른다고 설명하였다. 유엔문서 A/HRC/43/29 (2020. 3. 4), 각주1 참고.

단에 접근할 수 있도록 규정하는 법률, 규제 및 정책을 도입하여야 하고, 여성, 아동, 취약한 상황 또는 소외된 집단에 특정한 영향이 미치는 등 디지털 시대 프라이버시권의 침해나 남용에 대하여 예방 조치 및 구제 수단을 추가로 개발하거나 유지하여야 한다. 기업은, 유엔 <기업과 인권 이행지침>에 따라 프라이버시권을 비롯한 인권에 대한 존중이 자동화된 의사 결정 및 머신 러닝 기술의 설계, 운영, 평가 및 규율에도 반영되도록 보장하여야 하며, 그 원인이 되었거나 원인에 기여한 인권 침해에 대하여 보상하여야 한다. 기업의 운영, 제품 및 서비스에 직접적으로 연결된 부정적인 인권 영향을 방지하거나 완화하는 적정한 안전장치를 마련하고, 제품 및 서비스의 오용이 감지된 경우 그 남용 또는 위반에 대하여 관련 기관에 즉시 통지하여야 한다. 이 유엔 인권이사회 결의의 주요 내용은 2020년 12월 유엔 총회 결의로 채택되었다.[71]

더불어 인권이사회는 2019년 제41차 회기 및 2021년 제47차 회기에서 인공지능을 비롯한 '신기술과 인권'(New and emerging digital technologies and human rights)' 제하에서 인공지능의 인권 문제를 검토하기로 하고 관련 결의를 채택하였다.[72] 이 두 건의 결의는 한국 정부가 주도적으로 참여하였으며, 이 주제 관련하여 유엔 기구가 전문가 및 다양한 이해관계자가 참여하는 전체적, 포용적, 포괄적 접근방식(holistic, inclusive and comprehensive approach)을 취할 것을 촉구하였다.[73] 특히 2021년 결의에서는 유엔 사무총장이 회원국이 디지털 기술의 개발 및 이용에 관한 규제체계와 법안을 도입할 때 인권을 최우선시해야 함을 요청한 바 있고, 유엔 인권최고대표가 신기술 사용에 있어서 시스템 전반을 아우르는 인권실사 및 영향평가 지침 개발을 요청한 바 있다는 사실을 주목하였다.

인공지능과 인권 문제를 해결하기 위한 이상의 여러 권고들을 최근에 종합한 것으로는 유엔 인권최고대표의 보고서가 있다(부록 Ⅱ 참조).[74] 유엔 인권최고대표는 2021년 9월

---

71) Resolution adopted by the General Assembly on 16 December 2020, 75/176. The right to privacy in the digital age. 유엔문서 A/RES/75/176 (2020. 12. 28).

72) Resolution adopted by the Human Rights Council on 11 July 2019, 41/11. New and emerging digital technologies and human rights. 유엔문서 A/HRC/RES/41/11 (2019. 7. 17.); 유엔문서 A/HRC/RES/47/23 (2021. 7. 16).

73) 제41차 유엔 인권이사회, 우리 정부 주도로 "신기술과 인권" 결의 채택. 외교부 보도자료 (2019. 7. 12.); 제47차 유엔 인권이사회, 우리 정부 주도로 "신기술과 인권" 결의 채택. 외교부 보도자료 (2021. 7. 16).

74) 유엔문서 A/HRC/48/31 (2021. 9. 13).

유엔 인권이사회 제48차 회기에서 인공지능과 프라이버시권에 대한 보고서를 발표하였다. 최고대표는 보고서에서 인공지능 시스템에 대한 인권 기반 접근법을 강조하고 국가와 기업에 대하여 프라이버시권을 비롯한 인권 침해를 방지하는 안전장치의 설계 및 구현을 요구하는 권고를 담았다. 최고대표는 이러한 조치들이 인공지능이 제공할 수 있는 편익을 최대한 누리면서 유해한 결과물은 방지하고 완화할 수 있다고 보았다.

우선, 최고대표는 인공지능 시스템이 프라이버시권을 침해하게 되는 기능적 요인을 살펴보았다. 새로운 인공지능 애플리케이션이 프라이버시권 등 인권을 침해할 수 있지만, 기존 시스템 역시 개인정보의 수집 및 사용을 증가시켜 침해를 낳을 수 있다.

인공지능 시스템은 일반적으로 개인정보를 포함한 대용량 데이터셋에 의존한다. 이는 광범위한 데이터 수집, 저장 및 처리를 촉진한다. 데이터 수집은 온라인 서비스와 사물인터넷을 비롯하여 친밀한 공간, 사적인 공간 및 공공 장소를 가리지 않고 이루어진다. 데이터에 포함된 개인정보가 수집, 결합, 분석되고 여러 기관에 공유되는 규모가 전례없는 비율로 증가하였다.

이러한 데이터셋의 유통은 기업과 국가에 사람들의 사생활을 노출하는 것 외에도, 여러 가지로 개인들을 취약하게 만든다. 우선 계속된 개인정보 유출로 수백만 명의 민감정보가 노출되었다. 대용량 데이터셋은 개인에 대한 무수한 분석과 제3자 공유를 가능하게 하며, 이는 종종 추가적인 프라이버시 침해로 이어지고 인권에 부정적인 영향을 끼친다. 예를 들어, 정부 기관이 기업이 보유한 데이터셋에 직접 접근할 수 있도록 하는 방식은 관련된 개인의 프라이버시권에 자의적이거나 불법적인 간섭의 가능성을 증가시킨다. 다양한 출처의 데이터를 결합함으로써 기존에 익명이었던 개인이 식별될 가능성이 증가하기도 한다. 정량적 데이터셋 설계는 개인에게 특정한 성별 등 정체성을 강요하기도 한다. 또한 개인정보의 장기간 보관은 개인정보 수집 당시 예상하지 않았던 장래의 악용 가능성을 낳기 때문에 특별한 위험이 수반된다. 시간이 지남에 따라 데이터가 부정확해지거나 부적절해지거나 역사적인 편견에 따른 오인을 초래하여 향후 개인정보 처리의 편향 또는 오도된 결과를 초래할 수도 있다.

인공지능 시스템은 개인정보가 관련되지 않았더라도 인권에 부정적인 영향을 미칠 수 있다. 인공지능 도구는 개인의 정신적, 신체적 상태를 비롯하여 개인에 대해 광범위한 추론을 할 수 있으며 특정 정치적 또는 개인적 성향을 가진 사람들 집단을 식별할 수

있다. 인공지능은 또한 미래의 행동이나 사건이 일어날 가능성을 평가하는 데 사용된다. 인공지능이 만든 추론과 예측은 그 확률적 특성에도 불구하고 완전히 자동화된 방식으로 사람들 권리에 영향을 미치는 의사결정의 기반이 될 수 있다. 인공지능의 추론과 예측은, 사람들의 자율성과 자신의 정체성에 대한 세부사항을 확립할 권리를 포함하여, 프라이버시권의 향유에 깊은 영향을 미친다. 사상과 의견의 자유에 대한 권리, 표현의 자유, 공정한 재판 관련 권리 등 다른 권리에도 문제를 야기할 수 있다.

인공지능 기반 결정은 오류에서 자유롭지 않다. 인공지능 솔루션의 확장성은 작아 보였던 오류율의 부정적인 영향을 극적으로 증가시킬 수 있다. 인공지능 알고리즘의 출력에는 확률적 요소가 있으며, 이는 그 결과물에도 불확실성이 포함되어 있음을 의미한다. 사용된 데이터의 관련성과 정확성 또한 종종 의문스럽다. 비현실적인 기대로 원하는 목표를 달성할 채비를 갖추지 않은 인공지능 도구를 배치하는 경우도 있다. 이렇게 결함이 있는 데이터에 의존하는 인공지능 시스템의 출력물은 예를 들어, 한 개인을 테러범으로나 부정 수급을 저지른 것으로 오지목함으로써 여러 가지 인권 침해의 원인이 될 수 있다. 편향된 데이터셋이 인공지능 시스템에 기반한 차별적 의사결정으로 이어지는 경우가 특히 우려된다.

무엇보다 많은 인공지능 시스템의 의사 결정 과정이 불투명하다는 점이 큰 문제이다. 인공지능 시스템의 개발 및 운영을 뒷받침하는 정보 환경, 알고리즘, 모델의 복잡성은 물론 정부와 민간 행위자들의 의도적인 비밀주의가 일반 대중으로 하여금 인공지능 시스템의 영향을 이해하는 과정을 방해한다. 머신러닝 시스템은 설명하기 어렵거나 불가능한 패턴 또는 결과물을 도출할 수 있다. 이를 흔히 ‘블랙박스’ 문제라고 한다. 이러한 불투명성으로 인공지능 시스템을 유의미하게 조사하는 것이 어려워지고 인공지능 시스템이 위해를 야기하는 경우에도 효과적인 책무성 확보가 어려울 수 있다.

이어 최고대표는 수사/국가안보/형사사법/출입국관리, 공공서비스 제공, 고용, 온라인 정보 관리 등 주요 분야에서 인공지능이 프라이버시권 등 인권에 영향을 미친 사례를 검토하였다.

인공지능이 법집행(수사), 국가안보, 형사사법, 출입국관리 분야에서 사용될 경우 인권 침해 위험에 대하여 다음과 같이 요약하였다. 첫째, 사용된 데이터셋에는 다수의 개인에 대한 정보가 포함되어 있으므로 프라이버시권이 관련된다. 둘째, 인공지능 평가는 그 예

측에서 확률적 특성을 보유하고 있으므로 합리적인 의심의 근거로 생각해서는 안 되는데도 불구하고, 수색, 검문, 체포 및 기소 등 국가의 개입을 촉발할 수 있다. 셋째, 인공지능 기반 결정에 내재된 불투명성은, 특히 인공지능이 강제적인 조치의 기반이 될 때 국가의 책무성에 대한 중요한 문제를 제기하며, 대테러기관 활동처럼 일반적으로 투명성의 결여를 지적받아온 분야에서는 더욱 그러하다. 넷째, 예측 도구는, 특정 소수자 집단에 편중된 치안 집중의 사례에서처럼, 사용된 데이터셋에 내재된 역사적인 인종적 및 민족적 편견을 반영하여 차별을 영구화하거나 강화할 위험을 내재하고 있다. 특히 최고대표는 원격 생체 인식의 인권 침해 문제에 대하여 깊은 우려를 밝혔다. 한 개인의 생체인식 정보는 다른 사람과 구별되는 독특한 특성을 나타내기 때문에 그 사람의 인격의 핵심 특성 중 하나를 구성한다. 게다가, 원격 생체 인식은 공공 장소에서 체계적으로 개인의 신원을 확인하고 추적할 수 있는 국가의 능력을 극적으로 증가시켜, 사람들이 관찰되지 않고 자신의 삶을 영위할 수 있는 능력을 약화시키고, 이동의 자유 뿐 아니라 표현의 자유, 평화로운 집회 및 결사의 자유에 대한 권리 행사에 직접적으로 부정적인 영향을 미친다.

인공지능이 공공서비스 전달에서 사용되는 경우, 인공지능의 학습과 의사결정에 사용되는 데이터셋은 국가 보유 데이터와 출처가 불분명한 민간 기업 데이터를 사용하거나 결합시킬 수 있다. 이는 프라이버시권에 대한 우려 뿐 아니라 데이터에 내재된 역사적 편견이 공공 기관의 의사결정에 미치는 영향에 대한 우려를 낳는다. 공공서비스에 인공지능을 사용하는 것에 대한 가장 큰 우려는 그것이 차별적일 수 있다는 것이며, 특히 소외된 집단에 대한 차별이 우려된다.

더불어 인공지능이 고용 환경에서 사용되면서, 직무와 관련이 없는 노동자 행동 및 데이터에 대한 수집과 감시가 확대되었다. 수집된 데이터는 처음 공지된 목적외 다른 목적으로 사용될 수 있으며, 시스템 기능이 확장될 수도 있다. 또한 회사가 남성, 백인, 중년 남성을 선호하는 과거 데이터셋으로 학습된 인공지능 채용 알고리즘을 사용하는 경우, 결과 알고리즘은 구인에 적합한 자격을 동등하게 갖춘 여성, 유색인종 및 젊은이나 노년층을 선호하지 않을 것이다. 반면 노동자를 보호하기 위한 책무성 구조와 투명성은 결여된 경우가 많으며, 노동자들은 인공지능 기반 모니터링 실시에 대한 설명을 거의 또는 전혀 듣지 못하고 있다.

이러한 문제를 해결하기 위한 최고대표의 권고는 ‘인권 기반 접근’으로 요약할 수 있다. 우선 최고대표는 데이터 기반 인공지능 시스템에서 개인정보보호법으로 정보주체를 효과적으로 보호해야 할 필요성을 강조하였다. 설명에 대한 권리와 완전 자동화 의사결정에 반대할 권리 등 개인의 권리를 강화하고 인공지능 기술의 발전에 부응하는 더 많은 안전장치를 개인정보보호 체계 내에 수립할 필요가 있다. 방대한 정보 비대칭성을 비롯한 글로벌 데이터 환경의 복잡성과 불투명성 증가에 대응하기 위하여 독립적인 개인정보보호 감독기구의 중요성 또한 더욱 커졌다.

최고대표는 개인정보보호법에서 더 나아가 인공지능의 문제를 인권 기반 접근법으로 해결하기 위해서는 추가적인 입법의 필요성이 있다고 지적하였다. 이때 입법은 인공지능 애플리케이션, 시스템 및 용도의 다양성을 고려하여, 여러 부문에 적용할 수 있을 만큼 구체적이어야 한다. 인권에 대한 위험이 높을수록 인공지능 기술의 사용에 대한 법적 요구사항이 엄격해져야 한다. 따라서 법 집행, 국가안보, 형사사법, 사회보장, 고용, 보건의료, 교육 및 금융 등 개인의 이해관계가 특히 높은 분야가 우선되어야 한다. 위험기반 접근방식을 취하는 법률과 규제의 경우 국제 인권법 하에서 정당화되지 않는 잠재적 또는 실제적 영향을 초래하는 특정 인공지능 기술, 애플리케이션 또는 사용 사례에 대한 금지를 요구해야 한다. 정부의 사회신용점수 등 차별 금지와 본질적으로 충돌하는 인공지능의 사용은 허용되어서는 안 된다. 인권에 부정적인 영향이 발생할 가능성이 있을 때 인간의 감독 및 의사결정의 의무적 개입이 규정되어야 한다. 원격 실시간 얼굴 인식과 같은 잠재적인 고위험 기술의 경우, 그 사용의 인권 준수가 보장될 때까지 유예(모라토리엄)하여야 한다.

최고대표는 입법에서 인공지능에 대한 적정하고 독립적이고 공정한 감독 체계를 포함해야 할 필요성을 강조하였다. 이러한 감독은 행정적, 사법적, 준사법적 및 의회 감독 기관의 조합으로 수행될 수 있고, 개인정보보호 감독기구, 소비자 보호 기관, 부문별 규제기관, 차별 방지 기구 및 국가 인권 기구가 감독 시스템의 일부를 구성해야 한다. 인공지능 사용을 감독하는 부문별 규제 기관은 기본 표준을 수립하고 정책 및 집행의 일관성을 보장하는 데 도움이 될 수 있다.

한편, 최고대표는 인권 기반 접근법의 일환으로 인권 실사를 강력히 요구하였다. 국가와 기업은 인공지능 시스템의 구입, 개발, 배치 및 운영 시 뿐 아니라 개인에 대한 빅데

이터를 공유하거나 사용하기 전에 포괄적인 인권실사를 실시해야 한다는 것이다. 실사 절차에서 인공지능의 사용이 인권과 양립할 수 없는 것으로 드러나는 경우 그 사용을 중지하여야 한다. 이때 인권영향평가는 인권 실사 과정의 필수적인 요소이다. 인권 실사는 인공지능 시스템의 수명주기 전반에 걸쳐 실시하여야 하며, 여성과 십대 여성, 레즈비언, 게이, 양성애자, 트랜스젠더 및 성소수자, 장애인, 소수자 집단에 속하는 사람, 노인, 빈곤층 및 기타 취약한 상황에 처해 있는 사람들에게 불균형한 영향을 미치는 것에 각별한 주의를 기울여야 한다. 인권실사 과정에서 잠재적으로 영향을 받는 권리 주체 및 시민사회와 유의미한 협의를 수행해야 하며, 여러 학제간 전문성을 갖춘 전문가도 완화 방법 개발 및 평가를 비롯한 영향평가에 참여해야 한다. 국가와 기업은 그들이 사용하는 인공지능 시스템이 인권에 부정적인 영향을 미치는지 여부를 확인하기 위해 지속적으로 모니터링해야 한다. 인권영향평가의 결과 및 인권위험을 해결하기 위해 취해진 조치와 협의 사항은 그 자체로 공개되어야 한다.

국가와 기술 기업 사이에 긴밀한 연합 관계가 있는 상황에 대해서는 집중적인 주의가 요구된다. 국가가 공공재 또는 서비스를 제공하기 위해 민간 인공지능 기업에 의존하는 경우, 국가는 인공지능 시스템의 개발과 배치를 감독할 수 있도록 보장해야 한다. 이는 인공지능 애플리케이션의 정확성과 위험에 대한 정보를 요구하고 평가함으로써 이루어질 수 있다. 위험을 효과적으로 완화할 수 없는 경우, 국가는 공공재 또는 서비스 전달에 인공지능을 사용하지 말아야 한다.

마지막으로 최고대표는 투명성 확보를 강조하였다. 국가, 기업 및 기타 인공지능 사용자는 그 시스템의 종류, 사용 목적, 시스템 개발자와 운영자의 신원에 대한 정보를 일반적으로 제공하여야 하며, 의사결정이 자동으로 이루어졌을 때 영향을 받는 개인에게 이를 체계적으로 알려야 한다. 인권에 중대한 위험이 있는 인공지능 애플리케이션에 대하여 국가는 등록제도를 도입해야 한다. 개인정보를 인공지능 데이터셋으로 사용하는 데 대한 통지와 이에 대한 정보주체의 열람·삭제·정정권을 보장하고, 개인이 프로파일링을 더 잘 이해하고 통제할 수 있도록 주의를 기울여야 한다. 여기서 더 나아가 특별한 투명성 증진이 필요하다. 소위 ‘블랙박스’ 문제를 극복하기 위하여 설명가능성의 개발과 배치를 위한 적극적인 노력이 필요하다. 지적재산권 보호가 인권에 영향을 미치는 인공지능 시스템에 대한 유의미한 조사를 방해하지 않도록 조치하여야 하며, 조달규칙 또

한 인공지능 시스템에 대한 감사가능성 등 투명성을 확보하여야 한다. 국가는 인권에 중대한 부정적인 영향을 미칠 수 있음에도 유의미한 감사 대상이 될 수 없는 인공지능 시스템의 사용을 피해야 한다.

이러한 검토를 거쳐 최고대표는 국가에 대하여 (a) 인공지능 개발, 사용 및 거버넌스에서 인권을 보호하고 (b) 인공지능의 사용에서 발생하는 인권 간섭이 정당성, 필요성, 비례성을 충족하는 법률에 의해 제한되도록 하며 (c) 국제인권법을 준수하며 운영될 수 없는 인공지능 애플리케이션을 명시적으로 금지하고 고위험 인공지능의 경우 인권보호 안전장치가 마련되기 전까지 그 판매 및 사용에 모리토리엄을 부과하고 (d) 인권 보호를 위한 기준과 권고가 시행되기 전까지 공공장소 원격 생체 인식 기술 사용에 모리토리엄을 부과하며 (e) 개인정보보호법을 독립적이고 공정하게 시행하고 (f) 공공과 민간 인공지능 사용의 부정적인 인권 영향을 방지하고 완화하는 입법 및 규제 체계를 도입하며 (g) 인공지능 인권 침해 피해자가 효과적인 구제수단을 이용할 수 있도록 보장하고 (h) 특히 공공부문에서 인권에 중대한 영향을 미칠 수 있는 모든 인공지능 의사결정에 설명가능성을 보장하며 (i) 인공지능 시스템 사용과 관련된 체계적인 평가와 모니터링을 실시하는 등 차별을 해소하기 위한 노력을 강화하고 (j) 인공지능 기술의 제공 및 사용에 대한 민관 파트너십은 투명해야 하고 독립적인 인권 감독의 대상이 되어야 하며, 인권에 대한 정부 책무를 포기하는 결과로 이어지지 않도록 보장할 것을 권고하였다.

또한 모든 국가와 기업은 (a) 인공지능 시스템의 수명주기 전반에 걸쳐 체계적으로 인권실사를 수행할 것을 강력히 요구하였다. 인권 실사의 핵심 요소는 정례적이고 포괄적인 인권영향평가여야 한다. (b) 인공지능 사용의 투명성은 크게 증가시켜 일반인과 영향을 받는 개인들에게 적절히 알리고 자동화 시스템에 대한 독립적이고 외부적인 감사를 가능하게 하여야 한다. (c) 인공지능 사용 등에 대한 결정에 모든 관련 이해당사자, 특히 영향을 받는 개인 및 집단의 참여를 보장하며, (d) 인공지능 의사결정의 설명가능성을 증진하기 위하여 노력하여야 한다.

특히 기업의 경우 (a) 기업과 인권 이행지침을 준수하여 인권 존중 책임을 다하고 (b) 체계적인 평가 및 모니터링을 실시하는 등 차별을 해소하기 위한 노력을 강화하며 (c) 인공지능 개발 인력의 다양성을 보장하고 (d) 인권에 부정적인 영향을 미치거나 초래하였을 때 고충처리 등 정당한 구제 수단을 제공할 것을 권고하였다.

제2절 유럽연합의 기준 및 제도

1. 유럽연합 개인정보보호 일반규정

2016년 제정된 유럽연합 개인정보보호 일반규정(General Data Protection Regulation, 이하 ‘GDPR’)은 인간의 개입이 전혀 없는 완전 자동화 의사결정에 대하여 정보주체가 그 대상이 되지 않을 권리를 규정하였다. 유럽 각국은 이러한 규정과 공통의 해석에 근거하여 인공지능의 자동화된 개인정보 프로파일링 및 완전 자동화 의사결정으로부터 정보주체의 권리를 보호하고 위반 행위를 제재하는 조치를 취해 왔다.

GDPR은 원칙적으로 완전 자동화 의사결정을 금지한다(제22조). GDPR은 ①정보주체에게 법적 효력을 초래하거나 이와 유사하게 본인에게 중대한 영향을 미치는 의사결정이고 ②자동화로 처리되는 의사결정을 ③오로지 자동화된 처리 방식에만 의존하여 이루어지는 경우를 일반적으로 금지하였으며, 이때의 금지는 정보주체의 능동적인 반대를 요하지 않는다. ‘유사하게 중대한 영향’의 경우란, 법적 권리나 의무에 변화가 없더라도 ‘인간의 개입 없이 이루어지는 전자채용’ 등 개인의 상황, 행동 또는 선택에 중대하게 영향을 미치거나, 정보주체에 지속적이거나 영구적인 영향을 미치는 경우, 또는 개인이 배제되거나 차별을 받게 되는 경우를 포함한다. 또 완전 자동화된 처리가 아니기 위해서는 의사결정에 대한 인간의 감독이 형식적이 아니라 유의미하게 이루어져야 하며, 의사결정을 바꿀 수 있는 권한과 능력을 가진 사람이 개입하여야 하고, 분석 단계에서는 관련된 모든 데이터를 사람이 검토하여야 한다.

다만 계약의 체결 또는 이행을 위해 필요한 경우, 법률이 허용하는 경우, 정보주체의 명시적인 동의에 근거한 경우는 예외적으로 완전 자동화 의사결정이 허용된다. 특히 계약의 체결 또는 이행의 사유로 완전 자동화 의사결정을 실시하기 위하여는 자동화가 목적 달성에 필요최소한 처리여야 하고, 동일한 목표를 달성할 수 있는 덜 침해적인 수단이 있는 경우는 해당하지 않는다.[75] 완전 자동화 의사결정이 민감정보에 근거하여 이루어질 수 있는 경우로는 정보주체의 명시적인 동의에 의하거나, 법률에 기반한 상당한 공

75) Article 29 Data Protection Working Party (2017). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation, 2016/679. 23p.

익상의 이유로 처리가 필요하며 정보주체를 보호 조치가 존재하는 경우 뿐이다.

개인정보를 처리하는 자가 예외적으로 완전 자동화 의사결정을 실시하는 경우 정보주체의 권리와 자유 및 정당한 이익을 보호하기 위한 보호 조치가 반드시 마련되어야 한다. 이때의 보호 조치는 프로파일링 및 완전 자동화 의사결정 유무, 관련된 로직에 관한 구체적이고 유의미한 정보, 처리의 중대성과 이로 인해 발생할 수 있는 결과 등을 정보주체에게 사전적으로 설명하고, 그에 대하여 정보주체가 인간의 개입을 요구할 권리, 본인의 의견을 피력할 권리, 결정에 대한 설명을 들을 권리 및 결정에 이의를 제기할 권리 등을 보장하는 것을 포함한다. 더불어 개인정보를 처리하는 자는 처리한 데이터셋에서 편향성이 있는지 확인하고, 이를 해결할 수 있는 방법을 개발해야 한다. 또 알고리즘을 검사하고 자동화 의사결정의 정확성과 관련성을 주기적으로 검토하여 그 개선에 반영하여야 한다.

네덜란드에서는 플랫폼 노동자들이 GDPR 규정들을 토대로 완전 자동화 의사결정에 대한 권리를 인정받았다.[76] 2021년 3월 네덜란드 암스테르담 지방법원은 차량공유 플랫폼 올라와 우버에 대하여 노동자들이 열람을 요구한 개인정보를 각각 공개하라고 판결하였다. 법원은 올라 사건에서 운전자의 소득 프로필, 부정 행위 감지 시스템, 주행 할당 시스템 등 다양한 알고리즘과 자동화된 의사결정 절차를 검토하였고, 올라의 자동화된 '벌점 및 공제 시스템'이 법적 효력과 유사하게 본인에게 중대한 영향을 미치는 완전 자동화 의사결정이라고 결론을 내렸다. 이에 법원은 올라에 대하여 주행 업무에 대한 익명 평점, '부정행위 위험성 점수'를 생성하는 데 사용된 개인정보, 업무 할당에 영향을 미치는 소득 프로필을 생성하는 데 사용된 정보를 노동자에게 공개하라고 판결하였다. 또 우버에 대하여는 부정행위를 이유로 노동자를 차단하는 데 사용된 개인정보와, 주행 업무의 평균 평점이 아닌 각각의 익명 평점을 노동자에게 공개하라고 판결하였다. 다만 법원은 우버의 경우 노동자에게 중대한 영향을 미치는 완전 자동화 의사결정은 아니었다고 판단하였다.

---

76) Dutch court rulings break new ground on gig worker data rights. Financial Times 보도 (2021. 3. 13.); The Ola & Uber judgments: for the first time a court recognises a GDPR right to an explanation for algorithmic decision-making. EU Law Analysis (2021. 4. 28) <http://eulawanalysis.blogspot.com/2021/04/the-ola-uber-judgments-for-first-time.html (검색일: 2021. 9. 1.)>.

2021년 6월 10일에는 이탈리아 개인정보보호 감독기구가 배달 플랫폼인 푸디뉴가 라이더에게 차별적인 알고리즘을 사용한 데 대하여 2,600,000유로(약 35억원)의 과징금을 부과했다.[77] 푸디뉴가 라이더들에게 평점/벌점 시스템에서 사용하는 로직, 중요성, 예상 결과 등을 알리지 않았고, 완전자동화 의사결정에서 라이더의 인적 개입 요구권, 의견제시권, 이의제기권을 보장하지 않았다는 이유였다. 더불어 벌점 알고리즘의 정확성과 공정성을 보장하지 않고 배정 제외 등으로 라이더를 차별하고 평점의 부적절하고 차별적인 사용을 방지하는 조치 역시 위법하다고 인정되었다.

2. 유럽연합 인공지능법(안)

가. 인공지능법(안)의 내용

2021년 4월 21일 유럽 집행위원회는 인공지능법(안)을 발의하면서, 그 배경에 대하여 인공지능의 잠재적 이점은 의료 개선부터 더 나은 교육에 이르기까지 다양하지만 어떤 인공지능 시스템은 바람직하지 않은 위험을 발생시킨다고 설명하였다. 예를 들어, 많은 알고리즘의 불투명성은 불확실성을 유발하고, 안전 및 기본권리에 대한 기존 법률의 효과적인 집행을 방해할 수 있다. 이러한 과제에 대응하여 이점과 위험이 모두 적절하게 처리되는 인공지능 시스템을 위한 내부 시장이 제대로 작동하게 하는 입법 조치가 필요하다는 것이다.[78]

유럽연합 인공지능법(안)은 최초로 인공지능에 관하여 포괄적 규제프레임워크를 제시하였으며 ‘위험 기반 접근법(risk-based approach)’을 취하고 있는 것이 큰 특징이다.

---

77) Garante per la protezione dei dati personali (Italy) - 9675440. GDPR hub. <https://gdprhub.eu/index.php?title=Garante_per_la_protezione_dei_dati_personali_(Italy)_-_9675440 (검색일: 2021. 9. 1.)>.

78) Proposal for a Regulation laying down harmonised rules on artificial intelligence (2021. 4. 21). <https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence>; New rules for Artificial Intelligence - Questions and Answers. 유럽 집행위원회 보도자료 (2021. 4. 21). <https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683 (검색일: 2021. 9. 1.)>.

1) 입법배경

유럽 의회 및 유럽 이사회는 유럽연합 인공지능 시스템 시장에서 인공지능으로 인한 혜택과 위험이 유럽연합 수준에서 적절히 잘 다루어질 수 있도록 입법조치가 필요하다고 수차례 요구했다. 이에 유럽 집행위원회는 2021. 4. 21. 인공지능 활용, 투자 및 혁신을 강화하는 동시에 인공지능 개발 및 활용에 대한 신뢰, 기본권 및 사용자 안전을 보장하기 위하여 총 85개 조문으로 구성된 인공지능법(안)을 마련하였다.

2) 제정목적

유럽연합 시장에 출시·사용되는 인공지능 시스템의 안전성을 보장하고 기본적 권리와 유럽연합의 가치를 존중하고, 인공지능에 대한 투자·혁신이 촉진되도록 법적 안전성을 보장하며, 인공지능 시스템에 적용될 수 있는 기본권 및 안전기준 관련 법령의 효율적 집행을 강화하는 한편으로, 합법적이고 안전하며 신뢰할 수 있는 인공지능의 개발을 촉진하는 것이 주요 제정목적이다.

3) 제1편 일반조항

가) 적용범위(제2조)

인공지능법(안)은 유럽연합 내에서 설립되었는지 또는 제3국에서 설립되었는지 여부에 관계없이 유럽연합에서 인공지능시스템을 출시하거나 서비스를 개시하는 제공자, 인공지능시스템 사용자, 인공지능시스템 산출물이 유럽연합에서 사용되는 경우, 제3국에 위치한 인공지능시스템 제공자와 사용자 등이 적용대상에 포함된다. 이는 유럽연합을 시장으로 비즈니스를 하는 모든 기업이 규제대상이 되는 것을 의미하며, 고위험 인공지능의 경우 구매 후 사용시 사용자에게도 의무사항이 부과된다. 그러나 이 법은 오직 군사 목적으로만 개발되거나 사용되는 인공지능에는 적용되지 않는다.

나) 정의(제3조)

인공지능 시스템이란 기계 학습, 논리・지식기반 또는 통계적 접근방식으로 개발되고, 인간이 정의한 목표를 위해 그것이 상호 작용하는 환경에 영향을 미치는 콘텐츠, 예측, 추천, 결정 등의 아웃풋을 생성할 수 있는 소프트웨어를 뜻하며, ‘제공자’는 인공지능 시스템을 개발하거나 의뢰하는 자연인, 법인, 공공기관, 기구, 단체 등을 뜻하며, ‘사용자’는 인공지능 시스템을 사용하는 자연인, 법인, 공공기관, 기구, 기타 단체 등을 의미한다.

4) 인공지능 시스템의 분류

유럽연합은 인공지능 시스템을 위험 기반 접근법에 따라서 분류한 것이 가장 큰 특징이다. ① 금지되는 인공지능(용인할 수 없는 인공지능) ② 고위험 인공지능 시스템 ③ 제한된 위험 인공지능 시스템 ④ 최소위험 인공지능 시스템 총 4가지로 나누어서 각 인공지능 시스템별로 요구되는 의무를 규정하였다.

가) 금지되는 인공지능(용인할 수 없는 인공지능)

금지되는 인공지능이란 유럽연합의 가치, 인간의 존엄성, 자유, 평등, 민주주의, 기본권을 침해 하고 사람들의 안전, 생계 및 권리에 대한 명백한 위협으로 간주되는 인공지능 시스템으로서, 원칙적으로 금지된다. 다만, 법령에 따라 공공의 안전을 보호하기 위하여 적절한 안전조치를 취하여 공권력을 행사하는 경우는 예외이다.

금지되는 인공지능으로는 ① 사람의 의식을 뛰어넘는 잠재의식기술을 배치, 인지하지 못하는 방식으로 인간의 행동, 의견 또는 결정을 조작하여 자신 또는 타인에게 신체적・정신적 위험을 가져올 수 있는 인공지능 시스템 ② 개인·단체에 대한 정보 및 예측을 악용하여 아동・장애인 등의 취약성 또는 특수 상황을 표적으로 삼는 인공지능 시스템 ③ 공공기관이 사회적 행동 또는 알려지거나 예측된 개인의 특성을 기반으로 자연인의 신뢰도를 평가하거나 사람의 특성을 분류하여 불리한 대우를 하는 인공지능 시스템 ④ 경찰 등이 공개된 장소에서 실시간으로 생체정보를 활용하여 신원확인을 하는 인공지능 시스템. 다만, 실종아동 수색, 테러 대응 등 특정 범죄인 검거를 위해 사법당국의 사전승인을

받은 경우는 제외된다.

나) 고위험 인공지능 시스템

(1) 고위험 인공지능 시스템 분류

고위험 인공지능 시스템은 부속서Ⅱ의 제품자체 또는 제품의 안전 요소인 경우와 부속서Ⅲ의 기본권에 영향을 미칠 수 있는 독립형 인공지능 시스템으로 분류하고 있다. 먼저, ① 유럽연합 조화법령이 적용되는(부속서Ⅱ[79]) 제품의 안전 구성요소로 사용되거나 그 자체가 제품인 경우, 인공지능 시스템이 안전 구성요소인 제품 또는 제품으로서 인공지능 시스템 자체를 부속서Ⅱ에 열거된 유럽연합 조화법령에 따라 출시하거나 서비스 개시하려면 제3자 적합성 평가를 거쳐야 한다. 그리고 ② 자연인의 '실시간' 및 '사후' 원격 생체 인식에 사용되는 인공지능 시스템(자연인의 생체 인식 및 분류), ③ 도로 교통, 수도, 가스, 난방 및 전기 공급의 관리 및 운영에서 안전 구성 요소로 사용되는 인공지능 시스템(중요 인프라의 관리 및 운영), ④ 교육 및 직업 훈련 기관에 대한 접근을 결정하거나 자연인을 할당할 목적으로 사용되는 인공지능 시스템 및 교육 및 직업 훈련 기관의 학생을 평가하고 교육 기관 입학을 위해 일반적으로 요구되는 시험 참가자를 평가하기 위한 목적으로 사용되는 인공지능 시스템(교육 및 직업 훈련), ⑤ 자연인 모집 또는 선정, 특히 결원 광고, 지원자 심사 또는 선별, 인터뷰 또는 테스트 과정에서의 후보자 평가에 사용하는 인공지능 시스템, 업무 관련 계약 관계의 촉진 및 종료에 대한 결정, 작업 할당, 이러한 관계에 있는 사람들의 성과 및 행동을 모니터링 및 평가하는 데 사용되는 인공지능(고용, 근로자 관리 및 자영업에 대한 접근), ⑥ 공공 기관에 의해 또는 공공 기관을 대신하여 공공 지원 혜택 및 서비스에 대해 자연인의 적격 여부를 평가하고, 그러한 혜택 및 서비스를 부여, 축소, 취소 또는 회수하기 위해 사용하는 인공지능 시스템, 소규모 공급 업체가 자체 목적을 위해 사용하는 인공지능 시스템을 제외하고 자연인의 신용도를 평가하거나 신용 점수를 설정하기 위해 사용하는 인공지능 시스템, 소방관 및 의료 지원을 포함하여 긴급 대응 서비스를 파견하거나 우선 순위를 설정하기 위해 사용하는 인공지능 시스템(필수적인 민간 서비스, 공공 서비스 및 혜택에 대한 접

79) 장난감, 자동차, 의료기기, 기계류, 항공 등 - 제품의 안전과 관련된 요소.

근 및 향유), ⑦ 법집행과 관련하여 자연인의 범죄 또는 재범 위험 또는 형사 범죄의 잠재적 피해 위험을 평가하기 위해 법 집행 기관에서 자연인에 대한 개별 위험 평가를 위해 사용하는 인공지능 시스템, 법 집행 기관에서 거짓말 탐지기 및 유사한 도구로 사용하거나 자연인의 감정 상태를 감지하기 위해 사용하는 인공지능 시스템, 딥 페이크 감지를 위해 법 집행 기관이 사용하는 인공지능 시스템, 범죄 수사 또는 기소 과정에서 증거의 신뢰성을 평가하기 위해 법 집행 기관이 사용하는 인공지능 시스템, 자연인의 프로파일링을 기반으로 실제 또는 잠재적인 범죄 행위의 발생 또는 재발을 예측하거나 또는 자연인 또는 그룹의 성격 및 특성 또는 과거 범죄 행위를 평가하는데 사용하는 인공지능 시스템, 형사 범죄의 탐지, 조사 또는 기소 과정에서 자연인의 프로파일링을 위해 법 집행기관에서 사용하는 인공지능 시스템, 자연인에 대한 범죄 분석에 사용되어 법 집행기관이 다양한 데이터 소스 또는 다양한 데이터 포맷에서 사용 가능한 관련성이 있거나 없는 복잡한 데이터를 검색하여 데이터의 알 수 없는 패턴을 식별하거나 숨겨진 관계를 발견하는데 사용되는 인공지능 시스템, ⑧ 이주, 망명 및 국경 통제 관리와 관련하여 관할 공공 기관에서 거짓말 탐지기 및 유사한 도구로 사용하거나 자연인의 감정 상태를 감지하기 위해 사용하는 인공지능 시스템, 회원국의 영토에 입국을 시도하거나 입국한 자연인에 의해 발생되는 보안 위험, 비정규 이민 위험 또는 건강 위험을 포함한 위험을 평가하기 위해 관할 공공 기관이 사용하는 인공지능 시스템, 관할 공공 기관이 여행 문서의 진위 여부를 확인하고, 자연인의 문서를 지원하고, 보안 기능을 확인하여 비 인증 문서를 감지하기 위해 사용하는 인공지능 시스템, 망명, 비자 및 거주 허가 신청, 신분 신청 자연인의 적격성과 관련된 민원 등의 조사를 위해 관할 공공 기관을 지원하는 인공지능 시스템, ⑨ 사법 당국이 사실과 법을 조사 및 해석하고 구체적인 사실에 법을 적용하는데 도움을 주기위한 인공지능 시스템(사법 행정 및 민주적 절차) 등을 고위험 인공지능 시스템으로 분류하고 있다.

(2) 고위험 인공지능시스템에 요구되는 사항

① 위험관리 시스템.

고위험 인공지능시스템과 관련한 위헌 관리 시스템을 구축, 시행, 기록, 유지해야 한다. 위험관리 시스템은 (i)각 고위험 인공지능 시스템과 관련된 알려지고 예측 가능한

위험의 파악 및 분석, (ii)합리적으로 예측 가능한 오용의 조건 하에서 원래의 목적으로 사용할 때 발생할 수 있는 위험의 추정 및 평가, (iii)출시 후 모니터링 시스템에서 수집한 데이터의 분석에 근거한 발생 가능한 다른 위험의 평가, (iv)적합한 위험 관리 수단 등으로 구성된다. 또한 고위험 인공지능 시스템이 원래 목적에 따라 또는 합리적으로 예측 가능한 오용의 조건 하에서 사용되도록 하여야 하고, 이러한 위험을 사용자에게 통지해야 한다. 가장 적합한 위험 관리 수단을 모색하는 과정에서 (a) 적합한 설계와 개발을 통해 최대한 위험 제거 또는 완화 (b) 적절한 경우, 제거할 수 없는 위험에 대해 적합한 완화 및 통제 조치 시행 (c) 충분한 정보 제공, 및 적절한 경우 사용자 교육 등의 사항이 보장되어야 한다.

② 데이터 및 데이터 거버넌스

데이터를 통한 모델의 학습을 수반하는 기법을 사용하는 고위험 인공지능 시스템은 학습, 검증, 테스트 데이터셋을 기반으로 개발되어야 하고, 학습, 검증, 테스트 데이터셋에는 적절한 데이터 거버넌스 및 관리 관행[80]이 적용되어야 한다. 학습, 검증, 테스트 데이터셋은 관련성이 있고, 오류가 없고, 완전해야 한다. 데이터셋은 고위험 인공지능 시스템의 사용 대상인 개인 또는 집단과 관련성을 포함하여 적절한 통계적 특성을 가져야 한다. 데이터셋의 이러한 특성은 개별 데이터셋 또는 데이터셋 조합의 수준에서 충족될 수 있다. 학습, 검증, 테스트 데이터셋은 원래 목적이 요구하는 한에서, 고위험 인공지능 시스템이 사용되는 지리적, 행동적, 기능적 환경에 특유한 특성 또는 요소를 고려해야 한다. 고위험 인공지능 시스템과 관련된 편향 모니터링, 탐지, 시정의 목적을 위해 절대적으로 필요한 경우, 특수한 범주의 개인 데이터를 처리할 수 있다. 단, 가명화 또는 익명화가 추구하는 목적에 상당한 영향을 줄 수 있는 경우, 암호화와 같은 첨단 보안 및 개인정보보호 수단의 사용 및 재사용에 대한 기술적 제한을 포함하여 자연인의 기본권과 자유를 보호할 적절한 수단이 확보되어야 한다.

---

80) 이러한 관행은 특히 다음과 관련된다.
(a) 설계 선택
(b) 데이터 수집
(c) 주석, 레이블링, 정리, 보강, 집계 등 데이터 준비 처리 작업
(d) 특히 데이터가 측정하고 표시해야 하는 정보와 관련한 가정의 공식화
(e) 필요한 데이터셋의 가용성, 품질, 지속가능성에 대한 사전 평가
(f) 가능한 편향을 고려한 조사
(g) 가능한 데이터 갭 또는 부족 및 그러한 갭과 부족을 해소하는 방법의 파악.

③ 기술문서

고위험 인공지능 시스템의 시장출시 또는 서비스 개시 전에 법률 요구사항을 준수하고 있음을 입증하기 위해 최신의 기술문서를 작성·유지해야 한다(제11조 기술문서). 고위험 인공지능 시스템이 요구사항을 준수한다는 것을 입증하는 방식으로 작성되어야 하며, 인공지능 시스템이 동 요구사항을 준수하는지 평가하는 데 필요한 모든 정보를 국가관할 당국과 인증 기관에 제공해야 한다.

④ 기록 유지

모든 단계에서 인공지능 시스템의 기능을 추적할 수 있는 수준으로 로그기록이 저장되어야 한다. 이러한 로깅 기능은 공인 표준 또는 공통 규격을 준수해야 한다. 로깅 기능은 인공지능 시스템의 라이프사이클 전반에 걸쳐 그 기능에 대해 시스템의 원래 목적에 적합한 수준의 추적 가능성을 보장해야 한다. 자연인의 '실시간' 및 '사후' 원격 생체 인식에 사용되는 고위험 인공지능 시스템의 경우 (a) 시스템의 각 사용 기간(각 사용의 시작 날짜 및 시간과 종료 날짜 및 시간)의 기록 (b) 시스템이 인풋 데이터를 확인하는 근거가 된 참조 데이터베이스 (c) 검색이 일치로 이어진 인풋 데이터 (d) 결과의 검증에 관여한 자연인의 신원 항목을 제공해야 한다.

⑤ 투명성과 정보 제공

고위험 인공지능 시스템은 사용자가 시스템의 아웃풋을 해석하고 적절히 사용할 수 있을 만큼 충분히 투명하게 운영되도록 설계·개발되어야 한다. 투명성을 보장하기 위해 필요한 수단은 다음과 같다.

(a) 제공자 및 적절한 경우 공인 대리인의 신원과 연락처 세부사항

(b) 다음을 포함한 고위험 인공지능 시스템의 수행 특성, 기능 및 제한

(i) 원래 목적

(ii) 고위험 인공지능 시스템을 테스트·검증한 기준이 되고 예상될 수 있는 정확성, 견고성 및 사이버 보안의 수준, 및 그와 같이 예상되는 정확성, 견고성 및 사이버 보안의 수준에 영향을 미칠 수 있는 알려지고 예측 가능한 상황

(iii) 원래 목적에 따라 또는 합리적으로 예측 가능한 오용 조건 하에서 고위험 인공지능 시스템을 사용하는 데 따른, 건강과 안전 또는 기본권에 대한 위험으로 이어

질 수 있는 알려지거나 예측 가능한 상황

(iv) 시스템의 사용 대상인 개인 또는 집단과 관련된 수행

(v) 인공지능 시스템의 원래 목적을 고려한 인풋 데이터에 대한 규격 또는 사용되는 학습, 검증, 테스트 데이터셋과 관련된 기타 모든 정보

(c) 고위험 인공지능 시스템과 초기 적합성 평가 시에 제공자가 사전 결정한 그 성능의 변경

(d) 사용자가 인공지능 시스템의 아웃풋을 해석할 수 있도록 해주는 기술적 수단을 포함한 인간의 감독 수단

(e) 고위험 인공지능 시스템의 예상 수명, 및 소프트웨어 업데이트를 포함하여 동 시스템의 올바른 기능을 보장하는 데 필요한 유지 관리 수단

⑥ 인간의 감독

고위험 인공지능 시스템은 사용되는 기간 동안 인간이 효과적으로 감독할 수 있는 방식으로 설계·개발되어야 한다. 감독책임을 맡은 인간이 상황에 따라 다음과 같은 작업을 수행할 수 있도록 해야 한다. (a) 고위험 인공지능 시스템의 능력과 한계를 충분히 이해하고 그 운영을 적절히 모니터하여 이상의 징후, 기능 장애 및 예기치 않은 작동을 탐지하고 가능한 한 신속히 해결한다. (b) 특히 자연인이 내리는 의사결정을 위한 정보 또는 권고를 제공하는 데 사용되는 고위험 인공지능 시스템의 경우, 시스템이 산출한 아웃풋에 자동적으로 의존하거나 지나치게 의존하는 경향(‘자동화 편향’)을 인지한다. (c) 특히 시스템의 특성과 가용한 해석 도구 및 방법을 고려하여 고위험 인공지능 시스템의 아웃풋을 정확하게 해석한다. (d) 특별한 상황에서 고위험 인공지능 시스템을 사용하지 않거나 고위험 인공지능 시스템의 아웃풋을 무시 또는 번복하기로 결정한다. (e) 고위험 인공지능 시스템의 운영에 개입하거나 “중지” 버튼 또는 유사한 절차를 통해 시스템을 중단시킨다.

⑦ 정확성, 견고성, 사이버 보안

고위험 인공지능시스템은 정확성, 견고성, 사이버 보안을 갖추어야 한다. 고위험 인공지능 시스템의 정확성 수준 및 정확성 척도는 첨부한 사용 지침에 명시되어야 한다. 고

위험 인공지능 시스템은 특히 자연인 또는 다른 시스템과의 상호작용으로 인해 시스템 내에서 또는 시스템이 운영되는 환경에서 발생할 수 있는 오류, 고장, 불일치에 대해 복원력을 가져야 한다. 고위험 인공지능 시스템은 허가받지 않은 제3자가 시스템 취약성을 이용하여 그 사용 또는 수행을 변경하려는 시도에 대해 복원력을 가져야 한다.

(3) 고위험 인공지능시스템 제공자의 의무

고위험 인공지능시스템 제공자는 위에서 언급한 고위험 인공지능시스템의 요구사항을 준수해야 한다. 또한 품질관리 시스템 배치, 기술문서 작성, 로그기록 유지, 시스템 출시 전 적합성 평가, 등록의무, 요구사항들을 준수하지 않을 경우 필요한 시정 조치 등을 취해야 할 의무가 있다.

① (품질관리 시스템) 고위험 인공지능 시스템 제공자는 품질 관리 시스템을 배치하여야 한다. 정책, 절차, 지침의 형태로 체계적이고 정연한 방식으로 기록되어야 하며, 구체적으로 고위험 인공지능시스템의 설계・관리・검증에 사용되는 기법, 절차 및 체계적 조치, 시스템 개발・품질 관리・품질 보증에 사용되는 기법, 절차 및 체계적 조치 등의 사항들을 포함해야 한다.

② (기술문서 작성 의무) 제공자는 앞에서 언급했던 고위험 인공지능 시스템에 관한 기술 문서를 작성해야 한다.

③ (적합성 평가 의무) 제공자는 시스템이 출시되거나 서비스가 개시되기 전에 제43조에 따른 적합성 평가 절차를 거쳐야 한다. 이에 따라 요구사항들이 준수된 것으로 입증된 경우, 제공자는 법 제48조에 따른 EU 적합성 선언을 작성하고, 제49조에 따른 CE 적합성 마크를 부착해야 한다.

④ (자동으로 생성하는 로그 유지 의무) 제공자는 로그가 사용자와의 계약 또는 법률에 의해 제공자의 통제하에 있는 경우에 한하여 자동으로 생성하는 로그를 유지해야 한다. 이러한 로그는 고위험 인공지능시스템의 원래 목적과 유럽연합법 또는 국가법에 따라 적용되는 법적의무에 비추어 적절한 기간 동안 보관해야 한다.

⑤ (시정조치 의무) 제공자는 시스템 준수를 이행하거나 회수, 리콜하는데 필요한 시정조치를 취해야 한다.

⑥ (통지의무) 고위험시스템이 위험을 야기할 경우, 제공자는 시스템을 제공한 국가, 인증기관 등에게 시정 조치를 즉시 통지할 의무가 있다.

⑦ (관할 기관과 협력 의무) 국가 관할 기관이 요구할 경우, 제공자는 고위험 인공지능 시스템의 요구사항을 준수한다는 정보와 문서를 제공해야 한다.

(4) 고위험 인공지능시스템 수입업자의 의무

수입업자는 제공자가 시장출시 전에 적합성 평가 절차를 수행하고 기술문서를 작성했는지, 시스템에 요구되는 적합성 표시가 있는지, 필수 문서 및 사용안내서가 함께 제공되는 여부를 확인해야 한다. 고위험 인공지능시스템이 규정을 준수하지 않는다고 판단될 경우 이행하기 전까지 해당 시스템을 시장에 출시해서는 안된다. 국가 기관의 요청이 있는 경우, 고위험 인공시스템이 요구사항을 준수하고 있음을 입증하는 데 필요한 모든 정보와 문서를 쉽게 이해할 수 있는 언어로 제공해야 한다.

(5) 고위험 인공지능시스템 유통업자의 의무

유통업자는 시장출시전 고위험 인공지능시스템에 CE 적합성 표시가 있는지, 필수 문서 및 사용 지침이 포함되어 있는지, 제공자와 수입업자가 명시된 의무를 준수하였는지 확인해야 한다. 수입업자와 마찬가지로 고위험 인공지능시스템이 요구사항을 준수하지 않는다고 판단될 경우, 이행될 때까지 해당 시스템을 시장에 출시해서는 안 된다. 위험 징후가 보이는 경우 제공자, 수입업자에게 그 영향을 알려야 하며, 관할 당국에 즉시 통보하고 미준수 및 시정조치 등 세부 사항에 관한 정보를 제공해야 한다. 고위험 인공지능시스템이 요구사항을 준수하지 않을 경우, 시정조치, 회수 또는 리콜 등을 취하거나, 제공자, 수입업자, 관련 운영자가 해당 시정조치를 취하도록 해야 한다.

(6) 고위험 인공지능시스템 사용자의 의무

사용자는 사용지침에 따라 고위험 인공지능시스템을 사용하여야 하며, 고위험 인공지능시스템 운영을 모니터해야 한다. 위험을 야기할 수 있다고 간주할 만한 이유가 있는 경우, 제공자, 유통업자에게 통지하고 시스템 사용을 중단해야 한다. 또한 사용자는 각자 고위험 인공지능시스템이 자동으로 생성하는 로그를 유지해야 한다.

다) 특정 인공지능 시스템(혹은 제한된 위험 인공지능 시스템)의 투명성 의무

금지되는 인공지능 시스템이나 고위험 인공지능 시스템이 아니나, 명의도용 또는 사기의 위험을 초래할 수 있는 특정 인공지능 시스템에는 투명성 의무가 적용된다.

① 제공자는 자연인과 상호 작용하는 인공지능 시스템이 해당 자연인이 인공지능 시스템과 상호 작용하고 있다는 것을 고지하는 방식으로 설계·개발되도록 보장한다. 단, 사용 상황과 맥락에서 이것이 명백한 경우는 예외로 한다. 이 의무는 범죄 행위의 탐지, 방지, 수사, 기소를 위해 법률이 허가한 인공지능 시스템에는 적용되지 않는다. 단, 일반인이 이러한 시스템을 범죄 행위의 신고에 이용할 수 있는 경우는 예외로 한다. ② 감정 인식 시스템 또는 생체 인식 분류 시스템의 사용자는 그에 노출되는 개인에게 시스템의 작동에 대해 고지해야 한다(이 의무는 범죄 행위의 탐지, 방지, 수사를 위해 법률이 허용하는 생체 인식 분류에 사용되는 인공지능 시스템에는 적용되지 않는다). ③ 기존의 사람, 물체, 장소, 기타 실체 또는 사건과 현저히 유사하고 마치 진본처럼 보이는 이미지, 오디오 또는 비디오 콘텐츠를 생성하거나 조작하는(‘딥 페이크’) 인공지능 시스템의 사용자는 해당 콘텐츠가 인공적으로 생성 또는 조작되었음을 공개해야 한다. 단, 범죄 행위의 탐지, 방지, 수사를 위해 법률이 사용을 허가하거나 유럽연합 기본권 헌장에 보장된 표현의 자유 또는 학문과 예술의 자유에 대한 권리를 행사하기 위해 필요하고 제3자의 권리와 자유에 대한 적절한 보호 조치가 적용되는 경우에는 적용되지 않는다.

5) 혁신을 지원하는 조치

가) 규제샌드박스

하나 이상의 회원국 관할 기관 또는 유럽 개인정보보호 감독관(European Data Protection Supervisor, EDPS)에 의해 설립된 인공지능 규제 샌드박스는 특정한 계획에 따라 혁신적인 인공지능 시스템이 출시되거나 서비스 개시되기 전에 제한된 시간 동안 개발, 테스트, 검증할 수 있는 통제 환경을 제공한다. 그리고 인공지능 규제 샌드박스는 관할 기관의 감독 및 시정 권한에 영향을 주어서는 안된다. 시스템의 개발 및 테스트 과정에서 건강과 안전 및 기본권에 대한 중대한 위험이 확인되는 경우 이를 즉시 완화해

야 하며 그에 실패할 경우 완화가 이루어질 때까지 개발 및 테스트 프로세스가 일시 중지된다.

또한 규제샌드박스 내에서는 안보, 범죄의 예방·조사·탐지·기소 또는 형 집행, 공공 안전 및 공중 보건, 환경보호 및 품질 개선 등 공공의 이익 목적이 있는 경우, 익명화·가명화로는 효과적인 목적달성이 어려운 경우, 개인정보 침해위험 모니터링 및 대응 메커니즘 마련해야 하는 경우, 개인정보에 대한 통제와 접근제한이 이루어진 경우 인공지능 시스템의 개발· 테스트 목적으로 개인정보 처리가 가능하다. 그러나 처리되는 개인 데이터가 기능적으로 분리·보호되는 데이터 처리 환경에 있고, 참가자의 통제 하에 오로지 허가된 사람만 해당 데이터에 접근할 수 있다. 처리되는 개인 데이터를 다른 당사자가 전송, 이전 또는 접근하지 않고, 개인 데이터의 처리가 정보주체에게 영향을 미치는 조치 또는 결정으로 이어지지 않으며, 샌드박스에 대한 참여가 종료되거나 개인 데이터의 보존 기간이 만료되면 샌드박스의 맥락에서 처리되는 모든 개인 데이터가 삭제되는 조치를 취하였기 때문에 일정한 목적 하에 개인정보 처리가 가능하더라도 이를 보호할 수 있는 장치를 마련하였다.

나) 중소사업자 지원

소규모 제공자와 스타트업 기업들이 자격조건을 충족하는 경우 인공지능 규제샌드박스에 우선 접근할 수 있도록 한다. 이들과 의사소통을 위한 전담채널을 구축하여 본 법안의 지침을 제공하고 질문에 응답한다. 적합성 평가 수수료 책정시 이들의 요구를 고려하여 규모와 시장 규모에 비례하도록 수수료를 경감한다.

6) 거버넌스 - 유럽 인공지능 위원회, 감독기관

회원국들 전문 지식과 모범사례 수집하고 공유하고, 규제 샌드박스 기능을 포함하여 일관성 있는 행정실무에 기여하며, 기술 표준, 지침서를 작성하는 유럽 인공지능 위원회를 설립한다. 따라서 회원국들은 법의 적용과 이행을 감독하고 법 이행에 필요한 지침을 제공하고 지원하는 감독기관을 지정하거나 설립해야 한다. 감독 기관은 시장 감시 활동 과정에서 파악된, 경쟁 규칙에 관한 유럽연합법의 적용에 중요할 수 있는 모든 정보를

유럽연합 집행위원회와 관련 국가 경쟁당국에 지체 없이 보고해야 한다. 유럽 연합 기관, 기구, 단체가 본 규정의 범위 내에 속하는 경우 유럽 데이터 보호 감독관이 이들의 시장 감독 기관 역할을 수행한다.

또한 활동의 맥락에서 데이터 및 문서에 대한 접근이 허용되어야 한다. 시장 감독기관은 애플리케이션 프로그래밍 인터페이스('API') 또는 원격 접근을 지원하는 기타 적절한 기술적 수단 및 도구를 포함하여 제공자가 사용하는 학습, 검증 및 테스트 데이터셋에 전면적으로 접근할 수 있어야 한다. 고위험 인공지능 시스템이 법안에 명시된 요구사항을 준수하는지 평가하는 데 필요하고 합리적으로 요청하는 경우, 시장 감독기관은 인공지능 시스템의 소스 코드에 대한 접근이 허용되어야 한다.

7) 처벌규정

금지되는 인공지능시스템을 준수하지 않거나, 데이터 및 데이터 거버넌스 요구사항을 준수하지 않을 경우 최대 3,000만 유로 또는 위반자가 기업인 경우 전년도 전세계 연매출의 최대 6% 중에서 더 높은 금액의 과징금이 부과된다. 요구사항 또는 의무(금지되는 인공지능시스템 및 데이터 및 데이터거버넌스 제외)를 위반한 경우, 최대 2,000만 유로, 또는 위반자가 기업인 경우 전년도 전세계 연매출의 최대 4% 중에서 더 높은 금액의 과징금이 부과된다. 인증 기관과 국가 관할 기관의 요청에 응답하여 부정확하거나 불완전하거나 오도하는 정보를 제공하는 경우 최대 1,000만 유로, 또는 위반자가 기업인 경우 전년도 전세계 연매출의 최대 2% 중에서 더 높은 금액의 과징금이 부과된다.

나. 인공지능법(안)에 대한 의견

1) 유럽 개인정보보호 기관의 의견[81]

유럽연합의 주요 개인정보보호 감독기구로는 EDPS와 EDPB를 들 수 있다.

81) Euproean Data Protection Board, European Data Protection Supervisor (2021). Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (2021. 6. 18.).

EDPS(European Data Protection Supervisor)는 독립적인 유럽 개인정보보호 감독관으로 유럽연합 기관과 기구들의 개인정보 처리를 감독한다. EDPB(Euproean Data Protection Board)는 유럽연합 GDPR을 해석하고 유럽 집행위원회의 개인정보보호 정책을 자문하는 독립적인 유럽 개인정보보호이사회로서 EDPS 및 회원국 개인정보보호 감독기구 대표들로 구성되어 있다.

EDPB와 EDPS는 2021년 6월 18일 집행위원회가 의회에 발의한 인공지능법(안)에 대하여 공동의견을 발표하며, 이 법안이 개인정보보호와 중요한 관련성이 있다고 지적하고 개인정보보호 관점에서 견해를 밝혔다.

우선 EDPB와 EDPS는 인공지능법(안)이 인공지능 위험성에 대한 규제를 추진하고 있다는 점을 긍정적으로 평가하였다.

인공지능 시스템의 위험성이 큰 것은 우리가 대부분 인공지능 시스템이 개인 및 사회에 미치는 영향에 대한 경험이 없기 때문이다. 기계학습 기법이나 로직, 확률론적 추론 규칙을 통해 콘텐츠를 생성하거나 자동적으로 예측하거나 의사결정을 하는 것은, 인간이 결과에 대한 모든 책임을 지고 창의적 또는 이론적 추론을 통해 활동을 수행하는 것과 같지 않을 수밖에 없다는 것이다. 인공지능은, 인간의 눈에는 보이지 않지만 기계에는 보이는 데이터 간 측정 가능한 상관관계로부터 시작해 많은 분야에서 수행할 수 있는 예측의 양을 늘려 우리의 삶을 더 수월하게 만들고 수많은 문제를 해결할 수 있을 것이다. 그러나 이와 동시에 결과에 대해 인과적 해석을 할 수 있는 인간의 능력을 약화시켜 투명성, 인간의 통제, 책무성 및 결과에 대한 책임이라는 개념이 심각하게 도전받는다고도 지적한다. 또한 데이터에 기반하여 의사결정 업무를 기계에 할당하는 것은 개인의 권리와 자유에 위험을 발생시키며 개인의 사생활에 영향을 미치고 집단이나 사회 전체에 해를 끼칠 수 있다. 특히 사생활과 개인정보보호에 대한 권리는, 인공지능 개념의 기초가 되는 기계의 의사결정 자율성이라는 가정과 상충하는 측면이 있다. 따라서 EDPS와 EDPB는 인공지능 법안이 인공지능 성장 전망과 기계에 대한 인간의 중심성과 우선성을 조화시키는 데 기여할 수 있기를 기대한다.

EDPB와 EDPS는 인공지능법(안)이 유럽연합 시민과 거주자의 기본권을 보장하기 위해 필요하다고 생각하지만, 그 적용 가능성과 효율성을 보장하기 위해 몇 가지 쟁점에서 법안을 수정할 것을 제안하였다. 특히 EDPB와 EDPS는 인공지능법(안)과 기존 개인정보보

호 체제와의 일관성과 조화를 주문하였다. 유럽연합의 기존 개인정보보호 체제는 GDPR, 유럽연합 기관·기구 및 청의 개인정보보호 규정(EUDPR) 및 경찰 지침(Law Enforcement Directive, LED)으로 구성되어 있다.

EDPB와 EDPS은 인공지능법(안)을 쟁점별로 살펴보았는데, 이는 크게 (1)법안의 주요 원칙에 있어서 쟁점과 (2)개인정보보호 체제와의 관계에 있어서 쟁점으로 나누어볼 수 있다. 다시 (1)법안의 주요 원칙에 있어서 △법안의 적용 범위 및 기존 법률 체제와의 관계 △위험 기반 접근법 △사용 금지 인공지능 △고위험 인공지능 △유럽 인공지능위원회 등 거버넌스에 대한 쟁점을 살펴 보았고, (2)개인정보보호 체제와의 관계에 있어서는 △기존 개인정보보호 법률들과의 관계 △샌드박스 및 추가처리 △투명성 △민감정보 △인증 및 행동지침 등 준수 구조에 대한 쟁점을 검토하였다.

우선 법안의 적용 범위와 관련하여, 이 법안은 일부 법규의 범위에 속하는 고위험 인공지능 시스템[82]에 대해서는 제84조(평가 및 검토)만을 적용하고 있다(제2조(2)). 또한 오직 군사 목적으로만 개발되거나 사용되는 인공지능 시스템에는 법안이 적용되지 않으며(동조(3)), 국제적인 법집행 및 사법 협력을 위한 국제협약 차원에서 인공지능 시스템을 사용하는 경우에도 법안이 적용되지 않는다(동조(4)). EDPB와 EDPS는 법안 적용범위에 국제 형사사법 협력 부문이 배제되는 것에 대하여 우려하였다. 또한 이 법의 적용에 있어 전문(recital)은 개인정보보호 법률들을 함께 준수해야 한다고 명시하고는 있지만,

82) (a) Regulation (EC) 300/2008 on common rules in the field of civil aviation security (민간항공보안 공통규칙 관련 규정)
(b) Regulation (EU) No 167/2013 on the approval and market surveillance of agricultural and forestry vehicles (농업 및 임업용 차량 승인 및 시장 감독 관련 규정)
(c) Regulation (EU) No 168/2013 on the approval and market surveillance of two- or three-wheel vehicles and quadricycles (이륜, 삼륜 차량 및 사륜 오토바이 승인 및 시장 감독 관련 규정)
(d) Directive 2014/90/EU on marine equipment (해양장비 관련 지침)
(e) Directive (EU) 2016/797 on the interoperability of the EU's rail system (EU 철도시스템 상호운용성 관련 지침)
(f) Regulation (EU) 2018/858 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles (자동차 형식 승인 및 시장 감독 관련 규정)
(g) Regulation (EU) 2018/1139 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency (무인항공분야 공통규칙 및 EU항공안전청 설립 관련 규정)
(h) Regulation (EU) 2019/2144 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users (자동차 형식 승인 요건과 안전 및 보호 관련 규정)

EDPB와 EDPS는 법안 조문에서 이를 명확히 할 것을 요구하였다.

다음으로 위험 기반 접근법과 관련하여, EDPB와 EDPS는 법안의 위험 기반 접근법을 환영하였다.

나아가 인공지능 시스템에 의해 야기되는 사회적/집단별 위험성(예: 집단적 차별이나 공공장소 의사 표현 등)을 동등하게 평가하고 완화시킬 것을 권고하였다. 또한 개인정보보호와 관련된 측면이 적용되는 만큼, '기본권에 대한 위험'의 개념이 GDPR 및 EUDPR과 일치되어야 한다고 지적하였다. 특히 법안 본문에 인공지능 시스템의 영향을 받는 개인에 대한 어떠한 언급도 없다는 사실을 사각지대로 보았다. 사람들에게 영향을 미친 행위자에게 부과되는 의무는 더 구체적으로 도출되어야 하기 때문에, 입법자들은 인공지능 시스템의 대상자가 이용할 수 있는 권리와 구제 수단을 법안에 명시적으로 언급하여야 한다는 것이다.

더불어 법안의 부속서에 열거된 고위험 인공지능 목록에 현재 보험료 결정이나 의료적 처치 또는 건강 연구 목적의 사용과 같이 상당한 위험을 수반하는 일부 유형의 사용 사례가 누락되어 있다고 지적하고, 해당 목록이 정기적으로 갱신되어야 한다고 강조하였다.

한편, 사전 적합성평가 의무가 부과된 제공자는 인공지능 시스템의 모든 사용이나 개인정보처리에 따른 위험성을 평가하는 것이 불가능하다. 따라서 제공자의 초기적인 위험 평가가 '고위험'으로 평가하지 않았더라도, 시스템의 사용자인 개인정보처리자가 개인정보보호 법률들에 따른 개인정보보호 영향평가를 독립적으로 실시할 필요가 있다. 현재는 '고위험'으로 평가된 인공지능 시스템의 사용자의 경우 개인정보보호 영향평가 실시 의무만 언급되어 있다(제29조(6)).

EDPB와 EDPS는 인공지능 시스템을 고위험으로 분류하는 것만으로 반드시 합법적이고 따라서 사용자가 이를 도입할 수 있다는 의미는 아니라는 언급에 동의하면서(전문31), 특히 유럽 시장 진입을 위한 개인정보보호 법률들의 준수 문제는 이 법 이전의 전제조건이라고 강조하였다. 이에 법안에서 고위험 시스템은 금지된 시스템과 달리 원칙적으로 허용될 수 있다고 받아들여질 수 있는 문구를 보완할 것과, 법안 Ⅲ편 2장(고위험 인공지능 시스템에 대한 요구사항)에 GDPR 등의 준수를 보장하는 요구사항을 포함할 것을 권고하였다. 또 제공자가 실시하도록 되어 있는 위험 평가에서 기술적 특성과 사용 사례를

고려할 것을 권고하였다.

사용 금지 인공지능과 관련하여서, EDPB와 EDPS는 법집행 목적의 인공지능(부속서 Ⅲ. 6.) 중 침입적 형태의 인공지능(특히 인간의 존엄성에 영향을 미칠 수 있는 인공지능)은 단순히 '고위험'으로 분류되는 것이 아니라 법안 제5조에 따라 금지된 인공지능 시스템이라고 보았다. 특히 경찰의 감시를 받을 이유가 없거나 경미한 사람에게도 영향을 미치는 대규모의 데이터 비교, 또는 개인정보보호 법률들에 따른 목적 제한 원칙을 훼손하는 개인정보 처리를 수행하는 경우가 그렇다. 치안과 법 집행 분야에서 인공지능을 사용하려면 관련된 당사자의 이익과 민주 사회의 기능에 미치는 영향을 지역별로 고려하고, 정확하며, 예측 가능하고, 비례적인 규칙이 필요하다. 이러한 고려가 없는 법안 제5조(금지되는 인공 지능 관행)는 실제로는 의미가 없는 '립서비스'에 그칠 수 있다고 비판하였다(예: 제5조(1)(a) 및 (b)의 "물질적 또는 정신적 피해를 주거나 줄 가능성이 있는 방식" 부분, 제5조 (1)(c)를 공공기관으로 한정한 부분, 제5조(c)(i) 및 (ii)의 모호한 문구, 제5조(d)에서 명확한 정의 없이 '실시간' 원격 생체인식으로 한정한 부분 등).

특히 제5조(1)(c)는 '일정 기간 이상' 또는 '공공기관에 의해 또는 그 기관을 대신하여' 수행되는 경우에 '사회신용점수 부여'를 금지한다. 하지만 소셜미디어나 클라우드 서비스 제공업체 등 민간 기업도 방대한 양의 개인정보를 처리하고 사회신용점수를 부여할 수 있기 때문에, 인공지능법은 어떠한 형태의 사회신용점수 부여도 금지하여야 한다는 것이 EDPB와 EDPS의 의견이다. 또한, 공개적으로 접근할 수 있는 공간에서 개인에 대한 원격 생체인식은 감시대상 뿐 아니라 모든 사람의 사생활을 침해할 위험성이 높으며, 공공장소 익명성에 대한 사람들의 기대에도 심각한 영향을 미쳐 결국 의사 표현의 자유, 집회의 자유, 결사의 자유 및 이동의 자유에 직접적으로 부정적인 영향을 미친다. 이에 EDPB와 EDPS는 얼굴 뿐 아니라 걸음걸이, 지문, DNA, 음성, 키보드 입력 및 기타 생체인식 신호 또는 행동 신호와 같은 인적 특성을 자동으로 인식하는 인공지능을 공개적으로 접근할 수 있는 공간에서 사용하는 것에 대하여, 실시간과 법집행 목적에 한정하지 않고 온라인을 비롯한 어떤 상황에서든 전면 금지할 것을 요구한다.

생체인식으로 개인을 인종, 성별, 정치적 또는 성적 지향, 기타 차별을 금지하고 있는 기준에 따라 집단으로 분류하는 인공지능 시스템도 마찬가지로 금지되어야 한다. 법안이 고위험으로 분류한 거짓말탐지기의 경우도 과학적 타당성이 입증되지 않았거나 유럽연합

의 본질적 가치와 상충하기 때문에 공공기관과 민간 기업 모두에서 금지할 것을 권고한다(부속서 Ⅲ.6.(b) 및 7.(a)). 더불어 인간의 자유 의지에 따라 독립적으로 존재하는 장래 행동에 대해 컴퓨터가 결정하거나 분류하는 것은 인간의 존엄성에 영향을 미치기 때문에, 법집행 기관이 자연인의 재범 위험성을 평가하기 위해(부속서 Ⅲ.6.(a)) 또는 실제 또는 잠재적인 범죄의 발생 또는 재발을 예측하기 위해(부속서 Ⅲ.6.(e)) 자연인의 프로파일링이나 습관 및 성격적 특성 또는 과거 범죄행위에 대한 평가에 기반하여 자연인의 개인적 위험 평가를 하기 위해 인공지능 시스템을 사용하는 것은 금지되어야 한다. 또한 자연인의 감정을 추론하기 위하여 인공지능을 사용하는 것은 매우 바람직하지 않기 때문에 감정 인식이 중요한 환자의 경우나 연구 목적 등을 제외하고는 금지되어야 한다.

고위험 인공지능과 관련하여서, EDPB와 EDPS는 우선 고위험 인공지능 시스템이 시장에 출시되거나 유럽연합에서 운영되기 전에 사전 적합성평가를 받도록 한 것을 환영하였다. 다만 공공 서비스 기관이나 중요 인프라의 의사결정 과정과 같은 특정 환경에서 사용되기 위해서는 전체 소스 코드를 조사할 수 있는 방안이 마련되어야 한다고 덧붙였다.

또한 고위험 인공지능 시스템의 사전 적합성평가는 항상 제3자가 수행할 수 있도록 적합성평가 절차를 조정할 것을 제안하였다. 고위험 인공지능 시스템이 관련 부문별 제품의 안전 법규의 적용을 받는 경우 해당 법규에 따른 제3자 적합성평가 절차를 받는데, 법안은 그와 별도로 이 법이 부속서에서 열거하는 독립형 고위험 인공지능 시스템의 경우 원격 생체인식 시스템에만 제3자 적합성평가를 받도록 하고 나머지 독립형 고위험 인공지능 시스템은 내부적인 평가를 실시하도록 하였다(제43조).

더불어, 법안 발효 시점에 이미 구축되어 운영 중인 인공지능 시스템도 적용범위에 포함할 것을 요구하였다.

유럽 인공지능위원회 등 거버넌스 문제와 관련하여서, EDPB와 EDPS는 유럽연합 기관, 기구 및 청의 감독을 위한 소관기관으로 EDPS를 지정한 것을 환영하면서도, 법안이 개인정보보호 감독기구인 EDPS를 ‘시장 감독기관’으로서 설정한 데 대하여 그 역할과 업무가 좀더 명확해져야 한다고 지적하였다.

무엇보다 인공지능법은 그 감독업무 및 집행업무의 수행에 있어 감독기관의 독립성을 명확히 확립해야 한다. 시장 감독기관 및 개인정보보호 감독기구 모두 기존 법률에서 독립성을 보장받는데, 이 법안은 감독기관이 독립적일 것을 요구하지 않으며, 이들 기관에

집행위원회 보고를 요구하고 있다(제63조(2)). 특히 법안은 국가 개인정보보호 감독기구들이 법집행 목적의 인공지능에 대한 국내 시장 감독기관으로 지정될 수 있도록 하였는데(제63조(5)), 현재의 집행위원회 보고 구조는 이들 기관의 독립성과 양립되지 않는다. 법안은 인공지능법에 따른 감독기관이 업무 수행에서 완전한 독립성을 명확히 확립해야 한다. 이것이 향후 인공지능법의 적절한 감독 및 시행을 보장할 수 있기 때문이다.

또한, 이미 각국에서 국가 개인정보보호 감독기구들이 개인정보 관련 인공지능 시스템을 감독하고 있고, 개인정보보호 법률들과의 조화로운 해석과 집행을 보장하기 위하여 이들 개인정보보호 감독기구들이 국가 인공지능 감독 기관으로 지정되는 것이 바람직하다는 것이 EDPB와 EDPS의 의견이다. 최소 공공장소에서 법집행 목적으로 실시간 원격 생체인식을 수행하는 인공지능의 경우 개인정보처리를 포함하기 때문에 유럽 기본권헌장(제8조) 등에 따라 독립 기구의 감독을 받아야 한다.

또한 법안은 '유럽 인공지능위원회' 구조에서 유럽 집행위원회에 중요한 역할을 부여한다. EDPB와 EDPS는 집행위원회의 비중이 인공지능 유럽 기구가 어떠한 정치적 영향으로부터도 독립할 필요성이 있는 상황과 상충된다고 보았다. 향후 인공지능 규제는 인공지능위원회의 독립성을 보장하기 위해 인공지능위원회의 권한과 업무에 더 많은 자율성을 부여하고, 인공지능위원회가 독자적으로 활동할 수 있도록 보장해야 한다는 것이다. 집행위원회의 부속서 개정에 있어서도 인공지능위원회가 제안과 협의 권한을 가져야 한다. 특히 EDPB와 EDPS는 인공지능 분야에서의 업무 경험과 인권 전문성 측면에서 유럽 인권기구인 유럽연합 기본권청을 인공지능위원회 참관 기관으로 고려할 것을 권고하였다.

기존 개인정보보호 법률들과의 관계와 관련하여, EDPB와 EDPS는 개인정보보호 법률들은 이 법안으로 영향을 받거나 간섭받지 않는 전제조건으로 고려되어야 하고, 서로간의 불일치와 상충 가능성을 회피하여야 한다고 강하게 주장하였다. 특히 GDPR 제22조 등 프로파일 등 자동화된 개별 의사결정에 대한 권리나 개인정보보호 법률들에 따른 통지권, 열람권, 반대권, 처리제한권, 삭제권 등 정보주체의 권리는 인공지능 기법이나 기술아키텍처에 상관없이 처음부터 인공지능 시스템에서 제공하여야 한다.

샌드박스 및 추가처리와 관련하여, EDPB와 EDPS는 개인정보보호 법률들을 침해하지 않도록 그 범위와 목표, 요구사항 준수 및 개인정보처리자의 책임을 명확히 할 것을 권

고하였다. 특히 법집행 목적의 개인정보 처리는 2차적인 목적의 처리를 전제로 하는 샌드박스와 추가처리의 대상이 될 수 없다.

투명성과 관련하여, EDPB와 EDPS는 고위험 인공지능 시스템을 공공데이터베이스에 등록하도록 한 것을 환영하였다(제51조 및 제60조). 이 데이터베이스는 인공지능 시스템의 적용 범위와 기능을 저해할 수 있는 알려진 결함 및 사고, 이를 해결하고 교정하기 위해 제공자가 채택한 구제 수단에 대한 정보를 일반 대중에게 제공하는 기회가 될 수 있다.

다만 범죄를 탐지・예방・수사・기소하는 데 사용되는 인공지능 시스템에 투명성 의무가 적용되지 않는다는 사실은 너무 광범위한 예외이다. 무죄추정의 원칙이 적용되는 범죄 탐지 및 예방 용도의 인공지능 시스템과 범죄 기소를 위한 수사를 목표로 하는 인공지능 시스템을 구분해야 하고, 법안이 이런 문제를 신중하게 다루어야 한다. 비밀유지 대상이 되는 인공지능의 경우에는 일반 대중이 아니라도 소관 감독기관에 등록할 필요가 있다.

민감정보와 관련하여, EDPB와 EDPS는 이 법안이 기존 개인정보보호 법률들에 따른 민감정보 처리의 일반적인 근거가 되지 않는다고 하였으면서도(전문41), 일부 조항들이 민감정보 처리의 예외를 규정하고 있다는 점을 지적하였다(제10조(5) 등). 이는 개인정보보호 법률들들과 불일치와 침해 문제를 야기한다.

준수 구조와 관련하여 법안은 인증과 행동지침(codes of conduct)을 제안하고 있다.

법안에 개괄된 인증 시스템은 개인정보보호 법률들 뿐 아니라 고위험 인공지능 시스템의 각 '영역'에 적용되는 다른 법률과의 명확한 관계가 누락되어 있으며, 적합성마크를 획득하기 전에 고려해야 할 요소의 하나로 데이터 최소화 및 개인정보보호 설계 원칙을 고려하지 않고 있다. 향후 개인정보보호 감독기구들이 조화 표준 및 공통 규격의 준비 및 수립에 참여해야 한다. 또한, 적합성평가 예외사유(제47조)가 공공안전상의 예외적 사유, 개인의 생명 및 건강의 보호, 환경 보호, 주요 산업 및 인프라 자산의 보호 등 너무 광범위하기 때문에 범위를 좁힐 필요가 있다.

비고위험 인공지능에 대한 자발적 행동지침과 관련하여, EDPB와 EDPS는 이들 행동지침이 다룰 수 있는 '추가 요구사항' 중 개인정보보호를 고려해야 하는지 명확히 할 필요가 있고, '기술 규격 및 솔루션'이 기존 개인정보보호 체제의 규정 및 원칙들과

상충되지 않도록 보장할 필요가 있다고 지적하였다.

2) 시민사회의 의견

시민사회는 유럽연합의 인공지능 규제 구상을 전반적으로 환영하면서도, 인공지능법이 인권을 보다 강력하게 보호할 것을 요구하고 여러 예외로 인한 규제 공백을 보완할 것을 요구하였다.

독일에 기반을 두고 활동하는 정보인권 시민단체 알고리즘왓치는, 법안이 2020년 2월 발표된 유럽연합 <인공지능 백서>에 기반을 두고 있으면서도, 경쟁력을 기본권 보호보다 앞세웠던 백서에 비하여 유럽연합이 보호하는 기본권과 가치를 침해하는 인공지능을 금지하고 있다며 환영하였다. 그럼에도 법안은 기본권과 공익을 앞세운 취지와 상충되는 심각한 결함을 가지고 있다고 지적하였다.[83)]

첫째, 인권 침해가 큰 인공지능들이 금지대상에서 제외되었다.

그간 유럽 시민들은 인공지능 규율에 있어 기본권 보호와 생체인식 대량감시 금지를 요구하여 왔다. 2021년 3월 8일 유럽의회 의원 116명이 기본권 보호를 촉구하는 공개 서한을 발표하였고, 47,000명 이상의 유럽 시민들이 생체인식 대량감시 금지 서명에 참가하였다. 이러한 강력한 요구에도 불구하고 법안은 실시간 원격 생체인식을 금지를 법집행기관이 운용하는 경우로 한정하여 다른 공공기관과 민간회사를 제외하였다. 이러한 금지에서도 "자연인의 생명 또는 신체적 안전에 대한 구체적이고 실질적이며 임박한 위협 또는 테러 공격의 방지"의 경우에는 예외로 하였으며, 예외 사유의 경우 법원 등의 사전 허가를 받도록 하였음에도 '긴급 상황'에서는 또다시 예외로 하였다. 사회신용점수의 금지 또한 공공기관에 한정되어 있다. 한편, '예측 치안'용 인공지능 시스템은 인권 침해 우려가 큼에도 불구하고 금지 대상이 아니라 고위험으로 분류되어 있다.

둘째, 고위험 인공지능 규제에서 모호하고 한계가 있다.

법안의 고위험 규제는 백서보다 구체화되어 채용, 신용평가, 사회 급여, 예측 치안, 출입국관리 및 사법적 지원을 포함하고 판단 기준도 보다 명확해졌다. 또 고위험 인공지능

83) AlgorithmWatch (2021). AlgorithmWatch's response to the European Commission's proposed regulation on Artificial Intelligence : A major step with major gaps (2021. 4. 22.). <https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021/ (검색일: 2021. 9. 1.)>.

시스템 제공자에 대한 요구사항도 엄밀해졌다는 점에서 진전된 부분이 있다. 그럼에도 “악마는 디테일에 있다.” 법안 제7조 및 부속서 Ⅲ에서 고위험 인공지능의 기준이 해석상 모호하다. 또한 원격 생체인식을 제외한 예측 치안, 출입국 관리, 채용 등 다른 고위험 인공지능에 대한 사전 적합성평가를 이해관계가 있는 회사 및 기관 자체적으로 수행하도록 한 점이 부적절하다. 한편, ‘제한된 위험’에 속하는 감정 인식, 생체인식 분류, 딥페이크는 개인 뿐 아니라 사회적으로 중대한 위험이 될 수 있고 특히 감정 인식의 과학적 근거는 논쟁 중이다. 또 이 분야 인공지능에 요구되는 최소한의 투명성 요구가 범죄 탐지, 방지, 수사에서 또다시 예외가 된다(제52조).

셋째, 모든 고위험 인공지능에 대한 등록과 공개 제도는 매우 고무적이지만, 모든 공공기관 인공지능은 위험성 여부와 무관하게 일반에 공개되어야 한다. 이때 공개되는 정보에는 시스템의 목적, 알고리즘(로직) 설명, 개발자, 영향평가 결과가 반드시 포함되어야 한다. 또한 현재 공개대상에서 법집행 및 출입국 관리 분야의 고위험 인공지능 시스템을 일부 면제시키는 것은 바람직하지 않다(부속서Ⅷ(11) 등).

넷째, 집행 체계와 관련하여 유럽연합 수준의 인공지능위원회와 집행위원회가 각국에 대한 감독 권한과 개입 권한을 가지고 있다. 유럽연합 전체적으로 일관된 규제가 가능할 것인지 지켜볼 필요가 있다.

다섯째, 사람 중심의 인공지능 체계를 제안했던 유럽 집행위원회 법안에서 인공지능 시스템으로 영향을 받는 사람들에 대한 책무성 체계를 누락하고 있다는 점은 매우 큰 결함이다. 사람들의 삶에 영향을 미치는 결과를 낳는 인공지능에는 투명성을 보장하는 것에서 더 나아가 그 영향을 받은 사람이나 집단이 결과에 이의를 제기하고 번복하거나 재고를 요청할 수 있는 구제 수단에 수월하게 접근하고 법적인 권리를 행사할 수 있도록 보장하여야 한다.

한편, 네덜란드에 기반을 둔 유럽 법률 정책 전문 비정부기구인 ECNL은 인공지능 규제에 있어 인권 중심 접근(rights-based approach)을 요구해 왔다. ECNL은 인공지능 규제 법률의 도입을 환영하면서도 인권 보장 강화를 요구하였다.[84]

ECNL은 우선, 전반적인 접근방식에 있어 법안이 인공지능 제품에 대한 내부 시장 조

84) European Center for Not-for-Profit Law (2021). ECNL Position Statement on the EU AI Act (2021. 7. 26.). <https://ecnl.org/news/ecnl-position-statement-eu-ai-act (검색일: 2021. 9. 1.)>.

화 및 기업과 기관들의 이해관계 만큼 인권을 고려하고 있지 않다고 비판하였다. 특히 권력관계의 불균형 문제에 충분히 대처하지 않고 인공지능의 대상이 되는 취약 집단을 보호하고 있지 않다.

소수의 인공지능 시스템만이 약소한 법적 요구사항의 대상이 되며, 대다수의 인공지능 시스템은 영향평가나 규제 대상이 아니다. 법안상으로는 국가별 추가 규제도 지장을 받을 수 있다. 이에 ECNL은 부문별 추가 규제를 요구하면서, 인공지능 위험 수준과 관계없이 모든 인공지능 시스템에 최소한의 투명성 및 인권 준수를 촉구하고, 제공자는 물론 사용자 역시 의무를 준수할 것을 요구하였다.

또한, ECNL은 인공지능 시스템에 대한 인권영향평가의 실시를 요구하였다. 현재는 제공자에게만 평가 의무가 부여되어 있는데, 인공지능 시스템의 사용자가 그 사용 전에 인권영향평가를 비롯한 인권 실사를 실시하여야 한다는 것이다. 더불어, 인공지능 시스템으로부터 영향을 받는 집단이나 사람의 피해에 대한 배상권 및 구제 수단이 법안에 반드시 추가되어야 한다.

무엇보다 영향을 받는 사람들을 비롯한 이해관계자들에 대한 통지와 의미있는 참여가 보장되어야 한다. 시정조치, 통지 등의 집행 절차 뿐 아니라 영향평가에 시민단체를 포함한 이해관계자 참여를 보장하여야 하고, 규제기구의 결정사항에 대한 제3자 이의제기도 가능하여야 한다.

금지대상 인공지능이 협소하다는 점도 문제로 지적되었다. 이들은 금지대상 인공지능을 위 EDPB 및 EDPS가 권고한대로 확장하고, “물질적 또는 정신적 피해”를 입증하도록 한 조건을 삭제하며, 공공장소 원격 생체인식에 대한 예외를 축소할 것을 요구하였다. 형사사법이나 난민 심사에서 위험성 예측도 금지되는 것이 바람직하다. 경찰 등 공권력에 의한 인공지능 감시는 막대한 해악으로 이어질 수 있다. 소수자나 차별받는 집단, 인권 활동가 및 언론인 등이 위협에 처할 가능성이 있으며, 빈곤하고 이주민과 유색인종이 밀집한 지역에 과잉 치안이 집중될 우려가 있다.

고위험 인공지능의 경우, 그 기준에 인권에 미치는 위험성이 포함되어야 하며, 제품 설계 의도, 영향의 심각도, 실사 메커니즘, 인과 관계, 구제 가능성 및 인공지능이 도입되는 상황 등과 관련된 요소들을 포함하여야 한다. 예를 들어 경찰, 출입국관리, 사법접근성과 관련된 ‘상황’은 항상 인권 측면에서 중대하다. 공개 등록 제도는 큰 의미가

있지만, 대상 정보를 확대할 필요가 있다. 시스템의 도입 기관 및 목적은 물론 누가 인공지능 시스템의 정확성, 견고성, 사이버보안 수준을 정하는지, 누가 평가지표를 정하는지, 영향을 받는 집단이 어떻게 참여할 수 있는지 등에 대한 정보도 포함하여야 한다. 적합성 평가는 자체적인 실시가 아니라 외부적으로 실시하여야 하며, 공공안전상의 사유로 적합성평가의 예외를 둔 조항은 삭제되어야 한다. 고위험 시스템의 중대사고나 오작동에 대하여 국가기관에 통보하는 내용은 일반에도 공개되어야 한다.

저위험 인공지능으로 분류되어 있는 감정 인식 기술, 인종·성별·정치적 또는 성적 지향에 따른 생체인식 분류는 인권과 양립할 수 없으므로 전적으로 금지되는 것이 바람직하다. 챗봇 일부와 딥페이크 사용은 고위험으로 분류되어야 한다.

인공지능법이 규제하지 않는 상당수 인공지능의 경우에도 인공지능 제공자인 민간회사는 해당 기업은 물론 그 공급망에서 유엔 <기업과 인권 지침>에 따른 인권 준수 책임을 다해야 한다.

마지막으로, ECNL은 국가 및 유럽의 거버넌스 기구에 인권전문가, 연구자, 영향을 받는 집단의 참여를 보장할 것을 요구하였다. 또한 유럽표준위원회, 유럽전기표준화위원회 등 표준 기구가 사실상 인공지능법 관련 표준을 수립할 권한을 부여받았다는 사실을 우려하였다. 사실상 이 기준에 따른 적합성평가가 이루어질 것이라는 점에서 시민단체, 학계 및 영향을 받는 집단을 비롯한 이해관계자들이 기준 개발에 참여하여야 한다.

인공지능 시스템은 빈곤, 인종, 젠더 등 역사적인 취약 집단에 불균형적으로 영향을 미쳐 기존의 불평등을 더욱 악화시킬 수 있다. 그러나 차별에 대한 대응으로 기술적 편향 점검에만 의존하는 것은, 이러한 수단으로 시스템에서 편향성을 성공적으로 제거할 수 있다는 더욱 문제적인 구상을 강화할 수 있다. ECNL은 영향을 받는 집단 및 개인의 참여 보장이 핵심적이라고 강조한다.

제3절 인공지능 (인권)영향평가의 기준 및 제도

1. 인권 기반 접근법

인공지능에 대한 위험 기반 접근법과 또 다르게, 인공지능에 대하여 인권 기반 접근법을 취하고 있는 국제 규범들이 존재한다. 인권 기반 접근법의 주요 요소 중 하나는 인권영향평가이다.

유럽평의회 인공지능 특별위원회(AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE, CAHAI) 정책개발단은 2021년 5월 21일 인공지능 인권, 민주주의 법치 영향평가에 대한 보고서 초안에서 인공지능 인권영향평가에서 참고할 수 있는 기존의 규범적 문헌을 개괄하고 인공지능에 대한 적용을 검토하였다.[85)]

인권영향평가 일반에 대한 기존 국제 규범으로는 유엔 <기업과 인권 이행 지침>, 최종 사용 관련 인권 위험 식별 및 평가, 다국적 기업을 위한 OECD 지침이 있다.

우선 유엔 <기업과 인권 이행 지침>[86)]은 2011년 채택되었으며, a) 인권을 존중, 보호, 이행해야 할 국가의 의무, b) 법률 및 인권 준수를 보장하기 위해 기업이 수행하는 중요한 역할, c) 인권에 부정적인 영향을 미치는 경우 법적 보호 및 사법적 구제를 실행해야 할 필요성에 대해 강조하였다.

유엔 인권최고대표는 2020년 <최종 사용에서 인권 위험 식별 및 평가>[87)]를 발표하였다. 이 문서는 제품과 서비스에서 인권 위험을 식별하고 평가할 때 <기업과 인권 이행 지침>의 기본적인 요구를 이해하고자 하는 기술 기업 경영진을 대상으로 한다. 관련하여

---

85) Ad Hoc Committee on Artificial Intelligence Policy Development Group (2021). Human Rights, Democracy and Rule of Law Impact Assessment of AI systems(CAHAI-PDG(2021)05). 7p. <https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3 (검색일: 2021. 9. 1.)>. 같은 기구에서 같은 주제로 발표한 보고서 중 가장 최근본이지만 최종본은 아니다.

86) Guiding Principles on Business and Human Rights: Implementing the United Nations “Protect, Respect and Remedy” Framework. 유엔문서 A/HRC/17/31 (2011. 3. 21). <https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf (검색일: 2021. 9. 1.)>.

87) Office of the High Commissioner for Human Rights (2020). Identifying and Assessing Human Rights Risks related to End-Use. <https://www.ohchr.org/Documents/Issues/Business/B-Tech/identifying-human-rights-risks.pdf (검색일: 2021. 9. 1.)>.

<기업과 인권 이행 지침>은 기업에 대하여 1) 발생할 수 있는 영향에 대하여 광범위한 관점을 갖고, 2) 가장 심각한 해악에 초점을 맞추고, 3) 이해관계자들과 의미 있게 관계를 맺고 소통할 것을 요구하였다.

OECD는 2011년 <다국적 기업을 위한 OECD 지침>[88]을 발표하였다. 가입국에서 또는 가입국 기반으로 운영되는 다국적 기업에 대한 권고를 담은 이 지침은, 국제법이 규제하는 지구적 상황에서 책임 있는 사업 수행에 대한 원칙과 표준을 제시하였다. 지침은 정부가 민간 행위자들에 대하여 법적 규범 존중을 촉진할 것을 요구하였다. 특히 기업은 1) 인권, 2) 고용 및 산업 규제, 3) 환경, 4) 독점 금지, 5) 소비자 이익 6) 납세에 대한 법적 규범을 준수해야 한다.

한편, 인공지능과 관련된 인권영향평가와 관련한 국제 규범으로는 유럽평의회 인권위원장, 유럽평의회 권고, 유럽 집행위원회 인공지능 고위전문가그룹을 비롯하여 학계 및 시민사회에서 발표한 사항들을 참고할 만 하다.

유럽평의회 인권위원장은 2019년 <인공지능 블랙박스 개봉: 인권 보호를 위한 10단계>[89]에 대한 권고를 발표하였다. 이 권고는 인공지능이 인권에 미치는 부정적인 영향을 방지하고 완화하기 위한 방안을 제시하였다. 권고가 초점을 맞춘 10가지 조치 영역은 1) 인권영향평가, 2) 공개적인 의견 수렴, 3) 민간부문의 인권기준 이행을 촉진해야 하는 회원국 의무, 4) 정보와 투명성, 5) 독립적 감독, 6) 차별금지와 평등, 7) 개인정보보호 및 프라이버시, 8) 표현의 자유, 집회 및 결사의 자유, 노동권, 9) 구제 수단, 10) 인공지능 리터러시 증진 등이다.

유럽평의회는 2020년 4월 8일 알고리즘 시스템의 인권 영향에 대한 각료위원회 권고 CM/Rec (2020)1[90]를 채택하였다. 이 권고와 부록 <알고리즘 시스템의 인권 영향에 대한 대응 지침>은 알고리즘 시스템의 설계 및 개발과 관련하여 국가 및 민간의 행위자들에게 지침을 제시하여 기술 개발로부터 유럽인권협약에 규정된 인권과 개인의 자유를 보호하는 것을 목표로 한다. 특히 지침은 국가는 물론 민간 행위자 역시 알고리즘 시스템이 법을 준수하고 인권을 존중할 것을 명시하였다.

---

88) OECD (2011). OECD Guidelines for Multinational Enterprises.
<http://www.oecd.org/daf/inv/mne/48004323.pdf (검색일: 2021. 9. 1.)>.
89) Council of Europe Commissioner for Human Rights (2019).
90) Council of Europe (2020).

유럽 집행위원회 인공지능 고위전문가그룹은 2019년 <신뢰할 수 있는 인공지능을 위한 윤리 지침>[91]을 마련하였다. 이 지침은 신뢰할 수 있는 인공지능의 조건을 수립하는 것을 목표로 하였다. 신뢰할 수 있는 인공지능은 1) 합법적, 2) 윤리적, 3) 견고해야 한다. 더불어, 인공지능 시스템이 인권 원칙으로부터 도출되는 신뢰성을 확보하기 위해서는 개인정보보호, 투명성과 책무성 보장의 요구사항을 만족해야 한다. 인공지능 시스템이 이러한 요구사항을 만족하는지 평가하기 위하여, 지침은 "신뢰할 수 있는 인공지능 평가 목록"을 지침에 포함하였다.

영국의 에이다 러브레이스 연구소는 2020년 <블랙박스 검사> 보고서[92]를 발표하였다. 이 보고서는 알고리즘 감사와 알고리즘 영향평가의 조건을 명확히 하고 관련 연구 및 실무 현황을 설명하였다. 알고리즘 감사와 관련한 두 가지 주요 접근법으로는 1) 알고리즘 시스템의 편향성 평가에 초점을 맞춘 표적 접근법으로서 '편향성 감사'와 2) 알고리즘 시스템의 법규 또는 규범 준수에 초점을 맞춘 광범위한 접근법으로서 '규제 검사'를 들었다. 이들 알고리즘 감사는 다양한 도구와 방법론이 요구되며, 통상 규제기관이나 감사 전문가들이 실시한다. 알고리즘 영향평가와 관련하여서는 1) 시스템이 사용되기 전에 알고리즘 시스템의 사회적 영향 가능성을 평가하는 '알고리즘 위험 평가'와 2) 알고리즘 시스템이 사용자나 인구집단에 미칠 수 있는 사회적 영향을 평가하는 '알고리즘 영향평가'의 두 가지 접근법을 살펴 보았다.

미국 시민단체 미국시민자유연합(ACLU)은 2020년 AEKit으로도 알려진 <알고리즘 형평성 툴킷(Algorithmic Equity Toolkit)>을 제작하여 발표하였다.[93] 이 툴킷은 정부가 사용하는 감시 및 의사결정 기술을 식별하고, 이들 기술이 어떻게 작동하는지 이해하며, 그 영향, 효과 및 감독에 대한 문제제기가 가능하도록 설계되었다. 툴킷은 1) 순서도, 2) 시스템 맵, 3) 빈칸 채우기, 4) 설문지 등 4가지 요소로 구성되어 있다.

존스 홉킨스 대학교 정부 우수성 센터 (GovEx), 샌프란시스코 시 및 카운티, 하버드 케네디 스쿨 데이터스마트 프로젝트 및 비영리단체 데이터 커뮤니티 DC는 2018년 공동

91) European Commission's High-Level Expert Group on Artificial Intelligence (2019).
92) Ada Lovelace Institute (2020). Examining the Black Box.
<https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf (검색일: 2021. 9. 1.)>.
93) American Civil Liberties Union (2020). Algorithmic Equity Toolkit.
<https://www.aclu-wa.org/AEKit (검색일: 2021. 9. 1.)>.

으로 <윤리 및 알고리즘 툴킷(Ethics & Algorithms Toolkit)>[94]을 개발하였다. 이 툴킷은 정부 부문에서 알고리즘을 구축하거나 도입하는 모든 사람에게 초점을 맞추었다. 툴킷은 사용자가 1) 알고리즘 사용으로 인한 윤리적 위험을 이해하고 2) 이러한 윤리적 위험을 최소화하기 위해 무엇을 할 수 있는지 파악하는 데 도움이 될 수 있도록 일련의 질문지를 제시하였다. 공공 부문에서 의사결정을 지원하기 위해 알고리즘을 사용할 때마다 이 툴킷을 사용할 것이 권장된다.

캐나다 정부는 2019년부터 <캐나다 알고리즘 영향평가 도구(Canadian Algorithmic Impact Assessment Tool)>[95]를 사용하고 있다. 이 평가 도구는 캐나다 재정위원회의 <자동화된 의사결정 훈령(Directive on Automated Decision-Making)>을 지원하기 위하여 개발되었으며 공공기관이 의무적으로 실시하여야 하는 위험 평가 도구이다. 평가 도구는 48개의 위험성과 33개의 완화 질문들로 구성되어 있으며, 답변 결과에 따라 공공기관 자동화된 의사결정 시스템의 영향 수준이 결정된다. 해당 질문들은 캐나다 재정위원회 사무국이 학계, 시민 사회 및 기타 공공 기관과의 협의를 통해 수립하였으며, 정부 정책, 윤리 및 행정법적 고려 사항에 따라 자동화된 의사결정 시스템의 위험성 영역별로 구성되어 있다. 전반적으로 이 평가 도구는 각 부처 및 공공기관들이 자동화된 의사결정 시스템과 관련된 위험을 더 잘 이해하고 관리할 수 있도록 설계되었다. 알고리즘 영향평가는 프로젝트 설계 단계의 초기 시점에 실시되어야 하며, 그 결과는 캐나다 공식 언어로 접근 가능한 형식으로 발표되어야 한다.

뉴욕대학교 AI Now 연구소는 2018년 <알고리즘 영향평가: 공공 기관 책임을 위한 실용적인 프레임워크(Algorithmic Impact Assessments: A practical framework for public agency accountability)>[96] 제하의 영향평가 모델을 개발하였다. 이 모델은 다음을 권장한다. 1) 공공 기관은 자동화된 의사결정 시스템에 대하여 기존에 존재하거나 제안되고 있

94) GovEx, the City and County of San Francisco, Harvard DataSmart, Data Community DC (2018). Ethics & Algorithms Toolkit. <https://ethicstoolkit.ai/ (검색일: 2021. 9. 1.)>.
95) Algorithmic Impact Assessment Tool. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html (검색일: 2021. 9. 1.)>.
96) AI NOW (2018). Algorithmic Impact Assessments: A practical framework for public agency accountability. <https://ainowinstitute.org/aiareport2018.pdf (검색일: 2021. 9. 1.)>.

는 평가를 자체적으로 수행하고, 영향을 받는 집단 전반에게 미칠 수 있는 공정성, 사법 정의, 편향성 또는 기타 영향 정도를 평가하여야 한다. 2) 공공 기관은 시간대에 따라 영향을 발견, 측정 또는 추적하기 위해 유의미한 외부 연구자 검토 절차를 개발하여야 한다. 3) 공공 기관은 시스템을 도입하기 전에 '자동화된 의사결정 시스템'의 정의, 기존 및 제안 중인 시스템, 관련 자체 평가 및 연구자 검토 절차를 공개해야 한다. 4) 공공 기관은 우려 사항을 명확히 하고 미해결 질문에 대응하기 위해 공개적으로 의견을 구해야 한다. 5) 정부는 영향을 받는 개인이나 집단이 기관이 완화하거나 시정하지 못한 부적절한 평가 또는 불공정, 편향적, 기타 유해한 시스템의 사용에 이의를 제기할 수 있도록 강화된 적법 절차 메커니즘을 제공해야 한다.

특히 최근 유엔 인권최고대표는 인권 기반 접근법의 일환으로 인권 실사를 강력히 요구하였다.[97] 국가와 기업은 인공지능 시스템의 구입, 개발, 배치 및 운영 시 뿐 아니라 개인에 대한 빅데이터를 공유하거나 사용하기 전에 포괄적인 인권실사를 실시해야 한다는 것이다. 실사 절차에서 인공지능의 사용이 인권과 양립할 수 없는 것으로 드러나는 경우 그 사용을 중지하여야 한다. 이때 인권영향평가는 인권 실사 과정의 필수적인 요소이다. 인권 실사는 인공지능 시스템의 수명주기 전반에 걸쳐 실시하여야 하며, 여성과 십대 여성, 레즈비언, 게이, 양성애자, 트랜스젠더 및 성소수자, 장애인, 소수자 집단에 속하는 사람, 노인, 빈곤층 및 기타 취약한 상황에 처해 있는 사람들에게 불균형한 영향을 미치는 것에 각별한 주의를 기울여야 한다. 인권실사 과정에서 잠재적으로 영향을 받는 권리주체 및 시민사회와 유의미한 협의를 수행해야 하며, 여러 학제간 전문성을 갖춘 전문가도 완화 방법 개발 및 평가를 비롯한 영향평가에 참여해야 한다. 국가와 기업은 그들이 사용하는 인공지능 시스템이 인권에 부정적인 영향을 미치는지 여부를 확인하기 위해 지속적으로 모니터링해야 한다. 인권영향평가의 결과 및 인권위험을 해결하기 위해 취해진 조치와 협의 사항은 그 자체로 공개되어야 한다.

이하에서는 언급된 상기 국제 문헌들 가운데, 인권적 측면에서 국제적인 규범력이 높다고 판단되는 문헌의 내용을 좀더 상세하게 살펴본다.

97) 유엔문서 A/HRC/48/31.

2. 유엔 인권최고대표 <최종 사용에서 인권 위험 식별 및 평가>[98]

유엔 인권최고대표는 2020년 기술 기업 경영진을 대상으로 <기업과 인권 이행 지침>의 기본적인 요구를 해설하는 자료로서 <최종 사용에서 인권 위험 식별 및 평가>를 발간하였다.

기업이 인권을 이행하기 위하여 실시하여야 하는 인권 실사 절차는 ① 인권 위험의 성격과 정도를 측정하기 위하여 영향을 식별하고 평가하는 단계, ② 내부 기능 및 절차에 대한 반영을 비롯하여 사람에 대한 위험을 방지하고 완화하기 위하여 조치하는 단계, ③ 위험 완화 대응의 시간 경과에 따른 효과성을 추적하는 단계, ④ 인권 영향을 해결하기 위한 활동에 대하여 적절하게 소통하는 단계의 4단계로 구성되며, 인권 위험을 식별하고 평가하는 것은 인권 실사의 첫 단계이다.

이 자료는 기술 기업이 인권에 미치는 영향을 식별하고 평가함에 있어 다음과 같은 점을 유념하도록 안내하였다.

첫째, 이때 인권 위험을 식별하고 평가하는 분석 범위는 포괄적으로 자사 제품과 서비스의 설계, 개발, 판촉, 배포, 판매/라이선스 및 사용이 인권에 부정적인 영향을 미칠 수 있는지 여부 및 그 방법을 식별하는 활동이 포함된다. 외부 자문단 또는 기타 외부 이해관계자 및 예정 사용자 참여 방식의 수립은 기업이 제품 및 서비스의 실제적 또는 잠재적 위험을 식별하고 이해하고자 할 때 또한 도움이 될 수 있으며 지속적이고 누적적인 참여를 가능케 할 수 있다.

둘째, 기술 기업이 대량의 제품, 서비스 및 사용자를 보유하고 있는 경우, 인권 관점에서 어떤 제품, 서비스, 솔루션, 사업 관계 또는 사용 환경이 더 높은 위험인지 알기 위해 첫 평가를 실시해야 할 수 있으며, 따라서 보다 상세한 인권 실사를 위해 우선순위를 정해야 한다. 어떤 경우에 특히 심각한 위험이 분명히 존재하는 경우, 전체적인 분석을 먼저 수행하지 않고 명백한 고위험 영역부터 시작하는 것이 필요하고 현명할 수 있다. 어느 쪽이든, 기업은 최소 수준의 평가가 지속적이고 동적일 수 있도록 보장해야 하며, 이는 시간이 지남에 따라 회사의 실사 초점을 변경하고 넓히는 결과로 이어질 수 있다.

셋째, 기업이 보다 심층적인 인권 실사를 위해 특정 제품/서비스, 사용자 유형 또는 사

98) 전문은 부록 I 참조.

용 환경에 우선순위를 부여한 경우, 기업은 이와 관련하여 인권에 미칠 수 있는 실제적 또는 잠재적 부정적인 영향에 대한 분석을 실시하여야 한다. 보다 세분화된 수준에서 실제적 및 잠재적 인권 영향을 식별하는 활동은 몇 단계 계층적 분석이 필요하며, 특정 제품을 대상으로 하거나 특정 사용자 또는 특정 지역적 환경에 대한 분석 시작점에 따라 다양할 수 있다.

넷째, 기업이 초점을 맞출 지점의 우선순위를 부여해야 할 때마다, 기업은 사람에 대한 위험의 '심각성'에 초점을 맞추는 원칙적인 접근법을 사용해야 한다. <기업과 인권 이행지침>에 따르면 영향의 심각도는 규모, 범위 및 구제불가능성에 의해 판단된다. 이때 규모는 그 영향이 얼마나 중대하거나 심각한지와 관련이 있으며, 범위는 영향을 얼마나 광범위하게 미치는지 또는 영향을 받는 사람의 수와 관련이 있고, 구제가능성은 상황에 영향을 받은 사람들을 적어도 영향 이전 상황과 같거나 동등한 수준으로 회복시킬 수 있는 능력을 의미한다. 취약하거나 소외될 위험이 높은 개인이나 집단에 따라 영향의 규모, 범위 및 구제가능성이 어떻게 다를 수 있는지, 그리고 남성과 여성 등 서로 다른 집단이 직면하는 위험이 다를수 있다는 것을 기업이 고려하는 것이 중요하다.

3. 유럽평의회 인권위원장 <인공지능 블랙박스 개봉: 인권 보호를 위한 10단계>[99]

유럽평의회 인권위원장[100]은 2019년 5월 <인공지능 블랙박스 개봉: 인권 보호를 위한 10단계>를 발표하고 유럽평의회 회원국들에 대하여 공공기관과 민간 기업이 개발, 도입, 사용하는 인공지능 시스템의 인권 준수를 촉구하였다. 권고가 제시한 10가지 조치 영역은 1) 인권영향평가, 2) 공개적인 의견 수렴, 3) 민간부문 인권기준 이행을 촉진하는 국가의 의무, 4) 정보 및 투명성, 5) 독립적 감독, 6) 차별금지 및 평등, 7) 개인정보보호 및 프라이버시, 8) 표현의 자유, 집회 및 결사의 자유, 노동권, 9) 구제 수단, 10) 인공지능 리터러시 증진을 아우른다. 특히 이 권고는 인공지능에 대한 공공기관 인권영향평가 및

99) 전문은 부록 III 참조.

100) 유럽평의회는 독립적이고 불편부당한 비사법적 인권기구로서 1999년부터 인권위원장(Commissioner for Human Rights)을 두고 있으며, 48개 회원국 내에서 인권에 대한 인식과 존중을 증진하는 임무를 수행하고 있다.

민간 기업 인권 실사, 국가인권기구를 비롯한 독립적 감독 체계에 대하여 자세한 권고를 담았으며, 이 권고 이행을 위한 체크리스트를 제시하였다. 이하에서는 각 영역에 대한 인권위원장의 체크리스트를 중심으로 내용을 살펴본다.

첫째, 국가는 공공기관이 구입, 개발 또는 배치하였거나 예정인 인공지능 시스템에 대하여 인권영향평가 실시를 요구하는 법률과 규제를 도입하는 조치를 취해야 한다. 이때 인권영향평가는 규제 영향평가 및 개인정보보호 영향평가 등 공공기관이 수행하는 다른 영향평가와 유사한 방식으로 구현 및 운영되어야 한다. 인권영향평가 관련 법률 체계가 도입되는 시점에 공공기관은 이미 배치하였거나 사용하고 있는 인공지능 시스템에 대하여 인권영향평가를 즉시 수행하여야 한다. 또는, 인공지능 시스템의 구입 및 개발 전에 인권영향평가를 먼저 수행하여야 한다. 이때 인권영향평가는 인공지능 시스템에 대하여 독립된 감독 기관이나 관련 전문지식을 갖춘 외부 연구자/감사관이 실시하는 유의미한 외부 검토를 포함해야 한다. 공공기관은 이러한 유의미한 외부 검토를 수행할 때 국가인권기구를 참여시킬 것을 고려해야 한다. 인권영향평가에 관한 법적 체계를 도입할 때는 시민사회단체와 인공지능 및 인권에 관련된 전문가를 비롯한 관련 이해관계자와 유의미하게 협의하고 의견을 수렴하여야 한다. 인권영향평가는 투명한 방법으로 시행하여야 하고, 인권영향평가의 수행이나 공개를 방해하기 위한 목적으로 기밀, 사생활, 영업 비밀, 국가 기밀 또는 지적 재산에 관한 법률을 사용하거나 사용을 유도하지 말아야 한다. 인권영향평가의 적용을 받지 않았거나, 인권영향평가가 인공지능 시스템의 실제적인 인권 침해 위험을 나타냈음에도 식별된 위험을 방지하거나 완화하기 위한 조치, 안전장치 또는 방법을 채택하지 않은 상황에서는 인권을 간섭할 잠재성이 있는 인공지능 시스템을 구입, 개발, 배치 및 사용하지 말아야 한다.

둘째, 국가는 인공지능 시스템의 사용에 공개 조달 기준과 투명한 절차를 적용하여야 한다. 최소한 조달 및 인권영향평가 단계에서는 공개적인 의견 수렴에 영향을 받는 집단 또는 공동체를 비롯한 모든 이해관계자를 포함하여야 한다. 공개적인 의견 수렴은 인공지능 시스템과 관련된 모든 관련 정보를 적시에 사전에 공개하고 모든 관련 이해관계자의 참여를 적극적으로 추진하는 등 유의미한 적절한 조치를 취하여야 한다. 국가인권기구는 시민사회와 국가 기관 사이의 가교 역할을 맡아 유의미한 협의 수행에 도움을 줄 수 있다.

셋째, 국가는 인공지능 관련 인권 침해에 대해 인공지능 행위자가 책무를 지는 데 있어 간극 또는 장벽을 파악하여야 하고 이를 위하여 기존 형법 및 민법은 물론 기타 동등한 법적 책임 체제에 대한 감사를 수행하여야 한다. 인공지능 행위자의 침해로부터 개인의 인권을 보호하기 위해 국가의 의무를 이행할 필요가 있는 경우 기존 법률을 시행할 필요도 있다. 인공지능 행위자가 인권 존중에 대한 책임을 다하고 있음을 "알고 보여" 주도록 조치를 취해야 한다. 여기에는 인공지능 시스템과 관련된 인권 위험을 식별하고 그러한 시스템이 초래하는 피해를 방지 및 완화하기 위하여 효과적인 조치를 취하는 투명한 인권 실사 절차가 포함된다. 다만, 인공지능 부문에 적용되는 법률, 정책 및 규정을 회원국에 대한 인권 의무로부터 분리하거나 정보가 없는 것으로 취급하여서는 안 되며, 인공지능 부문에서 인권 기준의 시행과 집행을 차별적인 방식으로 실시해서는 안 된다.

넷째, 국가는 특히 공공 서비스에서 인공지능 시스템이 언제 어떻게 사용되고 있는지 개인이 이해할 수 있도록 필요한 모든 정보를 제공하여야 한다. 적절한 투명성 및 책무성 기준에 따라 사람이 검토하고 조사할 수 없을 정도로 복잡한 인공지능 시스템은 사용하지 말아야 한다.

다섯째, 국가는 국가인권기구를 포함한 기존 감독 기관을 활용하여 인공지능 시스템의 인권 준수에 대한 독립적이고 효과적인 감독을 위한 체계의 수립을 입법화하여야 한다. 모든 관련 감독 기관이 충분한 전문 지식에 접근하고, 인공지능 시스템 및 그 인권 영향에 대한 적절한 교육을 받으며, 기능을 효과적으로 수행하는 데 적정한 재정 및 기타 자원을 제공받도록 보장하기 위해 조치도 취해야 한다. 인공지능 시스템의 인권 침해(개발, 검사 및 사용 중에 발생하는 행위 포함)에 책임이 있을 수 있는 모든 행위자를 공공기관 또는 민간기관 여부를 불문하고 조사 및 모니터링하려는 목적에 적정하도록 관련 감독 기관의 기능이 보장되어야 한다. 더불어, 인권 침해 발생(위험)을 식별하는 상황에서 유의미하게 개입할 수 없을 정도로 감독 기관의 기능과 권한이 제한되어서는 안 되며, 인공지능 시스템의 인권 준수에 대한 조사 및 모니터링을 담당하는 감독 기관의 제도적, 운영적, 재정적 및 개인적 독립성을 손상해서는 안 된다. (학습 및 검사용) 데이터셋, 인공지능 입력/출력물, 모델/알고리즘, 운영 지침 및 인권 실사에 대한 접근권 박탈을 비롯하여, 감독 기관이 효과적으로 기능을 수행하는 데 필요한 정보를 박탈하거나 제3자가

박탈하도록 허용되어서도 안 된다.

여섯째, 국가는 인공지능 시스템으로 인해 그 권리가 불균형적으로 영향을 받을 위험이 높은 집단에 대하여 인공지능 시스템 사용의 차별 위험을 방지하고 완화하여야 한다. 특히 법 집행 상황에서 인공지능 시스템을 사용할 때 특정 집단에 속한 개인의 프로파일링을 방지하기 위하여 가장 높은 수준의 정밀 조사를 실시하여야 한다. 차별적이거나 차별적인 결과를 초래하는 인공지능 시스템을 사용하거나 제3자가 사용하도록 허용해서는 안 된다.

일곱째, 국가는 기존 개인정보보호법에 대한 검토 및 평가를 수행하여 인공지능 시스템 환경에서 사생활권과 개인정보보호권을 충분히 보호하는지 판단하고, 그렇지 않은 경우 법적 개혁을 실시하여야 한다. 더불어, 인공지능 시스템의 개발, 배치 및 사용과 관련된 민간 및 공공 기관이 정보주체의 권리를 존중하고 해당 개인정보보호법에 따른 의무를 이행하도록 사전에 조치를 취해야 한다. 반면, 인공지능 시스템을 개발, 배치 또는 사용하는 사람들에 대하여 개인정보 처리에 관한 의무의 광범위하고 불균형적인 예외 또는 면제를 규정해서는 안 된다. 그 학습 또는 검사에서 사생활권 및 개인정보보호권을 위반하여 수집되거나 처리된 데이터셋에 의존하는 인공지능 시스템을 개발 또는 사용하거나, 이러한 권리를 위반하는 개인정보를 그 입력 또는 출력 데이터로 처리하는 인공지능 시스템의 개발 또는 사용은 허용하지 않아야 한다.

여덟째, 국가는 인공지능의 사용에 의해 잠재적으로 관련될 수 있는 국제 인권 기준의 모든 범위를 고려하여야 한다. 특히 인공지능 중심의 콘텐츠 관리 및 큐레이션이 표현의 자유, 정보에 대한 접근 및 의견의 자유에 미칠 수 있는 영향을 유념하고, 집회의 자유를 효과적으로 행사할 수 있도록 얼굴 인식 기술의 사용을 엄격하게 규제하여야 한다. 더불어 인공지능이 노동권에 미칠 수 있는 부정적인 영향을 모니터링하고 학교 교육을 비롯하여 이를 완화할 계획을 수립하여야 한다. 국가는 전반적으로 국제 인권 기준에 따라 보호되는 인권을 침해하는 인공지능 시스템의 개발, 배치 또는 사용을 허용하거나 촉진하지 말아야 한다.

아홉째, 국가는 민사, 형사 및 행정법을 포함한 기존 법률에 대한 평가를 수행하고 해당 법률이 인공지능 시스템의 개발, 배치 또는 사용으로 인해 발생한 인권 침해의 피해자임을 주장하는 사람들에게 효과적인 구제수단을 제공하지 않는 경우 개정을 추진하여

야 한다. 더불어 인공지능 시스템의 수명 주기 각 단계에서 발생할 수 있는 모든 범위의 인권 침해에 대해 법적 책임이 있는 사람을 명확하게 규정하도록 책임 체제를 보장하여야 한다. 국가인권기구는 소관에 따라 조사와 결정 업무를 수행함으로써 구제 조치의 근거를 제공할 수 있다. 사법부 및 기타 관련 국가 기관은 인공지능 시스템의 가정/인식된 정확성 또는 객관성에 부적절한 비중을 부여하지 않고, 인공지능 시스템으로 인한 인권 침해에 문제를 제기하는 피해자와 대상자 간에 동등한 힘의 제공을 보장하여야 한다. 전적으로 인간의 통제 범위 밖에서 작동하는 인공지능 시스템의 개발, 배포 또는 사용을 허용해서는 안 된다. 유의미한 인적 개입을 구할 기회가 제공되지 않고 의사결정이 실행되기 전에 자신의 의견이 검토되지 않은 상황에서 개인이 자신에게 중대한 영향을 미치는 자동화된 의사결정의 대상이 되어서도 안 된다.

열째, 국가는 인공지능 관련 문제에 대해 협의하는 정부 자문 기구를 설치하여야 한다. 더불어 인공지능 시스템 개발 관계자부터 일반 대중에 이르기까지 모든 사람의 인공지능 및 인권에 대한 지식과 이해를 증진하는 리터러시 활동을 수행하여야 한다. 인공지능 리터러시 활동에서 인권에 대한 잠재적 영향을 포함하지 않고 기술적 측면으로 한정하지 말아야 한다.

4. 유럽평의회 <알고리즘 시스템의 인권 영향에 대한 대응 지침>[101]

유럽평의회는 2020년 4월 8일 47개 회원국에 대하여 알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고(Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems)를 발표하였다.[102] 이 권고는 유럽연합 인공지능법(안)과 같이 고위험 인공지능에 대한 규제를 요구하면서도 인권영향평가 및 영향을 받는 당사자에 대한 권리구제를 강조하는 인권 기반 접근법을 취하고 있으며, 부록인 <알고리즘 시스템의 인권 영향에 대한 대응 지침(Guidelines on addressing the human rights impacts of algorithmic systems)>에서는 구체

101) 전문은 부록 IV 참조.

102) Algorithms and automation: Council of Europe issues guidelines to prevent human rights breaches. Council of Europe 보도자료 (2020. 4. 8.); Council of Europe (2020).

적인 국가의 의무와 기업의 책임을 명시하였다.

대규모 데이터셋에서 패턴을 감지하여 작동하는 알고리즘 시스템은 특히 정밀도, 타겟팅 및 일관성 향상을 통해 디지털 정보의 분류 및 검색 가능성을 크게 향상시켰고 업무 및 시스템의 효율성 및 효과성을 증진시켰다. 권고는 알고리즘 프로세스가 혁신과 경제 발전을 촉진하는 방대한 잠재력을 보유하고 있다는 사실을 인정하면서도, 알고리즘 사용과 관련하여 공정한 재판을 받을 권리, 프라이버시 및 개인정보보호, 사상·양심 및 종교의 자유, 의사 표현과 집회의 자유, 평등권, 경제적 및 사회적 권리와 관련된 인권 문제가 야기되고 있다고 우려하였다. 알고리즘 시스템의 기능은 종종 온라인 및 오프라인에서 개인 및 집단의 신원 및 행동에 대한 대규모 디지털 추적으로 수집된 데이터의 체계적인 집계 및 분석을 기반으로 하기에 사생활 침해와 조작 가능성이 우려된다. 또한 대부분의 알고리즘 시스템은 오류가 불가피한 통계 모델을 기반으로 하며 때로는 기존의 편향, 오류 및 가정을 유지, 복제 및 강화하는 피드백 순환구조를 가지고 있다. 대규모로 적용되는 알고리즘 시스템은 많은 사람들에게 영향을 미치며, 특히 관련 데이터가 처리되지 않거나 고려되지 않은 개인 및 집단에게 중대한 영향을 미칠 수 있다. 알고리즘 시스템의 작동은 일반적으로 명시적이고, 투명하며, 책임성 있고, 영향을 받는 개인이 통제할 수 있는 방식으로 이루어지지 않으며, 특히 소수자 및 소외집단 또는 취약집단에 부정적인 영향을 미칠 수 있다.

이에 유럽평의회는 정부가 알고리즘 시스템의 사용, 개발 또는 조달을 통해 인권을 침해하지 않도록 보장하기 위한 <알고리즘 및 자동화에서 인권 침해 방지를 위한 지침(Guidelines on addressing the human rights impacts of algorithmic systems)>을 제시했다. 지침은 국가의 의무(B장)와 민간 부문 행위자의 책임(C장)으로 나누어 인권 침해 방지를 위한 원칙을 담았다. 국가는 알고리즘 시스템 사용에서 인권 침해를 억지하고 모든 행위자가 인권을 존중 및 증진하고 침해 가능성을 방지하도록 입법 및 규제 체제를 개발해야 할 의무가 있다. 이와 별개로, 공공 및 민간 부문 행위자는 국제적으로 인정된 인권을 존중할 책임이 있다. 특히 지침은 알고리즘 및 자동화의 인권 침해 방지 쟁점으로 △일반 적용 원칙, △데이터 관리, △분석 및 모델링, △투명성·책무성 및 효과적인 구제수단, △예방적 조치, △연구·혁신 및 대중적 인식 제고의 문제를 꼽고 이에 대한 국가와 민간의 의무를 각각 제시하였다.

우선 국가는 일반 원칙으로, 인권 침해를 예방, 탐지, 금지 및 구제하는 효과적이고 예측 가능한 입법, 인권영향평가 등 지속적인 검토 체계, 알고리즘 시스템의 위험에 대응하는 민주적 참여 및 인식 제고, 규제 및 감독을 위한 제도적 체계를 수립해야 한다. 국가는 관련 행위자의 법적 준수 여부를 확인하기 위해 적절한 문서 제출을 요구하는 등 법의 집행 가능성 및 집행을 보장해야 한다. 공공 및 민간 부문 행위자가 그 법적 의무를 이행하지 않는 경우 책임을 져야 한다. 국가는 알고리즘 시스템의 전체 수명 주기 동안 개별 시스템의 인권 영향 및 다른 기술과의 상호 작용을 정기적으로 평가해야 한다. 평가는 영향을 받거나 영향을 받을 가능성이 있는 사람들과의 광범위하고 효과적인 협의를 기반으로 수행되어야 한다. 특히 국가는 독립적 감독 기관, 평등 기구, 국가인권기구, 대학, 표준 수립 기구, 서비스 운영자, 알고리즘 시스템 개발자 및 특히 인권 옹호 등 다양한 분야의 관련 비정부 기구와 긴밀하게 협력해야 한다.

데이터 관리와 관련하여, 국가는 알고리즘 시스템에서 개인의 정보적 자기결정권을 보장하고 개인이 자신의 개인정보를 통제할 수 있는 방안을 제공하여야 한다. 또한 알고리즘 시스템에서 사용되는 데이터는 종종 편향을 포함하고 성별, 인종, 종교, 정치적 의견 또는 사회적 출신과 같은 분류 기준의 대리변수 역할을 할 수 있기 때문에 국가는 데이터 품질의 결과로 인권 및 차별금지 원칙이 영향을 받는 상황을 평가해야 한다. 또한 데이터의 탈익명화 가능성, 부적절하거나 탈맥락적 사용, 자동화 수단을 통해 새롭고, 추론적이며, 민감할 수 있는 데이터의 생성과 같은 내재적 위험뿐만 아니라, 데이터셋의 품질과 출처를 더욱 신중하게 고려해야 한다.

분석 및 모델링에 있어서, 국가는 알고리즘 시스템이 설계에서부터 안전, 프라이버시, 개인 정보 보호 및 보안 안전장치를 포함하도록 보장해야 한다. 중대한 인권 영향을 유발할 가능성이 있는 컴퓨팅 실험은 인권영향평가 후에만 수행하도록 보장해야 하며, 실시간 환경에서 검사되고 실시간 효과를 생성하는 알고리즘 시스템에 대해서는 더욱 엄격한 검사, 평가, 보고, 감사가 이루어져야 한다. 국가는 시스템이 달성하거나 최적화하려는 목표의 합법성과 정당성, 그리고 인권과 관련하여 미칠 수 있는 영향에 대하여 공개적이고 협의적이며 독립적으로 평가를 수행하기 위하여 노력하여야 한다. 또한 그러한 평가가 조달 절차의 일부를 구성해야 한다. 시스템 검사 중 인권에 대한 중대한 제한이 식별될 경우 즉각적인 시정이 이루어져야 하며, 시정이 이루어질 때까지 시스템이 정지

되어야 한다. 또한 데이터셋 및 시스템의 외부 효과에 대한 평가도 요구된다. 알고리즘 시스템이 인권에 미칠 수 있는 영향에 따라 검사는 가능한 한 실제 개인정보를 사용하지 않고 수행되어야 하며, 제안된 시스템이 인구집단 및 그 환경에 미치는 외부 효과를 충분히 고려하여 배치 전과 후에 다양하고 대표적인 이해관계자 절차를 포함하는 검사가 이루어져야 한다.

투명성, 책무성 및 효과적인 구제 수단과 관련하여, 국가가 직접 구현하거나 민간 부문 행위자들이 국가를 위해 구현하는 알고리즘 시스템의 공공 조달, 사용, 설계에 관련하여 적절한 수준의 투명성과 설명 가능성 등 기본적인 처리의 기준 및 방법이 설정되어야 한다. 지적 재산 또는 영업 비밀에 대한 입법 체제는 이러한 투명성을 배제해서는 안 된다. 알고리즘 시스템에 의해 수행되거나 지원되는 모든 선택 절차 또는 결정은 인권 행사에 중대한 영향을 미칠 수 있기 때문에, 국가는 알고리즘 선택 또는 결정에 대하여 명확하고 접근 가능한 방식으로 식별 가능하고 추적 가능하도록 보장해야 한다. 또한 알고리즘 관련 결정과 판단으로 영향을 받는 개인 및 집단은 이의를 제기할 수 있는 효과적인 수단을 제공받아야 한다. 또 국가는 국가적 알고리즘 시스템으로 영향을 받는 개인 또는 집단이 제기하는 이의제기에 대해 적절한 자원을 갖춘 국가인권기구 등 독립기관이 적정한 감독을 수행하도록 보장해야 한다. 이의제기와 그에 대한 시정조치 등 후속 조치에 대한 정보는 정기적으로 문서화되고 공개되어야 한다.

예방적 조치와 관련하여, 무엇보다 국가는 인권영향평가를 실시하기 위해 관련 표준, 체계, 지표 및 방법 등에 대한 적절한 지침을 개발하고 구현하여야 한다. 인권영향평가를 통하여 중대한 인권 영향을 미칠 수 있는 모든 알고리즘 시스템에 대하여 그 수명주기의 모든 단계에서 잠재적 위험을 평가하고 그러한 위험을 예방하거나 완화하기 위한 조치, 보호장치 및 메커니즘을 수립하여야 한다. 인권영향평가는 인권에 높은 위험을 수반하는 모든 알고리즘 시스템에 대해 의무화되어야 한다. 공공 조달되는 알고리즘 시스템은 정기적이고 협의적으로 인권영향평가를 수행하도록 해야 한다. 고위험 알고리즘 시스템과 관련된 모든 인권영향평가가 독립적인 전문가 검토 및 검사를 위해 제출되어야 하고, 국가 시스템과 관련된 인권영향평가는 공개되어야 한다. 인권영향평가에서 식별된 인권 침해는 즉시 해결 및 구제되어야 하며, 완화할 수 없는 중대한 인권 위험이 식별된 경우 공공기관은 해당 알고리즘 시스템을 구현하거나 사용해서는 안 된다. 국가가 공공

서비스 전달에서 민간 서비스의 조달 또는 참여 메커니즘을 활용할 때, 국가는 알고리즘 시스템의 사용과 상호 작용에 대한 감독, 노하우, 소유권, 통제를 유지해야 할 필요성을 충분히 고려해야 한다.

마지막으로 국가는 인권에 대한 평등한 접근과 향유를 향상시키는 권리 촉진 기술을 지원하고, 공익을 증진하며, 인간 중심적이고 지속 가능한 혁신을 장려하는 한편, 독립적인 연구를 촉진하여야 한다. 적절한 경우 상업적으로 가장 실행 가능한 최적화 프로세스를 절대적으로 선호하려는 영향력에 대한 억제가 필요할 수 있으며, 알고리즘 시스템의 불투명성, 설명 불가능성 및 이의제기 불가능성과 관련하여 현존하는 책임성 격차에 대응하는 효과적인 책무성 메커니즘 및 솔루션의 개발을 연구하는 독립적 연구가 이루어져야 한다.

한편, 민간 부문 행위자는 일반 원칙으로, 자신의 알고리즘 시스템으로 영향을 받는 당사자들의 인권을 존중해야 할 책임이 있다. 알고리즘 시스템의 설계, 개발, 판매, 배치, 구현 및 서비스에 관여하는 민간 부문 행위자는 인권을 존중하고 차별을 조장하거나 고착화하는 것을 방지하기 위하여 인권 실사를 실시하여야 한다. 이러한 활동은 문서화되어야 한다. 또한 앞서 국가의 의무로 설명된 바 있는 지속적인 검토, 민주적 참여 및 인식 제고, 정보적 자기 결정권, 컴퓨팅 실험, 검사 및 알고리즘 의사결정의 식별 가능성의 경우, 민간 부문 행위자 역시 준수하여야 한다.

데이터 관리와 관련하여, 민간 부문 행위자는 알고리즘 시스템의 영향을 받는 개인에게 개인정보의 모든 사용에 관하여 동의하거나 철회할 수 있음을 알려야 한다. 프라이버시 선택 옵션은 쉽게 볼 수 있고 중립적이며 이해하기 쉬운 방식으로 제시되어야 하며 프라이버시 증진 기술을 사용해야 한다. 기본 설정은 개인정보 처리의 구체적이고 합법적인 목적에 필요하고 비례적인 개인정보를 수집해야 하며 추적 설정은 옵트아웃 모드가 기본값으로 설정되어야 한다.

분석 및 모델링에 있어서는, 민간 부문 행위자는 알고리즘 시스템 학습에 사용하는 데이터의 품질, 특성 및 출처와 관련된 위험을 알고 있어야 한다. 알고리즘 시스템의 개인정보에 대한 평가 및 검사는 충분히 다양하고 대표적인 표본 집단으로 수행되어야 하며 특정 인구통계학적 집단을 의존하거나 차별하지 않아야 한다. 알고리즘 시스템의 개발, 검사 또는 배치가 특정 개인, 집단, 인구집단 및 환경에 대한 위험 또는 손실의 외부 효

과를 수반하는 경우 그 시스템의 개발을 중단하거나 조정해야 한다.

투명성, 책무성 및 효과적인 구제 수단과 관련하여, 민간 부문 행위자는 중대한 인권 영향을 유발할 수 있는 알고리즘 시스템의 사용 및 그 특성과 기능에 대하여, 일반 대중은 물론 영향을 받는 당사자 모두에게 이용 약관으로 명확하게 알려야 한다. 변경 사항이 있는 경우에도 영향을 받는 모든 당사자, 고객 및 사용자에게 변경 사항을 통지하고 해당되는 경우 동의를 요청하여야 한다. 이의제기가 가능하도록 민간 부문 행위자는 인간 검토자에 대한 접촉이 가능하도록 보장해야 한다. 민간 부문 행위자는 그들이 제공하는 제품 및 서비스와 관련하여 영향을 받는 개인 및 집단이 제기한 불만에 대하여 공개하여야 한다. 민간 부문 행위자는 알고리즘 시스템에 이의를 제기하거나 권리 침해를 구제하려는 개인, 집단 및 법인이 온라인과 오프라인에서 효과적인 구제 수단 및 분쟁 해결 시스템을 이용할 수 있도록 보장해야 한다.

예방적 조치와 관련하여, 민간 부문 행위자는 알고리즘 시스템의 기술적 오류뿐 아니라 잠재적인 법적, 사회적, 윤리적 영향을 탐지하기 위해 지속적으로 평가 및 검사가 이루어지도록 내부 절차를 개발하고 문서화해야 한다. 알고리즘 시스템이 인권에 고위험을 수반하는 경우, 민간 부문 행위자는 위험을 관리하는 방법에 대하여 관련 소관 감독 기관에 통지하고 협의할 수 있어야 한다. 민간 부문 행위자는 정기적이고 독립적인 전문가 검토 및 감독을 위해 이러한 알고리즘 시스템을 제출해야 한다. 인권영향평가는 영향을 받는 개인 및 집단의 적극적인 참여와 함께 가능한 한 공개적으로 수행되어야 한다. 인권영향평가의 후속 조치로서, 식별된 오류를 가능한한 신속하게 해결하고, 적절한 경우 관련 활동을 일시 중단시켜야 한다. 이를 위해서는 알고리즘 시스템의 설계, 검사 및 배치 단계 전반에 걸쳐 정기적이고 지속적인 품질 보증 검사와 실시간 감사가 필요하다. 영향을 받는 개인과의 정기적인 협의가 추가적으로 필요하며, 이는 부정적인 인권 영향을 악화시키고 고착화할 수 있는 피드백 순환구조의 위험을 고려할 때 특히 중요하다.

마지막으로 민간 부문 행위자는 긍정적인 인권 영향을 창출하고 공익을 증진할 수 있는 알고리즘 시스템에 대한 연구를 지원하여야 한다. 특히 알고리즘 시스템의 영향을 평가하기 위한 메커니즘을 개발하고 취약하고 소외된 인구집단의 요구를 해결하기 위한 알고리즘 시스템을 개발하는 독립적인 연구를 지원하여야 한다.

5. 캐나다 정부 <알고리즘 영향평가 도구>[103]

2019년 캐나다 정부는 재정위원회 훈령으로 <자동화된 의사결정 지침>을 제정하여 공공기관에 대한 인공지능 도입 요건을 법규화하였다. 이 훈령은 공공기관이 의사결정에 사용하는 인공지능 알고리즘에 대하여 위험 기반 접근법을 취하고 있으며, 공공기관 인공지능에 대하여 영향평가를 의무적으로 실시하도록 하였다. 알고리즘 영향평가는 프로젝트 설계 단계의 초기 시점에 실시되어야 하며, 그 결과는 캐나다 공식 언어로 접근 가능한 형식으로 발표되어야 한다.

공공기관 인공지능은 영향평가 결과별로 위험성 수준을 수준 Ⅰ, 수준 Ⅱ, 수준 Ⅲ, 수준 Ⅳ의 4단계로 나누고 요구사항을 차등 적용한다.[104] 위험성 수준이 높아지면 전문가 검토, 공지, 의사결정에 대한 인간의 개입, 설명 요구사항, 검사, 모니터링, 교육훈련, 비상계획, 시스템 구동 승인에 대한 요구사항도 높아진다.

전문가 검토(peer review)의 경우, 위험 수준에 따라 비해당(수준 Ⅰ)부터 1개 이상을 수행하거나(수준 Ⅱ, 수준 Ⅲ) 2개 이상을 수행(수준 Ⅳ)하여야 하며, 공지의 경우, 위험 수준에 따라 비해당(수준 Ⅰ)부터 프로그램이나 웹사이트에 단순 게시하거나(수준 Ⅱ) 구성 요소 작동 방식, 감사 결과, 학습 데이터 등 상세한 사항을 문서로 안내(수준 Ⅲ, 수준 Ⅳ)하여야 한다. 의사결정에 대한 인간의 개입(Human-in-the- loop for decisions)의 경우, 위험 수준에 따라 의사결정이 인간의 직접적인 개입 없이 내려질 수 있거나(수준 Ⅰ, 수준 Ⅱ) 특정 시점에 인적 개입이 없으면 의사결정이 내려질 수 없고 최종 의사결정은 사람에 의해 이루어져야 하며(수준 Ⅲ, 수준 Ⅳ), 설명의 경우, 위험 수준에 따라 의사결정 결과에 대한 공통 설명이 제공되는 경우(수준 Ⅰ)로부터 거부 처분에 대하여 요청에 따른 설명을 제공(수준 Ⅱ)되거나 모든 거부 처분에 대하여 설명을 제공하도록(수준 Ⅲ, 수준 Ⅳ) 요구하였다.

검사의 경우, 모든 위험 수준에서 공통적으로 생산에 착수하기 전 학습 데이터가 의도하지 않은 데이터 편향 및 결과에 부당하게 영향을 미칠 수 있는 요소에 대해 검사할 수 있는 적절한 절차를 개발하여야 하며, 그 후로도 정기적으로 자동화된 의사결정 시스

103) 전문은 부록 Ⅴ 참조.
104) Directive on Automated Decision-Making.

템에서 사용 중인 데이터가 여전히 관련성이 있고 정확하며 최신인지 확인하기 위해 검사를 실시하여야 하고, 모니터링의 경우에도 역시 모든 위험 수준에서 공통적으로 시스템의 결과를 지속적으로 모니터링하여 의도하지 않은 결과로부터 보호하고 이 지침과 관련 법률의 준수를 보장하여야 한다. 교육훈련의 경우, 위험수준에 따라 비해당(수준 Ⅰ)부터 시스템의 설계 및 기능에 대한 문서화를 할수 있어야 하거나(수준 Ⅱ) 이에 추가하여 교육과정 이수가 필수적(수준 Ⅲ)이고, 나아가 반복적 이수 및 확인(수준 Ⅳ)이 이루어지도록 요구하고 있다. 비상 계획의 경우, 수준 Ⅲ과 수준 Ⅳ에서 자동화된 의사결정 시스템을 사용할 수 없는 경우를 대비한 비상 계획 및 백업 시스템을 요구하였고, 시스템 구동 승인의 경우, 수준 Ⅲ(부서장 승인)과 수준 Ⅳ(재정위원회 승인)에서 요구되고 있다.

캐나다 정부는 수준별로 요구사항이 달라지는 공공기관 인공지능의 위험성 수준을 측정하기 위하여 영향평가 도구도 개발하였다.

평가 도구는 질문지 형식으로 시스템의 위험 영향 정도를 판단하며, 48개의 위험성과 33개의 완화 조치에 대한 질문들로 구성되어 있다. 위험성 영역은 (1) 프로젝트의 단계, 동기, 위험성 지표, 권한 등에서 프로젝트의 위험성에 대하여 검토하고, (2) 이미지 인식 등 시스템 사양 등으로 시스템의 위험성을 살펴보고, (3) 알고리즘의 투명성, 설명가능성 등으로 알고리즘의 위험성을 검토하며, (4) 보건의료서비스 등 자동화된 의사결정의 영역별 위험성을 살펴보며, (5) 영향의 지속성, 가역성, 영향을 미치는 분야에 따른 위험성을 검토하고, (6) 데이터의 출처나 보안 등급 등의 소스와 유형에 따른 위험성을 살펴본다. 완화성 영역은 (7) 내부 및 외부의 전문가 및 이해관계자로부터 자문을 받는지 여부, (8) 데이터가 대표성을 가지고 편향적이지 않도록 보장하는 절차 및 투명성 조치를 가지고 있는지, 시스템과 그 의사결정을 감사하는 절차 및 회복 조치를 가지고 있는지, 개인정보를 보호하는 조치를 가지고 있는지 등 위험성 제거 및 완화 조치를 살펴본다.

위험성에 따라 부과된 영향평가 점수는, 완화성 조치에 따라 차감될 수 있다. 완화성 점수가 달성 가능한 최대 점수의 80% 이상일 경우 15%를 차감한다. 달성 가능한 최대 점수 대비 현재 점수가 0% ~ 25%일 경우 영향이 거의 없거나 전혀 없는 수준 Ⅰ로 정의되고, 26% ~ 50% 일 경우 중간 영향의 수준 Ⅱ, 51% ~ 75% 일 경우 높은 영향의 수준 Ⅲ, 76% ~ 100% 일 경우 아주 높은 영향의 수준 Ⅳ로 정의된다.

제4절 공공기관 모범 기준 및 제도

세계 각국은 특히 공공부문 인공지능에 대하여는 강력한 투명성 기준과 조달 등의 공공 절차를 적용하고 있다.

먼저, 각국 법원이 공공부문 인공지능에 대하여 투명한 정보공개 및 적법절차 보장을 강력히 요구하는 판결을 내렸다.

2017년 폴란드 법원은 정부의 실업자 점수 알고리즘이 국회에 입법한 법률에 근거를 두고 있지 않은데 대하여 위헌이라고 결정하였다.[105] 미국 텍사스 휴스턴의 연방지방법원은 민간 기업에서 조달한 교육청의 교사 평가 알고리즘에 대하여 투명성과 적법절차 부족을 이유로 운영을 중단시켰다.[106] 특히 법원은 민간 기업의 영업 비밀과 국민의 헌법상 권리인 적법절차를 모두 충족하기 위해서는 공공기관의 중요한 의사결정에 비밀 알고리즘을 사용해서는 안된다고 설시하였다. 2020년에는 네덜란드 헤이그 지방법원이 사회복지급여 부정수급 탐지 시스템에 대하여 영업 비밀을 이유로 한 투명성 부족과 개인정보보호법 위반을 이유로 운영을 중단하라는 취지의 판결을 내렸다.[107]

일부 국가는 조달 지침으로 공공부문에 도입되는 인공지능에 대하여 더욱 엄격한 기준과 절차를 요구하고 있다.

유럽 집행위원회는 2020년 <인공지능 공공조달 백서>[108]에서 "데이터 윤리, 민주주의 및 기본권에 부합하는 공공조달을 구현"하고, 특히 위험성에 따른 체계적 규율을 추진하고자 하였다. 특히 유럽연합이 지향하는 '신뢰가능 인공지능'은 "책임성, 기술적 안전성, 지속가능성에 대한 요구 뿐 아니라 데이터 윤리 요소를 포함한 공공조달 체계를 수립하고 이를 현행 법적 의무에 적용함으로써 달성될 수 있다"고 지적한다. 백서는 공공조달에 있어 위험기반·체계적 접근법을 취하고, 이를 위한 5단계 실사 절차를 권장하였다. 첫째, 사전적인 위험 영향평가를 실시하여 사람과 집단, 권리와 자유, 민주적 조직

105) Poland: Government to scrap controversial unemployment scoring system. Algorithm Watch (2019. 4. 16).
106) Federal Suit Settlement: End of Value-Added Measures in Houston. Education News 보도 (2017. 10. 10.); HOUSTON FED. OF TEACHERS v. HOUSTON INDEPENDENT. https://www.leagle.com/decision/infdco20170530802# 참조.
107) 앞의 The Guardian 보도 (2020. 2. 5.); Welfare Surveillance on Trial in the Netherlands. Human Rights Watch (2019. 11. 8).
108) European Commission (2020b).

과 절차, 사회와 환경에 부작용을 미치는지를 살핀다. 둘째, 공급자 예비 심사를 실시하여 설계 절차의 최초 단계서부터 인공지능 관련 데이터 윤리 요구사항을 고려하고 정의하고 구현하도록 요구한다. 이때 요구되는 데이터 윤리 요구사항으로는 △ 인공지능이 이용자와 직접적으로(챗봇, 가상비서 등) 또는 간접적으로(자동화된 의사결정) 상호작용한다면 이는 필히 인간이 아니라는 점을 밝혀야 하고, △ 인공지능 시스템이 추적가능하고, 설명가능하고 이해관계자를 수용해야 하며, △ 인공지능 시스템이 편향을 방지하고 보편적 설계를 따라야 하며 검토 절차를 포함해야 하고, △ 기술적 안전성은 문서화되어 설명가능성, 공정 커뮤니케이션 및 감사를 보장해야 한다. 셋째, 공급자를 선정하고 계약하는 품질 기준으로 정보보안, 데이터 윤리, 환경 측면, 프라이버시, 보편적 설계 등에 적용되는 표준 및 관리시스템에 관한 기술사양을 반영해야 한다. 넷째, 발주 공공기관은 계약 이행 조건에 지속가능성, 기본권 존중, 데이터 윤리에 대한 조항을 포함하고 제재 조항 및 문서화 요구사항을 명시해야 한다. 다섯째, 공급자는 계약을 집행할 때 데이터 윤리, 법적 준수, 책임, 기술적 안전성 및 지속가능성의 다섯 가지 표제에 따라 공공 계약에 명시된 요구사항을 충족해야 한다.

유럽평의회 인권위원장은 2018년 이미 공공기관 인공지능 조달 절차에 인권영향평가의 실시를 권고한 바 있다.[109] 공공기관은 인권영향평가의 공표나 수행이 가능하지 않는 공급자로부터 인공지능 시스템을 조달해서는 안 된다는 것이다.

영국 정부는 2020년 6월 인공지능 조달지침을 발표하고 공공조달을 통하는 인공지능에서 10대 원칙을 따르도록 요구하였다.[110] 이 지침은 첫째, 공공기관에 대하여 인공지능 도입 계획에 조달을 포함할 것을 요구하였다. 둘째, 관련 의사결정을 위하여 다학제 팀을 구성할 것을 요구한다. 기관 내부 뿐 아니라 낙찰 공급업체에 대해서도, 적합한 기술력을 갖춘 팀을 구성하고 인공지능 시스템의 편향성을 완화하기 위해 다양성 수용을 요구한다. 셋째, 조달 절차 개시 전에 데이터 평가 실시를 요구한다. 이는 △조달 절차의 개시 단계부터 데이터 거버넌스 메커니즘이 가동될 수 있도록 확보하고, △프로젝트에 관련 데이터를 사용할 수 있는지 여부를 평가하며, △시장에 출시하기 전에 데이터 내부의 결함 및 편향 가능성을 해결할 것을 요구한다. 데이터 문제를 직접 해결할 수 없는

109) Council of Europe Commissioner for Human Rights (2019).
110) Guidelines for AI procurement.

경우 이를 해결하기 위한 계획을 수립해야 한다. 또한, △조달 계획 및 후속 프로젝트를 위해 공급업체와 데이터를 공유할 것인지 여부 및 방법을 정의할 것 또한 요구한다. 넷째, 인공지능 도입의 혜택과 위험성에 대한 평가를 요구한다. 이러한 평가를 위하여 △ 제안서 평가와 의사결정에 있어 공익이 주요 요소라는 점을 조달 문서에 설명하고, <사회적 가치 지침>에 따라 인공지능 시스템이 인간과 사회 경제에 미치는 영향 및 편익을 고려하도록 하며, 해당 조달 사업이 공익적 목표와 관련이 있고 차별금지, 동등한 대우 및 비례성의 원칙을 준수할 것을 요구한다. 더불어 △ 당면한 문제와 관련하여 인공지능을 고려한 배경을 조달 문서에 명확히 설명하고 대안적 솔루션에 대하여 열린 태도를 취해야 하며, △ 조달 절차 개시 단계에서 인공지능 영향평가를 수행하고, 중간 조사 결과가 조달에 반영되는지 확인하도록 요구하였다. 주요 의사결정 단계에서는 평가 결과를 재차 살펴보아야 한다.

다섯째, 기관은 시장 형성 초기단계부터 효과적으로 개입해야 한다. 다양한 인공지능 공급자들과 다양한 방식으로 관계를 맺고, 인공지능 생태계에 개방적인 경쟁 환경을 구축해야 한다. 여섯째, 기관은 올바른 시장 경로를 구축하고 특정 솔루션보다 해결하고자 하는 공익적 과제를 제시해야 한다. 일곱째, 거버넌스 및 정보인증을 위한 계획이 수립되어야 한다. 인공지능 시스템에 대한 철저한 검사를 위한 관리감독 메커니즘을 구축하고, 기존 법과 표준을 준수하도록 요구해야 한다. 인공지능 의사결정의 투명성을 최대화하여 사용자에게 인공지능 시스템이 잘 기능한다는 확신을 부여해야 한다. 여덟째, 특히 이 지침은 조달되는 인공지능 시스템의 "'블랙박스' 및 공급업체에 대한 종속(lock-in) 방지"를 요구한다. 알고리즘의 설명/해석 가능성을 중요 기준으로 설정하고, 특정 공급업체에 고착되지 않도록 여러 다른 공급자들의 인공지능 시스템 참여를 유도해야 한다. 아홉째, 평가 단계에서는 인공지능 도입의 기술적/윤리적 한계를 해소 필요성에 집중할 것을 요구하였다. 데이터 편향 문제는 없는지, 기존 서비스/기술과 통합 과정에 충분한 검토가 이루어졌는지, 적절한 기술적 표준을 준수하고 있는지 등을 살펴야 한다. 열째, 인공지능 시스템의 수명주기 전체에 대한 관리를 고려하여야 한다. 인공지능 관리는 △ 인공지능 조달 과정에서 일회성이 아니라 수명주기 전체에 걸친 검사 필요성을 고려하여야 하며, △ 지식 이전 및 교육훈련을 요구사항에 포함하여야 하고, △ 인공지능 시스템을 이해해야 하는 비전문가 대상 교육훈련 및 설명을 요구사항에 포함하며, △

적절하고 지속적인 고객지원 및 호스팅 협의를 보장하여야 한다.

한편, 해외 일부 지방자치단체는 자치단체가 도입한 인공지능에 대한 사항을 시민들에게 공개하는 정책을 시행 중이다. 네덜란드 암스테르담과 핀란드 헬싱키 시는 2020년 시민들에 알고리즘 등록부를 시범적으로 공개하였다.[111] 이 알고리즘 등록부는 시가 도입·운영하는 인공지능 시스템의 △학습 데이터셋에 대한 정보 △데이터 처리에 대한 정보 △차별 방지에 대한 정보 △인간 감독에 대한 정보 △위험성에 대한 정보 등을 읽기 쉬운 평문으로 공개하는 한편 이를 책임지는 공직자의 이름, 부서 및 연락처도 공개하여 시민들이 의견을 제출할 수 있도록 하였다. 프랑스 앙티브시는 「디지털 공화국법」에 따라 2021년 2월부터 알고리즘 주민공개 제도를 실시하고, 각 알고리즘의 △행정 정보(관계 행정부서명, 관리/위탁업체, 행정부서 연락처, 정보 업데이트일자) △알고리즘 관련 사업 및 의사결정에 대한 정보(알고리즘명, 배경, 알고리즘 목적, 의사결정 절차, 의사결정 자동화 수준, 법적 근거, 관련자료) △의사결정 영향에 대한 정보(연간 이루어진 행정결정 수, 결정 범위, 결정의 영향을 받는 사람들) △알고리즘 작동원리에 대한 정보(처리 데이터, 처리 데이터 출처, 처리 데이터 수집방법, 알고리즘 유형, 알고리즘의 수행 작업)를 공개하였다.[112]

111) Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI. VentureBeat 보도 (2020. 9. 28).
112) Antibes publie l'inventaire de ses algorithmes. laGazette 보도(2021. 2. 11).

제5절 시사점

인공지능 등 신기술과 관련하여 인권 보장을 요구하는 국제적인 기준 및 제도에서 비교적 공통적이고 인권적 함의가 높은 기준을 추려보면 다음과 같다.

첫째, 인공지능 개발과 활용에서 인권과 책임성을 보장하는 법률과 감독 체계 수립을 요구한다. 유엔 사무총장은 신기술의 인권 보장을 위하여 국가가 민간 부문 활동에 관한 조치를 포함하여 입법 조치를 취해야 할 의무를 재확인하였으며, 법률로써 신기술이 사용되는 상황에서 책임성을 완전하게 보장하고 적절한 감독 체계와 구제 수단을 수립할 것을 요구하였다. 유엔 인권최고대표는 2021년 보고서에서 인공지능 문제를 권리 존중 방식으로 해결하기 위한 새로운 법률 체계의 마련을 촉구하였다(44문). 유럽평의회는 각 회원국에게 인권 침해를 예방, 탐지, 금지 및 구제하는 효과적이고 예측 가능한 입법을 요구하고, 공공 및 민간 부문 행위자가 그 법적 의무를 이행하지 않는 경우 책임을 져야 한다고 강조하였다. 더불어 국가는 관련 행위자의 법적 준수 여부를 확인하기 위해 적절한 문서 제출을 요구하는 등 법의 집행 가능성을 보장해야 한다. 캐나다 정부는 2019년 공공기관 인공지능에 대한 영향평가를 의무화하고 위험 수준별 요구사항을 법규화하였으며, 뉴질랜드 정부 역시 2020년부터 공공기관 인공지능에 대하여 영향평가와 위험수준별 규율을 적용하고 있다.[113] 영국 정부는 2020년 공공 조달 지침을 발표하고 시행 중이며, 조달되는 인공지능 시스템의 "'블랙박스' 및 공급업체에 대한 종속(lock-in) 방지"를 요구하였다.[114] 특히 유럽연합은 2021년 4월 21일 공공과 민간 부문 고위험 인공지능에 요구사항을 적용하는 인공지능법(안)을 발의하였는데, 유럽 집행위원회는 많은 알고리즘의 불투명성이 불확실성을 유발하고 안전 및 기본권리에 대한 기존 법률의 효과적인 집행을 방해할 수 있다는 점을 지적하고, 이러한 문제를 해결하기 위하여 입법 조치에 이르렀다고 설명하였다.

둘째, 인공지능 개발과 활용에서 투명성과 자기결정권 보장을 요구한다. 이러한 투명성은 인공지능이 사용된다는 사실을 공개하는 것으로부터 공공부문 인공지능의 개발과

113) Algorithm charter for Aotearoa New Zealand. <https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/ (검색일: 2021. 9. 1.)>.

114) Guidelines for AI procurement.

활용에서 그 영향을 받는 사람들의 참여를 보장하고 설명할 수 없는 의사결정을 금지하는 데에도 이른다. 유엔 사무총장은 신기술의 개발 및 도입에 관한 의사결정에 모든 관련 이해관계자의 참여를 보장하고 특히 공공부문에서 인공지능이 지원하는 의사결정에 대하여 적절한 설명가능성이 보장될 필요가 있다고 권고하였다. 유엔 인권최고대표는 2021년 보고서에서 일반 사항 공개, 영향을 받는 당사자에 대한 설명가능성, 공공조사 가능성과 책무성 확보 등 다차원의 투명성을 소개하였다(55~56문). 유럽평의회 역시 국가가 개발하거나 민간으로부터 조달하는 인공지능의 경우 적절한 수준의 투명성과 설명 가능성을 요구하였고, 유럽평의회 인권위원장은 최소한 조달 및 인권영향평가 단계에서는 영향을 받는 당사자 집단 등으로부터 공개적인 의견 수렴을 권고하였다. 캐나다 자동화된 의사결정 지침 훈령에서는 공공부문 인공지능에 대하여 의사결정 전에 공지하도록 하고 의사결정 후에는 설명할 의무를 법규화하였다. 민간의 경우에도 고위험 인공지능에 대하여는 주요 요인에 대한 공개와 설명이 요구된다. 유럽연합 인공지능법(안)에서는 고위험 인공지능 시스템에 대하여 사용자가 그 결과물이 어떻게 만들어지는지 이해하고 적절하게 사용할 수 있도록 조치하도록 하였다.

특히 자동화된 의사결정의 대상이 되는 사람들에게는 사전적으로 통지받고 의견을 피력하거나 인적 개입을 요구하며 일정하게 거부할 권리가 보장될 필요가 있으며, 오로지 자동화로만 이루어지는 중요 의사결정의 대상이 되는 사람들에게는 그 사유에 대한 설명을 듣고 이의를 제기할 수 있는 권리가 보장되어야 한다. 유럽연합 GDPR에서는 모든 완전 자동화 의사결정에서 통지권, 의견의 피력권, 인적 개입 요구권, 설명요구권 및 이의제기권을 보장하였다.

한편, 공공과 민간 고위험 인공지능에 대해서는 문서화 보관 및 감독이 요구된다. 유엔 인권최고대표는 특히 국가는 감사 대상이 될 수 없는 인공지능 시스템의 사용을 피해야 한다고 강조하였다. 인공지능에 대한 감독 기관의 검사/감사가능성 보장과 기술적 설명가능성 확보에 대한 각국의 노력도 계속되고 있다.[115] 유럽평의회 인권위원장은 어

115) 예를 들어 미국국립표준기술연구소는 2020년 8월 <설명가능 인공지능 시스템을 위한 원칙(초안)>을 발표하였다. 이는 △설명성: AI시스템은 모든 산출물에 대해 증거 또는 이유를 제시할 수 있어야 함 △의미성: AI시스템은 개별 이용자가 이해할 수 있는 설명을 제공할 수 있어야 함 △설명의 정확성: 설명은 산출물을 생성하는 시스템의 프로세스를 올바르게 반영할 수 있어야 함 △지식의 한계성: 시스템은 그 산출물에 대해 충분하게 신뢰할 수 있도록 설계되었거나 이에 도달하는 조건 하에서만 운영할 수 있어야 한다는 내용을 담고 있다. National Institute of Standards and Technology (2020). Four Principles of Explainable

떠한 인공지능 시스템도 인적 검토와 정밀 감사를 허용하지 않을 정도로 복잡한 수준이어서는 안 되며, 투명성과 책임성의 적절한 기준을 따를 수 없는 시스템은 사용하지 말아야 한다고 강조하였다. 캐나다 훈령의 경우 캐나다 정부가 특별 감사, 조사, 검사, 심사, 집행 조치 또는 사법 절차에 필요한 경우 공공기관 인공지능 시스템에 대하여 접근하고 시험할 권리를 보유한다고 명시하였다. 유럽연합 인공지능법(안)은 공공과 민간을 불문하고 고위험 인공지능에 대해서 기술문서를 작성·유지하고, 모든 단계에서 인공지능 시스템의 기능을 추적할 수 있는 수준으로 로그기록을 저장하도록 의무화하였다. 만약 고위험 인공지능의 오작동으로 심각한 부상 또는 재산 피해가 발생한 경우에는 15일 이내에 관할 당국에 보고하고 조사를 받아야 한다.

셋째, 여러 인권 영역 중에 특히 인공지능 개발과 활용에서 문제가 되는 개인정보 권리를 보호하기 위한 규범 준수가 요구된다. 학습 및 검사에 사용되는 데이터셋을 비롯하여 인공지능 개발과 활용에서 사용되는 모든 데이터셋의 개인정보 처리는 개인정보보호법을 준수하여 합법적으로 이루어져야 하고 정보주체의 권리를 보장하여야 한다. 유엔 인권최고대표는 2021년 보고서에서 개인정보보호법과 개인정보보호 감독기구들이 인공지능에서 사용되는 데이터를 규율할 필요성을 강조하였다(42~43문). 유럽평의회 권고에서는 데이터의 탈익명화 가능성, 부적절하거나 탈맥락적 사용, 자동화 수단을 통해 새롭고, 추론적이며, 민감할 수 있는 데이터의 생성과 같은 내재적 위험뿐만 아니라, 데이터셋의 품질과 출처를 더욱 신중하게 고려해야 한다고 강조하였다.

넷째, 인공지능의 개발과 활용에서 편향과 차별을 금지하기 위한 조치로서, 데이터셋의 사전적인 품질 관리와 사후적인 모니터링을 요구하고 있다. 유엔 인권최고대표는 2021년 보고서에서 국가 및 기업이 인공지능 시스템 사용과 관련된 차별을 해소하기 위한 노력을 강화하여야 하며, 인공지능 시스템의 결과물 및 그 배치로 인한 영향에 대하여 체계적인 평가와 모니터링을 실시하고, 요구하고, 지원할 것을 권고하였다(59(i)문). 캐나다 훈령의 경우 공공기관 인공지능이 생산에 착수하기 전 학습 데이터에 대한 검사 절차를 요구하였으며, 배치 후로도 정기적으로 데이터의 관련성, 정확성, 최신성을 검사하고 시스템의 결과를 모니터링하도록 하였다. 유럽평의회 권고는 인공지능 시스템에서 사용되는 데이터가 종종 편향을 포함하고 성별, 인종, 종교, 정치적 의견 또는 사회적 출

Artificial Intelligence.

신과 같은 분류 기준의 대리변수 역할을 할 수 있기 때문에 국가는 데이터 품질의 결과로 인권 및 차별금지 원칙이 영향을 받는 상황을 평가해야 한다고 지적하였다. 민간 인공지능의 경우에도 특정 인구통계학적 집단을 의존하거나 차별하지 않아야 하며, 인공지능 시스템의 개발, 검사 또는 배치가 특정 개인, 집단, 인구집단 및 환경에 대한 위험 또는 손실의 외부 효과를 수반하는 경우 그 시스템의 개발을 중단하거나 조정해야 한다. 유럽연합 인공지능법(안)은 공공과 민간을 불문하고 고위험 인공지능의 학습·검증·테스트 데이터셋은 적절한 데이터 거버넌스 및 관리에 따라야 하고, 관련성과 대표성을 갖고 오류없이 완전할 것을 요구한다.

다섯째, 인공지능이 인권과 안전에 미치는 위험을 식별하고 완화시키기 위하여 다양한 영향평가가 제안되고 있으며, 특히 인권영향평가에 대한 요구가 두드러진다. 유엔 사무총장과 유엔 인권최고대표는 인권에 중대한 영향을 미칠 수 있는 인공지능 시스템에 대하여 전체 수명주기 동안 인권영향평가를 비롯한 체계적인 인권 실사를 요구하였다. 유럽평의회 권고는 물론 유럽평의회 인권위원장, 유럽연합 기본권청 및 호주 국가인권위원회 등 인권기구들 역시 인공지능에 대한 인권영향평가의 실시를 권고하고 있으며, 특히 공공기관에는 의무적인 인권영향평가를 실시하고 그 결과를 공개하는 한편, 조달 절차에도 적용할 것을 요구하고 있다. 민간 인공지능의 경우에도 유엔 〈기업과 인권 이행지침〉에 따른 인권 실사 의무가 요구되며, 인권 실사는 영향을 받는 개인 및 집단이 참여하고 공개적으로 실시되어야 한다. 인권에 미치는 위험이 나타났음에도 이를 완화시키거나 해결하지 못한 인공지능 시스템은 개발 또는 활용해서는 안 된다. 특히 유럽평의회는 회원국에게 인권영향평가를 실시하기 위한 표준, 체계, 지표 및 방법 등 적절한 지침을 개발하고 구현할 것을 요구하였다.

여섯째, 인공지능 개발과 활용의 인권 준수 보장을 위하여 독립적이고 효과적인 감독 체계의 마련이 요구된다. 유엔은 사무총장, 인종차별철폐위원회가 인공지능에 대한 국가적 감독 체계의 수립을 반복적으로 권고하였다. 특히 유엔 인권최고대표는 인공지능 시스템의 개발, 배치 및 사용에 대한 적정하고 독립적이고 공정한 감독이 필요하다고 강조하면서 이러한 감독은 행정적, 사법적, 준사법적 및 의회 감독 기관의 조합으로 수행될 수 있다고 제안하였다. 예를 들어, 개인정보보호 기관, 소비자보호 기관, 부문별 규제 기관, 차별방지기구 및 국가인권기구가 감독 시스템의 일부를 구성해야 한다. 더불어 인공

지능 사용을 감독하는 부문 간 규제 기관은 기본 표준을 수립하고 정책 및 집행의 일관성을 보장하는 데 도움이 될 수 있다고 보았다(47문). 유럽연합 기본권청은 인공지능 시스템이 기본권에 미치는 부정적인 영향을 감독하고 이를 해결할 수 있는 효과적인 책임 체계의 구축이 필요하다고 지적하면서, 국가인권기구를 비롯한 기존의 감독 전문 조직을 더 잘 활용할 것을 유럽연합과 회원국들에 권고하였다. 유럽평의회 인권위원장은 독립적인 기구가 인권 준수 여부를 조사하고 영향을 받은 개인의 진정을 접수하고 정기적인 검토를 수행할 수 있어야 한다고 보았다. 특히 인공지능에 대한 감독에서 앞서 살펴 본 투명성이 보장되어야 한다고 강조하면서 감독 기관에 문제의 시스템에 대한 정보가 제공되거나 독립적이고 효과적인 감사가 실시되어야 한다고 하였다.

일곱째, 인공지능의 인권 침해로 인한 피해를 구제할 수 있는 조치가 요구된다. 유엔은 사무총장 등 여러 기구가 인공지능으로 인한 피해에 제도적인 권리 구제를 보장할 것을 반복적으로 요구해 왔다. 특히 인공지능의 자동화된 의사결정에 대한 권리 구제 절차는 적절한 시점에 효과적인 인간의 개입을 요구한다. 유엔 인권최고대표는 인공지능 시스템의 사용과 관련된 인권 침해 및 남용의 피해자가 효과적인 구제수단을 이용할 수 있도록 보장할 것을 각국에 권고하였다(59(g)문). 유럽평의회는 인공지능 결정과 판단으로 영향을 받는 개인 및 집단이 이의를 제기할 수 있는 효과적인 수단을 보장할 것을 회원국에 권고하였다. 국가는 국가적 인공지능 시스템으로 영향을 받는 개인 또는 집단이 제기하는 이의제기에 대해 적절한 자원을 갖춘 국가인권기구 등 독립 기관이 적정한 감독을 수행하도록 보장해야 한다. 유럽평의회 인권위원장은 인공지능 시스템의 개발, 도입 또는 사용으로 발생하는 인권 침해에 대한 책임과 책무는 항상 자연인 또는 법인 측에 있다고 지적하며 공공기관 혹은 민간 기업 인공지능 시스템의 개발, 도입, 혹은 사용으로부터 인권 침해 피해를 주장하는 누구라도 소관 국가기관에서 효과적인 구제 수단을 제공 받을 수 있어야 한다고 강조하였다.

제4장 인공지능 관련 국내 기준 및 제도

제1절 국내 기준

1. 인공지능 국가전략

2019년 12월 17일 정부는 대통령 주재로 열린 국무회의에서 과학기술정보통신부를 비롯한 전 부처가 참여하여 마련한 <인공지능(AI) 국가전략>을 발표하였다.[116]

이 국가전략은 'IT 강국을 넘어 AI 강국으로'라는 제목 하에 3대 분야의 9대 전략과 100대 실행과제를 배치하였다. 3대 분야는 첫째, 연구개발 역량과 기반을 강화하는 '인공지능 생태계'를 구축하고, 둘째, 산업과 사회 분야에서 인공지능 활용을 넓히며, 셋째, 일자리 안전망을 구축하고 인공지능 윤리를 정립하는 사람 중심의 인공지능을 구현하겠다는 것이다. 분야별 실행과제로는 첫째, "세계를 선도하는 인공지능 생태계 구축"을 위한 인공지능 인프라 확충, 기술 경쟁력 확보, 규제 혁신과 법제도 정비, 스타트업 육성 분야에서 27개 과제가, 둘째, "인공지능을 가장 잘 활용하는 나라"를 위한 인재 양성과 국민 교육, 산업 전반의 활용, 인공지능 디지털 정부 구현 분야에서 60개 과제가, 셋째, "사람 중심의 인공지능 구현"을 위한 포용적 일자리 안전망 구축, 역기능 방지와 윤리체계 마련 분야에서 13개 과제가 선정됐다. 전체적으로 보면, <인공지능 국가전략>은 앞쪽에서 산업과 사회 전반의 문명사적 변화를 부각시키고 이후에는 뚜렷하게 기술과 산업 정책 중심으로 구성되었다.

<인공지능 국가전략>에 대한 비판적 지적 중 하나는 혁신을 위해 국가적 역량을 총결집할 것을 선언하면서도 인공지능으로부터 복합적인 영향을 받게 되는 다양한 시민사회 이해당사자의 목소리를 반영하고 참여를 보장하기 위한 계획을 포함하고 있지 않다는 것이다.[117] 국가 계획에 동원되는 존재로서 함의가 큰 '국민'이라는 총칭으로는 인공지능으로 인한 위험 또는 기회 상실의 위기 앞에 처해 있는 시민사회 이해당사자들이 인

116) "IT 강국을 넘어 AI 강국으로!" : 범정부 역량을 결집하여 AI 시대 미래 비전과 전략을 담은 'AI 국가전략' 발표. 부처합동 보도자료 (2019. 12. 17).
117) 오철우 (2020). '사람 중심'의 시선에서 본 인공지능 국가전략. 과학잡지 에피 11호(2020. 3).

권과 민주적 가치 속에 함께 성장과 복지를 추구하는 과정을 반영하기 어렵다.

무엇보다 인권 보장의 의무를 이행하여야 할 국가로서 국가전략을 수립함에 있어 인권에 대한 고려와 언급을 전혀 하지 않은 것은 중대한 결함이라 할 것이다. “사람 중심의 인공지능 구현”을 위한 역기능 방지와 윤리체계 마련은 ‘안전한 인공지능 이용환경 조성’이라는 실용적 목적의 하위 개념으로 제시되었다. 영국 세계디지털파트너와 미 스탠퍼드대학교 세계디지털정책인큐베이터는 2020년 4월 각국의 인공지능 국가 전략을 분석한 공동보고서에서 한국을 인권 준수 체계와 다른 윤리적 또는 사람중심 접근(Ethical or human-centric approaches as alternatives to the human rights framework)을 취한 국가로 분류하였다.[118]

2. 과학기술정보통신부

국내 인공지능 기준과 관련하여 대표적으로 언급되는 것은 과학기술정보통신부의 <인공지능(AI) 윤리기준>이다.

과학기술정보통신부와 정보통신정책연구원의 협의와 대통령 직속 4차산업혁명위원회 전체회의를 거쳐 2020년 12월 22일 발표된 <인공지능 윤리기준>은,[119] 윤리적 인공지능을 실현하기 위해 정부·공공기관, 기업, 이용자 등 모든 사회구성원이 인공지능 개발~활용 전 단계에서 함께 지켜야 할 주요 원칙과 핵심 요건을 제시하는 기준으로 마련되었다. 제정 과정에서 학계·기업·시민단체를 아우르는 주요 전문가들이 자문과 의견수렴 과정에 참여하였고, 초안 발표 이후 공개 공청회 등으로 시민 의견수렴을 거쳤다.

‘사람 중심의 인공지능’을 위하여 <인공지능 윤리기준>은 지향하는 최고가치를 ‘인간성(Humanity)’으로 설정하고, ‘인간성을 위한 인공지능(AI for Humanity)’을 위한 3대 원칙·10대 요건으로 제시하였다. 이때 3대 기본원칙은 ‘인간성(Humanity)’을 구현하기 위해 인공지능의 개발 및 활용 과정에서 ① 인간의 존엄성 원칙, ② 사회의 공

118) Global Partners Digital, Global Digital Policy Incubator (2020). "National Artificial Intelligence Strategies and Human Rights: A Review" <https://www.gp-digital.org/wp-content/uploads/2020/04/National-Artifical-Intelligence-Strategies-and-Human-Rights%E2%80%94A-Review_.pdf (검색일: 2021. 11. 1.)>.

119) 과기정통부, 사람이 중심이 되는 「인공지능(AI) 윤리기준」마련. 과학기술정보통신부 보도자료 (2020. 12. 22).

공선 원칙, ③ 기술의 합목적성 원칙을 지켜야 한다는 것이다. 또한 10대 핵심요건은 3대 기본원칙을 실천하고 이행할 수 있도록 인공지능 개발~활용 전 과정에서 ① 인권 보장, ② 프라이버시 보호, ③ 다양성 존중, ④ 침해금지, ⑤ 공공성, ⑥ 연대성, ⑦ 데이터 관리, ⑧ 책임성, ⑨ 안전성, ⑩ 투명성의 요건이 충족되어야 한다는 것이다.

과학기술정보통신부는 이 윤리기준이 구속력 있는 '법'이나 '지침'이 아닌 도덕적 규범이자 자율규범으로서 '자율적'인 준수를 목표로 하며, 기업 자율성을 존중하고 인공지능 기술발전을 장려하며 기술과 사회변화에 유연하게 대처하고자 하는 지향점을 명확히 밝혔다. 이는 인권 보호에 대한 국가의 의무와 인권 존중에 대한 기업의 책임, 피해자 구제의 실현을 요구하는 국제 인권 규범과는 차이가 있다. 또한 <신뢰가능 인공지능 윤리 가이드라인>으로부터 시작하여 최근 고위험 인공지능을 규제하기 위한 인공지능법(안)을 발의한 유럽연합의 접근법과도 차이를 보인다. 윤리 기준에 '인권 보장'을 포함하면서 국가 인권 기준을 주무하는 국가인권위원회와의 협의 과정 또한 제대로 이루어지지 않았다.

특히 <인공지능 윤리>는 준비 과정에서부터 산업·경제 분야의 자율규제 환경을 조성함으로써 인공지능 연구개발과 산업 성장을 제약하지 않고 기업에 부담을 지우지 않겠다는 목표 하에 마련되었다(서문). 이러한 접근법은 의견 수렴 과정에도 영향을 미쳐 초안에서 독자적인 실행원칙으로 제시되었던 '견고성' 원칙은 기술의 불확실성 등 한계를 고려해야 한다는 학계와 기업의 요구로 삭제되었으며, 초안보다 투명성 원칙 강화를 요구하는 시민단체의 의견은 "현재의 기술 수준을 고려해야 한다는 학계·기업 다수의견과 충돌되어 미수용" 되었다.[120] 그 결과 <인공지능 윤리>의 투명성은 '인공지능 활용 상황에 적합한 수준의 투명성과 설명 가능성'만을 의미하며, 발생할 수 있는 위험 등의 유의사항을 사전에 고지하도록 요구하는 데 그쳤다. '견고성' 원칙의 삭제는 한국의 과학기술정보통신부가 주도적으로 참여하였다고 밝힌 경제협력개발기구(OECD)의 인공지능 권고안[121]이 견고성(robustness), 보안성(security), 안전성(safety)을 동등하게 요구하고 있으며, 유럽연합의 경우 '견고성(robust)'을 <신뢰가능 인공지능 윤리 가이드라인>의 3대 속성 중 하나로 제시하였고 이후 인공지능법(안)에서도 이를 고위험 인공지능의 의

120) 제19차 4차산업혁명위원회 심의안건 제2호 (2020. 12. 23).
121) OECD (2019). OECD Principles on AI.
<https://www.oecd.org/going-digital/ai/principles. (검색일: 2021. 11. 1.)>.

무적 요구사항의 하나로 규정한 것과 차이를 보였다.

과학기술정보통신부는 <인공지능 윤리> 발표 직후 2020년 12월 23일 <인공지능 법·제도·규제 정비 로드맵>을 발표하였다. 인공지능 산업 진흥·활용 기반을 강화하고 역기능을 방지하기 위해 관계부처 합동으로 총 30개의 과제를 도출한다는 내용이다.[122] 과학기술정보통신부는 로드맵 수립에 있어서 '민간자율 우선'을 그 추진 방향의 하나로 설정하였다. 특히 알고리즘의 투명성·공정성 확보 과제에서 "기업의 알고리즘 개발이 위축되지 않도록 기업 자율적으로 알고리즘 편향성·오류를 평가·관리하는 체계를 우선 유도해나갈 계획"임을 밝혔다.

과학기술정보통신부는 2021년 5월 14일 <신뢰할 수 있는 인공지능 실현전략>을 발표하였다. 이 실현전략은 특히 「지능정보화 기본법」에 기반하여 인공지능 영향평가와 고위험 인공지능에 대한 기준 도입 방침을 밝혔다는 점이 두드러진다.

우선 국가와 지방자치단체가 국민의 생활에 파급력이 큰 지능정보서비스 등의 활용과 확산이 사회·경제·문화 및 국민의 일상생활 등에 미치는 영향에 대하여 영향평가를 할 수 있도록 한「지능정보화 기본법」제56조에 따라 인공지능이 국민생활 전반에 미치는 영향을 체계적으로 분석하여 대응하기 위해 사회적 영향평가를 도입하겠다는 계획이다. 또한 국민의 생명·신체안전 등에 밀접한 지능정보기술에 관련된 사업자는 과기정통부장관이 정하여 고시하는 기준에 적합하도록 지능정보기술을 개발·관리·활용하도록 한「지능정보화 기본법」제21조에 따라 고위험 분야 기술기준을 마련하겠다는 계획이다. 과학기술정보통신부는 2021년까지 생명·신체 등에 밀접한 의료 등 고위험 분야에서 사업자가 준수해야 할 인공지능의 안전성·신뢰성 등에 관한 기술기준을 제시할 예정이다.

「지능정보화 기본법」제56조의 경우 시행 주체가 국가와 지방자치단체로 명시되어 있으므로 그 주무기관으로는 산업진흥부처인 과학기술정보통신부보다는 인권에 미치는 영향을 객관적으로 평가하고 조치할 수 있는 규제부처가 더 적절할 것으로 보인다. 고위험 인공지능 기준 수립에 있어서도 인공지능법(안)을 발의한 유럽연합의 경우 그 주무기관으로 독립적인 시장 감독 기관과 개인정보보호 감독기구를 상정하고 있다는 점을 참고할 필요가 있다.

---

122) 인공지능 시대를 준비하는 법·제도·규제 정비 로드맵 마련. 과기정통부, 국무조정실 공동보도자료 (2020. 12. 23).

전반적으로 과학기술정보통신부가 추구해 온 인공지능 기준은 산업 성장의 가치를 가장 우선시하고 있다고 평가할 수 있다. 인공지능이 인권에 부정적인 영향을 미칠 수 있다는 사실을 인정하면서도 인공지능 환경에서 직접적인 영향을 받게 될 사람들의 인권과 안전을 구체적으로 보장하기 위하여 노력하는 모습은 찾아보기 어려웠다. 인공지능 윤리의 경우 시민사회 의견을 수렴하는 절차를 마련하였으면서도 산업 선장에 저해될 수 있는 기준이나 규제에 대한 의견을 회피하고, 기준의 이행 역시 실천력이 부족한 자율적인 규범으로 안주하는 경향을 보였다. 과학기술정보통신부는 <신뢰할 수 있는 인공지능 실현전략>에서 2025년까지 인공지능 관련 기준과 제도를 관계부처・민간과 소통하며 추진할 일정과 계획을 밝혔으나, 인권 기준 및 인권기구와의 협업에 대한 내용은 포함되어 있지 않다. 이는 과학기술정보통신부가 정보통신산업 진흥을 주무하고 인공지능산업과 데이터산업을 육성 및 지원해야 하는 소임을 맡은 부처라는 한계가 반영된 것으로 볼 수 있다.

3. 방송통신위원회

방송통신위원회는 2019년 11월 11일 <이용자 중심의 지능정보사회를 실현하기 위한 원칙>을 발표했다. 이 원칙의 주요 내용은 △사람 중심의 서비스 제공 △투명성과 설명가능성 △책임성 △안전성 △차별금지 △참여 △프라이버시와 데이터거버넌스 등이며, 유럽연합, OECD 등 국제사회와 구글, 카카오, MS, IBM 등 주요 기업에서 먼저 발표된 인공지능 윤리를 기초로 삼았다.[123]

그러나 이 원칙은 지능정보서비스가 초래할 수 있는 '피해'에 대하여 인지하면서도 그에 대한 대비체계를 제공자와 이용자가 자율적으로 수립하고 운영하도록 하는 등(안전성 원칙) 지능정보서비스 제공자와 이용자에게 동등한 책임을 부여하고, 업계 등 지능정보사회 구성원들의 '자율적인 노력'에 방점을 두었다. 이는 고위험 인공지능 제품과 서비스로부터 소비자의 안전과 기본권을 보호하기 위한 규제 도입을 예비하며 <신뢰가능 인공지능 윤리>를 마련한 유럽연합 등 타 국가기관과 다른 접근법이었으며, 소관법률에

123) 방통위,「이용자 중심의 지능정보사회를 위한 원칙」발표. 방송통신위원회 보도자료 (2019. 11. 11).

따라 이용자 보호 업무를 집행하고 부족한 제도적 공백을 보완하기 위한 시책을 마련하여야 할 국가기관의 의지와 기능에 대한 의문을 낳을 수밖에 없었다. 2021년 5월 31일 <인공지능 자율점검표>를 발표한 개인정보보호위원회의 경우 소관하는 개인정보보호법에 의거하여 법률에 따른 '의무사항'과 자율적인 준수가 요구되는 '권장사항'으로 나누어 제시함으로써 비교적 규범력을 확보하였던 바와도 대조된다. 무엇보다 '이용자 중심'의 원칙을 마련하는 과정에서 이용자의 참여가 없었다는 점은 중대한 흠결로 보인다. 방송통신위원회가 이 원칙 자문단으로 밝힌 참여자 명단은 구글, 페이스북, 카카오, 통신3사 등 사업자 13명, 학계 6명으로 구성되어 있었을 뿐, 인공지능 서비스로부터 영향을 받는 소비자나 이용자를 대표할 수 있는 참여자는 없었다. 이러한 한계를 배경으로 방송통신위원회의 <이용자 중심의 지능정보사회를 실현하기 위한 원칙>은 실제 규범적 권위 또는 효력을 발휘하지 못하였다고 평가할 수 있다.

이후 방송통신위원회는 2021년 6월 30일 보다 구체적인 <인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙>을 발표하였다. 이는 방송통신위원회가 주무하는 이용자 보호 시책의 성격을 분명하게 드러내며 '디지털 미디어 플랫폼'에서 제공되는 인공지능 기반 추천 서비스의 불투명성, 편향성으로부터 이용자의 정보접근권과 선택권을 보장하려는 취지를 담았다. 이 추천 서비스 기본원칙은 3대 '핵심 원칙'으로 투명성, 공정성, 책무성을 제시하고 5대 '실행 원칙'으로 △이용자를 위한 정보공개 △이용자의 선택권 보장 △자율검증 실행 △불만 처리 및 분쟁 해결 △내부 규칙 제정 등을 제시함으로써 서비스 제공자의 책임성을 보다 명확히 하였다. 아울러 서비스 제공자에 대한 지원과 이용자 역량 강화를 중심으로 한 정부의 역할도 명시하였다. 방송통신위원회는 이 추천 서비스 기본원칙이 시민사회와 이해관계자들의 광범위한 의견수렴을 거쳐 수립되었다고 밝혔다.

그러나 추천 서비스 기본원칙 또한 전적으로 '자율적인' 규범으로 서비스 제공자에게 제시되었다는 점에서, 그 규범적 효력에 여전히 의문이 남는다. 예를 들어 추천 서비스가 이용자에게 미치는 위험성을 상시 관리하는 '자율검증' 체계를 마련하도록 하였으나, 그 검증 대상이 되는 '위험'이나 '이용자에게 미치는 영향'의 내용이나 기준을 전혀 제시하지 않은 채 서비스 제공자가 '실행 가능한 범위에서' 운영하도록 하였을 뿐이다. 또한 서비스 제공자에게 추천 시스템 사용 및 관리에 관한 내부 규칙을 제정

하고 이용자에게 공개하도록 하였으나 이 또한 자율에 맡겼다.

방송통신위원회는 플랫폼 이용자 보호 업무의 내용을 보다 구체적으로 법률에 명시하는 「온라인플랫폼 이용자보호에 관한 법률(온라인플랫폼 이용자법)」 제정안(의안번호: 2106369)을 여당 의원 발의로 독자 추진하였으나, 정부가 발의한 「온라인 플랫폼 중개거래의 공정화에 관한 법률」 제정안에서 이 법률의 집행을 소관하도록 한 공정거래위원회와 갈등을 빚기도 하였다.[124]

4. 공정거래위원회

공정거래위원회는 인공지능 환경에서 소비자를 보호하기 위하여 법률 정비에 착수하였다.

우선 공정거래위원회는 온라인플랫폼 중개거래 투명성·공정성 제고를 취지로 「온라인 플랫폼 중개거래의 공정화에 관한 법률」 제정법률안을 2020년 9월 28일 입법예고하고 2021년 1월 28일 국회에 발의하였다(의안번호: 2107743).

디지털 경제의 가속화, 코로나19로 인한 비대면 거래의 급증으로 온라인 플랫폼의 역할과 비중이 급속히 확대된 반면 이용자 소상공인들의 피해사례가 현실화되자, 공정거래위원회는 현행 법안으로 플랫폼 기업들을 규제하는 데 어려움이 있다며 입법의 필요성을 설명하였다.[125] 법안은 온라인 플랫폼 거래 관계에서 분쟁 예방을 위해 계약서 작성·교부에 대한 의무를 부여하면서, 계약서 필수 기재 사항에 플랫폼 알고리즘에 의하여 상품이 노출되는 순서, 형태 및 기준을 포함하였다. 계약 내용 변경 시에도 해당 내용 및 사유를 이용자에게 미리 통보하도록 하였으며, 다만 알고리즘을 직접 공개하도록 한 것은 아니다.

더불어 공정거래위원회는 소관하는「전자상거래 등에서의 소비자보호에 관한 법률」에 대한 전부개정법률안을 2021년 3월 5일 입법예고하였다. 이 법안에서는 알고리즘 검

124) 온라인플랫폼법 갈등 지속에 방통위 "플랫폼 규제는 설립 근거". 연합뉴스 보도 (2021. 3. 21.); '온플 보고서 돌연 삭제'…공정위·방통위' 다툼에 등 터진 '국회 입법조사처'. 아이뉴스24 보도 (2021. 5. 14.); [이슈분석]한국 온라인플랫폼 규제, 갑론을박 지속. 전자신문 보도 (2021. 7. 6.); 방통위-공정위, 플랫폼 규제 놓고 영역다툼. 동아일보 보도 (2021. 8. 13).

125) 「온라인 플랫폼 공정화법」 제정안 국무회의 통과. 공정거래위원회 보도참고자료 (2021. 1. 25.); 6개월째 논란…'온라인플랫폼 공정화법'. 중소기업투데이 보도 (2021. 7. 8).

색 결과 및 순위 등에서 소비자의 합리적 선택을 위한 정보 제공을 강화하여, 소비자가 광고 제품을 순수한 검색 결과로 오인하여 구매하는 것을 예방하기 위해 전자상거래 사업자가 이를 구분하여 표시하도록 하였다. 또한, 조회수, 판매량, 상품 가격, 광고비 지급 여부 등 검색·노출 순위를 결정하는 주요 기준도 표시하도록 하였다. 또한 후기 게시판과 관련하여 이용후기에 대한 소비자 신뢰도 확보를 위해, 전자상거래 사업자가 이용후기의 수집·처리에 관한 정보를 공개하도록 하였다. 한편 맞춤형 광고에서 전자상거래 사업자가 개별 소비자의 기호, 연령, 소비 습관 등을 반영한 광고를 할 경우 소비자가 인기상품으로 오인하여 구매하지 않도록, 맞춤형 광고 여부를 별도 표시하고 일반 광고도 선택할 수 있도록 하였다.[126]

126) 전자상거래 등에서의 소비자보호에 관한 법률 전부개정법률(안) 입법예고, 공정거래위원회공고 제2021-14호(2021. 3. 5.)
제16조(정보의 투명성 확보 조치) ① 온라인판매사업자는 소비자에게 재화등의 거래와 관련된 검색결과를 제공할 때 광고를 구분하여 표시하여야 한다.
② 온라인판매사업자는 재화등과 관련된 검색결과에 순위를 정하여 표시하는 경우 해당 순위를 결정하는 데 이용되는 주요 결정 기준을 표시하여야 한다.
③ 제2항에 따른 주요 결정 기준 및 표시방법 등에 관하여 필요한 사항은 대통령령으로 정한다.
④ 온라인판매사업자는 거래되는 재화등에 대한 소비자의 이용후기를 게시하는 경우 사용후기의 수집, 처리에 관한 정보를 공개하여야 한다.
⑤ 제4항에 따른 구체적인 내용과 범위, 방법 및 절차 등에 관하여 필요한 사항은 대통령령으로 정한다.
제18조(소비자에 관한 정보의 이용 등) ① 사업자는 전자상거래등을 위하여 소비자에 관한 정보를 수집하거나 이용(제3자에게 제공하는 경우를 포함한다. 이하 같다)할 때는 「개인정보보호법」 등 관계 규정에 따라 이를 공정하게 수집하거나 이용하여야 한다.
② 사업자는 재화등을 거래함에 있어서 소비자에 관한 정보가 도용되어 해당 소비자에게 재산상의 손해가 발생하였거나 발생할 우려가 있는 특별한 사유가 있는 경우에는 본인 확인이나 피해의 회복 등 대통령령으로 정하는 필요한 조치를 취하여야 한다.
③ 온라인판매사업자는 소비자의 기호, 연령, 성별, 소비습관, 구매내역 등의 특징에 따라 소비자에게 상품이나 서비스의 검색결과를 제공하거나 재화 등을 추천(이하 “맞춤형 광고”라 한다)하는 경우에는 그 내용과 방법을 사전에 소비자에게 고지하고, 맞춤형 광고의 수신 여부를 소비자가 선택할 수 있도록 하여야 한다.
④ 온라인판매사업자는 제3항에 따라 소비자가 맞춤형 광고 대신 해당 소비자의 특성이 반영되지 않은 일반적인 검색결과 또는 광고의 수신을 선택하는 경우에는 해당 정보를 소비자에게 제공하여야 한다.
⑤ 공정거래위원회는 제3항 및 제4항에 따른 맞춤형 광고의 고지의 내용과 방법, 맞춤형 광고에 대해 소비자가 선택할 수 있도록 하는 방법을 정하여 고시할 수 있다.

5. 개인정보보호위원회

2021년 4월 28일 개인정보보호위원회는 인공지능 챗봇 ‘이루다’ 개발사 (주)스캐터랩에 대하여 총 1억 330만원의 과징금과 과태료 등을 부과했다. 이 사건은 국가 개인정보보호 감독기구인 개인정보보호위원회가 인공지능 학습 및 서비스 데이터셋에 대하여 직권으로 조사하고 제재한 첫 사례로서, 인공지능 개발과 활용에 있어 올바른 개인정보 보호 처리 방향을 제시하였다.[127]

스캐터랩은 2013년 출시한 ‘텍스트앳’과 2016년 출시한 ‘연애의과학’의 회원 개인정보를 이용하여 인공지능 챗봇 이루다를 학습시키고 출시하였다. 카카오톡 대화 메시지에 대한 감정분석서비스를 내세운 ‘텍스트앳’이 수집한 데이터는 이메일 등 로그인 아이디, 닉네임, 성별, 직업(초중고/대학생/일반인) 등 회원정보와 카카오톡 대화문장 원본과 대화 상대방의 닉네임 및 그 관계(친구, 연인, 배우자, 소개팅상대, 아는 사람 중 하나)이었으며, 스캐터랩은 이용자별로 회원번호를 자체부여하여 관리하였다. 카카오톡 대화 메시지에 대한 연애 심리 검사를 내세운 ‘연애의 과학’이 수집한 데이터는 이메일/카카오톡/페이스북/애플 식별자 중 하나로 지정된 로그인아이디, 닉네임, 성별, 출생연도 등 회원정보와 카카오톡 대화문장 원본과 대화 상대방의 닉네임 및 그 관계(친구, 연인, 배우자, 소개팅상대, 아는 사람 중 하나)이었으며, 스캐터랩은 이용자별로 회원번호를 자체부여하여 관리하였다.

회사가 ‘텍스트앳’과 ‘연애의과학’에서 수집한 회원정보는 회원을 탈퇴하거나 1년 이상 서비스를 이용하지 않았어도 파기되지 않고 가공되어 2020년 12월 출시된 인공지능 챗봇 ‘이루다’ 서비스에 이용되었다. ‘이루다’는 수집된 회원정보를 이용하여 인공지능 학습을 위한 ‘학습 DB’와 이루다 서비스를 위한 ‘응답 DB’를 구축하였다. ‘학습 DB’에는 회원 60만 명의 회원정보와 회원번호와 더불어 회원들의 카카오톡 대화문장 94억 건이 포함되었다. 회원정보의 경우 로그인아이디, 닉네임은 포함되지 않았으나 성별, 나이, 대화 상대방과의 관계가 포함되었으며, 회원번호도 일방향암호화(SHA-256 해쉬함수)되어 회사가 알수 있는 형태로 포함되었다. 카카오톡 대화문장의 경

127) 개인정보위, '이루다' 개발사 (주)스캐터랩에 과징금·과태료 등 제재 처분. 개인정보보호위원회 보도자료 (2021. 4. 29.); 개인정보보호위원회 2021. 4. 28. 결정 제2021-007-072호 심의·의결서.

우 본인의 프로필은 ‘SEND’, 대화상대방의 프로필은 ‘RECV’로 치환되었으나, 대화 문장은 변경 없이 원문 그대로 포함되어 학습에 사용되었다. ‘응답 DB’는 ‘학습 DB’에 저장된 카카오톡 대화문장 약 94억건 중 20대 여성이 발화한 대화문장을 추출한 후 대화문장에서 실명, 장소명, 숫자/영문, 선정적 표현이라고 보여지는 단어 등이 포함된 대화문장을 반복하여 제거하는 방식으로 구축되었다. 그밖에 회사는 1,431건의 대화문장을 포함하는 100건의 테스트 샘플을 프로그램 소스코드 공유플랫폼인 Github에 공개로 게시하였다. Github에 공개된 데이터셋의 경우 이름을 <NAME>으로 숫자를 <NUM>으로 치환한 대화문장 원본이 포함되었고, 회원정보의 경우 성별, 대화 상대방과의 관계(friend, lover) 및 직업(student, collegian, civilian)이 포함되어 있었다.

인공지능 학습이나 서비스에 사용되는 데이터셋이 개인정보를 사용하여 구축된 경우에는 개인정보보호법을 준수하여야 한다. 그러나 스캐터랩은 현행 개인정보보호법 다수 조항을 위반하였으며, 이를 쟁점별로 요약해 보면 다음과 같다. 첫째, 회사는 개인정보를 수집하면서 정보주체에게 명확하게 인지할 수 있도록 알리고 동의를 받지 않았다. 텍스트앳과 연애의과학은 개인정보를 수집하면서 정보주체의 동의를 받을 때 수집·이용 동의, 제3자 제공 동의 등 각각의 동의 사항을 구분하여 정보주체에게 명확히 알리고 각각 동의를 받아야 하는데도, 서비스 초기화면에서 “로그인함으로써 이용약관 및 개인정보처리방침에 동의합니다”라는 안내에 그치는 등, 정보주체가 각각의 동의 사항을 명확하게 인지할 수 있도록 알리거나 각각 동의받지 않았다(법 제22조제1항 위반). 둘째, 개인정보처리방침은 이처럼 이용자들에게 명확히 알리지 않은 채 수집된 개인정보를 “신규 서비스 개발 및 마케팅·광고에의 활용”하겠다는 내용을 포함하고 있었다. 개인정보보호위원회는 “개인정보처리방침에 ‘신규 서비스 개발’이 명시되어 있다는 이유만으로, 이용자가 ‘이루다’와 같은 기존 서비스와 전혀 다른 신규 서비스의 개발과 서비스 운영에 자신의 개인정보가 이용될 것을 예상하고 이에 동의하였다고 보기 어렵다.”고 보았다. 이는 개인정보를 서비스에 필수적이지 않은 마케팅에 이용할 때 이용자가 선택할 수 있도록 필수 동의와 분리하여 선택 동의를 받도록 한 개인정보보호법을 위반한 것이며, 이용자로부터 동의 받은 목적 내에서 개인정보를 이용한 것이라고 할 수 없다(법 제18조제1항 위반). 셋째, 텍스트앳, 연애의과학, 이루다 모두 법정대리인의 동의 없이 만 14세 미만 아동의 개인정보를 수집하였다(법 제22조제6항 위반). 넷째, 민감정보를 처리

하면서 별도 동의를 받지 않았다. 법정 민감정보는 ①사상・신념, ②노동조합・정당의 가입・탈퇴, ③정치적 견해, ④장애 등 건강상태, ⑤성생활 등에 관한 정보, ⑥유전정보, ⑦범죄경력자료에 해당하는 정보, ⑧생체인식 정보(특징정보), ⑨인종이나 민족에 관한 정보를 의미한다(법 제23조제1항 위반). 다섯째, 탈퇴한 회원의 개인정보를 파기하지 않았다(법 제21조제1항 위반). 여섯째, 1년 이상 서비스 미사용자의 개인정보를 파기하거나 분리・보관하지 않았다(법 제39의6 위반).

한편, 개인정보보호위원회는 이루다 사건에서 2020년 8월 5일 개정시행된 가명조항의 적용을 해석하였다. 우선 가명정보란, 시간・비용・기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 익명정보(법 제58조의2)와 달리, 개인정보의 한 유형으로서 기존의 개인정보에 대하여 그 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 가명처리한 것이다(법 제2조 제1의2호). 즉, 개인정보처리자는 원래의 상태로 복원하기 위한 추가정보의 사용・결합이 가능한데, 이러한 추가정보가 없으면 특정 개인을 알아볼수 없도록 가명처리를 안전하게 수행한 가명정보(법 제2조 제1호다목)의 경우 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 처리할 수 있도록 허용되어 있다(법 제28조의2). 가명처리에 대하여는 정보주체의 헌법상 기본권인 동의권이 제한되는 것이고, 가명처리를 마친 가명정보의 경우 개인정보 수집출처 고지권, 파기권, 이전 제한권, 유출 통지권, 열람권, 정정・삭제권, 처리정지권 등 일체의 정보주체 권리가 제한된다(법 제28조의7). 따라서 광범위한 오남용이 우려되는 가명처리의 안전성이나 정보주체의 인지나 동의 없이 가명처리가 허용되는 과학적 연구의 범위가 그간 쟁점이 되어 왔다.

이루다 사건에서는 가명정보가 이루다 개발과 서비스에 이용되는 한편, Github에도 공개되었다. 스캐터랩은 이루다 개발과 서비스가 가명정보의 과학적 연구라고 주장하는 한편, Github에 공개된 카카오톡 대화내용의 경우에는 특정 개인을 알아볼 수 없으므로 개인정보가 아니라고 주장하였다. 개인정보보호위원회는 ‘학습 DB’의 경우에는 94억 건의 카카오톡 대화내용에 가명처리를 전혀 하지 않았고, ‘응답 DB’의 경우 일반 이용자에게 그대로 발화되도록 서비스하였으므로 과학적인 연구를 위한 것이라 할 수 없다고 보았다. 또한 Github에 공개된 카카오톡 대화내용의 경우 가명정보에 해당하지만 가명정

보를 복원할 수도 있는 불특정 다수에게 가명정보를 공개하는 행위는 허용되지 않는다고 보았다(법 제28조의2 제2항 위반).

개인정보보호위원회는 이루다 사건 이후 2021년 5월 31일 <인공지능(AI) 자율점검표>를 발표하고 인공지능의 개발·운영에 참여하는 자의 개인정보보호에 대한 인식을 제고하고 개인정보보호법 준수를 요구하였다.[128] 특히 이 점검표는 인공지능 개인정보보호의 6대 원칙으로 적법성, 안전성, 투명성, 참여성, 책임성, 공정성을 제시하고 현행 개인정보보호법에 따라 단계별 또는 상시로 준수해야 할 법령상 의무 또는 권장하는 내용에 대한 점검항목 16개와 확인사항 54개로 구성되어 있다.

점검표는 인공지능의 개발과 활용에 대한 수명주기를 ① 인공지능의 기획·설계, ② 개인정보 수집, ③ 개인정보 이용·제공, ④ 개인정보 보관·파기, ⑤ 인공지능 서비스 관리·감독, ⑥ 인공지능 서비스 이용자 보호 및 피해구제, ⑦ 개인정보 자율보호 활동, ⑧ 인공지능 윤리 점검 단계로 구분하고 각 단계에서 의무적으로 준수해야 할 개인정보보호법상의 기준을 제시하였다. 특히 검검표는 우리 법률에 아직 도입되어 있지 않았으나 국제적으로 통용되는 개인정보보호 중심설계 원칙 등은 권장사항으로 반영하였다. 인공지능 서비스 특성상 예상치 못한 개인정보 침해가 발생할 수 있으므로 기획 단계부터 사전 점검과 예방을 위해 개인정보보호 중심 설계(PbD) 원칙을 적용하고, 침해가 우려되는 경우 개인정보 영향평가를 수행하도록 한 것이다.

더불어 개인정보보호위원회는 개인정보보호법의 개정을 통하여 완전 자동화 의사결정에서 정보주체의 권리를 보장하기 위한 방안을 추진하고 있다. 개인정보보호위원회가 2021년 9월 28일 국회에 발의한 개인정보보호법 개정법률안(의안번호: 2112723)은 '자동화된 결정에 대한 정보주체의 권리 등'에 대한 조항을 신설하고 정보주체가 완전 자동화 의사결정을 거부하거나 설명 등을 요구할 수 있도록 하였다(안 제37조의2).[129] 다만

---

128) 개인정보위, 인공지능(AI) 자율점검표 발표. 개인정보보호위원회 보도자료 (2021. 5. 31.); 개인정보보호위원회 (2020). 인공지능(AI) 개인정보보호 자율점검표.

129) 제37조의2(자동화된 결정에 대한 정보주체의 권리 등) ① 정보주체는 완전히 자동화된 시스템(인공지능 기술을 적용한 시스템을 포함한다)으로 개인정보를 처리하여 이루어지는 결정이 자신의 권리 또는 의무에 중대한 영향을 미치는 경우에는 해당 개인정보처리자에 대하여 해당 결정을 거부하거나 해당 결정에 대한 설명 등을 요구할 수 있다. 다만, 자동화된 결정에 대한 거부는 개인정보가 제15조제1항제3호 또는 제5호부터 제7호까지의 규정에 따라 처리되는 경우에만 할 수 있다.
② 개인정보처리자는 제1항에 따라 정보주체가 자동화된 결정을 거부하거나 이에 대한 설명 등을 요구한 경우에는 정당한 사유가 없는 한 그에 따라야 한다.

이 조항은 정보주체가 사전적으로 프로파일링은 물론 자동화 의사결정의 로직 및 그 결과에 대한 정보를 통지받고 자신의 의견을 피력하거나 인적 개입을 요구할 수 있는 권리를 보장하고 있지 않다는 점에서 다소 한계가 있다. 국가인권위원회는 입법예고안에 대한 의견에서 정보주체에게 완전 자동화 의사결정을 받지 않을 권리가 있음을 원칙적으로 규정하고, 자동화 의사결정을 바로 정보주체에 적용하기보다 예외적으로 합리적이고 정당한 범위 내에서만 허용되는 것이 바람직하다고 보았다. 또한 자동화 의사결정에 의해 생성된 민감정보를 처리할 때 '중대한 공익상의 목적을 위해 법률이 허용하는 경우'와 '정보주체의 명백한 동의' 등 더 엄격한 조건을 규정할 필요가 있다고 지적하였다.[130]

6. 금융위원회

금융위원회는 2021년 7월 8일 <금융분야 인공지능(AI) 가이드라인>의 시행을 발표하였다.[131] 이 가이드라인은 금융회사가 인공지능 기반 금융서비스를 개발 및 활용할 때 준수하여야 할 기준과 절차를 제시하였다. 또한 2020년 2월 4일 개정된 「신용정보의 이용 및 보호에 관한 법률」(이하 '신용정보법')에서 신설된 '자동화평가 결과에 대한 설명 및 이의제기 등(제36조의2)' 조항[132]을 금융 실무에 적용할 수 있도록 "고객에 대

---

③ 개인정보처리자는 자동화된 결정의 기준과 절차를 정보주체가 쉽게 확인할 수 있도록 공개하는 등 필요한 조치를 하여야 한다.
④ 제1항부터 제3항까지에서 규정한 사항 외에 자동화된 결정의 기준·절차의 공개 등에 필요한 사항은 대통령령으로 정한다.

130) 국가인권위원회 2021. 4. 26. 결정. 「개인정보보호법 일부개정법률안」 의견 요청에 대한 회신.

131) 「금융분야 인공지능(AI) 가이드라인」이 시행됩니다 : 금융권 AI 활용을 활성화하고 AI 기반 금융서비스에 대한 신뢰를 제고하기 위한 모범규준 마련·발표. 금융위원회 보도자료 (2021. 7. 8).

132) 신용정보법 제36조의2(자동화평가 결과에 대한 설명 및 이의제기 등) ① 개인인 신용정보주체는 개인신용평가회사 및 대통령령으로 정하는 신용정보제공·이용자(이하 이 조에서 "개인신용평가회사등"이라 한다)에 대하여 다음 각 호의 사항을 설명하여 줄 것을 요구할 수 있다.
1. 다음 각 목의 행위에 자동화평가를 하는지 여부
가. 개인신용평가
나. 대통령령으로 정하는 금융거래의 설정 및 유지 여부, 내용의 결정(대통령령으로 정하는 신용정보제공·이용자에 한정한다)
다. 그 밖에 컴퓨터 등 정보처리장치로만 처리하면 개인신용정보 보호를 저해할 우려가 있는 경우로서 대통령령으로 정하는 행위

한 설명의무가 있는 금융서비스 등에 AI 시스템을 활용하는 경우 또는 고위험 서비스에 AI 시스템을 활용하는 경우 설명가능 인공지능 기술 등 적절한 인공지능 기술을 투명하게 적용하여 맥락에 맞는 설명이 도출되는지 여부를 확인하고, AI 시스템의 안정성·신뢰성 등을 훼손하지 않는 범위 내에서 설명가능성을 합리적인 수준으로 개선하기 위해 노력" 할 것 등을 규정하였다.

가이드라인의 주요 내용으로는 첫째, 인공지능 윤리원칙-인공지능 전담조직-위험관리 정책 수립의 3중 내부통제장치 마련하고, 둘째, 인공지능 학습 데이터에 대한 조사·검증 강화로 개인신용정보 오·남용을 방지하며, 셋째, 불합리한 소비자 차별 등이 없도록 시스템 위험관리 및 공정성을 제고하고, 넷째, 소비자에 인공지능 서비스에 대한 충분한 설명 및 권리행사를 보장한다는 것이다.

우선 이 가이드라인은 인공지능을 금융거래 및 대고객서비스에 적용한 전 금융업권을 대상으로 한다. 다만, 비금융업이라도 인공지능 활용의 결과가 금융거래에 미치는 영향이 큰 경우(예 : 신용평가회사의 개인신용평점 개발) 확대 적용하겠다는 계획이다. 가이드라인은 3중 내부 통제장치 중 인공지능 윤리의 경우 금융회사 자체적으로 수립하도록 하였다. 금융회사는 고객군, 서비스 내용 등 인공지능 활용 상황에 따라 인공지능 서비스 개발·운영시 준수해야할 원칙·기준을 마련한다. 또한 금융회사에 인공지능 전담조

---

2. 자동화평가를 하는 경우 다음 각 목의 사항
가. 자동화평가의 결과
나. 자동화평가의 주요 기준
다. 자동화평가에 이용된 기초정보(이하 이 조에서 "기초정보"라 한다)의 개요
라. 그 밖에 가목부터 다목까지의 규정에서 정한 사항과 유사한 사항으로서 대통령령으로 정하는 사항
② 개인인 신용정보주체는 개인신용평가회사등에 대하여 다음 각 호의 행위를 할 수 있다.
1. 해당 신용정보주체에게 자동화평가 결과의 산출에 유리하다고 판단되는 정보의 제출
2. 자동화평가에 이용된 기초정보의 내용이 정확하지 아니하거나 최신의 정보가 아니라고 판단되는 경우 다음 각 목의 어느 하나에 해당하는 행위
가. 기초정보를 정정하거나 삭제할 것을 요구하는 행위
나. 자동화평가 결과를 다시 산출할 것을 요구하는 행위
③ 개인신용평가회사등은 다음 각 호의 어느 하나에 해당하는 경우에는 제1항 및 제2항에 따른 개인인 신용정보주체의 요구를 거절할 수 있다.
1. 이 법 또는 다른 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우
2. 해당 신용정보주체의 요구에 따르게 되면 금융거래 등 상거래관계의 설정 및 유지 등이 곤란한 경우
3. 그 밖에 제1호 및 제2호에서 정한 경우와 유사한 경우로서 대통령령으로 정하는 경우
④ 제1항 및 제2항에 따른 요구의 절차 및 방법, 제3항의 거절의 통지 및 그 밖에 필요한 사항은 대통령령으로 정한다.

직을 두어 인공지능의 잠재적 위험을 평가·관리할 구성원의 역할·책임·권한을 기획·설계·운영·모니터링 등 서비스 전단계에 걸쳐 구체적으로 정의하도록 하였다. 위험관리정책 수립의 경우에도 금융회사 자체적으로 마련한다. 이때 개인권리에 중대한 위험을 초래할 수 있는 서비스에는 강화된 위험관리를 적용하도록 하였다. 특히 인공지능 의사결정이 신용평가, 대출심사, 보험심사, 카드발급 심사 등 개인의 금융거래계약의 체결·유지에 중대한 영향을 가져오는 경우, 이에 대한 내부통제·승인절차 등을 마련하고 별도의 책임자를 지정하도록 하였다. 또한 통제가능성을 확보하기 위하여 인공지능이 사람의 의사결정과정을 대체하는 경우, 필요시 사람에 의한 감독·통제가 가능하도록 시스템을 설계하도록 하였다. 이자율 산정 등 금융관련 중요 의사결정을 대체한 인공지능에 대해 결과의 설명 및 사후검증을 할 수 없다면 금융거래 투명성 및 신뢰성 저하가 발생할 우려가 있다는 것이다. 더불어 모니터링 절차를 두어 인공지능 시스템 성능을 주기적으로 모니터링하고, 인공지능 개발 환경의 보안취약성 점검 시스템 마련 등 위험완화 조치를 시행하도록 하였다.

다음으로 인공지능 학습 데이터 관리와 관련하여, 인공지능 학습에 사용되는 데이터의 출처·품질·편향성·최신성 등을 조사·검증하고 개선노력을 지속하도록 요구하였다. 인공지능 시스템의 성능은 학습데이터에 좌우되는데, 예를 들어 인공지능 챗봇 '이루다'의 사례에서처럼 학습 데이터에 오류·불균형·편향성이 존재할 경우, 인공지능 시스템이 잘못된 학습을 통해 여성·장애인·동성애 등에 대한 다양한 차별·혐오발언을 생산하는 등 오류·불균형·편향성을 확대 재생산할 우려가 있다. 따라서 금융회사는 인공지능 학습 데이터의 품질을 지속조사하고 개선하는 시스템을 구축할 필요가 있다. 즉, 데이터 출처 파악, 규격 문서화, 품질 검증(누락, 중복, 불일치 등), 편향되지 않은 충분한 데이터 확보, 무결성 검증, 주기적 갱신 등의 조치를 취하는 체계를 마련하여야 한다. 금융회사는 이와 같은 체계를 통하여 인공지능 시스템의 오류·불균형·편향성 발생 가능성을 시스템적으로 통제하고, 인공지능 결과에 대한 객관성·신뢰성을 제고할 수 있을 것이다. 한편, 인공지능 개발 과정에서 개인신용정보 오·남용을 방지하고, 불필요한 개인신용정보 처리를 최소화하는 시스템 또한 마련하여야 한다. 특히, 사생활정보 등 민감정보를 활용하고자 하는 경우, 비식별 조치 등 안전한 정보 활용을 위한 충분한 조치를 거친 뒤, 해당 정보 미사용시 인공지능 시스템에 성능 저하가 발생하는지, 해당 정보 사

용시 효과와 혜택을 보는 집단은 누구인지 등 정보활용 필요성을 면밀히 평가하고 재식별·유출 등을 방지하여야 한다.

소비자 차별을 방지하는 위험관리 및 공정성과 관련하여서는, 금융회사에 인공지능 시스템 운영과정에서 나타날 수 있는 다양한 위험요인을 서비스 특성에 맞게 합목적적으로 통제하는 성능평가를 요구하였다. 또한 금융회사는 공정성을 제고하여 인공지능에 따른 집단간 차별 등 기본권 침해행위가 발생하지 않도록 서비스 특성별 공정성 기준을 설정·평가하여야 한다. 금융소외계층에 대한 금융접근성 제고를 목적으로 한 정책금융상품 등 결과적 평등이 중요한 금융상품의 경우, 집단간 대출 승인율 등 인구통계적 동등성 지표를 기준으로 공정성을 평가하여야 한다. 또한 소속 집단 등과 무관하게 자격이 있는 소비자를 판단해야 하는 신용평가, 카드발급심사 등의 경우, 집단간 재현율 등 기회의 균등 기준으로 공정성하게 서비스가 개발되었는지 평가하여야 한다.

금융소비자에 대한 설명 및 권리행사와 관련하여서는, 금융소비자에게 인공지능이 활용된다는 사실을 사전고지하고, 소비자의 권리 및 이의신청·민원제기 등 권리구제 방안 등을 알기쉽게 안내하도록 하였다. 특히, 인공지능을 통해 신용평가, 보험가입 등 금융거래·계약체결 여부 결정 등을 한 경우 '설명요구·정정요구권'이 있음을 고지하도록 하였다. 소비자는 인공지능 평가결과 및 주요 평가기준, 사용된 기초정보에 대한 설명을 들을 수 있고, 인공지능 평가에 활용된 자신에 대한 정보의 정정·삭제요구, 결과 재산출 등을 요구할 수 있다.

금융위원회는 준비기간을 거쳐 연내 가이드라인을 시행하고, 금융업권 및 기능·서비스별 특성을 고려하여 가이드라인을 구체화한 세부 실무지침도 마련할 계획이라고 밝혔다.

전반적으로 이 가이드라인은 금융회사의 자율적 조치에 의존하고 있다. 국회입법조사처는 이 가이드라인의 운영원칙과 점검기준이 모호하다는 점을 지적하였다. 금융회사 등의 자율적 준수를 유도하기 위해서는 「금융소비자 보호에 관한 법률」에서 규정하고 있는 금융소비자의 기본적 권리와 영업규제 등과 연계하여 인공지능 관련 금융소비자 보호 원칙과 평가 목록을 구체적으로 마련할 필요가 있다는 것이다.[133]

133) 이수환·박소영 (2021). 금융분야 AI 가이드라인 도입 추진과 시사점. 국회입법조사처 <이슈와 논점> 제1878호 (2021. 10. 12).

무엇보다 신용정보법은 인공지능 개발과 활용에서 공정성, 책무성, 투명성 등 데이터 윤리의 구현을 구체적으로 규율하고 있는 사항이 없고, 자동화평가 결과에 대하여 동법 제36조의2에서 규정하고 있는 개인신용정보주체의 설명권 및 이의제기권 보장에 있어서도 설명하지 아니한 자에게 3천만원 이하의 과태료를 부과하고 있을 뿐이다. 반면, 유럽연합의 인공지능법안의 경우 개인의 신용도를 평가하거나 신용 점수를 설정하기 위해 사용하는 인공지능 시스템을 고위험으로 분류하여 법적 의무를 부과하고자 한다. 금융 소비자를 보호하기 위하여는 정보주체의 기본권과 이익에 중대한 위험을 야기할 수 있는 인공지능을 규제하고 위반시 제재하도록 법제도적 개선이 필요해 보인다.

7. 서울특별시교육청

2021년 9월 서울특별시교육청은 <인공지능(AI) 공공성 확보를 위한 현장 가이드라인>을 발표하였다. 교육청은 학교에서 인공지능(AI)을 활용한 교육이 활발해짐에 따라 전국 최초로 학생의 개인정보보호와 데이터 처리 과정의 투명성 및 안전성을 확보하기 위해 현장 가이드라인을 개발하였다고 밝혔다.[134] 특히 가이드라인은 학교에서 인공지능을 도입할 때 '인공지능(AI) 등급 평가 매트릭스'와 '인공지능(AI) 영향평가 체크리스트'를 사용해 인공지능에 기반한 결정의 영향을 평가하도록 하였다. 이들 매트릭스와 체크리스트 등 인공지능 영향평가 방안은 인공지능에 대한 해외 위험기반 접근법을 참고하여 개발되었다.[135]

이 가이드라인은 공교육에서 인공지능을 적용할지 여부를 결정하는 기준의 하나로 개인정보보호를 들었다. 즉 공교육에서 인공지능을 도입하는 주체는 인공지능이 학습하거나 수집/이용하는 개인정보가 윤리와 인권의 관점에서 보호되어야 하고, 그 적용 결과를 평가하고 필요한 조치를 취하기 위한 절차와 체계를 만들어야 한다. 이를 위하여 인공지능 등급 평가 매트릭스의 기준으로 '의사결정 영향 정도'와 함께 '개인정보 민감 정도'를 제시하고 인공지능 영향 등급을 평가하도록 하였다. 이때 개인정보 민감 정도는

134) 서울시특별시교육청, 인공지능(AI) 공공성 확보를 위한 현장 가이드라인 공청회 개최. 서울특별시교육청 보도자료 (2021. 7. 30.); 서울특별시교육청 (2021). 인공지능(AI) 공공성 확보를 위한 현장 가이드라인.

135) 김기중, 임완철, 장여경 (2021). 공교육에 적용되는 인공지능 알고리즘의 공공성 확보방안 연구. 서울특별시교육청 2020 위탁연구 보고서.

특정 개인을 식별할 수 없거나 매우 어려운 정보는 민감도가 ‘낮음’으로, 개인을 쉽게 식별할 수 있는 정보는 민감도가 ‘중간’으로, 법령상 별도 동의가 필요하거나 사생활 침해 우려가 있는 매우 민감한 개인정보는 민감도가 ‘높음’으로 구분하였다. 매트릭스상으로 영향 정도가 높아지는 수준에 따라 단위학교에 도입되는 인공지능을 바로 사용 가능하거나(4등급), 교내 AI위원회(3등급) 또는 외부위원을 포함하는 학교 AI위원회(2등급)에서 심의 및 조치를 거쳐 사용하거나, 교육청 AI위원회(1등급) 심의를 거쳐 채택 또는 조치하도록 하였다.

한편, 가이드라인은 인공지능 등급 평가 후 인공지능의 도입 가능성에 대한 점검 목적으로 인공지능 영향평가 체크리스트를 제공하였다. 체크리스트는 교내 또는 학교 인공지능위원회에서 심의 시 업체 담당자로부터 체크리스트 항목별 내용을 확인받고, 부적합한 부분이 있는 경우 업체와 조율을 통해 최종 도입 여부를 결정하기 위한 목적으로 사용된다. 개인정보보호에 대한 체크리스트는 개인정보보호위원회의 <인공지능(AI) 개인정보보호 자율점검표>를 준수하도록 하였다.

제2절 국내 제도

1. 인공지능 관련 법령

‘인공지능’이라는 용어는 다양한 법령에서 사용하고 있다. 예를 들면, 「행정기본법」은 “행정청은 법률로 정하는 바에 따라 완전히 자동화된 시스템(인공지능 기술을 적용한 시스템을 포함한다)으로 처분을 할 수 있다.”(제20조)는 규정을 두고 있고, 「바둑진흥법」은 “전통놀이, 인공지능, 과학기술 등 다양한 문화콘텐츠와의 융합·연계를 통하여 발전”이라고 규정하고 있으며(제12조), 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」은 “정보통신망을 통하여 유통되는 정보 중 인공지능 기술을 이용하여 만든 거짓의 음향·화상 또는 영상 등의 정보를 식별하는 기술의 개발·보급”이라는 규정(제4조 제2항 제7의2호)을 두고 있으며, 중앙행정기관 등의 각종 직제 규정에 인공지능 관련 직제를 두거나 인공지능을 기존 직제의 업무로 두는 내용이 있다.[136] 하지만, ‘인공지능’의 개념을 정의하거나 인공지능에 대한 대응방안을 규정한 법령 규정은 없다.

다만, 「지능정보화 기본법」은 “전자적 방법으로 학습·추론·판단 등을 구현하는 기술”을 ‘지능정보기술’의 하나로 정의(제2조 제4호 가목)하고 ‘지능정보기술’에 대한 진흥과 규제 정책을 규정하고 있고, 이 ‘지능정보기술’의 개념은 일반적으로 이해되는 ‘인공지능’의 개념과 가장 유사하며, 실제로 위 개념은 인공지능 기술을 염두에 두고 입법된 것[137]이므로, 현재의 인공지능에 대한 일반법은 과학기술정보통신부 소관 법률인 「지능정보화 기본법」으로 볼 수 있다.

「지능정보화 기본법」은 ‘지능정보사회 기본원칙’으로 다음과 같이 국가와 지방자치단체에게 인간의 존엄·가치를 보장하고 차별을 방지할 책무를 부여하면서 동시에 인간의 존엄·가치 보장, 지능정보기술의 역기능 방지, 개인정보의 보호/사생활 비밀·자유

136) 개인정보보호위원회 직제, 과학기술정보통신부와 그 소속기관 직제, 국토교통부와 그 소속기관 직제, 4차산업혁명위원회의 설치 및 운영에 관한 규정 등을 들 수 있다.

137) “최근 우리 사회는 인공지능, 데이터, 5G 등 첨단기술의 혁신적 발전으로 초연결·초지능 기반의 4차 산업혁명 패러다임에 접어들고 있는바, 4차 산업혁명에 따른 사회·경제적 변화에 선제적으로 대응하기 위한 범국가적 추진체계 구축과 기술혁신을 위한 규제체계 정비가 필요함. … 데이터·인공지능 등 핵심기술 기반과 산업생태계를 강화 … ”.「지능정보화 기본법」2020. 6. 9.자 ‘전부개정이유’ 참조.

보장의 책무는 “사회의 구성원 전체”에 부여하는 안전장치를 마련하였다는 특징이 있다(제3조). 「지능정보화 기본법」은 이어 제4조에서 국가, 지방자치단체 및 공공기관에게 “지능정보기술을 개발·활용하거나 지능정보서비스를 제공·이용할 때 안전성·신뢰성 및 공정성 확보를 위하여 노력할 의무”와 “지능정보화로 발생·심화될 수 있는 불평등을 해소하고 노동환경 변화에 대하여 적극적으로 대응하기 위하여 노력할 의무”를 부과하였다.

지능정보사회 기본원칙 (「지능정보화 기본법」 제3조)

1. 국가와 지방자치단체 및 국민 등 사회의 구성원은 인간의 존엄·가치를 바탕으로 자유롭고 개방적인 지능정보사회를 실현하고 이를 지속적으로 발전시킨다.

2. 국가와 지방자치단체는 지능정보사회 구현을 통하여 국가경제의 발전을 도모하고, 국민생활의 질적 향상과 복리 증진을 추구함으로써 경제 성장의 혜택과 기회가 폭넓게 공유되도록 노력한다.

3. 국가, 지방자치단체와 국민 등 사회의 모든 구성원은 지능정보기술을 개발·활용하거나 지능정보서비스를 이용할 대 역기능을 방지하고 국민의 안전과 개인정보의 보호, 사생활의 자유·비밀을 보장한다.

4. 국가와 지방자치단체는 지능정보기술을 활용하거나 지능정보서비스를 이용할 때 사회의 모든 구성원에게 공정한 기회가 주지도록 노력한다.

5. 국가와 지방자치단체는 지능정보사회 구현시책의 추진 과정에서 민간과의 협력을 강화하고, 민간의 자유와 창의를 존중하고 지원한다.

6. 국가와 지방자치단체는 지능정보기술의 개발·활동이 인류의 공동발전에 이바지할 수 있도록 국제협력을 적극적으로 추진한다.

「지능정보화 기본법」은 대부분의 규정에서 부문별 추진계획의 수립, 이를 종합한 종합진흥계획의 수립, 지능정보기술의 개발, 보급, 표준화, 인력양성 등의 진흥정책에 할애하고 있고, 인공지능 개발 주체에게 안전성, 신뢰성, 공정성을 확보하도록 하고 불평등 해소와 노동환경 변화에 대한 대응 방안을 주문하고 있기는 하나, 그러한 부담을 ‘노력할 의무’에 그치므로 선언적 성격의 의미를 벗어나기는 어렵다. 물론 「지능정보화 기

본법」은 “대통령령으로 정하는[138] 국민의 생명 또는 신체안전 등에 밀접한 지능정보기술에 관련된 사업자는 과학기술정보통신부장관이 정하여 고시하는 기준에 적합하도록 지능정보기술을 개발·관리·활용하여야 한다.”는 원칙 규정이 포함되어 있고(제21조 제2항), 위 고시 기준에 적합하지 않은 기술을 개발하거나 활용한 자에 대해서는 500만원 이하의 과태료를 부과할 수 있도록 하였으나(제70조 제2항), 위 규정은 “생명 또는 신체안전”에 위협 또는 문제를 일으킨 경우에 적용되는 것이 아니라 단지 과학기술정보통신부장관이 정하는 ‘고시 기준’에 위반되는 경우에 한하여 적용되며 그 위반의 효과도 낮은 금액의 과태료 처분에 불과하여 실효성이 있다고 할 수는 없다.[139]

2021년 3월 23일 개정 「전자정부법」은 행정기관이 인공지능을 활용하여 전자정부서비스를 제공할 수 있도록 하였다(제18조의2).[140] 그러나 이 규정 또한 행정부문에서 인공지능의 활용에 대한 근거만을 두었을 뿐, 그 활용과 결과물의 적용을 제한하거나 국민의 권리와 이익을 보호 또는 구제하기 위한 조치들은 별도로 규정하지 않았다.

---

138) 시행령 제16조 제2항
법 제21조 제2항에서 “대통령령으로 정하는 국민의 생명 또는 신체안전 등에 밀접한 지능정보기술”이란 지능정보기술을 이용하는 사람의 생명·신체를 보호하는데 현저한 지장을 줄 우려가 있는 지능정보기술로서 다음 각 호의 어느 하나에 해당하는 지능정보기술을 말한다.
1. 군사적 목적으로 개발·관리·활용하려는 지능정보기술
2. 의료법 제24조의2 제1항에 따른 수술등의 의료행위에 직접 이용되거 사람의 신체에 영향을 미칠 수 있는 지능정보기술
3. 지능정보기술이 오작동될 경우 사람에게 중대한 위해를 끼칠 우려가 있는 지능정보기술

139) 게다가 「지능정보화 기본법」 제21조 제1항(지능정보기술의 안정성·신뢰성·상호운용성 등을 확보하기 위하여 필요한 기술기준의 고시)에 위반되는 행위에 대해서는 아무런 조치를 규정하지 않았다.

140) 전자정부법 제18조의2(지능형 전자정부서비스의 제공 등) ① 행정기관등의 장은 인공지능 등의 기술을 활용하여 전자정부서비스를 제공할 수 있다.
② 행정안전부장관은 행정기관등의 장이 인공지능 등의 기술을 효율적으로 활용할 수 있도록 행정적·재정적·기술적 지원 등 필요한 지원을 할 수 있다.
③ 제1항 및 제2항에 따른 인공지능 등의 기술의 종류, 활용 및 지원에 필요한 사항은 국회규칙, 대법원규칙, 헌법재판소규칙, 중앙선거관리위원회규칙 및 대통령령으로 정한다.

2. 인공지능을 직접 규율 대상으로 하는 법률안

「지능정보화 기본법」으로는 다양하고 급격하게 발전하고 있는 인공지능에 대응하기 부족함이 분명하므로, 제21대 국회에는 인공지능을 직접 규율 대상으로 하는 제정법률안들이 다수 제안되어 있다.

<표 6> 인공지능 관련 제정법률안 (제21대 국회)

제정법률안 의안명 (의안번호)	대표 발의자	제안일자
인공지능 연구개발 및 산업 진흥, 윤리적 책임 등에 관한 법률안 (의안번호: 2101823)	이상민 의원	2020. 7. 13.
인공지능산업 육성에 관한 법률안 (의안번호: 2103515)	양향자 의원	2020. 9. 3
인공지능 집적단지의 육성에 관한 특별법안 (의안번호: 2104564)	송갑석 의원	2020. 10. 19.
인공지능 기술 기본법안 (의안번호: 2104772)	민형배 의원	2020. 10. 29.
인공지능교육진흥법안 (의안번호: 2110148)	안민석 의원	2021. 5. 17.
인공지능 육성 및 신뢰 기반 조성 등에 관한 법률안 (의안번호: 2111261)	정필모 의원	2021. 7. 1.
인공지능에 관한 법률안 (의안번호: 2111573)	이용빈 의원	2021. 7. 19.

제안된 법률안 대다수는 인공지능 산업의 기반이나 기술 개발을 진흥하는 것을 주요 목표로 하고 있다. 규정들 역시 산업 육성에 치우쳐 인공지능이 추구해야 할 원칙들을 담보할 내용에 대한 고민이 부족해 보이며, 이에 따라 인공지능과 상호작용하는 국민에 대한 구체적인 내용은 찾아보기 어렵다. 양향자 의원안에서 포괄적 인권보호 의무를 규율하고 있는 것과 이상민 의원안에서 윤리원칙 제정과 관련한 내용 정도가 예외이다. 양향자 의원안과 민형배 의원안에서 제한적으로 인권과 관련한 의제도 다루고 있으나 개인정보보호, 소비자 보호, 인권 보호를 위한 개인정보보호위원회, 국가인권위원회, 공정거래위원회의 역할은 언급되어 있지 않다. 또한 인공지능 대응체계, 즉 거버넌스 구조 역시 산업 진흥에 관한 내용을 주로 규율하고 있어 그 구성 역시 인공지능산업 육성과 관련된 정부부처와 민간위원만이 참여하는 위원회로 구성되어 있다.

<표 7> 인공지능 관련 제정법률안 중 인권 관련 규정 (제21대 국회)

대표 발의자 (의안번호)	인권 관련 규정
이상민 의원 (의안번호: 2101823)	제3조(국가 및 지방자치단체 등의 책무) ② 국가 및 지방자치단체, 인공지능사업자 등은 인공지능 산업에서 이용자보호를 위한 인공지능 윤리원칙을 제정하고 인간의 기본적 인권과 존엄성이 보호되도록 하여야 한다. 제4조(인공지능정책심의위원회) ① 인공지능 기술개발 및 산업진흥과 인간의 기본적 인권과 존엄성을 논의하기 위하여 과학기술정보통신부장관 소속으로 인공지능정책심의위원회(이하 “정책심의위원회”라 한다)를 둘 수 있다. 제6조(기본계획의 수립 등) ① 과학기술정보통신부장관은 인공지능 기술개발 및 산업 진흥을 위하여 중장기적인 기본계획(이하 “기본계획”이라 한다)을 수립하여야 한다. ② 기본계획에는 다음 각 호의 사항이 포함되어야 한다. 10. 인공지능산업에서의 인권보호 및 차별과 편향을 예방하기 위한 윤리강령에 관한 사항 제17조(인공지능산업협회의 설립) ① 인공지능사업자는 인공지능 산업의 발전적인 생태계 조성 및 인공지능 사업자의 공동이익을 도모하기 위하여 인공지능산업협회(이하 “협회”라 한다)를 설립할 수 있다. ③ 협회는 다음 각 호의 업무를 수행한다. 6. 인공지능 산업에서의 인권보호 및 윤리강령기준 마련을 위한 연구
양향자 의원 (의안번호: 2103515)	제3조(인공지능산업에서의 인권보호) ① 국가, 지방자치단체 및 인공지능사업을 영위하는 자는 인공지능산업을 육성함에 있어 인간의 존엄성이 보호되도록 하여야 한다. ② 누구든지 인공지능기술의 개발, 생산, 유통, 활용 등 모든 단계에서 차별과 편향이 발생하거나 인권이 침해되지 아니하도록 하여야 한다. 제6조(기본계획 및 시행계획의 수립·시행) ① 과학기술정보통신부장관은 인공지능산업을 효율적이고 체계적으로 육성하기 위하여 3년마다 인공지능산업 육성 기본계획(이하 “기본계획”이라 한다)을 수립·시행하여야 한다. ② 기본계획에는 다음 각 호의 사항이 포함되어야 한다. 7. 인공지능산업에서의 인권보호에 관한 사항

송갑석 의원 (의안번호: 2104564)	제3조(기본원칙) ① 국가, 지방자치단체 및 인공지능사업자 등은 인공지능 산업에서 인간의 기본적 인권과 존엄성이 보호되도록 하여야 한다.
정필모 의원 (의안번호: 2111261)	제12조(인공지능사회를 위한 정책 및 인공지능기술 개발) ① 과학기술정보통신부장관은 인공지능사회 정립에 필요한 정책 개발 연구를 위하여 다음 각 호의 사업을 추진할 수 있다. 4. 인공지능 및 인공지능기술로 인한 차별 및 인권침해 등에 관한 조사, 관련 문제의 예방 및 해결 지원 제16조(민간자율인공지능윤리위원회의 설치 등) ① 윤리원칙을 준수하기 위하여 다음 각 호의 기관 또는 단체는 민간자율인공지능윤리위원회(이하 “민간자율위원회”라 한다)를 설치할 수 있다. 1. 인공지능 및 인공지능기술 연구 및 개발을 수행하는 사람이 소속된 교육·연구 기관 2. 인공지능사업자 3. 그 밖에 대통령령으로 정하는 인공지능기술 관련 기관 등 ② 민간자율위원회는 다음 각 호의 업무를 자율적으로 수행한다. 2. 인공지능기술 연구·개발·활용의 안전 및 인권침해 등에 관한 조사·연구

다만 정필모 의원안의 경우, 의료, 전기·가스·수도·핵시설 등 에너지·기간서비스, 범죄수사 생체인식, 개인에 대한 평가·의사결정, 공공기관 사용 부문에서 ‘사람의 생명·신체에 위험을 줄 수 있거나 부당한 차별 및 편견의 확산 등 인간의 존엄성을 해칠 위험이 있는 인공지능’을 정의(제2조 제2호)하고 규율하고자 했다는 점에서 유럽연합 등에서 추진하고 있는 위험 기반 접근법과 유사한 지향을 가지고 있다. 이 안에서는 이들 특수활용 인공지능을 사용하여 업무를 수행하는 자에게 해당 사실을 상대방이 쉽게 알 수 있도록 사전에 고지하고 요청이 있는 경우 그 의사결정 원리 및 최종결과 등을 설명하도록 하고(제20조), 인공지능사업자가 특수활용 인공지능을 개발·제조·유통하거나 이와 관련된 인공지능서비스를 제공하려는 경우 과학기술정보통신부장관에게 신고하도록 하여야 하고 기술적·관리적 조치를 갖추어야 한다(제21조). 과학기술정보통신부장관 및 관계 중앙행정기관의 장은 특수활용 인공지능에 대하여 사전 고지의무 및 신고의무 이행을 위하여 필요하다고 인정하는 경우에는 소속 공무원으로 하여금 조사하게 할 수 있도록 하였다(제24조). 그러나 법안의 주무부처가 과학기술정보통신부라는 점에서 이

법에 기반하여 인권 보호를 위한 규율이 이루질 것을 기대하기는 어려워 보인다.

한편, 이상민 의원이 대표발의한 평등에 관한 법률안(의안번호: 2110822)의 경우 ‘인공지능 디지털 기술 등에 대한 동일 적용’에 대한 조항을 두고 이 법이 인공지능, 빅데이터 등 디지털 기술을 기반으로 한 모든 영역에도 동일하게 적용함을 규정하였다(안 제8조).

제안된 법안 중 일부 내용은 일부 부처 소관업무와 충돌하는 측면이 있어 해당 부처들과 상임위원회 전문위원이 반대 의견을 밝히기도 하였다. 이상민 의원안의 전문인력 양성, 표준화 지원 등에 대하여 산업통상자원부, 중소벤처기업부가 소관업무와 충돌한다는 입장이고,[141] 양향자 의원안의 세제지원, 국유재산・공유재산의 대부・사용 규정에 대하여 기획재정부, 부담금 감면 시설에 대하여는 국토교통부, 농지보전부담금 감면에 대하여는 농림축산식품부. 연구개발 특화단지 중복 규정에 대하여는 과학기술정보통신부가 반대 의견을 밝혔다.[142] 민형배 의원안에 대하여는 법 제정 필요성에 산업통상자원부가 반대 의견을 밝혔다.[143]

「지능정보화 기본법」에도 불구하고 인공지능에 대한 독자적인 제정 법률을 추진하기 위하여는 「지능정보화 기본법」이 이미 그 목적으로 밝히고 있는 인공지능 관련 정책의 수립・추진 및 산업 경쟁력 여건 조성에 대한 내용을 반복적으로 규정하기 보다는, 최근 국제 규범에서 강조되고 있는 인공지능 규율을 포괄하는 것이 바람직할 것이다. 즉, 인공지능의 개발과 활용으로부터 국민의 안전과 인권을 보호하기 위한 원칙을 밝히고, 그 이행을 준수하는 감독 체계를 수립하며, 영향을 받는 사람들의 의견을 수렴하고 참여를 보장하는 거버넌스를 갖추고, 책임성 체계를 갖추어 피해 국민을 구제할 수 있는 규정이 마련되어야 할 것이다.

141) 인공지능 연구개발 및 산업 진흥, 윤리적 책임 등에 관한 법률안 검토보고 (2020. 9). 국회 과학기술정보방송통신위원회.
142) 인공지능산업 육성에 관한 법률안 검토보고 (2020. 11). 국회 과학기술정보방송통신위원회.
143) 인공지능 기술 기본법안 검토보고 (2021. 2). 국회 과학기술정보방송통신위원회.

제5장 인공지능과 국가인권기구

제1절 인공지능과 국가인권기구 관련 해외 논의

인공지능에 대한 인권적 감독의 중요성이 커질수록 국가인권기구의 이 분야 역할과 개입에 대한 요구도 커져 왔다. 앞서 유럽평의회 인권위원장은 인권영향평가, 공개적인 의견 수렴, 독립적 감독, 구제 수단 등에서 국가인권기구의 역할을 강조한 바 있다.

국가인권기구는 독립성, 전문성 및 진정사건 처리 경험을 보유하고 있다는 점에서 인공지능의 인권 준수에 대한 독립적인 감독은 물론 인권영향평가 등 관련 지침과 지원 역시 전문적이며 효과적으로 수행할 수 있을 것으로 기대되고 있다. 인공지능 기술의 발전은 국가인권기구의 권리구제 활동에도 변화를 가져올 수 있기 때문에 이 분야 국가인권기구의 대응 역시 필연적으로 이루어질 수밖에 없다. 또한 차별 시정 업무를 수행하는 국가인권기구는 관련 권리구제 뿐 아니라 인공지능 시스템으로부터 차별적인 영향을 받는 사람들과의 협의 역할을 수행할 수 있다.

다만 인공지능에 대한 전문지식 측면에서 국가인권기구의 역량과 자원이 현재까지 부족한 측면이 있기 때문에 이 관련된 인적, 재정적 자원을 보다 더 지원할 필요가 있다는 점 또한 지적되고 있다.

1. 유럽연합 기본권청의 검토

유럽연합의 인권기구인 유럽연합 기본권청(European Union Agency for Fundamental Rights)은 2020년 국가인권기구의 강화 및 효과적인 활동에 대한 보고서[144]에서 인공지능 기술의 발전으로 사생활권, 개인정보보호권, 차별금지 관련 조항이 동반하여 문제가 되고 있다고 지적하면서, 이 문제에서 독립성, 전문성 및 진정사건 처리 경험을 보유하고 있는 국가인권기구들이 알맞은 역할을 수행할 수 있다고 보았다. 예컨대 유럽연합 인공지능 윤리 가이드라인의 경우 “불공정한 편향이나 차별 등 인공지능 시스템에서 발생하

144) European Union Agency for Fundamental Rights (2020a). 3.4.1.절.

는 유해한 결과"를 확인할 수 있도록 인공지능 시스템에 대한 '감사 메커니즘'을 권고하고 국제인권법 의무 준수에 대한 감독 필요성을 제기하였는데, 이러한 상황에서 독립적인 지위 및 인권에 관한 전문성을 보유한 국가인권기구 및 평등기구가 기여할 수 있는 역할이 있을 것이라고 기본권청은 강조하였다.

인공지능 기술의 발전은 국가인권기구의 권리구제 활동에도 변화를 가져올 수 있기 때문에 이 분야 국가인권기구의 대응은 필연적이다. 우선 알고리즘 의사결정으로 영향을 받은 개인들의 국가인권기구 진정 접수가 증가할 수 있으며, 국가인권기구는 알고리즘에 의한 개인정보 처리에 대해 이해하지 못하고 구제 곤란을 겪는 개인들을 위해 이해도 증진 활동에 개입해야 할 수 있다. 또한 국가인권기구 내부적으로도 알고리즘 의사결정에 대한 디지털 리터러시 향상 및 전문가 자문 네트워크 구축이 필요하고, 인공지능에 대한 총체적 접근을 위해서 다양한 행위자, 기관, 학계에 걸쳐 광범위하게 협업할 필요가 있다. 여기에는 개인정보보호 감독기구를 비롯해 인공지능 관련 감독 기능을 가진 기존 기관과의 협력이 포함된다.

더불어 기본권청은 2020년 12월 <인공지능과 기본권> 보고서[145]를 발표하고 유럽연합 인공지능 규제에 포함될 기본권 보장 관련 문제를 살펴 보았다. 이 보고서에서 기본권청은 인공지능 시스템이 기본권에 미치는 부정적인 영향을 감독하고 이를 해결할 수 있는 효과적인 책임 체계의 구축이 필요하다고 지적하면서, 국가인권기구를 비롯한 기존의 감독 전문 조직을 더 잘 활용할 것을 유럽연합과 회원국들에 권고하였다(의견3).

보고서에서 기본권청은 인공지능에 대한 인권영향평가의 적용과 국가인권기구를 비롯한 감독 기관의 역할을 중점적으로 검토하였다. 우선 인공지능에 대한 인권영향평가의 필요성에 대한 그간의 논의를 간략히 개괄하면서 유럽연합과 유럽평의회의 개인정보보호 관련 법률들에서 권장하고 있는 개인정보보호 영향평가(data protection impact assessments)를 비판적으로 고찰하였다. 유럽연합 GDPR은 개인정보 처리가 개인정보보호권에 대한 영향 외에도 보다 폭넓은 '자연인의 권리와 자유에 큰 위험을 초래할 가능성'에 대한 평가를 요구한다. 따라서 원칙적으로 개인정보보호 영향평가가 알고리즘과 그 기본권에 미치는 영향을 추가로 조사하는 도구로 사용될 수 있다. 그러나 GDPR은 개인정보보호 영향평가를 '고위험' 사례로 제한하고 있다는 한계가 있다(제35조).[146] 따

145) European Union Agency for Fundamental Rights (2020b).

라서 개인정보 처리와 관련이 없는 다른 고위험 사례를 놓칠 수 있다. 개인정보보호 영향평가가 제도적으로 시행되어 옴에 따라 법률과 지침으로 제시되어 있는 체크리스트 등 영향평가의 기준을 참고할 만 하다. 개인정보보호 영향평가는 자체적으로 실시되지만 영향평가로 완화할 수 없는 위험에 대해서는 개인정보보호 감독기구와 사전 협의를 의무로 규정하였고, 유럽연합의 감독기구들은 영향평가 방법에 대한 다양한 지침도 마련하고 있다. 즉, 영향평가 제도에서 감독기관의 역할과 위상이 중요하다.

국가인권기구가 실시 중인 인권영향평가의 모범 사례로는 덴마크 국가인권기구의 사례가 제시된다. 덴마크 국가인권기구는 온라인에서 대화형으로 실시할 수 있는 인권 준수 '신속 점검(quick check)'을 개발하였다. 신속 점검은 인권 준수 평가 도구(Human Rights Compliance Assessment tool)에 기반하였으며, 이 평가 도구는 350개 이상의 질의와 그에 상응하는 인권 지표 1,000개를 데이터베이스로 보유하고 있다. 기준 지표는 국제 인권법상 기준들이다.

더불어 기본권청은 100개 기관 이상의 공공기관과 민간기업의 관계자 및 전문가를 인터뷰하여 인공지능 영향평가의 현재 관행을 살펴보았다. 많은 인공지능 시스템이 여러 유형의 영향평가나 검사(testing)를 실시하고 있었지만 주로 기술적인 평가나 개인정보보호 영향에 국한되어 있었으며, 기본권에 미치는 영향은 거의 다루어지지 않았다. 기본권 영향평가의 실시에 대하여 반대하는 응답자들은 시스템이 기본권에 부정적인 영향을 미친다고 생각하지 않았다. 예를 들어 트래픽 모니터링 업무에서 인공지능을 사용하는 응답자는 법정 개인정보보호법 준수 외에는 시스템의 정확성만을 검사하였고 기본권에 미

146) GDPR에서 개인정보보호 영향평가 의무를 부여하고 있는 고위험 개인정보 처리는 다음과 같다. △평가나 점수화. 특히 신용평가, 질병 예측을 위한 유전자 검사, 맞춤 마케팅 등 사람에 대한 프로파일링 및 예측의 경우 △법적 혹은 이와 유사한 중대한 효과를 미치는 자동화된 결정 △체계적인 감시. 특히 정보주체가 인지하지 못하는 사이에 공공장소 등에서 개인정보가 수집, 이용되는 경우 △민감정보 또는 통신비밀, 위치정보, 금융정보 등 매우 사적인 데이터 △정보주체의 수, 처리되는 데이터의 양과 범위, 데이터 처리 행위의 지속성 및 영구성, 처리행위의 지리적 범위 등에서 대규모로 처리되는 데이터 △데이터셋의 연계 또는 결합. 정보주체의 합리적 기대를 벗어나 다른 처리자에 의해, 다른 목적을 위해 처리되는 둘 이상의 데이터 처리의 경우 △취약한 정보주체에 대한 데이터. 아동, 노동자 및 정신질환자, 망명신청자, 노인, 환자 등 처리자와 정보주체의 불균등한 권력관계에 처한 경우 △신기술의 혁신적인 사용 또는 기술적, 조직적으로 새로운 솔루션의 적용 시에는 영향평가의 의무적 실시. 예를 들어 물리적 접근통제를 위해 지문이나 얼굴인식 기술을 사용하는 경우 △처리 자체가 정보주체의 권리 행사 및 서비스접근이나 계약체결의 중단을 낳는 경우. Article 29 Data Protection Working Party (2017). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 참조.

치는 영향은 검사하지 않는다고 답변하였다. 특히 많은 응답자들은 차별 문제에 대하여 일반적으로는 알고 있었지만, 자신의 시스템이 차별금지 속성에 따른 차별을 유발할 가능성을 배제하였다. 개인정보보호 영향평가는 알고리즘의 사용에 대하여 적용하기에 불분명한 부분이 있었다. 예를 들어 예측 치안이나 금융 분야에서 인공지능을 사용하는 응답자들은 시스템의 기본 구조나 데이터에 대하여는 개인정보보호 영향평가를 실시하였으나 머신 러닝 등 특정 인공지능 알고리즘 또는 데이터의 교차 사용에 대해서는 그 대상이 아니라고 보거나 적합하지 않다고 판단하여 해당 평가에서 제외하였다. 대체로 기술 분야 응답자들은 자신의 기술 업무에 적용되는 개인정보보호 영향평가와 분리되어 그 내용에 대하여 잘 알고 있지 못했으며 자신의 책임 영역이 아니라고 보았다.

외부 기관 감독과 관련하여 응답자들은 현행 예산 검사, 정보통신 시스템 검사나 보건의료 또는 금융 분야 등 특정 부문 인증 제도에 따른 검사를 받았거나 개인정보보호 감독기구 및 소비자 보호 당국과 협의하고 있었다. 인권기구 및 평등기구와 관련해서는 차별 감독 측면에서 전문성이 있지만 인공지능에 대한 전문지식이 요구된다는 반응이 있었다.

마지막으로 기본권청은 이상의 검토를 바탕으로 인공지능 관련 기술을 사용할 때 기본권 영향평가에 포함될 수 있는 주요 요소를 제안하였다. 우선 여러 인권 영역 중에서도 주요하게 평가할 대상으로는 개인정보보호, 차별 금지, 권리 구제를 꼽았다. 첫째, 개인정보에 대한 처리는 개인정보보호법을 준수하여 합법적이어야 한다. 둘째, 인공지능 프로세스는 차별 우려가 있는 집단에 대한 부당한 대우나 차별을 금지하여야 한다. 이때 불이익의 정도는 그 특성(위해 유형), 심각성(위해 정도), 중대성(다른 집단에 비해 불이익을 받을 위험이 높은지)에 따라 다르다. 셋째, 인공지능의 대상이 되는 사람들은 이의를 제기하고 효과적인 구제 수단에 접근할 수 있어야 한다.

또한 기본권청은 기본권에 미치는 잠재적인 영향을 평가하기 위해 사전에 제공이 필요한 최소한의 정보로서 △ 시스템의 목적, 환경 및 법적 근거에 대한 설명 △ 위양성, 위음성, 기타 자동화 및 사용 규모에 따른 위해성에 대한 의문점을 비롯하여 시스템 사용에 따른 잠재적 위해성에 대한 설명 △ 사용된 기술에 대한 설명. 여기에는 시스템 구축에 사용되는 데이터와 시스템 처리에 대한 법적 근거를 포함함 △ 인공지능 시스템의 정확성에 대한 증거 기반 설명. 학습 데이터 및 실제 환경 검사와 실험에서 도출된 결과

물에 기반하여야 함. 이때 위양성 및 위음성은 각각 검토되어야 함. 잠재적 차별을 검사할 수 있는 가능한 한 많은 집단에 대한 분류를 포함할 것(여성과 남성 간의 정확성 비교 등) △ 기존 표준 준수 및 인증에 대한 정보 등을 제시하였다.

이러한 영향평가는 인공지능 시스템의 배치 전에 의무적으로 실시되어야 할 뿐 아니라 배치된 후에도 반복적으로 실시될 필요가 있다. 인공지능 시스템의 대상이 되는 사람들에게 그 사실을 알리는 것이 중요하며, 인공지능 의사결정에 이의를 제기하기 위한 손쉬운 수단을 제공하여야 한다.

현행 개인정보보호 감독기구, 평등기구, 옴부즈만기구, 국가인권기구 등은 인공지능이 각자의 전문 영역에서 기본권에 미치는 잠재적 영향에 대하여 협의하고 감독하는 역할을 수행할 수 있다. 다만 이들 기관에 대하여 전문성과 자원 할당 측면에서 좀더 향상된 뒷받침이 필요하다.

한편, 유럽평의회 인권위원장도 2019년 보고서에서 인공지능의 인권보장 체계에서 국가인권기구의 역할이 중요하다는 점을 지적하였다.[147] 우선 외부 독립성과 전문성이 요구되는 인공지능 인권영향평가의 실시에 있어 공공기관에 대해서는 국가인권기구가 역할을 할수 있다고 지적하였고(권고 1), 공공기관 인공지능의 공공 의견 수렴 및 협의 절차와 관련하여서 국가인권기구가 기여할 바가 있다고 보았으며(권고 2), 민간 부문의 인권기준 이행을 위한 인권실사 등(권고 3)의 촉진에도 국가인권기구가 기여할 수 있을 것이며, 특히 인권위원장은 인공지능 시스템의 인권 준수에 대해 독립적이고 효과적인 감독기구로서 국가인권기구의 역할을 강조하고(권고 5), 국가인권기구는 인공지능의 인권침해에 효과적인 구제수단 제공 및 권고 기능(권고 9)을 수행할 수 있고, 인공지능에 대한 인식 증진(권고 10)에 있어서도 기여할 수 있다고 지적하였다.

2. 국가인권기구들의 활동 사례

이미 몇 년 전부터 인공지능에 대한 효과적이고 독립적인 감독 체계 수립에 있어 국가인권기구들의 역할과 관여가 요구되어 왔다. 국제적으로 권위 있는 개인정보보호 감독기구 국제협회는 2018년 발표한 <인공지능 윤리 및 개인정보보호에 대한 선언>[148]에서

147) Council of Europe Commissioner for Human Rights (2019).

개인정보보호 감독기구들이 인권기구들과 협력할 필요성이 있다고 강조하였다. 인공지능 분야에서 이루어지는 개인정보의 수집, 이용, 공개가 차별, 표현 및 정보의 자유 등 보다 광범위한 인권에 직접 영향을 미친다는 사실을 인식함에 따라, 개인정보보호 및 프라이버시 감독기구들이 인권을 보다 광범위하여 고려하면서 다른 인권기구들과 협력해야 한다는 것이다.

유럽 국가인권기구들의 네트워크 European NHRIs는 2020년 6월 30일 발표한 성명[149]에서 "국가인권기구들은 인공지능에 의해 기본권에 영향을 받는 이들을 효과적으로 지원하고 기본권 침해를 예방하고 평가하기 위해 충분한 자원, 힘, 특히 전문성을 지녀야 한다." 고 강조하였다. 더불어 국가인권기구들이 다른 기구들과 협업하여 수행할 수 있는 역할로서 △인공지능이 기본권에 미치는 영향을 파악하기 위한 모니터링 활동 및 법률 검토 △법률의 간극을 확인하고, 기본권을 보장하며 국가 또는 유럽연합 수준에서의 조치를 확정하기 위해 인공지능 시스템을 어떻게 규제해야 하는가에 대해 조언 △유럽연합 헌장과 법률을 고려하여 인권에 기반한 인공지능 접근에 대한 가이드라인을 작성하기 위해, 인공지능, 법률, 인권 전문가들을 한데 모으는 플랫폼 역할 △인공지능 전문가와 함께 논의하고 인공지능의 설계, 개발, 도입 전반에 있어 인권 안전장치에 대해 조언하는 등의 역할을 제시하였다.

특히 인공지능의 차별 문제와 관련한 국가인권기구들의 역할이 요구되어 왔다. 영국에서는 2020년 공직생활윤리위원회가 모든 공공기관 인공지능으로 하여금 현행 법률을 준수하고 그 내용을 공표하도록 권고하고, 영국 평등인권위원회에 공공부문 인공지능의 평등법 준수지침 개발과 역할을 요청하였다.[150]

최근 각국 국가인권기구들의 인공지능 정책 관련 개입도 증가하였다. 뉴질랜드 국가인권기구는 신기술 환경에서 프라이버시, 데이터, 기술 문제를 아우르는 새로운 법적 규제를 제안하면서[151] △국제인권기준 준수 △뉴질랜드 법제도 준수 △프라이버시 침해의 예

148) International Conference of Data Protection and Privacy Commissioners (2018). Declaration on Ethics and Data Protection in Artificial Intelligence.
149) European Network of National Human Rights Institutions (2020), ENNHRI's work to promote and protected fundamental rights related to artificial intelligence (2020. 6. 30).
150) The Committee on Standards in Public Life (2020). Artificial Intelligence and Public Standards.
151) NewZealand Human Rights Commission, A new legal framework for privacy, data

외 △감독 체제, 투명성, 목적 제한 등 안전조치를 요구하였다. 네덜란드 국가인권기구는 <전략 계획 2020-2023>의 핵심 주제 중 하나로 “디지털화와 인권”을 채택하고 디지털 채용 절차에서 나타날 수 있는 노동 시장 차별 가능성에 대해서 검토하면서 동등한 대우를 받을 권리를 강조하였다. 스웨덴 국가인권기구와 평등 옴부즈만은 인공지능의 차별금지법 준수를 감독하고 있다.[152] 독일의 연방차별금지국은 2019년 <알고리즘 사용에 관련된 차별 위험>에 대한 연구에서 알고리즘 차별 방지를 위하여 정부에 대한 정책권고를 요구하였다.[153]

특히 호주 국가인권위원회는 2018년부터 일련의 ‘인권과 기술’ 프로젝트 사업을 수행하고 2021년 6월 <인권과 기술> 최종보고서를 발간하였다.[154] 호주 국가인권위원회는 이 보고서에서 인공지능 의사결정에서 인권 보장을 위하여 호주 정부에 대한 여러 권고를 담았다. 우선 정부 인공지능 사용의 법적 책임성을 위하여, △공공기관 인권영향평가를 입법하고(권고2), △인공지능을 사용하는 행정적 의사결정에서 영향을 받는 개인들에 대한 고지를 입법하며(권고3), △인공지능의 지원을 받는 정부기관 의사결정에 대한 전문 감사기관을 지정하고(권고4), △행정적 의사결정에서 사유나 기술적인 설명이 불가능한 자동화/인공지능의 사용을 금지하며(권고5), △의사결정 책임자 명확화, 기술적인 설명 요구권 등 인공지능 의사결정에 관련된 법률을 개정하고(권고6), △인공지능 안전위원회 등 독립적인 전문기관을 설립하며(권고7), △행정재심재판소 등 인공지능 의사결정에 대한 이의제기에 독립적인 검토를 입법할 것(권고8)을 권고하였다. 한편, 민간 인공지능 사용의 법적 책임성을 위한 호주 정부에 대한 권고로는 △인공지능 윤리 원칙에서 인권영향평가 수행을 권장하고(권고9), △개인의 권리에 법적 또는 중대한 영향을 미치는 기업의 인공지능 의사결정에 대한 고지를 입법하며(권고10), △인공지능 의사결정에 대한 기업의 법률적 책임을 입법하는 한편(권고11), △규제기구가 기업의 자료 제출을 요구할 권한에 대한 입법(권고13)도 권고하였다. 그 밖에도 호주 정부에 대한 권고로, 인공지능 안전위원회 등 전문기관이 인공지능 인권영향평가 도구와 인권 기반 인공지능 조달 절차 및

---

and technology.

152) European Union Agency for Fundamental Rights (2020a). 92p.

153) Federal Anti-Discrimination Agency (2019). Risks of Discriminationthrough the Use of Algorithms.

154) Australian Human Rights Commission (2021). Human Rights and Technology. <https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf (검색일: 2021. 9. 1.)>.

표준을 개발하도록 조치할 것과, 인권위원회가 인공지능 사용 의사결정에서 차별금지법 준수를 위한 지침을 발간하는 것을 지원할 것, 중대한 프라이버시 침해 우려가 있는 얼굴인식 및 기타 생체인식 기술에 있어 규제 입법이 이루어지기 전까지 사용 유예(모라토리엄)할 것 또한 권고하였다.

제2절 인공지능과 국가인권위원회의 역할

인공지능의 개념을 정의하거나 인공지능을 직접 규율하는 현행 법령 상의 기준은 없으나, 「지능정보화 기본법」에 따라 과학기술정보통신부장관을 주무 부처로 하는 정부는 지능정보사회 종합계획을 수립하고 이 계획에 "지능정보사회윤리의 확립", "정보보호, 정보격차 해소, 역기능 해소, 이용자의 권익보호 및 지식재산권의 보호"하고 구체적인 실행계획을 수립해야 하므로(「지능정보화 기본법」 제6조, 제7조), 인공지능에 의한 위험 중 일부 내용은 지능정보사회 종합계획과 그 실행계획에 포함되는 정책과 기준에 의해 대응이 될 수 있다. 하지만, 인공지능에 의한 위험은 사회 전반과 사람의 일상 전반에 미치므로, 체계적이고 조직적인 규정을 갖춘 별도의 입법에 의한 대응이 필요하나, 별도의 입법이 없는 상태라도 인권 일반에 관한 독립 기관인 국가인권위원회의 적극적인 대응이 요구되는 상황으로 봐야 한다.

앞에서 살펴보았듯이, 유럽연합 기본권청과 유럽평의회 인권위원장은 인공지능 사회에서 인권위원회의 역할을 강조한 바 있다.

우선 유럽연합 기본권청은 인공지능 기술의 발전으로 사생활권, 개인정보권, 차별금지 관련 조항이 함께 문제가 되고 있어 독립성, 전문성 및 진정사건 처리 경험을 보유하고 있는 국가인권기구의 역할이 필요하다고 지적하였다. 유럽연합 인공지능 윤리 가이드라인의 경우, "불공정한 편향이나 차별 등 인공지능 시스템에서 발생하는 유해한 결과"를 확인할 수 있도록 인공지능 시스템에 대한 '감사 메커니즘'을 권고하고 국제인권법 의무준수에 대한 감독 필요성을 제기한 바 있다. 인공지능 기술의 발전은 국가인권기구의 권리구제 활동에도 변화를 가져올 수 있기 때문에 이 분야 국가인권기구의 대응은 필연적이다.[155] 나아가 유럽연합 기본권청은 유럽 각국에 대한 권고에서 "인공지능 시스템이 기본권에 미치는 부정적인 영향을 감독하고, 필요한 경우 이를 실질적으로 해결할 수 있는 효과적인 책임 시스템을 구축해야" 하며, 이때 "개인정보보호 기관, 평등기구, 국가 인권 기구, 옴부즈만 기관 및 소비자 보호기관 등 기존의 감독 전문 조직을 더 잘 활용해야 한다"고 지적하였다.[156]

155) European Union Agency for Fundamental Rights (2020a).
156) European Union Agency for Fundamental Rights (2020b). Opinion 3.

한편, 유럽평의회 인권위원장도 인공지능의 인권보장 체계에서 국가인권기구의 역할을 지적하였다.[157] 외부 독립성과 전문성이 요구되는 공공기관 인공지능 인권영향평가의 실시에서 국가인권기구가 역할을 할 수 있고, 민간 부문의 인권실사 촉진에도 국가인권기구가 기여할 수 있을 것이다. 더불어 국가인권기구는 인공지능 시스템의 인권 준수에 대해 독립적이고 효과적인 감독 역할을 수행하고 인공지능의 인권침해에 효과적인 구제수단을 제공하고 권고할 수 있다.

우리의 상황도 마찬가지라 할 수 있다. 국가인권위원회는 국가인권위원회법과 국제인권법에 의해 설립된 국가인권기구로서 대한민국 헌법과 국제인권규범에서 보호하는 모든 사람의 인권과 자유를 인공지능을 비롯한 지능정보사회에서 보호하고 향상시켜야 할 책무가 있다. 국가인권위원회는 국가인권위원회법에 따라 인공지능에 관한 법령・제도・정책 등 인권 개선에 관한 권고 또는 의견을 표명할 수 있고, 공공기관 인공지능의 인권침해 및 차별, 법인・단체・사인의 인공지능에 의한 차별에 대해 조사 및 구제 활동을 할 수 있으며, 인권침해의 유형, 판단 기준 및 그 예방 조치 등에 관한 지침의 제시 및 권고를 할 수 있으므로, 공공기관이 직접 또는 조달을 통하여 개발하거나 활용하려는 인권지능에 대한 인권 기준을 제시할 수 있다. 유엔 〈기업과 인권 이행 지침〉에 따른 기업의 인권경영이 인공지능에 대하여도 이행될 수 있도록 관련 인권 기준을 제시할 수 있다.

특히 국가인권위원회는 인공지능에 대한 인권영향평가가 시행되도록 지침을 마련하고 권고하는 것이 바람직하다. 앞에서 이미 살펴보았듯이, 유럽평의회 인공지능 특별위원회 정책개발단은 2021년 5월 21일 인공지능 인권, 민주주의 법치 영향평가에 대한 보고서 초안에서 인공지능 인권영향평가에서 참고할 수 있는 기존의 규범적 문헌을 개괄하고 인공지능에 대한 적용을 제안하였다. 유럽평의회 인권위원장과 유럽평의회의 영향평가 권고, 유럽 집행위원회 인공지능 고위전문가그룹의 “신뢰할 수 있는 인공지능 평가 목록” 지침 제안, 영국의 에이다 러브레이스 연구소의 알고리즘 감사와 알고리즘 영향평가 제안, 미국 시민단체 미국시민자유연합의 〈알고리즘 형평성 툴킷〉, 캐나다 정부의 〈캐나다 알고리즘 영향평가 도구〉 제안 등에서 알 수 있듯이, 인공지능에 대한 인권영향평가는 인공지능의 위험을 제어하는 유효한 수단으로 제안되고 있다. 이미 국가인권위원회는 「공공기관 인권경영 매뉴얼」적용 권고를 통해 공공기관에 대한 인권영향평가가

157) Council of Europe Commissioner for Human Rights (2019).

시행되도록 한 바 있으므로,[158] 적어도 공공기관에 대한 인권영향평가에서 인공지능을 도입할 경우의 영향평가 기준을 마련하고 시행을 권고할 필요가 있다.

또한 국가인권위원회는 인공지능에 대한 인권영향평가와 더불어 인공지능과 인권에 관한 정보와 이슈를 상시 관측하면서 우리 사회의 인공지능 거버넌스에도 적극 참여해야 한다. 인공지능 기술의 혜택과 위험 문제를 다양한 이해당사자들 사이에서 어떻게 다루고 조정해야 하는지의 문제는 최근에 여러 이해당사자가 참여하는 이른바 '인공지능 거버넌스'의 틀을 중심으로 논의되어 왔다. 이와 관련해 인공지능 거버넌스에서 논의될 만한 사회적 이슈로서 인공지능의 개발과 사용에 관한 원칙에 관해서도 국제사회에서 활발한 논의가 이뤄져 왔는데, 그중 하나로서 경제협력개발기구(OECD)는 2019년 5월 보고서 <사회 안의 인공지능 (AI in Society)>에서[159] "신뢰할 만한 인공지능의 책임 관리인(the responsible stewardship of trustworthy AI)"을 국제표준과도 같은 정부 역할로서 제시해 강조한 바 있다. 이 보고서는 같은 해 주요 20개국 협의체(G20) 선언문[160]에도 실린 바 있는 다섯 가지의 'OECD 인공지능 원칙'을 제시했는데, 이는 다음과 같다. 첫째, 인공지능은 포용적이며 지속가능한 성장과 복지에 기여해야 한다. 둘째, 인공지능은 인간 중심의 가치와 공정성을 존중해야 한다. 셋째, 인공지능 개발운영자는 사용하는 인공지능 시스템이 어떤 것이며 어떻게 작동하는지 투명하게 밝히고 설명할 수 있어야 한다. 넷째, 인공지능 시스템은 견고하고 안전해야 한다. 다섯째, 인공지능 개발운영자는 이에 책무성을 갖춰야 한다. "사람 중심 인공지능'(Human-centered AI)"이라는 표현으로 요약되는 이런 다섯 가지 원칙은 정부가 추구해야 하는 역할이자 민관의 인공지능 거버넌스가 추구할 방향이기도 하다. 국가인권위원회는 인공지능과 관련한 인권 문제의 국가기구로서 이런 인공지능 거버넌스에 책임 있게 참여할 수 있어야 하며, 이를 위해서는 인공지능과 인권에 관한 정보와 이슈를 상시적으로 관측하는 활동을 벌여야 한다.

인공지능과 인권에 관한 정보와 이슈를 상시 관측하는 활동은 인공지능 거버넌스에서 매우 중요하다. 우리 사회에서 잘 드러나지 않은 채 일어나고 있거나 앞으로 일어날

158) 사단법인 공익법센터 어필 (2020). 공공기관·공기업 인권영향평가 현황 실태조사 및 개선방안 연구. 국가인권위원회 인권상황실태조사 연구용역보고서.

159) OECD (2019). Artificial Intelligence in Society. OECD Publishing.

160) G20 Trade and Digital Econoy Ministers (2019). G20 Miniterial Statement on Trade and Digital Economy. published at The G20 Ministerial Meeting on Trade and Digital Economy held in Tsukuba, Japan on 8 and 9 June, 2019. <https://g20-digital.go.jp/asset/pdf/g20_2019_japan_digital_statement.pdf (검색일: 2021. 11. 1)>.

것으로 우려되는 인권 침해 문제를 폭넓게 관측하는 역할은 인공지능 거버넌스에서 기존의 하향 접근이나 상향 접근과는 구분되는 중간적인 접근으로서 이해될 수 있다. 이와 관련한 논의로서 유럽의회와 유럽집행위원회 지원을 받아 여러 이해당사자로 구성된 포럼 '사람을 위한 인공지능'(AI4people)이 2019년 11월 <좋은 인공지능 거버넌스에 관하여(On Good AI Governance)> 보고서에서 인공지능 윤리 원칙을 실현하는 데 필요한 우선 실행 과제(priority actions)로서 발표한 내용을 참조할 수 있다.[161] 이 보고서는 기술과 윤리, 시장, 사회규범 간의 균형을 강조하면서 법과 제도를 중심으로 하는 하향(top-down) 접근과 참여를 강조하는 상향(bottum-up) 접근의 우선 실행 과제들를 제안하면서,[162] 이와 함께 하향과 상향의 중간에 놓이는 의미 있는 활동으로서 '중간 접근(in-between)'의 실행 과제를 제안했다. 그중에는 이른바 '인공지능 유럽관측소(European Observatory for AI)' 같은 기구를 설립해야 한다는 제안도 포함됐는데, 가칭 인공지능 관측소는 변화 과정에 놓인 인공지능 사회에서 전례 없는 법률적, 도덕적 문제에 대처하기 위해 변화를 늘 관측하고 분석하고 알리고, 관련 논의를 중재하며 기술-사회 갈등을 줄이는 역할을 하는 지속적인 기구로서 제안되었다. 인공지능 기술과 관련해 현재 일어나고 있는 문제와 앞으로 잠재적으로 일어날 수 있는 문제를 분명하고 충분하고 다양하게 이해할 수 없는 현재의 상황에서, 이와 관련한 국내외 논의 동향과 정보, 이슈를 관측하고 수집하고 분류하고 분석하고 정책이나 연구개발에 반영하도록 노력하며 무엇보다 다양한 이해당사자들이 논쟁하고 대화할 수 있는 의제를 새롭게 발굴하려는 노

161) AI4People (2019). On Good AI Governance: 14 Priority Actions, a S.M.A.R.T. Model of Governance, and a Regulatory Toolbox. presented to the European Parliament and the European Commission on 6 November 2019. <https://www.eismd.eu/wp-content/uploads/2019/11/AI4Peoples-Report-on-Good-AI-Governance.pdf (검색일: 2021. 11. 1)>.

162) 이 보고서가 권고한 하향식, 상향식 접근의 실행과제들은 다음과 같다. 먼저 주로 법률과 제도와 관련한 하향(top-down) 접근의 실행 과제로서, (1)유럽연합 회원국 공동의 목표와 인간 번영을 지지하며 (2) 인공지능 기술의 여파에 관한 교육 커리큘럼을 개발하고 (3) 윤리적, 법률적, 사회적 고려사항을 통합하는 지속적인 인공지능 연구를 증진하며 (4) 포괄적인 혁신과 인간 중심 설계를 통해서 인간-기계의 협력을 통한 새로운 일자리로 부드러운 이행을 촉진하고 (5) 인공지능 기술의 윤리적 의미에 대한 기업 임원진의 책임 역량을 지원하는 활동을 우선 실행할 활동으로 꼽을 수 있다. 일반인과 전문가들이 참여하는 이해당사자들의 소통과 협력을 강화하는 상향(bottum-up) 접근의 활동으로는, (1) 사회적 가치와 대중 여론의 이해와 일치를 이루기 위한 이해당사자들 간의 참여적 대화를 꾀하고 (2) 논쟁과 합의와 인식 공유의 장을 넓히고 새로운 실행 방안을 숙의하며 3) 기술, 사회 이슈, 법률 연구, 윤리 간의 다학제 협력과 논쟁 프로그램을 열고 (4) 탈규제의 특별한 공간에서 인공지능의 경험적 시험과 개발을 위해 과학자와 일반인이 상호작용하는 기회를 미련하는 활동들이 제안됐다.

력이 이런 중간 접근 실행 방안에서 중요하게 다뤄질 수 있는 것이다. 국가인권위원회는 인공지능과 관련한 인권 문제를 전문으로 다루는 국가기구로서 이런 인공지능 관측소와도 같은 중간 접근 실행의 역할을 담당할 수 있을 것이다.

제6장 심층면접조사

제1절 심층면접조사 개요

1. 조사 대상

심층면접조사는 개발된 가이드라인(초안)에 대하여 개발자를 비롯한 기술자, 학계 연구자와 기관 연구원 등 전문가들과 인공지능 활용으로 영향을 받는 당사자인 노동자, 수급대상자, 여성, 장애인 등을 대표하는 개인 및 단체를 대상으로 진행되었으며, 구체적인 구성은 다음과 같다.

<표 8> 심층면접조사 대상

분야	대상(명)	분야	대상(명)
기술자	8	외국인이주민단체	1
법률가	4	인공지능의견발표단체	1
학술연구자	5	장애인단체	1
교원단체	1	지역인권단체	3
빈곤단체	1	평화단체	1
성소수자단체	1	학부모단체	1
소비자단체	1	배달노동자	3
여성단체	1		

코로나19 방역지침으로 인해 전체 심층면접조사는 비대면 서면 및 온라인으로 실시되었으며, 특히 배달노동자 4명에 대해서는 온라인 화상회의 도구를 이용해 집단심층면접(Focus Group Interview) 조사를 수행하였다.

2. 조사문항 설계

조사문항은 인공지능 개발과 활용에서의 인권 가이드라인(초안)의 각 항목 내용에 대한 질문과 함께 자기 분야 관련 사항에 대하여 자유롭게 답변할 수 있도록 구성하였으며, 더불어 신기술 환경에서의 인권 보호 방안, 국가인권위원회의 역할 등에 대해 질문하였다.

인공지능 개발과 활용에서의 인권 가이드라인(초안)

제1장 가이드라인의 의의

제1절 가이드라인의 제정경위

1. 사회 전반에 인공지능(AI: Artificial Intelligence) 활용이 증가함에 따라 채용, 노동, 금융, 행정, 복지 등 거의 모든 분야에서 인간의 기본적 삶과 인권에 영향을 미치는 사례가 증가하고 있습니다.

2. 앞으로도 막대한 양의 빅데이터를 분석하고 학습하는 과정을 통해 다양한 영역에서 사람의 판단을 대신하는 인공지능은 점차 적용영역을 넓혀 사회 전반과 개인의 삶에 강한 영향력과 파급력을 행사할 것입니다.

3. 인공지능의 발전과 확산은 국가경쟁력과 개인 삶의 질을 높일 것으로 기대되지만 기술 오·남용, 데이터 편향성, 개인정보 침해 등과 같은 인권을 침해하는 문제들도 대두되고 있습니다.

4. 그러나 인공지능으로 영향을 받는 개인들은 인공지능의 도입, 운영, 결정에 대하여 발언과 참여 기회를 보장받고 있지 못하며, 인공지능으로 인한 인권 침해가 발생한 경우에도 권리구제를 받을 수 있는 절차와 방법이 마땅하지 않은 상황입니다.

5. 이런 상황에서 인공지능을 개인 삶의 질과 사회적 공익 증진에 기여하도록 설계하며, 인간의 존엄성과 차별금지, 자기결정권 보장 등과 같은 인권적 가치와 기본적인 권리에 기반을 두도록 하는 것이 매우 중요합니다.

6. 이 가이드라인은 인공지능 개발과 활용 전 단계에서 인권침해를 예방하고, 인권

적 관점에서 인공지능의 개발과 활용 과정에서 준수해야 할 기본적이고 원칙적인 기준을 제시하고자 제정하게 되었습니다.

제2절 가이드라인의 목적 및 적용

7. 본 가이드라인은 인공지능의 개발과 활용에 의해 우리 사회의 인권적 가치가 훼손되지 않고 인간의 존엄성을 보장하며 기본권을 실현하는 방향으로 나아가도록 하는데 목적이 있습니다.

8. 본 가이드라인은 인공지능의 개발과 활용 과정에 적용할 인권원칙을 제시하고, 그 이용자, 영향을 받는 당사자 등에게 주어진 권리와 피해구제수단에 관한 사항 및 국회와 정부에게 적절한 법률과 정책을 수립할 수 있는 기준을 제시하고자 합니다.

제3절 가이드라인의 법적 의미

9. 국가인권위원회는 「국가인권위원회법」제19조 제6호에 따라 인권침해의 유형, 판단기준 및 그 예방조치 등에 관한 지침을 제시할 권한을 가지고 있습니다.

10. 본 가이드라인은 국가인권위원회에서 인공지능으로 인한 인권침해 여부와 차별 여부를 판단하는데 고려할 수 있는 기준이며, 개선 등의 권고나 구제 절차를 마련하는데 필요한 해석의 기준이 될 수 있습니다.

11. 다만, 가이드라인 자체가 법률은 아니므로, 가이드라인에 규정된 내용에 따라 새로운 권리가 창설되거나 의무가 부여되는 것은 아닙니다.

제2장 가이드라인의 적용범위 및 정의

제1절 적용 범위

12. 인공지능의 개발과 활용에서 보호해야 하는 인권은 헌법 및 법률에서 보장된 것 뿐만 아니라, 국가인권위원회법, 국제인권법 및 각종 인권규범에서 열거된 것을 모두 포함합니다.

13. 본 가이드라인은 인공지능의 개발부터 활용에 이르는 모든 과정에 참여하는 사

회구성원들을 대상으로 하며, 여기에는 인공지능 개발자, 이용자, 영향을 받는 개인, 정부 및 공공기관, 기업 등이 포함됩니다.

14. 인공지능 개발과 활용에 관한 법률을 제정하거나 개정할 때에는 본 가이드라인의 목적과 기본원칙에 맞도록 노력하여야 합니다.

제2절 정의

15. 일반적으로 4차 산업혁명과 신기술로 불리는 사물인터넷, 빅데이터, 인공지능, 플랫폼 기술 등은 각기 독자적인 기술로 이해할 수도 있으나 전체적으로 하나의 생태계를 구축하고 있습니다.

16. 이중에 핵심적인 위치를 차지하는 인공지능은 막대한 양의 빅데이터를 분석하고 학습하는 과정을 통해 추론과 판단을 하고 사람과 유사한 방식으로 주어진 과제 및 문제를 해결하는 기술로, 점차 적용영역을 넓혀 가고 있습니다.

17. 본 가이드라인에서 말하는 인공지능은 일차적으로는 학습과 추론, 판단을 전자적으로 구현하는 알고리즘과 해당 프로세스를 지칭하나, 이차적으로는 빅데이터, 사물인터넷, 플랫폼 등 인공지능을 기능하게 하는 일련의 기술들을 포함하고 있습니다.

제3장 가이드라인 기본 원칙

제1절 인간의 존엄성

18. 「헌법」제10조에서 보장하고 있는 인간의 존엄과 가치는 누구든지 인간이 누려야 할 불가침의 양도할 수 없는 기본적 인권이며, 이로부터 구체적인 기본권이 나옵니다.

19. 이러한 인간의 존엄과 가치를 바탕으로 외부의 강요 없이 누구나 스스로 판단에 따라 타인의 간섭 없이 결정하고 행동할 수 있으며, 동시에 선택하지 않을 자유도 보장됩니다.

20. 인공지능은 인간의 존엄과 가치를 침해하지 않는 범위에서 개발 및 활용되어야 하며, 개인의 선택과 판단 및 행동을 강요하거나 자율성을 침해할 수 없습니다.

제2절 알권리

21. 알권리란 의사형성에 필요한 정보를 자유롭게 수집하고, 수집된 정보를 취사·선택할 수 있는 자유를 의미합니다.

22. 개인은 알권리를 통해 충분한 정보를 획득하고 지식과 이해의 폭을 넓힐 수 있고, 이를 바탕으로 합리적인 판단과 자신의 인격을 발현시킬 수 있습니다.

23. 개인의 의사 판단과 인격 발현의 중요한 요소인 알권리는 인공지능으로 인해 간과될 수 있는 인간의 가치와 자율성을 보장하는데 중요한 권리이므로 알권리를 보장할 수 있도록 인공지능의 판단 과정 및 결과에 대한 합리적인 설명 등을 보장하고 이를 뒷받침하는 기술적, 제도적 장치를 마련해야 합니다.

제3절 자기결정권

24. 「헌법」제10조가 규정하고 있는 인간의 존엄과 행복추구권에서 도출되는 자기결정권은 결정의 주체인 개인이 중대한 사안에 대해 스스로 결정할 권리를 의미합니다.

25. 자기결정권은 인간으로서 존엄과 가치를 유지함과 동시에 인간이 여타 권리를 향유하기 위한 전제조건이기도 합니다.

26. 인공지능은 인간을 보조하여 서비스를 제공하는 것 뿐만 아니라 스스로 판단하고 추론하여 인간만이 가지고 있던 결정 권한을 부여받게 될 수준에 이르렀습니다.

27. 따라서 자연권적 권리이며 개인 인격의 기본적 가치를 의미하는 자기결정권의 보장은 인공지능을 개발하고 활용하며 자동화된 의사결정에 이르는 모든 과정에서 우선적으로 이행되어야 합니다.

제4절 평등과 차별 금지

28. 「헌법」제11조는 모든 국민에게 평등권을 보장하고 있으며, 「국가인권위원회법」제2조는 합리적인 이유 없이 성별, 장애, 나이, 출신 지역, 용모 등 신체조건, 피부색, 성적 지향 등 개인 특성에 따라 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위를 평등권 침해의 차별행위로 정의하고 있습니다.

29. 이렇듯 모든 사람은 평등하고, 어떠한 이유로도 생활의 모든 영역에서 구별, 배

척, 제한을 받지 아니하며, 인권의 인정과 향유를 보장받을 권리가 있습니다.

30. 인공지능의 개발과 활용은 개인의 행복과 사회적 공공성의 증진을 목표로 하며, 인공지능을 통한 경제·사회·문화적 권리의 향유에 있어 다양한 계층을 포용하고 참여 기회를 동등하게 보장해야 합니다.

31. 또한 인공지능의 결정이 특정집단이나 일부 계층에게 차별적이거나 부당한 영향을 초래하지 않기 위해 다양한 계층의 의견을 수렴하고, 당사자의 안전이나 권리에 영향을 미치는 인공지능의 경우 차별적 결과가 발생하지 않도록 하는데 필요한 조치를 취해야 합니다.

제4장 가이드라인 핵심 과제

32. 본 가이드라인은 국제인권법 및 「헌법」이 보장하는 인권적 기본 원칙을 구현하기 위해 다음의 5가지를 인공지능의 개발과 활용에 있어 적용되어야 하는 핵심과제로 선정하였습니다.

제1절 투명성과 자기결정권 보장

33. 학습, 추론, 판단의 과정과 결과에 이른 이유를 설명하기 어려운 인공지능의 특성은 이에 대한 대응의 불확실성과 영향을 받는 당사자의 불안감을 유발하고, 인권 및 안전에 관한 법령과 정책의 집행효과는 불분명할 수 있습니다. 인공지능 및 이를 이용한 의사결정이 개인에게 미치는 영향력이 갈수록 증가하므로 인공지능의 판단과정과 그 결과에 대한 적절하고 합리적인 설명이 보장되어야 합니다.

34. 공공기관은 직접 또는 조달을 통하여 개발하거나 활용하려는 인공지능에 대하여 원칙적으로 사전 공개하고 그 영향을 받는 사람들의 의견을 공청회 등으로 수렴하여야 합니다. 공공기관이 사람에 대한 의사결정을 할 때 설명할 수 없는 인공지능을 활용해서는 안 됩니다.

35. 공공기관이 개발하고 활용하는 모든 인공지능과 민간이 개발하고 활용하는 일부 인공지능(당사자의 안전이나 권리에 큰 영향을 미치는 인공지능, 이하 ‘고위험 인공지능’)의 경우 원칙적으로 그 학습데이터와 인공지능 알고리즘의 주요 요인에 대

하여 일반에 공개하고 설명하여야 합니다.

36. 인공지능 기반으로 자동화된 의사결정이 이루어질 경우, 영향을 받는 당사자는 사전에 의사결정이 예정되어 있다는 사실, 사용된 데이터, 의사결정의 주요 요인, 의사결정의 논리에 대한 정보 등을 통지받고 의견을 진술할 수 있어야 합니다. 영향을 받는 당사자는 원칙적으로 자동화된 의사결정 방식을 거부하거나 인적 개입을 요구할 수 있는 권리 또한 행사할 수 있어야 합니다.

37. 특히 인공지능으로 당사자에게 법적이거나 상당한 영향을 미치는 의사결정을 오로지 자동화된 방식으로 수행하는 일은 원칙적으로 제한되어야 합니다. 이러한 의사결정이 이루어진 경우 영향을 받는 당사자는 그 의사결정 사유에 대한 설명을 들을 수 있어야 하며, 이의를 제기하거나 구제를 받을 권리를 보장받아야 합니다.

38. 인공지능과 상호작용하는 사람에게는 언제나 그 상대방이 인공지능이라는 사실을 알려야 합니다.

39. 감독 당국은 공공기관과 민간의 위법한 인공지능 개발과 활용 여부를 조사하고 피해 구제 및 조치를 취하기 위하여 상세 정보에 접근할 수 있어야 합니다. 이를 위하여 공공기관 인공지능 및 민간 고위험 인공지능의 경우 사용된 데이터셋과 알고리즘의 주요 요소에 대하여 기록하고 문서화하여 일정 기간 보관하여야 합니다.

제2절 정보주체권리 보장

40. 인공지능은 학습용 데이터로 학습하고 다양한 데이터를 활용하여 새로운 데이터를 만드는 기술로서, 개인이 데이터가 사용되는 방법을 이해하고 그에 대한 통제권을 가지는 것이 중요합니다. 정보주체는 인공지능이 언제 어디서 자신의 데이터를 수집하고, 데이터가 어떻게 처리되어 사용, 보관, 삭제되는지에 대해 알고 참여할 권리가 있습니다. 대상 정보가 가명처리된 후라도 시간·비용·기술 등을 합리적으로 고려할 때 불가능하거나 처리 목적 달성을 명백하게 저해하지 않는 한, 원칙적으로 정보주체의 권리를 보장하는 것이 바람직합니다.

41. 인공지능의 개발과 활용에서 개인정보를 수집하고 이용하여 처리할 때는 적법하고 공정하고 투명하여야 합니다. 개인정보는 목적에 필요한 범위에서 최소한의 개

인정보만을 처리하여야 하며, 처리목적 달성에 필요한 기간 동안만 보관되어야 합니다. 이러한 개인정보에 대한 처리 방침은 정보주체가 언제든지 확인할 수 있도록 공개되어야 합니다.

42. 인공지능의 개발과 활용에서 민감정보를 처리할 때에는 특별한 주의를 기울여 보호하여야 합니다. 더불어 정보주체가 자신과 관련성이 없고, 부정확하며, 지나간 시점의 데이터에 기반한 의사결정의 대상이 되지 않도록 데이터의 정확성, 완전성 및 최신성을 보장해야 합니다.

43. 개인정보와 관련된 정보주체의 권리는 개인정보 열람권, 정정 및 삭제권, 처리정지권 등으로 인공지능의 개발 및 활용의 전 과정에서 보장되어야 합니다.

제3절 편향·차별 금지

44. 인공지능 개발과 활용은 반드시 모든 사람의 다양성과 대표성을 반영하고, 성별·장애·나이·출신 지역·용모 등 신체 조건·피부색·성적 지향 등 개인 특성에 따라 편향적이고 차별적인 결과가 나오지 않도록 해야 합니다.

45. 데이터의 수집·선정 및 시스템 설계, 활용 등 인공지능 개발 전반에 걸쳐 편향이나 차별을 배제하는 것을 우선순위로 두어야 하며, 이는 데이터 요소를 검사하고 조정하여 차별적인 데이터를 제거하는 등의 조치를 포함합니다.

46. 특히 훈련용 데이터가 인공지능의 판단에 직접적인 영향을 미치는 상황을 고려할 때, 훈련용 데이터의 수집 단계부터 차별적 요소를 통제하고 데이터 편향성을 최소화하여 인공지능을 통한 의사결정이 특정 집단에 부정적 영향을 미치지 않도록 해야 합니다.

47. 개발한 인공지능 알고리즘의 사용 전·후에 모니터링과 검사를 거쳐 데이터 품질과 위험을 관리하고, 차별적이거나 의도치 않은 결과에 대해 적극적인 조치를 주기적으로 수행해야 합니다.

48. 인공지능 기술 및 서비스에 대한 접근성과 인공지능이 주는 혜택은 사회적 약자와 취약계층을 포함하여 모든 사회구성원에게 동등하게 제공되어야 합니다.

제4절 인권영향평가 시행

49. 국가는 인공지능의 개발과 활용에 있어 인권적 가치가 우선시 되도록 하여야 하며, 인공지능으로 인해 발생할 수 있는 인권침해와 차별에 대하여 사전적 또는 사후적으로 관리 감독을 할 의무가 있습니다. 국가는 공공기관 인공지능 및 민간 고위험 인공지능에 대하여 인권에 미치는 영향을 평가하고 조치하는 제도를 개발하고 도입하여야 합니다.

50. 인권영향평가 내용에는 인공지능의 특성, 상황, 범위 및 목적을 고려하여 본 인권 가이드라인이 제시한 기본원칙과 핵심과제의 준수 여부가 포함되어야 합니다. 인권영향평가는 개발 및 출시 전에 실시하고 인공지능의 기능 또는 범위 변경 시 평가를 갱신하여야 합니다. 평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하기 위한 조치사항을 수립하여 적용하여야 하며, 원칙적으로 그 내용이 공개되어야 합니다. 공공기관 인공지능 및 민간 고위험 인공지능의 경우 독립적인 외부전문가의 평가와 정기적인 감사를 받아야 합니다.

51. 특히 공공기관은 직접 또는 조달을 통하여 개발하거나 활용하려는 모든 인공지능에 대하여 인권영향평가제도를 실시하고, 평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하는 조치를 취하기 전에는 그 개발과 활용을 중단해야 합니다.

제5절 인권을 보장하는 법제도 개선

52. 국회와 정부는 공공기관과 민간이 개발하고 활용하는 모든 인공지능에서 인권과 책임성을 완전하게 보장하도록 적절하게 규율하는 법률과 감독 체계를 마련하여야 합니다. 특히 정부는 공공기관이 직접 또는 조달을 통하여 개발하거나 활용하려는 인공지능에 대하여 인권을 보장하기 위한 구체적인 지침과 정책을 마련하여야 합니다.

53. 국가는 인공지능을 감독할 수 있는 체계를 수립하여 인권과 안전을 보장하고 피해를 구제하여야 합니다. 인공지능 국가 감독 체계는 독립적이고 효과적이어야 하며, 진정 또는 인지로 접수한 사건을 조사하기에 충분한 자원, 권한 및 전문지식을 구비해야 합니다.

54. 국가는 인공지능으로 인하여 피해를 입거나 인권을 침해당한 사람에게 진정을 접수하는 등 국가기관의 구제수단에 대한 접근을 보장하여야 합니다. 인공지능을 개발하고 활용하는 공공기관과 민간은 언제든지 구제가 가능하도록 그 책임자에 대한 정보는 물론, 이의를 제기할 수 있는 기관과 방법에 대한 정보를 일반에 공개하여야 합니다.

55. 국가는 인권과 안전에 미치는 위험성이 매우 높아 금지되는 인공지능 영역으로부터 조건이 필요한 고위험 인공지능 영역, 위험성이 거의 없는 영역을 구분하여 위험성 정도에 맞는 규제 수준을 정하고, 적절한 수준의 인적 개입을 위한 법제도를 마련하여야 합니다.

56. 특히 국가는 대량 감시와 차별로 이어질 위험이 높은 얼굴인식 등 원격 생체인식 기술의 사용을 원칙적으로 금지하고 예외적인 사용에 대한 법원의 통제 등 인권을 보호하기 위한 제도를 시급히 마련하여야 합니다. 또한 국가는 생명의 존엄성 및 인간의 역할과 책임 윤리를 훼손할 가능성이 높은 자율살상무기에 대하여 인도주의적으로 접근하고 그 연구, 개발, 생산 및 활용을 금지하는 국제 규범 및 이에 대한 논의에 적극적으로 참여해야 합니다.

제2절 개별서면조사 결과

1. 인권 가이드라인(초안)에 대한 의견

귀하께서는 인공지능 개발과 활용에서의 인권 가이드라인(안)의 내용에 대해 어떻게 생각하십니까? 보완 및 개선을 위한 내용을 작성해주십시오.

가. '제1장 의의'에 대한 의견

분야에 상관없이 다수의 응답자들은 가이드라인의 제정경위, 목적 등이 대체로 적절하다는 의견이었으나, 가이드라인의 범용성과 권고 활용가능성 등을 고려해 주요 용어와

적용 대상을 좀 더 명확히 할 필요가 있다는 의견도 보였다.

일부 기술분야 응답자는 현 시점에서 인공지능 기술이 사람의 판단을 대체하는 경우보다는 전통적인 컴퓨터 알고리즘으로 처리하던 부분을 대체하는 경우가 훨씬 많다는 점을 고려할 때 가이드라인에서 사용한 '사람의 판단을 대신하는 인공지능'이라는 용어 정의가 대상이 되는 인공지능을 축소 해석시킬 수도 있다는 우려를 보였다. 또 다른 기술분야 응답자는 지역 방언 음성인식 인공지능을 예시로 들며 편향적 결과의 해소를 위해 편향된 데이터 구축이 오히려 필요한 경우가 있기에 데이터 편향성의 용어 사용에 신중해야 한다는 의견도 주었다.

또한 인공지능이 활용되는 분야에 대한 예시적 열거에서 중요한 분야인 상거래, 의료, 군사, 교육 분야를 추가해야 한다는 의견도 각각 제시되었다.

한편, 이 가이드라인의 내용이 향후 국가인권위원회 권고 및 법령 제정 등으로 이어져 규범력을 갖출 방안이 모색되어야 한다는 의견도 다수 있었다.

"가이드라인이 구체적으로 어떻게 활용될 수 있을지에 대한 이야기가 좀 더 구체화되면 좋겠습니다. 통상 가이드라인이라고 하면, 사적 기관(기업)에도 적용될 수 있는 원칙을 말하기도 합니다. 이 가이드라인은 오로지 국회와 정부만 참조할 수 있는 것인가요? 지방자치단체 등 다른 공공기관도 포함될 수 있을 것입니다. (중략) 가이드라인이 목표로 하는 지점을 명확하게 하면 좋겠습니다. 인권위에서 활용한다면, 그 자체로 '정책 권고'에 활용될 수도 있을 겁니다. 그러니까 가이드라인을 활용해 필요한 법률을 제정하도록 국회에 권고할 수도 있고요. 이런 부분이 좀 더 명확히 기술되면 좋겠습니다." {학술연구자}

나. '제2장 적용범위 및 정의'에 대한 의견

다수의 법률 및 학술분야 응답자가 이용자, 영향을 받는 개인 등 모호한 표현을 명확히 해야 한다는 의견을 보였다. 관련하여 외국인이주민단체 응답자의 의견은 다음과 같다.

“인공지능은 사회 구성원 모두에게 영향을 끼칠 수 있다는 점에서 단순히 개발자나 그 과정에 참여하는 자만 아니라 그 영향을 받는 모든 이들의 권리 보호와 피해 구제 문제도 포함해야 한다고 봅니다. 특별히 스스로 목소리를 내기 어려운 국내 체류 이주노동자들은 편견을 학습한 인공지능에 의해 피해를 볼 수 있다는 점이 이미 드러나고 있다는 점에서 관련 규정을 명확히 할 필요가 있습니다.” {외국인이주민단체}

‘인공지능’ 용어 정의에 대한 의견도 다수 있었다. 먼저 주요 인공지능 기술 중 하나로 딥러닝을 예시하고, 빅데이터에 전혀 의존하지 않거나 비학습형 인공지능 또한 포함할 수 있도록 기술하는 등 정의의 확대가 필요하다는 의견이 있었다. 반면에 어느 법률분야 응답자 인공지능의 용어 정의에 있어 “빅데이터, 사물인터넷 플랫폼 등 인공지능을 가능하게 하는 일련의 기술들을 포함” 할 경우 범위가 너무 넓고 불확실성과 불투명성 등 인권에 위험한 인공지능 기술 고유의 속성과는 거리가 있다는 우려를 보이기도 하였다. 또 다른 법률분야 응답자는 최근 유엔 문서 등이 주로 인용하는 정의 (“The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages”)를 부분 차용할 것을 제시하였다.

그외 적용범위에 국가인권위원회법과 국제인권법 뿐만 아니라 군사적 인권 침해 문제를 고려하여 국제인도법을 명시할 것, 아동복지법과 장애인복지법 등 포괄할 수 있는 국내법을 명시할 것, 또는 국가인권위원회법상의 자구를 차용할 것 등의 구체적 의견이 있었다.

다. ‘제3장 기본 원칙’에 대한 의견

기본원칙의 각 항목들에 대해 다양한 의견이 있었다. 먼저 분야에 상관없이 ‘알 권리’와 ‘자기결정권’이 중요하다는 의견이 있었다. 관련해 어느 지역인권단체 응답자는 복지 영역에서 인공지능이 활용되는 경우 응급과 안전을 명분으로 신기술에 대한 이해가 미흡한 노인과 발달장애인에게 충분한 사전 설명 없이 형식적인 동의로 사업이 진행되는 예시를 들며 이에 집중되어야 할 필요성을 제시하기도 했다.

“매우 동의한다. 현재 복지영역에서 복지사업의 일환으로 시범적으로 진행되고 있는 사업들을 보면 ‘응급과 안전’을 명분으로 사전에 충분히 이익과 불이익이 설명되지 않는 점, 데이터 수집의 범위의 포괄성, 형식적인 동의(특히 신기술에 대한 이해가 미흡한 노인, 발달장애인) 가 문제되고 있다. 또한 시범사업 참여에 대한 결정권의 보장도 형식적이다. 복지서비스 대상자가 복지서비스를 선택할 수 없는 현실의 문제가 인공지능을 활용한 복지사업에는 더 두드러지게 나타나고 있다. 이는 권리의 주체가 아니라 관리의 대상으로 전락하는 것을 강화하는 측면이 있다.” {지역인권단체}

가이드라인의 기본 원칙에 사생활의 비밀과 자유, 표현의 자유, 실효성 있는 구제를 받을 권리 등을 추가해야 한다는 법률 분야 응답자의 의견이 다수 있었다. 가이드라인에 언급되지 않은 기본권의 관련성이 배제될 경우를 대비해 보완적 설명이나 언급이 필요하다는 의견이었다.

또한 평등과 차별 금지 원칙이 보다 구체화해야 한다는 의견이 있었다. 성소수자단체 응답자와 장애인단체 응답자는 각각 최근 국가인권위원회가 발의한 평등법 개정안의 기준을 반영할 것, 반대하는 차별적 결과가 무엇인지 설명할 것 등의 의견을 제시했다. 관련해 어느 학계 응답자는 기본 원칙 전반에서 국내법 뿐 아니라 국제적인 인권 규범 및 인공지능과 인권 관련 국제적인 가이드라인 등 국제적이고 보편적인 원칙을 포함해야 한다는 의견을 제시했다.

한편 인공지능의 개발과 활용의 목표와 차별 금지 조치와 관련하여, 방향성은 타당하지만 이를 실천하는 것이 빅테크 기업 등 거대 자본력을 갖춘 곳만이 가능하다면 인공지능 개발과 활용 기회가 오히려 박탈될 수 있으므로 보다 세밀한 접근이 필요하다는 의견이 있었다.

라. ‘제4장 핵심 과제 제1절 투명성과 자기결정권 보장’에 대한 의견

다수의 응답자가 투명성과 자기결정권 보장이 인공지능과 인권에 있어 핵심적이며 일정한 조치가 필요하다는 점에 공감하였다.

“ ‘영향을 받는 사람의 의견을 수렴해야 한다. 사람에 대한 의사결정을 할 때 설명할

수 없는 인공지능을 활용해서는 안 된다’ 는 내용은 인공지능을 개발하고 사용하고자 하는 주체라면 누구든 인지해야 하는 지점이라고 생각합니다.” {여성단체}

다만 구체적인 투명성 보장 방안에는 다소간의 입장 차이가 존재했다. 기술 분야 응답자들의 경우 ‘합리적인 설명’, ‘설명가능성’에 대하여 연구가 이루어지고 있긴 하지만 구현하기 어렵고 그 의미 또한 모호하다는 의견들이 다수 있었다. 반면 설명가능한 인공지능 기술을 긍정적으로 평가하는 기술 분야 응답자도 있었는데, 이 경우에도 ‘적절하고 합리적인 설명’ 또는 ‘의사결정’ 등의 개념을 구체화할 필요하다고 보았다. 한 기술 분야 응답자의 경우 안전하고 품질 높은 공공기관 인공지능을 위해 오픈소스로 기술을 공개할 것을 제안했다. 법률 분야와 학술 분야의 일부 응답자는 투명성, 설명 의무, 공개 의무에 있어 무엇을 어떻게 공개할 것인지 보다 적절한 수준을 제시해야 하다는 의견이었다.

한편 시민단체, 법률 및 학술 분야의 다수 응답자가 투명성과 자기결정권 보장이 공공기관을 넘어 민간을 포함시키는 등 보다 포괄적이어야 한다는 의견을 보였다. 이와 관련 공공기관의 경우 정부와 지방자치단체 포함을 명시해야 한다는 점, 공공기관과 국내 기술 기업의 협력 동향을 보았을 때 준공공적인 성격이 이미 존재하는 민간에도 이에 준하는 의무를 부여할 필요가 있다는 점, 구제조치를 고위험에 한정하는 것은 피해자의 권리를 제약할 우려가 있다는 점, 고위험으로 한정하는 것은 소비자 보호를 지나치게 협소화할 수 있다는 점 등이 지적되었다.

그 외에도 인공지능을 기반으로 자동화된 의사결정이 이루어질 경우 이를 거부하는 당사자를 위한 대안을 마련해야 실효적이라는 의견이 있었다.

마. ‘제4장 핵심 과제 제2절 정보주체권리 보장’에 대한 의견

정보주체의 권리 보장에 있어서는 적절하다는 의견이 다수였으며, 다만 정보주체, 학습용 데이터, 개인정보, 민감정보 등의 용어 사용에 있어 그 의미와 범위 해석에 논란이 없도록 직접 정의하거나 법령을 인용하는 등 구체화해야 한다는 의견이 있었다.

나아가 외국인이주민 단체와 장애인 단체 응답자는 개인정보 수집과 폐기 등 정보주

체의 권리가 형식적으로만 보장되는 상황을 고려할 때 실질적이고 예외 없는 보장을 명문화하는 등 정보주체 권리를 지금보다 강화해야 한다는 의견을 제시했다. 반대로 한 법률 분야 응답자는 가명 정보의 위험성에 대해서는 논란이 있을 수 있으나, 연구 목적 등 일정한 요건 하에서 활용이 가능한 현행 법령에 부합하는 완화된 서술이 바람직하다는 의견이었다.

한편 학부모단체의 응답자는 교육 목적으로 정보주체의 권리가 손쉽게 침해되는 상황을 예시로 들며 학교 교육에서 활용되는 인공지능의 경우 수집된 학생 개인정보를 교육 이외의 목적이나 동의 없는 평가 자료로 활용되는 일이 없도록 보호하여야 한다는 의견을 밝혔다.

"학교에서 교육하는 도구로 인공지능을 활용할 경우 무한정 데이터를 수집 활용할 우려가 있습니다. 학생에게 수학을 가리키기 위해 인공지능을 활용할 경우 학생의 취향, 버릇, 수학에 대한 이해도나 수준 같은 자료들을 수집해 활용할 것입니다. 이 자료가 수학교육 이외의 목적, 학생이나 보호자 동의 없이 상위 학교 진학 자료로 활용한다든가 성취도 수준을 매겨 서열화한다든가 해서는 안됩니다." {학부모단체}

바. '제4장 핵심 과제 제3절 편향·차별 금지'에 대한 의견

응답자들은 전체적으로 편향과 차별을 방지해야 한다는 데 동의했지만 그 구현에 대하여 다수의 기술 분야 응답자가 우려를 표했다. 데이터 통제의 한계, 자본력의 부족, 사회적 기준의 부족 등을 문제점으로 들었다. 기술적 어려움을 해소하기 위한 노력과 함께 차별을 이해하고 시정하기 위한 사회적 접근이 필요하다는 의견도 있었다.

"전체적으로 동의가 가고 필요한 조치들이라고 생각이 되는데 실무적으로 이 부분을 어떻게 처리해야할지가 관건일 것 같습니다. 차별성을 걸러내기 위해서 데이터를 통제하는 것이 어느 수준에서 가능한 것일지 감이 잡히지는 않습니다. 단순히 차별적 표현을 걸러내는 정도는 아주 어렵지 않게 구현을 할 수 있더라도, 데이터셋에서 나타나는 차별적 경향을 다시 인위적으로 보정해주는 것은 단순히 기술적인 차원을 넘어서는 문제일 것 같다는 생각도 듭니다. 사회 구조 속에 뿌리박힌 차별을 이해하고, 그러한

요인이 배제될 수 있도록 세밀하게 조정을 할 수 있는 역량은 기술 외에도 사회학적 접근도 필요해보여서요. 이런 고민이 들기도 하지만, 그럼에도 불구하고 가이드라인에서 차별금지 원칙을 명시해주는 것은 필요해보입니다." {기술자}

관련하여 어느 법률분야 응답자는 인공지능 개발, 모니터링 등 산업에 참여하는 사람들에 대한 인식교육 및 참여자들의 다양성 보장이 의무적이어야 한다는 의견이었다.

한편 편향과 차별을 방지하기 위한 구체적 방안을 제시하는 의견들도 있었다. 차별 발생 유무에 대해 당사자 집단이 참여하고 판단하는 방안, 인공지능 알고리즘의 사용 전후 모니터링과 검사 등을 조치하고 결과를 공표하는 방안, 기업이 다양한 공정성 기준을 설정하고 공개하는 방안 등이 제시되었다.

그 외 예시된 차별 목록에 '경제적 지위', '국적과 체류 자격' 등을 포함시켜야 한다는 의견, 아동과 노인 그리고 장애인 등 사회적 약자와 소수자의 권리가 보다 구체적으로 명시되어야 한다는 의견, 국가인권위원회법에 열거된 현행법적 기준은 물론, 유전, 성정체성 등 국제인권규범에서 인정된 기준을 포함해야 한다는 의견이 있었다.

사. '제4장 핵심 과제 제4절 인권영향평가 시행'에 대한 의견

인권영향평가의 구체적인 시행 내용에 있어 보완 및 수정, 문제점을 제기하는 다양한 의견이 있었다. 어느 기술분야 응답자의 경우 전체적으로 가능해보이지만 데이터를 학습해가며 결과가 개선되는 인공지능 기술의 특성 상 개발 전 인권영향평가가 제대로 된 평가 결과를 가져올 수 있을지 의문을 표했으며, 이에 대해 개발 및 출시 전 평가하는 것이 아닌 출시 전 평가로 대체할 것을 제시했다. 관련하여 인권영향평가 과정에서 데이터 편향성에 대한 구체적 검증이 필요하며 그렇지 않으면 현상에 대한 평가가 무의미할 여지가 있을 수 있다는 어느 법률분야 응답자의 의견도 있었다. 한편 인권영향평가가 사전예방적 성격만 띄고 있다는 점을 지적하고 개발 과정에서 드러나지 않았던 인권침해적 요소들이 운영과정에서 발생할 수 있으므로 수시 혹은 정기적 모니터링이 필요하다는 의견도 있었다.

인권영향평가의 실질적인 시행 방안이 부족함을 우려하며 이를 강제할 수 있는 방법

모색, 독립적인 외부전문가 확보 등의 의견을 제시하기도 했다. 관련하여 다음과 같은 의견이 있었다.

"인공지능에 의한 인권 침해 사례가 실제 발생하고 향후 증가될 가능성이 있으므로, 국가가 주도하여 인권영향평가를 시행할 필요가 있다고 생각합니다. 국가인권위원회가 중요한 역할을 할 수 있을 것 같습니다. 다만「지능정보화 기본법」제56조에서 '지능정보서비스 등의 사회적 영향평가'를 규정하고 있어, 조화로운 입법 및 정책 시행이 필요해 보입니다." {법률가}

인권영향평가의 시행과 그 책임을 묻는 내용에 있어 대상에 대한 구체적 설정과 구분이 필요하다는 의견도 있었다. 관련하여 어느 기술분야 응답자의 의견은 다음과 같다.

" '인공지능 기술'의 연구와, 타인이 연구발표한 '인공지능 기술'에 자신만의 데이터를 접목시켜 '인공지능 학습'을 시키는 행위와, 학습시킨 인공지능 모델을 활용하는 서비스를 출시하는 행위 등은 각각 밀접하게 얽혀 있으면서도 또 따로 떨어져 있습니다. 제4절은 이런 각각을 별도로 떼어놓지 않고 '인공지능'으로 통칭하고 있는 인상을 줍니다. 이런 부분에서 조금 더 엄밀한 구분을 하는 방향으로 보완이 되면 좋을 듯합니다." {기술자}

그 외 인권영향평가가 고위험 인공지능으로 한정되는 문제에 대한 우려, 자동화된 의사결정을 거부할 권리 및 인적개입을 요구할 권리를 위해 디지털이 아닌 옵션이 언제나 제공되어야 한다는 지적, 고위험인 군사분야 인공지능의 위험성과 불투명성을 감안했을 때 국가 차원에서 이를 점검하고 감시하기 위한 독립적인 언급이 필요하다는 의견이 있었다.

아. '제4장 핵심 과제 제5절 인권을 보장하는 법제도 개선'에 대한 의견

먼저 인공지능 감독 체계의 수립에 있어 소비자단체 응답자의 다음과 같은 의견이 있

었다.

"인공지능을 감독하기 위한 감독체계는 산업정책과는 독립적이고 이해상충의 소지가 없어야 한다는 점을 명시하는 것을 고려하기 바랍니다. 아울러 금융(금융위), 전기통신(방통위), 경쟁정책 및 소비자보호(공정위), 개인정보보호 등 세부분야에서도 인권보호를 위한 인공지능 정책을 세부적으로 추진 및 감독해야 한다는 점을 포함하는 것이 보다 구체적인 정책 실현의 수단이 될 것으로 보입니다." {소비자단체}

동시에 감독 체계에 있어 정보공개 청구가 가능하도록 명시할 것, 감독 체계가 사후적인 구제조치만을 담당하는 게 아닌 인권영향평가 등 사전 예방 감독 권한도 가져야 한다고 명시할 것을 제시하는 등 감독 체계의 보완 및 강화에 의견이 있었으며 관련한 어느 법률전문 응답자의 의견은 다음과 같다.

"구제의 권리에 요구되는 관점은 피해자 중심적 접근이고, 피해자의 권리는 진실, 정의, 배상의 권리로 구분할 수 있습니다. 배상의 권리에는 금전적 배상뿐만 아니라, 원상회복을 위한 제도적 지원, 관련 책임자들에 대한 적절한 형사적, 행정적 제재 및 재발방지 대책의 마련까지 포함이 되는데 이러한 부분들이 해당 부분에 반영이 될 수 있으면 좋을 것 같습니다." {법률가}

이어서 별도의 구제 접근 수단을 마련하는 것보다는 인권침해 진정 체계 내에서 이를 포함할 수 있도록 하는 게 적절할 것이라는 의견도 있었다.

한편 얼굴인식 기술 등 원격 생체인식 기술의 사용 금지에 대한 내용에 있어 모든 얼굴인식 기술을 금지시키는 것으로 오인되는 것을 막기 위해 명확한 정리가 필요하다는 의견과 원칙적 금지가 필요한 경우 사회적 합의와 실증적 연구의 뒷받침이 필요해 보인다는 의견이 있었다.

그 외 평화단체 응답자는 제5절처럼 다른 절과 항목에서도 인공지능의 군사 분야에서의 개발과 활용에 대한 가이드가 구체적으로 언급될 필요가 있다는 점을 제시했으며, 관

련해 어느 기술분야 응답자의 다음과 같은 의견도 있었다.

"자율살상무기의 경우, 인도주의적인 접근은 현실적으로 실효성이 없기 때문에, 특히 민에 의한 통제와 관리가 매우 필요해 보이는 영역이라고 생각됩니다. 자율살상 무기 사용과 관련된 규제입법 제안이나, 자율살상무기 사용에 있어 국회 또는 독자 독립기구를 통한 통제와 관리 감독이 필요하지 않을까요?" {평화단체}

2. 일반질문

인공지능의 개발과 활용 단계에서 고려해야 하는 인권 문제와 관련해서, 가이드라인(안)에 담기지 않은 다른 인권 기준이나 가치가 있다고 생각하십니까? 그렇게 생각하신다면 그것이 무엇이며 어떻게 반영되어야 한다고 생각하시는지를 자유롭게 작성해주세요.

다수의 법률가와 학술연구자 등이 사생활의 비밀과 자유에 대한 보호가 추가되어야 한다는 의견을 보였다. 가이드라인 내 개인정보보호에 대한 문항이 일부 있지만 이와 별개로 사생활권은 독자적인 의미에서 인정받은 권리인 점, 인공지능 기술과 이를 위한 데이터 수집이 사생활 침해를 강화할 수 있는 점을 고려하여 책임성을 강화하거나 지나치게 침해 위협이 큰 기술의 개발과 활용을 제한할 필요가 있다는 것을 제시하였다.

"사생활의 비밀과 자유 보호가 개인정보보호와 별개로 포함되어야 한다고 생각합니다. 사물인터넷, 빅데이터, 인공지능, 플랫폼 기술 등을 통해 기술이 개인의 사생활을 인지하게 되고, 기술을 개발하는 과정에서도 학습 데이터를 통해 사생활과 관련된 부분을 '예측' 하려는 노력을 기울이는데, 지나치게 사생활 관련 정보를 수집하거나 사생활을 침해할 가능성이 있는 기술은 개발과 활용을 제한할 필요가 있습니다." {법률가}

아울러 인공지능 기술 환경의 공적 책임과 이용에 대한 다수의 의견이 있었다. 평등하

고 접근성이 보장된 서비스 개발을 위한 개발자와 서비스 제공자에 대한 교육과 동시에 디지털 시대를 살아가는 시민으로서 필요한 지식을 평등하게 제공받을 수 있는 디지털 리터러시에 대한 권리를 제시하기도 했다. 관련한 논의가 다양한 재화와 서비스 접근에 대한 차별 해소 뿐만 아니라 인공지능에 의한 피해나 인권침해를 인지하고 구제 요구를 위해서도 필요하다는 의견도 있었다.

"디지털 기술과 모든 인간의 권리에 대한 교육, 이에 필요한 장비 혹은 프로그램에 접근하기 위한 비용과 시간을 보장해야 합니다. 이는 디지털 지식 격차가 이용 수준의 격차로 이어지고, 사회서비스를 비롯한 다양한 재화와 서비스 접근의 차이로 귀결되지 않기 위해 필요하고, 인공지능에 의한 피해나 인권침해를 인지하고 구제를 요구하기 위해서도 필요합니다." {빈곤단체}

또한 공적인 인공지능 서비스를 민간이 아닌 공공기관이 적극적으로 제공해야 한다는 의견, 인공지능으로 창출되는 경제적 이익이 널리 공유되어야 한다는 의견도 있었다.

"최근 AI 서비스 동향을 보며 가장 고민되는 부분은 '피해자가 돈을 내야 하는 AI 서비스'로 많이 기우는 듯한 모습입니다. 공공이 해야 할 영역을 민간 기업의 인공지능이 대체하면서, 오히려 보호 받아야 할 피해자들이 돈을 지불해 보호를 요청하는 듯한 상황이 되는 것 같아요. 이런 것을 윤리 가이드라인만으로 제재하긴 어려울 것이고, 나아가서는 공공에서 해야 할 기술을 민간에서 개발한 것이기 때문에 오히려 공공-민간의 협력이 더 중요하게 여겨집니다. 공공기관이 좀 더 적극적으로 이런 서비스를 공유재화하여 직접 서비스를 제공하는 것이 가장 이상적이지 않을까 생각합니다." {기술자}

한편 자율살상무기와 같은 물리력 행사, 노동권, 의료 정보 활용 등 인공지능이 활용되는 특정 영역의 보호 필요성에 대한 구체적인 의견도 있었다.

국가인권위원회는 인권 보호와 향상을 위한 업무를 수행하는 독립적인 국가기관입니다. 인공지능 등 새롭게 개발되고 사용되는 기술 환경에서 발생할 수 있는 기본권 침해를 방지하기 위해 국가인권위원회가 어떤 역할을 해야 한다고 생각하십니까? 그런 역할을 정립하기 위해 필요한 요건이 있다면 무엇이라고 생각하십니까?

분야에 상관없이 국가인권위원회가 인공지능에 대한 전문성과 역량을 강화해야 한다는 의견이 다수였다. 인공지능과 같은 신기술의 특성 상 기존에 알려진 유형의 인권 문제보다는 새로운 유형의 인권 문제가 발생하는 경우가 많다는 점을 고려할 때 국가인권위원회에서 선도적으로 문제를 발견하고 조치하거나, 가이드라인과 같은 구체적 대안을 제시하거나, 챗봇 이루다 사건 등 진정 사례를 보았을 때 차별과 폭력을 해소하기 위한 기준을 마련하고 신속한 피해 구조를 위해 인공지능 관련한 구제가 긴급구제 조치에 포함되어야 한다는 의견이 있었다.

"인공지능 기술 특성상 기존에 알려진 유형의 인권문제가 생기는 경우보다는, 기존에 알려지지 않았던 유형의 인권 문제가 창발적으로 발생하는 경우가 많습니다. AI 전문가들이 이를 지적하면 국가인권위원회가 이것을 쫓아가는 형태보다는, 국가인권위원회에서 자체적으로 인공지능 전문성을 갖추고 발생할 수 있는 문제를 앞서나가서 발견하고 조치하는 역할을 해야 한다고 생각합니다. 이를 위해선 국가인권위원회에서 소프트웨어, IT, 딥러닝 분야 전문성을 갖추는 것이 가장 필요하다고 생각됩니다." {기술자}

이에 대한 구체적 제안으로 데이터의 편향 문제를 식별하는 전문가 양성, 새로운 혐오표현을 식별하고 사례화하는 역량 구비, 각 기술 분야 전문성 확보를 위해 위원회 산하 인공지능과 인권 관련 전담부서를 신설하는 등 국가인권위원회의 기능과 조직 확대 개편의 필요성을 말하는 의견이 있었다.

"향후 가이드라인의 적용과 개선 등을 위해 국가인권위원회는 지속적으로 인공지능의 개발과 활용에 있어서 발생할 수 있는 인권 침해의 양상에 충분히 대처할 수 있는

전문성을 확보할 필요가 있습니다. 이를 위해 과학기술 분야의 전문가들이 중심이 되는 인공지능과 인권 관련 전담 부서를 신설하는 등 국가인권위원회의 기능과 조직을 확대 개편하는 것이 시급하다고 봅니다." {학술연구자}

특히 어느 법률분야 응답자는 국가인권위원회가 개인정보보호위원회와는 별개로 보다 넓은 관점과 다양한 인권 보장의 관점에서 인공지능 등 신기술 환경에 의한 인권침해를 조사하고 관련 의견을 표명할 수 있다고 밝히며, 독립적인 인권보호기구로서 정기적으로 신기술이 인권에 미치는 영향을 관리하고 감독하는 제도를 수립할 필요가 있다는 의견을 제시했다. 이어서 현행 법제 내에서도 적극적인 태도를 보여야 한다는 의견이었다.

"국가인권위원회의 역할을 적립하기 위해서는 관련 법령 등의 국가인권위원회의 권한을 보다 명확히 규정하는 방안이 있습니다. 그러나 현행 법제 내에서도 충분히 실현 가능한 지점들이 대부분이기 때문에, 구체적 권한을 규정한 법령이 없다는 이유가 소극적인 관 보호 체계를 정당화할 수 없습니다." {법률가}

국가인권위원회가 인공지능과 인권 관련 법 제도를 제안하고 의미 있는 권고를 수행하기 위해 인공지능 정책 체계에서 적극적으로 행동해야 한다는 의견도 다수 있었으며, 이를 위해 인권영향평가를 주도적으로 시행할 것, 공공과 민간에 가이드라인을 수립하도록 권고할 것, 상시적으로 인권의 측면에서 신기술의 위험성을 논의하고 공론화의 토대를 마련할 것을 제시하기도 했다.

동시에 인공지능과 인권 침해 사례 수집과 발간 등 인식 제고 사업, 알 권리와 자기결정권 교육 등 영향 받는 당사자에 대한 교육에 나서야 한다는 의견도 있었으며, 디지털 처리의 복잡성과 일방성에 의해 피해당사자가 스스로를 구명하기 어려운 점을 고려할 때, 개별 피해 조력을 넘어 정책 개입을 위한 적극적 자세가 필요하다는 의견도 있었다.

"국가인권위원회 차원에서 새로운 기술과 인권의 문제에서 여전히 유효한 인권원칙들을 확인할 필요가 있다고 생각함. 이러한 관점에서 국가인권위원회는 국제사회에서 논의되는 여러 가지 기준들을 적극적으로 소개하고 홍보할 필요가 있으며, 자체 기준

을 수립하기 위한 토론 등을 진행할 필요가 있습니다." {법률가}

아울러 평화단체와 학부모단체 응답자 등은 각각 인공지능 무기체계 및 군대 내 병사들의 개인정보 수집과 활용 문제 모니터링, 교육현장에서 이뤄지는 인공지능 교육에 대한 검토가 필요하다는 의견을 제시했다.

인공지능, 자동화, 가상현실, 유전자검사서비스 등과 같은 새로운 기술이 실제 생활 환경에 적용되면서 기존의 기본적 권리에 영향을 미치는 경우가 생기고 있습니다. 이러한 신기술 확장과 관련해 인권을 보호하기 위한 제도를 입법화하는 방안에 대해 어떻게 생각하십니까? 이런 입법화 방안이 필요하다고 생각하신다면 입법화에서는 어떤 점에 중점을 두어야 할까요? 자유롭게 작성해주세요.

"가이드라인 이상의 좀 더 실효성 있는 인권보장 정책을 실현하기 위해서는 근거 법률이 마련되는 것이 중요하다 할 것입니다. 해당 법의 구체적인 내용까지는 제가 제시하기는 어렵겠지만 기본적인 방향은 신기술 육성에만 초점을 두는 것이 아니라 기술이 가져올 수 있는 위험성을 인식하고 인권보장과 기술발전이 조화를 이루도록 한다는 것이 법의 목적이 되어야 할 것입니다. 또한 특히 인공지능에 있어서 차별문제는 단지 인공지능에만 초점을 둔 법으로만 규율할 수는 없고 복합적이고 구조적인 차별의 문제가 함께 이루어져야 하는 만큼, 기본법으로서의 차별금지법/평등법을 마련하고 이와 조화를 이루는 형태로 인공지능 관련 법이 제정되어야 할 것입니다." {법률가}

신기술의 확장과 관련해 인권을 보호하기 위해 입법화가 필요하다는 의견이 다수였다. 최소한의 투명성과 알 권리 등 이용자의 권리 보장, 기본권에 대한 위험 예측, 인권 침해 방지책과 사후 구제책 마련, 인권영향평가 수행을 위한 법률적 근거 마련, 정기적 감사제도, 개인정보 침해요인 평가제와 같은 인공지능과 인권 평가제 등 입법화가 중점이 되어야 할 내용은 다양했다.

"국제인권법상 국가가 사인에 의한 인권침해를 보호하고, 권리 향유를 실현하기 위한 행정적, 입법적 조치를 취할 법적의무를 부담하고 있기 때문에, 신기술의 확장과 관련

하여 인권을 보호하기 위한 입법은 필수적이라고 생각합니다. 입법화에 있어서는 신기술 확장으로 영향을 받는 사람들이 가지는 불가침의 인권을 확인하고 차별을 금지하며, 특히 취약한 사람들의 권리를 보장하기 위한 의무를 구체화하며, 인권침해 발생 시 피해자 중심적 관점에서 구제수단과 내용을 정하는 것이 필요 하다고 생각합니다. (권리중심적 접근) 나아가 안전장치, 정기적 감사제도, 영향평가 제도, 실사제도 등 인권침해를 사전적, 사후적으로 방지할 수 있는 정책의 마련이 법제화되는 것이 필요하다고 생각합니다." {법률가}

"기존에 우리나라에서는 기업 자율성을 존중, 인공지능 기술발전 장려, 사회변화에 대한 유연한 대처 등을 이유로 도덕적 규범, 자율규제 또는 모범규준의 형식으로 인공지능 규제에 관한 논의가 진행되어 왔습니다. 그러나 해외에서는 이미 인공지능 규제를 법규의 틀내에서 규제를 하고 있습니다. 예를 들면, (1) EU처럼 인공지능 관련 일반법적인 법적규제를 하거나, (2) (미국 처럼) 기존의 소비자법(FTC법) 또는 금융관련 법과 같은 구속력 있는 법규의 틀 내에서 인공지능을 활용한 결과를 규율되는 경우가 있습니다. 즉, 외국의 경우 인공지능을 사용한 결과가 기존의 법규 위반행위에 해당한다면, 인공지능에 관한 전문성을 가진 시장규제당국이 이를 발견하여 제재를 할 뿐만 아니라, 기존의 소비자보호 법제의 틀 내에서 인공지능의 준수사항에 관한 해석지침을 제시(미국의 ftc)하는 경우가 있습니다. 이러한 점에 비추어 보면 인공지능 규제는 적어도 일부 영역에서는 이미 윤리 또는 자율규제의 영역을 넘어서 법적규제의 영역으로 들어온 것으로 판단됩니다. 인공지능을 규율하는 법규제의 틀은 이원화되어야 하는데, 첫번째로는 인공지능에 관한 일반법의 제정과, 두번째로는 각 분야별 개별법에 인공지능에 관한 사항을 포함시키는 것입니다. 예를 들면, EU의 인공지능법(안)처럼 일반법으로서 인공지능에 관한 기술적 개념정의 및 준수사항, 공공/경찰/사법 및 인권보호 분야에서의 일반적인 규제 및 인공지능 관련 금지사항 등을 포함하고, 소비자 분야 또는 금융분야 등의 세부적인 분야에서는 개별법의 형태로 각 영역별로 규제의 틀 내에서 세부적인 내용을 담아야 할 것입니다. 후자의 경우에 예를 들면 불공정거래행위의 행위유형에 인공지능의 사용에 관한 내용 포함한다든지, 내부통제나 리스크 관리, 소비자보호 관련 규정에서도 인공지능에 관한 구체적인 사항을 포함하는 방안을 고려할 수 있습니다." {소비자단체}

동시에 고위험 영역의 인공지능에 대한 입법 통제, 아동 등 취약 집단에 대한 고려가

추가적으로 필요하다는 의견이 있었다.

반면 기술 분야와 법률 분야 응답자에서는 입법화는 아직 이르다는 의견도 다수 있었다. 이 경우 기존의 인권교육 전반이 부족했다는 점을 고려해 이에 대한 교육이 선행되어야 한다는 의견, 입법의 필요성은 공감하나 변화하는 신기술의 특성 상 세부적인 내용을 담기 어렵기에 입법보다는 신기술에 대한 인권영향평가를 의무화하는 것이 중요하다는 의견을 보였다.

> **"입법 자체는 필요할 것 같은데, 계속해서 변화하는 과정에 있는 신기술의 특성상 세부적인 내용을 세세하게 법에 담는 것은 쉽지 않을 것 같고, 신기술의 도입 시에 그러한 기술 체계가 미칠 수 있는 영향에 대한 평가 자체를 의무화하는 것이 중요할 것 같습니다. 구체적인 기술 체계로 인한 문제 해결은 개별법이 필요할 수도 있겠지만요."** {기술자}

아울러 현재 인공지능과 인권 문제가 대두됨에 있어 문제의 책임이 개발자 개인에게 환원되는 양상이 있어, 입법화를 한다면 인공지능 기술을 서비스에서 별도로 분리하는 것이 아니라 전체적인 맥락과 취지를 점검하고 그 피드백과 책임을 특정 개발팀이 아닌 서비스 제공 회사에 돌려주는 형태가 필요하다는 의견이 있었다. 나아가 입법을 한다면 인공지능을 다룬 서비스를 출시할 때 특정한 기관을 거쳐 서비스 검토를 받는 형태가 적절할 것이라는 제안이 있었다.

입법화의 구조에 대한 의견도 일부 있었는데, 인공지능에 관한 일반법 제정과 더불어 각 분야별 개별법에 인공지능에 관한 사항을 포함시키거나, 각 개별법에서 규정해야 할 기본원칙과 이행 체계를 잘 담을 수 있는 인공지능 관련 기본법이 필요하다는 의견을 제시했다. 또한 신기술에 따른 인권침해의 양상이 불투명해 새로운 인권 보호의 장치를 규정하는 데 단행의 법률로 규율하는 건 무리가 있으니 신기술과 인권침해에 대한 국가인권위원회의 적극적 대응을 명시하고 인권영향평가 실시, 피해 확산 방지를 위한 긴급구제 등에 대한 내용을 추가하여 국가인권위원회법을 개정하는 방안을 제시하는 의견도

있었다.

“신기술의 적용에 따른 인권 침해의 문제에 대처하기 위한 입법의 필요성에 공감합니다. 현재 「지능정보화 기본법」에서도 지능정보기술 및 지능정보서비스 이용의 안전성 및 신뢰성 보장에 대한 규정을 두고 있습니다(제6장 제2절). 하지만 동법은 지능정보기술의 고도화 및 지능정보서비스의 이용촉진, 지능정보화의 기반 구축, 지능정보사회의 기반 조성 등에 중점을 두고 있습니다. 이러한 점에서 신기술 확장과 관련하여 인권 침해를 예방하고 새로운 유형의 피해를 효과적으로 구제하기 위해 관련 입법을 추진하는 것은 나름의 의미가 있다고 봅니다. 다만 신기술의 전개 양상이나 그에 따른 인권침해의 양상이 불투명한 상황에서 새로운 인권 보호의 장치를 규정하는 데 어려움이 예상되므로, 이에 관한 내용을 단행의 법률로써 규율하는 것은 무리가 있다고 봅니다. 신기술 관련 인권 침해에 대한 적극적 대응을 국가인권위원회의 새로운 기능으로 명시하고, 신기술 도입에 따른 인권영향평가의 실시, 피해의 확산 방지를 위한 긴급구제 등에 대한 내용을 추가하는 정도로 국가인권위원회법을 개정하는 방안을 생각해볼 수 있을 것입니다.” {학술연구자}

그 외 인공지능과 관련한 차별 문제는 단지 인공지능에만 초점을 둔 법만으로 규율할 수 없고 구조적 차별 문제가 함께 다뤄져야 하기에 차별금지법, 평등법을 마련을 촉구하는 의견이 있었다.

귀한 시간 내주셔서 감사합니다. 마지막 질문으로, 인공지능을 개발하고 활용할 때 고려해야 하는 인권 보호 문제와 관련해, 또는 제시된 가이드라인(안)과 관련해 더 하시고 싶은 말씀이 있다면 자유롭게 작성해주세요.

먼저 가이드라인에 대한 접근성과 대중성 향상에 대한 의견이 있었다.

“이용자/영향을 받게 되는 사람들이 무엇이 문제인지, 어떤 권리가 침해받을 수 있는지를 조금 더 쉽게 이해할 수 있도록 사례나 예시들을 조금 더 담을 수 있는 방법이 있으면 어떨까 합니다. “여기서 OO란 고객응대, 서비스 이용자격에 대한 심사, OOO,

000 등과 같이 00한 행위를 말합니다"와 같은 식으로요. 그런 내용이 가이드라인의 성격에 맞지 않으면 가이드라인이 만들어지고 난 뒤에 별도의 해설서나 주석이 필요할 수도 있겠다는 생각도 듭니다. " {기술자}

"인공지능과 관련 구체적인 인권침해 사례 등을 잘 알지 못한 점 등으로 인해 기본적인 인권의 관점 안에서 중심으로 바라보는 것이 생기는 것 같습니다. 이러한 가이드라인이 다양한 사람들에게 접근하기 위해서는 가이드라인에 대한 구체적인 해설집이나 사례 제시 등이 있어야 좀 더 실질적인 활용과 접근이 가능해보입니다. 그것은 아무래도 가이드라인이기 때문에 이러한 인공지능을 활용하는 기업의 입장에 맞는 것이 있는 것으로 보이고요. 이후 이러한 가이드라인에 대해 실제로 정보를 접근하는 당사자 입장에서도 잘 이해할 수 있는 부분도 필요하지 않겠나 하는 생각도 듭니다." {성소수자단체}

" '인공지능이 개인의 선택과 판단 및 해동을 강요하거나 자율성을 침해' 하는 사례가 바로 떠오르지 않는다면 이해에 다소 어려움이 생길 수도 있다는 생각이 듭니다. 이용자가 본인이 이용하는 기술에 인공지능이 활용되는지 여부를 알지 못 하는 혹은 알 수 없는 사례가 많습니다. 그렇기에 차별이 발생하더라도 그것이 인공지능에 의한 차별임을 모르거나, 본인이 차별받고 있는지조차 모르는 상황이 발생합니다. 가이드라인 별첨으로 각 조항에서 다루고 있는 상황에 대한 구체적인 사례가 덧붙여진다면, 이용자에게도 훨씬 더 풍부하게 내용이 전달될 수 있을 거라 생각합니다." {여성단체}

또한 가이드라인 활용, 실제 실행을 위한 방안과 로드맵이 필요하다는 의견도 있었다.

"인권위 가이드라인이 갖는 의미에도 불구하고 현장에서 지침 준수 여부는 또 다른 문제입니다. 또한, 공적 영역에만 강제성을 갖다 보니 민간영역에 대한 강제성을 갖기 어려운 점도 한계입니다. 이런 한계를 법률 제정을 통해 보완하길 바랍니다." {외국인이주민단체}

"가이드라인이 제정된 후에 이 내용을 국가인권위원회에서 사회적으로 널리 알리고 법 제정까지 갈 수 있도록 노력했으면 합니다. 종교계에서도 본질적인 인간의 존엄성

에 도전하는 인공지능에 대해 우려하고 있음을 고려하면 같이 연대하는 것도 좋은 방법이 될 것 같습니다." {지역인권단체}

가이드라인이 포함하지 못한 내용을 지적하며 이를 보완하는 의견도 있었다.

"제시된 가이드라인(안)은 최소한으로 이해되어야 한다고 생각하고, 해당 안이 관련 국제인권규범 등이 제시하는 더욱 엄격한 기준을 외면하는 안으로서 활용되어서는 아니 될 것입니다." {법률가}

"시민사회가 거듭 요청한 바 있는, 인공지능 개발과 활용에서도 역시 노동법, 개인정보보호법, 제조물 책임법 등등 법률준수 의무가 있다는 내용도 포함되어야 할 것입니다. 특히 행정기본법 제정으로 완전히 자동화된 인공지능 시스템에 의한 행정처분이 가능하게까지 된 상황에서, 인공지능 행정처분에 따른 기본권 침해가 발생할 가능성이 커진 바, 특히 법률유보, 과잉금지 및 적벌절차 등 헌법원칙 준수에 대한 명시적 조문이 필요하다고 생각합니다." {인공지능의견발표단체}

"공공기관의 활용 외에, 안면인식 등 공공장소에서의 인공지능의 (무분별한) 사용금지가 들어갔으면 합니다. 부천시 사례도 그렇고, 최근 현대 엘리베이터 CF를 보니 엘리베이터 이용, 동선 관리를 위해 안면인식 기술을 도입하는데. 이런 부분은 개인들이 감지하기 어렵고, 거부하기 어려운 영역입니다. 개인들이 인지하기 어려운 영역, 공공영역에서 인공지능 사용을 할 수 없게 하고, 거부할 수 없는 영역에서 다른 선택지를 둬서 꼭 인공지능만을 경유하지 않게 했으면 합니다." {학술연구자}

제3절 집단심층면접조사 결과

주요 배달플랫폼 기업들은 2020년부터 현재까지 일감의 배정과 취소, 노동자 평가와 페널티 및 배달 요금 설정을 알고리즘을 통해 자동화시키는 인공지능 시스템을 적극적으로 이용하고 있다. 이에 영향을 받는 당사자에 대한 심층면접조사는 배달플랫폼의 인공지능 배차 등 인공지능 시스템을 경험 중인 배달노동자 4인을 대상으로 그룹 면접조사의 형태로 진행되었다. 조사 문항은 인공지능 시스템이 가져온 노동 환경과 조건의 변화, 권리 보장 여부 및 일반적인 경험에 대한 질문과 함께 국가인권위원회 인공지능 개발과 활용에서의 인권 가이드라인(초안)에 대한 내용 및 인권 보호 방안, 국가인권위원회의 역할 등에 대해 질문하였다.

1. 인공지능 시스템으로 인한 노동 환경 변화

우선 배달플랫폼 내 ‘AI추천배차’와 같은 인공지능 시스템의 적극적 활용은 배달노동자에 대한 통제 및 노동시간, 조건 및 환경의 변화와 함께 실질적 임금의 저하를 동시에 가져온 것으로 보인다. 특히 거리를 직선으로 산정해 실제 소요되는 시간보다 적게 배달제한시간을 측정하는 등 인공지능 배차가 제시한 배달거리 및 예상소요시간과 이에 따라 책정되는 배달 요금의 부정확도와 오류를 공통적으로 경험하였고 실시간으로 변경되는 배달 요금은 배달노동자에게 시간의 압박으로 다가오는 등 통제의 형태를 띄고 있다고 답했다.

“각 플랫폼 사에서 인공지능 알고리즘이 최적화되어서 라이더가 가까운 상점에 배차를 한다고 얘기는 하지만 실질적으로 일을 하다 보면 그와는 반대의 경향이 나타납니다. 굉장히 멀리 떨어진 상점에 배차를 해준다거나…” (배달노동자 A)

“인공지능이 도입되기 전에는 주문 여러개를 묶어서 배달을 다닐 수 있었고 그렇기 때문에 라이더의 숙련도에 따라 임금 차이가 있었습니다. 그러나 인공지능 도입 이후 숙련도에 따른 임금 차이가 없어지며 임금이 하향평준화되었습니다. 특히 오래 일해서

숙련도가 좋은 라이더에게는 인공지능 도입이 임금착취로 느껴질 정도였습니다.” (배달노동자 C)

“3배차, 4배차 등 배차를 많이 잡을 수 있었을 때에는 단가가 낮더라도 가까운 지역을 한번에 챙기며 효율이 있었는데, 단가는 더 내려간 상황에서 배차는 하나씩 수행하라 하니 타산이 안 맞는 부분이 많습니다. 숙련도에 따라 달라지긴 하지만 어떨 때에는 시급 1만원 이하로 나올 때도 있구요.” (배달노동자 B)

“작년 2월까지는 직접 골라서 가는 수동 콜 시스템이었습니다. 배차는 5개 제한이었구요. 제가 일했던 쪽은 7월에 인공지능이 도입되며 배차제한이 3개로 줄면서 인공지능이 이에 동선을 맞춰서 배차를 했었는데 이게 불필요한 동선을 많이 보여서 이게 인공지능이 맞나 사람들이 의심을 많이 했습니다. A지역에서 픽업해서 B까지 배달하는 경우 동선에 맞는 배차를 엮어서 2개 3개씩 배달하는 것이 보편화된 노동방식이었지만, 인공지능은 A지역 픽업과 B지역 배달이 여러 개 있더라도 무조건 A에서 B까지 하나의 배차를 완료한 후에 다시 A로 가서 픽업시키는 그런 불필요한 동선을 많이 보였죠.” (배달노동자 B)

또한 인공지능 배차 도입 초기 진행된 프로모션 및 인공지능 배차 거부에 대한 페널티 부여는 자동화된 의사결정을 거절할 수 없게 하는 등 강제성을 부여한 것으로 보인다.

“인공지능 배차가 나오고 배차 거절을 하면 라이더에게 불이익이 가는 경우가 많습니다. 강제성을 없잖아 잡아두는 거죠. 배차를 많이 거절하면 ‘배달에 문제가 있다’ 이런 문구가 뜨기 시작하면서 콜이 안 들어오기 시작합니다. 인공지능이 콜을 10분, 20분 딜레이를 갖고 주기도 하구요. 이런 식으로 딜레이 후에 콜이 들어오기 때문에 한 번 걸리기 시작하면 거절할 수 없는 상황이 옵니다. 그래서 아주 반강제적이라는 생각이 듭니다.” (배달노동자 B)

플랫폼 내 인공지능 활용 인지 및 통보 여부

배달 플랫폼 앱 내에서 AI추천배차의 도입에 대한 지역별 공지가 있었지만 확실한 통지는 미비했고 이에 영향을 받는 배달노동자의 동의 및 의견 수렴 과정은 없었던 것으로 보인다. 또한 당사자임에도 불구하고 AI추천배차가 실제 어떤 형태로 작동하여 일감을 배정하는지에 대해 언론보도 등을 통해 인지했음을 확인할 수 있었다. (배달노동자 C)

"인공지능에 대한 인지는 미리 했지만 어떤 형태로 콜을 주고 어떤 형태로 일을 하는지는 사전에 뉴스 같은 걸 통해 알게 되었습니다. 찾아보니 앱 내에 공지사항이 남아있긴 하지만 그게 다입니다." (배달노동자 D)

배차 인공지능의 효율 증가 또는 변화 체감 여부

데이터를 통해 고도화되는 등 시간이 지나며 인공지능 배차 시스템의 변화 또는 효율성의 증가가 있었는지와 관련하여, 1년이 지났지만 큰 개선이 없다는 의견이 공통적이었다.

"여전히 문제적이라고 생각합니다. 나아진 건 없구요 솔직히 말해서 알고리즘이 더 안 좋아지고 있는 것 같다고 체감합니다. 인공지능 알고리즘이 안정화됐다고 하지만 제가 납득하지 못할 루트를 준다던가 하는 말도 안되는 픽업 지역이 생겨났구요. 직선거리 4~5km라고 나오지만 실제 운행했을 때에는 8km, 9km, 10km 등 두 배 가량 차이가 나는 문제가 발생하고 있습니다." (배달노동자 B)

정보인권 침해

필수적 서비스 운영, 인공지능 시스템의 고도화 등을 위해 노동의 전 과정이 자동으로 전자적인 방식으로 기록되는 등 노동자의 다양한 개인정보가 생성되고 수집되는 플랫폼 노동의 특성과 관련해 정보인권 침해가 우려된다는 답변이 있었다. 특히 배달노동자의

동선이 소비자에게 공개되는 것이 노동 조건의 악화와 실질적인 사생활 침해로 이어지는 것으로 보인다.

"음식 픽업 후 중간에 배가 아파 화장실에 들렸다가 10분 정도 늦게 배달한 적이 있다. 이때 도착하기 전부터 고객센터에서 항의가 왔다. 앱 상에서 저희의 위치가 어디에 머물렀고 어디를 가고 있고 이런 게 다 보이지 않나. 라이더는 화장실 사용이 힘들어서 공영화장실을 이용하는 경우가 많은데, 지도에서는 공영화장실이라는 게 고객에게 다 보이니까 상당히 치욕스러웠다." (배달노동자 B)

"배달하다가 기름이 떨어져서 주유소에 잠깐 들렸는데 전화가 와서 왜 멈춰있냐는 항의를 받은 적이 있다. 고객이 라이더의 노동을 관리 감독하는 매니저 역할을 하기 시작한 것 같기도 하다." (배달노동자 D)

"과거에는 라이더 배차가 되면 고객에게 라이더 이름과 사진까지 보여줬다. 저희가 개인정보를 지켜달라고 요구해서 그나마 나아진 거다. 하지만 위치정보는 아직도 나오는 상황이고 감시 받는 느낌이 든다. 이것도 프라이버시를 지켜줘야 하는 게 아닌가 싶다." (배달노동자 A)

2. 인권 가이드라인(초안)에 대한 의견

전반적으로 가이드라인의 내용에 공감하며 필요성을 느끼지만 강제성이 없어 서로 신뢰할 수 있는 구조가 만들어지지 못할 것이라는 아쉬움을 느끼는 의견이었다. 특히 인공지능 시스템의 사용에 있어 배달의 민족 등 배달 노동 플랫폼 측에서는 이미 알고리즘을 공개할 의사가 없다고 밝혔기에 실제 적용을 위한 활용이 중요하다는 의견이었다.

"아무래도 가이드라인이다 보니 강제성이 없는 게 아쉽습니다. 일단 사측에서 알고리즘 공개 거부를 하면 있으나 마나 한 게 아닌가 하는 생각도 듭니다." (배달노동자 B)

또한 가이드라인 내용에 기반한 진정 등 국가인권위원회의 가이드라인 활용에 대한

의견도 있었다.

"국가인권위원회의 권한 자체도 문제가 있을 수 있습니다. 그러나 본문과 같은 내용으로 진정을 제기하면 국가인권위원회에서 받아주는 등 실질적으로 활용하여 진전을 이뤄냈으면 좋겠습니다." (배달노동자 D)

3. 신기술 환경에서 인권 침해를 방지하기 위한 국가인권위원회의 역할에 대한 의견

현재 인공지능 배차 시스템 등으로 영향을 받고 있는 당사자이기 때문에 시급하게 해결해야 할 문제이며 이를 위해 인공지능을 활용하는 기업을 감사하는 권한이 국가인권위원회에게 있어야 한다는 의견이 있었다.

"국가인권위원회에서 인공지능 알고리즘을 활용하는 기업을 감사하는 권한이 있어야 하지 않을까요? 알고리즘도 결국 사람이 만드는 건데, 이걸 위원회에서 감사하는 형태의 권한이 있어야 한다고 생각합니다. (배달노동자 A)

국가인권위원회가 진정 사례를 쌓아 인권을 고려한 법적 분쟁 절차에 도움을 줘야 한다는 의견도 있었다.

"위원회가 케이스를 쌓아야 합니다. 예를 들어 직장 내 성희롱 문제는 고용노동부로 가지 않고 국가인권위원회의 판단을 받아서 민사 절차를 밟는데, 이는 위원회에 대한 신뢰가 있기 때문입니다. 위원회가 인공지능과 노동 관련한 상징적인 사례를 발굴하여 도움을 줬으면 좋겠습니다." (배달노동자 D)

4. 신기술 환경에서 인권을 보호하기 위한 제도를 입법화하는 방안에 대한 의견

“현재 생활물류서비스산업발전법 일부개정법률안 (일명 라이더보호법)이 발의되어 있는 상태인데, 이것의 통과 등을 위해 국가인권위원회에서 정책적으로 노력해줬으면 하는 바람이 있습니다.” (배달노동자 C)

한편 기존의 노동법 체계 안에서 인공지능을 규제할 수 있도록 국가인권위원회가 노력해야 한다는 의견도 있었다.

“노동과 관련된 알고리즘을 국가인권위원회가 권고하고 이러한 권고가 취업규칙에 반영되는 등 노동법 규제로 이어지는 것이 노동자들에게 훨씬 유리할 것 같습니다.” (배달노동자 D)

제7장 인공지능 개발과 활용에서의 인권 가이드라인(안)[163]

제1절 의의

1. 제정 배경

1. 사회 전반에 걸쳐 인공지능(AI: Artificial Intelligence)의 활용이 증가함에 따라 고용, 금융, 행정, 복지 등 거의 모든 분야에서 인간의 기본적인 삶과 인권에 영향을 미치는 사례가 증가하고 있습니다.

인공지능 기술이 발전함에 따라 국가와 기업의 인공지능 활용이 크게 늘었다. 최근 인공지능의 발전은 업무와 시스템의 효율성을 크게 향상시켰고,[164] 중요한 의사결정을 지원하거나 직접 자동화된 결정을 수행하는 수준에 이르렀다.

교육과 사회복지, 치안과 군사 등 인간의 기본적인 인권에 중대한 영향을 미치는 공공부문에서 인공지능의 활용이 늘고 있다. 최근 관련 법률이 제·개정되면서 행정기관이 인공지능 기술을 활용하거나, 완전히 자동화된 방식으로 행정 처분을 내리는 것도 가능해졌다.

생활에서 널리 이용되는 제품과 서비스에 인공지능 기술이 적용되고 있으며, 많은 기업들이 고용과 상거래, 금융 분야 등 삶에 중요한 결정에 인공지능을 활용하고 있다.

163) 이 장의 내용은 제6장에서 개발한 초안에 대한 심층면접조사 검토 결과를 반영하고 제안기관인 국가인권위원회와 협의 하에 작성하였으나, 국가인권위원회가 실제 의결하는 가이드라인과 차이가 있음을 밝힌다.

164) Council of Europe (2020), A.3문.

행정기본법 [시행 2021. 9. 24.]

제20조(자동적 처분) 행정청은 법률로 정하는 바에 따라 완전히 자동화된 시스템(인공지능 기술을 적용한 시스템을 포함한다)으로 처분을 할 수 있다. 다만, 처분에 재량이 있는 경우는 그러하지 아니하다.

전자정부법 [시행 2021. 12. 9.]

제18조의2(지능형 전자정부서비스의 제공 등) ① 행정기관등의 장은 인공지능 등의 기술을 활용하여 전자정부서비스를 제공할 수 있다.

최근 공공기관과 민간 모두에서 인공지능 채용 시스템의 도입이 확산되고 있다. 도입 기관들은 인공지능이 채용 서류를 효과적으로 처리하고 공정한 의사결정에 도움을 줄 것을 기대한다. 그러나 인공지능 채용 절차의 투명성과 책임성이 부족하다는 지적이 이어지고 있다. 구직자들은 "관련 정보가 부족하다"며 인공지능 채용의 불투명성에 많은 부담감을 표했다. 일부 공공기관은 채용 당락을 결정하는 근거로 실제 인공지능 면접 자료를 이용하였지만 지원자들에게 이 사실을 알리지 않았다. 2020년 11월 국회의원의 검토 결과, A공공기관은 인공지능이 어떤 알고리즘 및 기준을 적용해 지원자를 불합격시켰는지 알고 있지 못했다. 또한 2021년 3월 감사원 감사 결과, B공공기관은 인공지능 면접을 참고자료로만 이용할 계획이었으나 실제로는 접속 오류를 겪은 지원자들 일부를 탈락 처리한 것으로 나타났다.

【참고】 투명성·공정성·신뢰성…AI면접 믿을 만할까?. 한겨레21 보도(2020. 10. 23.); 구직자 10명 중 6명, AI 채용 부담스러워!. 사람인 보도자료(2021. 1. 4.); AI 면접 접속 오류 이유로 신입 공채 불합격 처리한 코이카. 뉴스1 보도(2021. 3. 16.); 청년 앞길 막는 AI면접…AI윤리기준 투명성·공정성 필요. 국회의원 정필모 과학기술정보통신부 '21 예산심사 질의서(2020. 11. 5.)

2. 앞으로도 막대한 양의 빅데이터를 분석하고 학습하는 과정을 통해 다양한 영역에서 사람의 판단을 대신할 수 있는 인공지능은 점차 적용 영역을 넓혀 사회 전반과 개인의 삶에 강력한 영향력과 파급력을 행사할 것입니다.

3. 인공지능의 발전과 확산은 국가경쟁력과 개인 삶의 질을 높일 것으로 기대되지만 개인정보 및 사생활 침해, 차별 등과 같은 인권을 침해하는 문제들도 대두되고 있습니다.

인공지능 기술은 분석과 예측 등 디지털 정보의 활용 가능성을 크게 향상시켰다. 기존에 사람이 해왔던 판단이나 의사결정을 부분적으로나 완전히 자동화하기도 한다. 유엔 사무총장은 인공지능 등 신기술이 인권을 옹호하고 행사할 수 있는 새로운 수단이 되고 있다고 기대를 표했다. 동시에 이 신기술을 인권을 침해하고, 불평등을 심화시키고, 기존의 차별을 악화시키는 데 활용해서는 안 된다고 강조했다.[165] 유엔 인권최고대표는 머신러닝 기술을 비롯한 인공지능이 사회 문제를 극복할 수 있는 힘이 될 수 있지만, 인권에 미치는 영향을 충분히 고려하지 않고 이 기술을 도입할 경우 인권에 부정적이거나 심지어 치명적인 영향을 미칠 수 있다고 지적하였다.[166]

인공지능은 일반적으로 대규모 데이터셋에서 패턴을 감지하여 작동한다. 이때 인공지능이 개인정보를 이용하고 사람들의 삶에 영향을 미치는 결정을 내리게 되면서 개인정보 및 사생활의 권리가 영향을 받게 된다. 개인정보를 이용하지 않는 인공지능이라 하더라도 그 추론과 예측은 사생활의 권리에 깊은 영향을 미치며,[167] 인공지능이 편향적으로 개발되고 활용될 경우 차별받지 않을 권리에도 부정적인 영향을 미칠 수 있다.

일상 생활에서 알고리즘 시스템에 대한 의존도가 높아짐에 따라 공정한 재판을 받을 권리, 사생활 및 개인정보에 대한 권리, 사상·양심 및 종교의 자유, 의사 표현의 자유, 집회의 자유, 평등권, 경제적·사회적 권리 면에서 중대한 인권 문제가 제기되고 있다. 알고리즘 시스템의 기능은 대개 온라인 및 오프라인에서 대규모 디지털 추적을 통하여 개인 및 집단의 신원 및 행동에 대해 수집한 데이터를 체계적으로 집계하고 분석하는 데 기반을 두고 있다. 대규모 추적은 개인의 사생활 침해와 고도로 개인화된 조작에 대한 우려를 증가시키고, 알고리즘 시스템의 제안 단계부터

165) 유엔문서 A/HRC/48/31, 4문.
166) 같은 문서, 2문.
167) 같은 문서, 15~17문.

이후 모든 수명 주기에 걸쳐 고려되어야 하는 인권의 행사에 심각한 악영향을 미칠 수 있다.

【원문】 유럽평의회 <알고리즘 시스템의 인권 영향에 대한 대응 지침> 4문.

2020년 영국에서는 성적 예측에 이용된 인공지능 시스템이 공립학교와 가난한 지역의 학생들을 차별하는 결과를 낳아 사회적 논란이 일었다.

영국 시험감독청은 2020년 코로나19로 대학수학능력시험에 해당하는 A레벨 시험을 취소하는 대신 인공지능 알고리즘을 통해 학생 성적을 부여하기로 결정하였다. 이 알고리즘은 학생의 A레벨 예비 시험 점수와 교사의 예상 점수 등을 바탕으로 성적을 산출하고 소속 학교의 역대 학업능력을 고려한 가중치를 부과하였다.

그러나 평가 결과 부유한 지역 학생에 비하여 가난한 지역 학생이 상대적으로 차별을 받은 것으로 나타났다. 학교별 학업 성취도가 개별 학생들의 성적에 중대한 영향을 미쳤기 때문이다. 그러자 인공지능이 불평등을 강화한다며 전국에서 항의 시위가 일었고, 교육부 담당 공무원과 시험감독청장이 사임하였다. 2020년 8월 영국 교육부 장관과 시험감독청장은 A레벨 알고리즘 성적을 철회한다고 밝혔다.

【참고】 Who won and who lost: when A-levels meet the algorithm. The Guardian (2020. 8. 13.); 사는 곳으로 성적을 결정했다. 한겨레21 (2020. 9. 7.)

4. 반면 인공지능으로 영향을 받는 당사자들은 인공지능의 도입, 운영, 결정에 대하여 참여의 기회를 보장받고 있지 못하며, 인공지능으로 인한 인권침해가 발생한 경우에도 적절하고 효과적인 권리구제를 받을 수 있는 절차와 방법이 미흡한 상황입니다.
5. 따라서 인공지능을 개인의 삶과 사회적 공익에 기여할 수 있도록 설계하며, 인간의 존엄성과 차별금지, 자기결정권 보장 등 기본적 인권에 기반을 두도록 하는 것이 매우 중요합니다.
6. 본 가이드라인은 인권적 관점에서 인공지능의 개발과 활용의 전 과정

에서 인권침해를 예방하기 위하여 준수해야 할 기본 원칙과 주요 내용을 제시하고자 마련되었습니다.

학교, 직장, 금융, 공공 서비스 등 주요 생활영역에서 인공지능이 도입 또는 운영되거나 의사결정의 기반으로 활용되면 많은 사람들이 그 영향을 받을 수밖에 없다. 이렇게 영향을 받는 당사자들은 인공지능에 대한 정보를 제공받고 의견을 제시하거나 그 영향에 대한 평가에 참여할 수 있는 권리가 있다. 자신에 대하여 인공지능에 기반한 결정이 내려진 경우 그 이유에 대하여 설명을 듣고 이의를 제기할 권리도 있다. 인공지능으로 인한 인권침해와 남용 피해를 입은 경우, 적절하고 효과적으로 구제를 받을 수도 있어야 한다.

그러나 현재 우리나라 법령과 정책들은 인공지능으로 영향을 받는 당사자들에게 이러한 권리를 명확히 보장하고 있지 않다. 따라서 인공지능을 도입하는 모든 공공기관과 기업 들이 그 개발과 활용의 모든 단계에서 인권적 가치와 기본적 권리를 고려할 수 있도록 기준이 제시될 필요가 있다.

이 가이드라인은 인권 기반 접근에 기반하여 인공지능의 개발과 활용에 대한 기본 원칙과 주요 내용을 제시함으로써 인권침해를 방지하고자 마련되었다.

2. 목적 및 의미

7. 국가인권위원회는「국가인권위원회법」제19조 제6호에 따라 인권침해의 유형, 판단기준 및 그 예방조치 등에 관한 지침을 제시할 권한을 가지고 있습니다.
8. 이에 국가인권위원회는 인공지능으로 인한 인권침해와 차별을 판단하고, 개선 등의 권고나 구제절차를 마련하는 데 필요한 기준을 제시하고자 합니다.

9. 본 가이드라인은 인공지능의 개발과 활용에 있어서 우리 사회의 인권적 가치가 훼손되지 않고 인간의 존엄성을 보장하며 기본적 인권을 실현하는 방향으로 나아가도록 하는 데 목적이 있습니다.

10. 또한 인공지능의 개발과 활용 과정에 적용할 인권원칙을 제시하고, 인공지능 서비스 이용자, 영향을 받는 당사자 등에게 주어진 권리 및 피해구제수단을 제공하며, 정부에게 적절한 법령과 정책을 수립할 수 있는 가이드라인을 마련하고자 합니다.

국가인권위원회는 국가인권위원회법에 의하여 인권침해의 유형, 판단기준 및 예방조치 등에 관한 지침을 제시할 권한을 가지고 있고, 공공부문 전체 및 일부 민간 영역의 인권침해나 차별 여부를 조사하고 구제할 권한 또한 갖고 있다. 이 가이드라인은 국가인권위원회의 정책 권고로서 작성되었으며, 이후 국가인권위원회 권고나 구제절차의 해석 기준이 된다는 점에서 간접적인 실행력을 갖고 있다.

몇몇 국가가 인공지능에 대하여 특정 규범의 준수를 의무화한 법령을 시행하고 있지만, 우리나라의 경우 인공지능의 인권 규범 준수를 의무화한 입법 사례가 아직 없다. 따라서 이 가이드라인은 인공지능이 준수하여야 할 인권 기준을 제시하는 자율 규범의 성격을 갖는다. 국내·외에서 인공지능에 대하여 가이드라인보다 더 강한 인권 준수 규범이 마련될 경우 이를 우선적으로 고려하여야 한다.

한편, 각 분야를 소관하는 부처나 기관별로 자율 규범적인 인공지능 윤리 또는 점검표를 발간한 사례가 있다. 이러한 분야별 기준은 해당 분야 규범으로서 충분한 의미가 있고, 특히 관련 법령을 소관하는 기관에서 발간한 기준의 경우 실행력을 가질 수 있다. 다만 분야별 기준의 경우 그 적용이 해당 분야에 한정될 수밖에 없다는 점에서 한계가 있다. 유엔 인권이사회는 인공지능 등 신기술에 대응하는 인권의 증진 및 보호는 전체적이고, 포용적이며, 포괄적인 접근방식을 따를 필요가 있다고 지적해 왔다.[168]

168) 인권의 증진 및 보호에 관해 신기술이 갖는 영향력, 기회, 과제에 대응함에 있어 전체적, 포용적, 포괄적 접근방식을 취하는 것의 중요성과 모든 이해당사자들이 더욱 긴밀히 협력할 필요성을 재확인한다. 유엔문서 A/HRC/RES/47/23, 1문.

이에 이 가이드라인은 인간의 존엄과 인권적 가치의 존중이라는 헌법적 원칙을 기반으로, 인공지능을 개발하거나 활용하는 모든 기관이 전 과정에서 준수하여야 할 포괄적인 인권 원칙을 제시하는 것을 목적으로 한다. 더불어 인공지능 환경에서 이용자 및 영향을 받는 당사자가 보장받아야 할 권리를 밝히고 피해구제를 도모한다. 나아가 국가인권위원회의 권한이 허락하는 범위에서 국회와 정부가 인공지능 관련 법령과 정책을 수립할 때 이 가이드라인이 제시하는 인권 기준을 반영할 것을 권고한다.

제2절 적용범위 및 정의

1. 적용 범위

11. 인공지능의 개발과 활용을 위한 가이드라인에서 의미하는 인권은「대한민국 헌법」및 법률에서 보장하거나 대한민국이 가입·비준한 국제인권조약 및 국제관습법에서 인정하는 인간으로서의 존엄과 가치 및 자유와 권리를 의미합니다.
12. 본 가이드라인은 인공지능의 개발부터 활용에 이르는 모든 과정에 참여하는 사회구성원들을 대상으로 하며, 여기에는 인공지능 개발자, 이용자, 영향을 받는 당사자, 정부 및 공공기관, 기업 등이 포함됩니다.
13. 인공지능 개발과 활용에 관한 법률을 제정하거나 개정할 때에는 본 가이드라인의 목적과 기본 원칙 및 주요 내용을 참고할 수 있습니다.

인공지능을 개발하거나 활용하는 정부, 공공기관, 기업 뿐 아니라 모든 사회구성원이 이 가이드라인이 제시하는 인권 규범을 준수하기 위해 노력할 것을 권고한다. 더불어 인공지능 서비스를 이용하거나 자동화된 의사결정 등 인공지능의 결과물에 영향을 받는 개인과 집단이 자신의 권리를 이해하고 행사할 때 이 가이드라인을 참고할 수 있다.

2. 정의

14. 본 가이드라인에서 말하는 인공지능은 일차적으로는 학습과 추론, 판단을 전자적으로 구현하는 알고리즘과 해당 프로세스를 지칭하나, 이

차적으로는 빅데이터 등 인공지능을 기능하게 하는 일련의 기술들을 포함하고 있습니다.

15. 본 가이드라인에서 기본 원칙은 인공지능의 개발과 활용에 있어서 기본이 되는 인권적 가치를 도출한 것으로 본 가이드라인이 추구하는 방향을 제시한 원칙입니다.

16. 본 가이드라인에서 주요 내용은 기본 원칙을 바탕으로 구체적으로 적용해야 할 과제들을 제시하였으며, 향후 입법 및 제도화 과정에서 주요 내용을 보다 구체화할 필요가 있습니다.

이 가이드라인이 보호하고자 하는 인권은 국내·외 인권 규범에서 인정하는 인권 개념 전체를 아우른다. 다만 그 가운데 인공지능 환경에서 특히 문제가 불거져 온 개인정보에 대한 권리, 차별받지 않을 권리, 구제를 받을 권리에 더욱 주목하고자 한다.

인공지능에 대한 정의는 매우 다양하며, 빠르게 발전하는 중이다. 이 가이드라인은 국내·외에서 마련된 인공지능에 대한 정의 규정을 종합적으로 고려하되, 특정 기법을 명시하기보다 인권 보호의 대상으로서 폭넓게 정의하고자 하였다. 이 가이드라인의 적용대상이 되는 인공지능은 학습, 추론, 판단을 전자적 방법으로 구현하는 알고리즘 시스템뿐 아니라 빅데이터 처리, 사물인터넷, 클라우드컴퓨팅, 온라인플랫폼서비스 등 인공지능을 기능하게 하는 일련의 기술들을 포함하는 것으로 본다.

국가나 기업이 인공지능을 개발하거나 활용하는 행위 또는 그렇게 하도록 하는 규정은 개인 또는 집단의 인권에 영향을 미칠 수 있다. 인공지능 환경에서 영향을 받는 사람들은 개발자나 이용자와 또 다른 이해당사자로서, 인공지능 관련 인권 규범에서는 이들의 권리를 중요한 문제로 다루고 있다. 특히 사람에 대한 의사결정이 인공지능에 기반하여 이루어지는 경우 당사자의 권리에 미치는 영향이 중대할 수 있다. 인공지능의 개발 및 활용 환경에 오류, 편향 및 차별이 있을 경우에는 그 적용 대상이 되는 사람들이 개인적 또는 집단적 수준에서 부정적인 영향을 받을 수 있다.

지능정보화기본법의 정의 (제2조 제4호)

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

가. 전자적 방법으로 학습 · 추론 · 판단 등을 구현하는 기술

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

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

라. 「클라우드컴퓨팅 발전 및 이용자 보호에 관한 법률」 제2조제2호에 따른 클라우드컴퓨팅기술

마. 무선 또는 유 · 무선이 결합된 초연결지능정보통신기반 기술

바. 그 밖에 대통령령으로 정하는 기술

다른 나라의 정의 사례

캐나다 (자동화된 의사결정 정부 훈령)

인공지능 : 언어 이해, 행동 학습 또는 문제 해결과 같이 일반적으로 생물학적 지능이 필요한 작업을 수행하는 정보 기술이다.

자동화된 의사결정 시스템 : 인간 의사결정자의 판단을 지원하거나 대체하는 기술을 포함한다. 이러한 시스템은 통계학, 언어학, 컴퓨터 과학과 같은 분야에서 파생되며, 규칙 기반 시스템, 회귀 및 예측 분석, 기계 학습, 딥 러닝 및 신경망과 같은 기술을 이용한다.

유럽평의회 (알고리즘 시스템의 인권 영향에 대한 대응 지침)

‘알고리즘 시스템’이란 종종 수학적 최적화 기술을 이용하여 데이터 수집, 결합, 정리, 정렬, 분류 및 추론 뿐 아니라 선택, 우선 순위, 권장 사항 및 의사결정과 같은 업무를 하나 이상 수행하는 응용프로그램으로 이해된다. 적용되는 설정에서 요구 사항을 충족하기 위해 하나 이상의 알고리즘에 의존하는 알고리즘 시스템은 대규모 및 실시간으로 적응형 서비스를 생성하는 방식으로 업무를 자동화한다.

제3절 기본 원칙

1. 인간의 존엄성

17. 「헌법」 제10조에서 보장하고 있는 인간의 존엄과 가치는 누구나 누려야 할 불가침의 기본적 인권으로, 모든 권리의 출발점인 동시에 종국적으로 보장되어야 할 인권적 가치입니다.
18. 어떠한 활동도 인간의 존엄에서 유래하는 다양한 권리들의 희생을 강요해서는 안 되며, 궁극적으로 모든 활동은 인간의 존엄과 가치를 증대시키는 방향으로 수행되어야 합니다.
19. 따라서 인공지능은 인간으로서의 존엄과 가치 및 행복을 추구할 권리에 부합하는 방식으로 개발 및 활용되어야 하며, 개인의 선택과 판단 및 행동을 강요하거나 자율성을 침해해서는 안 됩니다.

인공지능을 개발하거나 활용할 때에는 인간의 존엄성이 우선되어야 사람 중심의 인공지능이 구현될 수 있다.

유엔 사무총장은 인공지능을 비롯한 신기술의 개발, 사용 및 거버넌스에 있어 모든 인권을 보호하고 강화하는 것을 목표로 하고, 온라인과 오프라인에서 모든 인권을 동등하게 존중하고 이행해야 한다고 지적하였다.[169] 국가인권위원회는 인공지능 환경에서 인권 및 인간존엄성 존중 원칙을 보장하고, 인권을 보호하기 위한 조치가 필요하다는 의견을 표명한 바 있다(국가인권위원회 2020. 5. 6. 「인공지능산업 진흥에 관한 법률안」에 대한 의견표명).

국가와 기업의 인공지능 활용은 여러 가지 편익을 가져올 수 있지만, 개인의 선택과 판단 및 결정에 대한 존중이 전제되어야 한다. 개인의 자율성을 침해하는 조작, 착취, 통제를 위해 인공지능을 개발하거나 활용하는 것은 금지되어야 한다. 유엔 의사 표현의 자

169) 유엔문서 A/HRC/43/29, 62(a)문.

유 특별보고관은 개인이 의견을 형성하고 보유할 수 있는 역량, 정보 환경에서 접근하고 표현할 수 있는 역량을 인공지능이 대체 또는 조작하거나 방해해서는 안 된다고 지적하였다. 또한 인공지능 관련 입법은 이용자가 알고 선택하고 통제할 수 있도록 보장하여 개인의 자율성을 존중해야 한다고 강조하였다.[170] 특히 인공지능의 추론과 예측은 사람들이 자신의 자율성과 정체성을 수립하는 데 깊은 영향을 미친다.[171]

인공지능은 어떻게 개발되고 활용되는지에 따라 개인정보 및 사생활의 권리, 차별받지 않을 권리 뿐 아니라 표현의 자유, 집회 및 결사의 자유, 효과적인 구제를 받을 권리 및 공정한 재판을 받을 권리, 노동권을 비롯한 여러 경제적·사회적 권리를 향유하는 데 영향을 미친다. 따라서 인공지능을 개발하거나 활용할 때에는 당사자의 인권에 미치는 영향 뿐 아니라 사회에 공공적으로 미치는 영향에 대한 깊은 고려가 필요하다.

2. 알 권리

20. 알 권리란 개인의 의사형성에 필요한 정보를 수집하고, 수집된 정보를 취사·선택할 수 있는 자유를 의미합니다.
21. 개인은 알 권리를 통해 충분한 정보를 획득할 수 있고, 지식과 이해의 폭을 넓힐 수 있으며, 이를 바탕으로 합리적인 판단과 자신의 인격을 발현시킬 수 있습니다.
22. 개인의 의사 판단과 인격 발현의 중요한 요소인 알 권리의 보장을 위해 인공지능의 판단 과정 및 결과에 대한 합리적인 설명 등을 보장하도록 하고, 이를 뒷받침하는 기술적·제도적 장치를 마련해야 합니다.

170) 유엔문서 A/73/348, 58문.
171) 유엔문서 A/HRC/48/31, 17문.

인공지능에 대한 알 권리 보장은 그 대상이 되는 당사자의 권리를 보호하고, 인권 및 안전에 대한 공공 정책을 집행할 수 있는 기초가 된다. 인공지능의 개발과 활용으로 영향을 받는 당사자들은 그 사실에 대하여 알 수 있어야 이의를 제기하거나 권리구제를 요구할 수 있다. 인권 및 안전 관련 규정 준수를 감독하는 기구들은 공공기관과 기업이 개발 및 활용하는 인공지능에 대하여 조사할 수 있어야 한다.

그러나 일부 인공지능 알고리즘 및 이에 기반한 의사결정은 학습, 추론, 판단의 과정과 결과에 이른 이유를 설명하기 어려운 특성이 있다. 유럽평의회는 인공지능 시스템이 투명하지 않고 영향을 받는 당사자가 알 수 없는 방식으로 작동하기 때문에, 그 책임성이 모호해지고 특히 소수자 및 취약 집단에 부정적인 영향을 미칠 수 있다고 지적하였다.[172] 따라서 국가와 기업은 인공지능의 개발과 활용에 대하여 투명성을 보장하기 위해 노력하여야 하며, 인공지능의 사용이 인권에 미치는 영향이 중대할수록 더 높은 투명성을 보장해야 한다.[173]

많은 인공지능 시스템의 의사결정 과정이 불투명하다. 인공지능 시스템의 개발 및 운영을 뒷받침하는 정보 환경, 알고리즘, 모델의 복잡성은 물론 정부와 민간 행위자들의 … 비밀주의는, 인공지능 시스템이 인권과 사회에 미치는 영향을 일반 대중이 이해할 수 있는 뜻깊은 여정을 방해하는 요인이다. 머신러닝 시스템은 불투명성의 핵심적인 요인이다. 이 시스템은 설명하기 어렵거나 불가능한 방식으로 패턴을 식별하고 설명하기 어렵거나 불가능한 처방을 내릴 수 있다. 이를 흔히 '블랙박스' 문제라고 한다. 불투명성으로 인해 인공지능 시스템을 유의미하게 조사하는 것이 어려워지고, 인공지능 시스템이 위해를 야기하는 경우 불투명성이 효과적인 책무성 확보에 장벽이 될 수 있다. 그럼에도 불구하고 이 시스템들이 전적으로 조사될 수 없는 것은 아니라는 점에 주목할 필요가 있다.

【원문】 유엔 인권최고대표 <디지털시대 프라이버시권 보고서> 20문.

172) Council of Europe (2020), A.7문.
173) 유엔문서 A/HRC/48/31, 60(b)문.

3. 자기결정권

23.「헌법」 제10조가 규정하고 있는 인간의 존엄과 행복추구권에서 도출되는 자기결정권은 결정의 주체인 개인이 중대한 사안에 대해 외부의 강요 없이 누구나 스스로의 판단에 따라 타인의 간섭 없이 결정하고 행동할 수 있으며, 동시에 선택하지 않을 자유를 의미합니다.

24. 인공지능은 인간을 보조하여 서비스를 제공하는 것뿐만 아니라 주어진 데이터를 바탕으로 스스로 학습하여 의사결정을 내릴 수 있는 수준에 이르렀습니다.

25. 따라서 자기결정권은 인공지능의 개발과 활용의 전 과정에서 우선적으로 보장되어야 합니다.

유엔 인권최고대표는 인공지능 환경에서 사생활에 대한 권리는 개인과 그 삶에 대한 정보와 해당 정보에 기초한 의사결정에 대한 권리, 그리고 개인의 정체성을 결정할 자유를 아우른다고 설명하였다.[174] 인공지능이 사용한 데이터셋에는 개인에 대한 정보가 포함되어 있으므로 개인정보자기결정권이 관련되며, 개인정보를 이용한 인공지능의 예측이 수색, 검문, 체포 및 기소 등 공권력의 강제적인 조치로 이어지면 다른 인권에도 영향을 미친다.[175]

개인정보자기결정권은 자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리를 말한다(헌재 2005. 5. 26. 결정 99헌마513). 이때 정보주체는 처리되는 정보에 의하여 알아볼 수 있는 사람으로서 그 정보의 주체가 되는 사람을 의미한다(개인정보보호법 제2조 제3호). 국가와 기업이 인공지능을 개발하거나 활용하기 위하여 개인정보가 포함된 데이터를 처리할 때에는 정보주체의 개인정보자기결정권을 보장하고 개인정보보호 원칙을 준수하여야 한다.

174) 같은 문서, 7문.
175) 같은 문서, 24문.

4. 평등과 차별 금지

26.「헌법」 제11조는 모든 국민에게 평등권을 보장하고 있으며,「국가인권위원회법」 제2조는 합리적인 이유 없이 성별, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향 등 개인 특성에 따라 특정한 사람을 우대・배제・구별하거나 불리하게 대우하는 행위를 평등권 침해의 차별행위로 정의하고 있습니다.

27. 인공지능의 개발과 활용은 개인의 행복과 사회적 공공성의 증진에 위배되어서는 안 되고, 인공지능을 통한 경제・사회・문화적 권리의 향유에 있어서 다양한 계층을 포용하고 참여의 기회를 동등하게 보장해야 합니다.

28. 또한 인공지능의 결정이 특정 집단이나 일부 계층에게 차별적이거나 부정적 영향을 초래하지 않기 위해 개발 단계부터 다양한 계층의 의견을 수렴하고, 차별적 결과가 발생하지 않도록 필요한 조치를 취해야 합니다.

인공지능의 개발과 활용 결과에서 국내・외 인권 규범에서 금지하고 있는 직간접적 차별이 나타났다는 우려가 커지고 있다. 인공지능이 사용한 데이터셋에 내포된 역사적인 편견이 그 결과물에 반영되어 차별을 영구화하거나 강화할 위험이 있다는 것이다. 유럽연합 <인공지능 백서>는 인간의 의사결정에도 오류와 편견이 있을 수 있지만, 인공지능 기반 의사결정에 작용하는 편향성은 이를 사회적으로 통제하지 않을 경우 훨씬 더 많은 사람들에게 장기간 영향을 줄 수 있다고 지적하였다.[176]

176) European Commission (2020a), p11.

독일 연방차별금지국 연구보고서가 소개한 주요 인공지능 차별 사례

채용 분야

미국 아마존사가 2014년부터 사용해온 채용 시스템이 지원자 이력서를 평가하면서 '여성체스클럽'과 여자대학교 이름 등 여성 관련 단어에 대하여 감점한 것으로 나타나 사용이 중단되었다. 지난 10년의 우수사원 데이터로 학습한 이 시스템은 주로 남성 데이터를 학습하였던 것으로 밝혀졌다.

맞춤 광고 분야

페이스북 맞춤 광고가 청소년이나 장애인을 둔 가정에는 주택 광고를 노출하지 않고 한부모나 노인에게는 구인 광고를 노출하지 않는 등, 이용자를 차별하는 기능을 제공했던 것으로 나타났다. 이에 통신노동조합, 공정주택연맹, 시민자유연합 등 여러 원고들이 미국 법원에 페이스북과 광고주들을 상대로 다수의 집단소송을 제기하였고, 2019년 페이스북은 일부 광고에서 맞춤 광고를 제한하기로 원고들과 합의하였다.

금융서비스 분야

2018년 핀란드 국가 차별금지 및 평등 심판원은 한 신용평가회사의 자동화된 신용평가 방식이 차별금지법을 위반했다며 10만 유로의 과징금을 부과하였다. 이 회사는 대출 신청인에 대하여 대출 연장을 거부하는 결정을 하면서 신청인의 언어, 성별, 거주지역, 연령 등 대리변수와 타인의 통계 데이터에 기반하였던 것으로 나타났다. 회사는 신청인의 개인 상환 능력을 평가하거나 그에 대해 해명할 기회도 부여하지 않았다.

형사사법 분야

2016년 미국 여러 법원이 피고의 재범 위험을 예측하기 위하여 사용하는 컴파스(COMPAS) 시스템은 흑인의 위양성률이 백인보다 두 배 높다는 언론의 비판을 받았다. 개발 회사는 위양성률의 경우 흑인의 재범률이 백인보다 높은 인구 구성 비율의 영향을 받았다고 해명하였으나, 양측이 서로 다른 공정성 기준을 사용하면서 사회적 논란이 커졌다. 위스콘신 주 대법원은 한 소송에서 컴파스가 판사가 고려하는 여러 요소 중 하나일 뿐이기 때문에 공정하게 재판받을 권리를 침해하지 않았다고 판단하였다.

생체인식 분야

2017년 한 연구에서 유튜브 음성인식 자동자막 서비스는 여성 음성과 사투리 인식에서 그 정확도가 현저히 낮아진 것으로 나타났다. 2018년 또다른 연구에서는 상용 얼굴인식 시스템들이 어두운 피부색의 여성 얼굴에 대한 오인식률이 밝은 피부색의 남성 얼굴에서보다 높은 것으로 나타났다. 연구자들은 이러한 문제의 원인으로 학습 데이터의 편향성을 들었다.

【원문】 Carsten Orwat (2020). Risks of Discrimination through the Use of Algorithms : A study compiled with a grant from the Federal Anti-Discrimination Agency. Chapter 4.

인공지능을 개발하거나 활용할 때에는 차별적인 영향을 받을 위험이 있는 당사자 개인 및 집단에 대하여 각별한 주의를 기울여 차별 위험을 방지하거나 완화하여야 한다. 특히 부분적으로나 완전히 인공지능에 기반하여 이루어진 자동화된 의사결정에서 차별을 당하는 것은 중대한 인권침해이므로 반드시 필요한 조치를 취해야 한다.

제4절 주요 내용

1. 투명성과 설명 의무

29. 인공지능 및 이를 이용한 의사결정이 개인에게 미치는 영향력과 중요성이 갈수록 증가하고 있으므로, 인공지능의 판단과정과 그 결과에 대한 적절하고 합리적인 설명이 보장되어야 합니다. 학습 및 추론, 판단의 과정과 결과에 이른 이유를 설명하기 어려운 인공지능은 이에 대한 대응의 불확실성과 영향을 받는 당사자의 불안감을 유발하고, 인권 및 안전에 관한 법령과 정책의 집행효과를 불분명하게 할 수 있습니다.

30. 인공지능과 상호작용하는 사람에게는 언제나 그 상대방이 인공지능이라는 사실을 알려야 합니다.

31. 공공기관은 인공지능의 개발과 활용 계획 등을 사전 공개하여야 하고, 관련 당사자들의 의견을 공청회 등으로 수렴하여야 합니다. 공공기관이 인공지능을 통한 의사결정을 할 때 설명할 수 없는 인공지능을 활용해서는 안 되며, 특히 조달의 경우 입찰 단계에서부터 설명가능성이 보장되어야 합니다.

32. 공공기관이 개발하고 활용하는 모든 인공지능과 민간이 개발하고 활용하는 인공지능 중 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능(이하 ‘고위험 인공지능’이라 한다.)은 사용된 데이터와 인공지능 알고리즘의 주요 요인을 일반에게 공개하고 설명하여야 합니다.

33. 또한, 인공지능에 의한 자동화된 의사결정이 예정되어 있는 경우, 영향을 받는 당사자들은 사전에 그 사실을 알아야 합니다. 자동화된 의사결정에 의하여 영향을 받는 당사자는 그 결정의 이유에 대하여 설명을 듣고, 당사자 진술을 할 수 있으며, 이의를 제기할 수 있어야 합니다.

34. 특히, 완전히 자동화된 의사결정으로만 개인에게 법적 효력 또는 생명·신체·정신·재산에 중대한 영향을 미치는 일은 제한되어야 하고, 이러한 의사결정이 이루어진 경우에는 당사자가 해당 방식을 거부하거나 인적 개입을 요구할 수 있는 권리를 보장받아야 합니다.

인공지능 인권 규범들은 인공지능 환경에서 인권을 보장하기 위하여 여러 투명성 조치들을 제안한다. 가장 기본적인 투명성은 인공지능을 이용하는 사람에게 그 상대방이 인공지능이라는 사실을 알리는 것이다. 유엔 인권최고대표는 모든 인공지능 사용자가 시스템의 종류, 사용 목적, 개발자와 운영자에 대한 정보를 제공해야 한다고 보았다.[177]

공공기관이 개발하거나 활용하는 인공지능에 대해서는 높은 투명성이 요구된다. 공공기관은 인공지능의 도입 계획을 사전에 공개하고, 공청회 등 다양한 방법으로 영향을 받을 가능성이 높은 당사자들의 의견을 수렴하고 참여를 보장하여야 한다. 공공기관의 의사결정에는 설명할 수 없거나 조사될 수 없는 인공지능이 활용되어서는 안 된다.[178] 특히 강제적인 조치로 이어지는 인공지능 기반 결정에 불투명성이 있다면 국가의 책무성과 적법절차 측면에서 심각한 문제가 될 수 있다.

캐나다의 경우 2019년부터 <자동화된 의사결정 정부 훈령>[179]에 의사결정에 사용되는 공공기관 인공지능 시스템의 투명성 보장을 위한 규정을 두고 있다. 자동화된 결정 전에는 일반에 그 사실을 공지하고, 자동화된 결정 후에는 영향을 받은 당사자들에게 설명하며, 의사결정 시스템의 구성 요소에 대한 공공기관의 조사를 보장하였다. 2020년부터

177) 유엔문서 A/HRC/48/31, 55문.
178) 같은 문서, 59(h)문.
179) Government of Canada (2019).

<인공지능 조달 지침>[180]을 시행중인 영국 정부는 공공 조달되는 인공지능 도구, 데이터 및 알고리즘에 대해 투명성을 요구하면서, 그 설계 기준으로 설명 가능성과 해석 가능성을 장려하고 블랙박스 효과 및 특정 공급업체 종속을 방지하고자 하였다.

시민들에게 알고리즘 등록부를 공개한 해외 도시

2020년 9월 네덜란드 암스테르담 시와 핀란드 헬싱키 시는 시에서 사용하는 인공지능 알고리즘에 대해 등록하고 공개하는 알고리즘 등록부를 도입하였다.

이 알고리즘 등록부는 각 알고리즘의 도입과 운영을 책임지는 공직자의 이름, 부서 및 연락처를 공개하고 시민들이 의견을 제출할 수 있도록 하였다. 더불어 각 인공지능 시스템의 △학습 데이터셋에 대한 정보 △데이터 처리에 대한 정보 △차별 방지에 대한 정보 △인간 감독에 대한 정보 △위험성에 대한 정보 등을 읽기 쉬운 평문으로 시민들에게 공개하였다.

【참고】 암스테르담시 알고리즘 등록부 <https://algoritmeregister.amsterdam.nl/en/ai-register/>; 헬싱키시 알고리즘 등록부 <https://ai.hel.fi/en/ai-register/>

공공기관 인공지능 알고리즘에 대한 각국 법원 판결

○ 2017년 미국 텍사스 휴스턴 소재 연방지방법원은 민간에 위탁한 교육청의 교사 평가 도구에 대한 교사연맹의 소송에서 이 도구의 불투명성이 적법절차를 위반하였다고 판단했다. 법원은 기업의 영업 비밀과 국민의 헌법상 권리인 적법절차를 모두 충족하기 위해서는 공공기관의 중요한 의사결정에 비밀 알고리즘을 사용해서는 안 된다고 지적하였다.

【참고】 Federal Suit Settlement: End of Value-Added Measures in Houston. Education News (2017. 10. 10.); Houston Federation of Teachers, Local 2415 v. Houston Independent School District, 251 F. Supp. 3d 1168 (2017).

○ 폴란드에서 정부의 실직자 점수 시스템이 불투명성, 차별, 이의제기 문제로

180) UK Government. Guidelines for AI procurement. <https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement (검색일: 2021. 11. 1)>.

많은 사회적 논란을 빚자 폴란드 인권위원회는 이 사건을 헌법재판소에 회부하였다. 2018년 폴란드 헌법재판소는 국회에서 입법한 법률적 근거를 갖추지 않고 정부가 이 시스템을 사용한 것이 위헌이라고 결정하였다.

【참고】 Poland: Government to scrap controversial unemployment scoring system. Algorithm Watch (2019. 4. 16.)

○ 2020년 네덜란드 헤이그 지방법원은 정부의 사회복지급여 부정수급 탐지 시스템에 대하여 운영을 중단하라고 판결하였다. 이 시스템은 여러 공공기관에 분산되어 있던 개인정보를 부당하게 결합하고 충분한 안전조치를 취하지 않았으며, 위험 평가 모델이나 지표를 공개하지 않았다는 것이다. 법원은 이로 인하여 시스템의 평가 결과로부터 개인을 방어하거나 불법적인 차별을 조사할 방법이 없었다고 지적하였다.

【참고】 Welfare surveillance system violates human rights, Dutch court rules. The Guardian (2020. 2. 5.)

인권에 중대한 영향을 미칠 수 있는 공공기관 인공지능 및 민간의 고위험 인공지능에 대해서는 더욱 높은 투명성이 요구되며,[181] 사용된 데이터와 알고리즘의 주요 정보를 일반에 공개할 필요가 있다.[182] 여러 나라가 신용평가, 플랫폼 노동 등 인권에 중대한 영향을 미치는 인공지능의 경우 그 순서나 평가에 영향을 미치는 주요 요소를 공개하는 제도를 마련하고 있다.[183]

개인에게 영향을 미치는 의사결정이 부분적으로나 완전히 인공지능에 기반하여 자동

181) Council of Europe (2020), B.4.1문.

182) 유엔문서 A/HRC/48/31, 55문.

183) 「온라인 플랫폼 중개거래의 공정화에 관한 법률안(2021. 1. 28. 정부 발의)」은 중개거래계약서 필수 기재 사항에 플랫폼 알고리즘에 의하여 상품이 노출되는 순서, 형태 및 기준을 공개하도록 하였다. 미국 연방거래위원회(FTC)는 「AI와 알고리즘 사용에 대한 지침(Using Artificial Intelligence and Algorithms)」(2020)에서 신용평가 등에서 알고리즘을 사용하여 소비자에게 위험도 점수를 부과할 경우, 점수 및 중요도 순위에 영향을 미치는 주요 요소(key factors)를 공개하도록 하였다. 유럽연합은 2020년 이른바 'P2B규칙'(온라인 플랫폼 공정성·투명성 규정)에서 온라인플랫폼 사업자가 웹사이트 화면에 배열되는 업체·상품 등의 우선순위를 결정짓는 주요 변수 및 고려되는 각 변수간의 상대적 중요도를 약관에 명시하고, 특정 업체의 경제적 대가 지급이 검색·배열순위의 결정에 영향을 미치는 경우 그러한 사실을 약관에 명시하거나 일반 대중에 공개하도록 하였다. 호주는 2021년 뉴스 미디어 협상법(News Media Bargaining Code)에서 플랫폼이 뉴스 노출 알고리즘을 변경할 때 그 해당 사항을 언론사에 고지하도록 하였다. 스페인은 2021년 이른바 '라이더법'(디지털 플랫폼 유통에 종사하는 개인의 고용 상태에 관한 법률)에서 플랫폼 노동자의 근로조건, 고용과 해고 결정에 영향을 미칠 수 있는 업무 배치와 평가 관련 알고리즘과 인공지능에 관한 정보를 근로자대표에게 공개하도록 하였다.

으로 이루어졌을 때에는 그 당사자의 권리를 보장하여야 한다. 인공지능에 기반한 결정의 대상이 되는 당사자는 그러한 결정이 예정되어 있다는 사실을 사전에 제공받고 본인의 의견을 제시할 권리가 있다. 또한 당사자는 의사결정 과정 및 결과에 대하여 합리적인 설명을 듣고 이의를 제기할 수 있어야 한다. 특히 당사자의 권리와 의무에 미치는 영향이 지대한 경우 완전히 자동화된 의사결정은 원칙적으로 제한되어야 한다. 예외적으로 이러한 방식의 결정이 이루어지는 경우에는 당사자가 이러한 방식을 거부하거나 인적 개입을 요구할 수 있는 권리가 명시적으로 보장되어야 한다. 유엔 인권최고대표는 국가가 인공지능 의사결정에 대한 개인의 권리를 강화하고, 완전히 자동화된 결정에 대하여 충분한 설명을 들을 권리와 거부할 권리 등을 보장해야 한다고 지적하였다.[184]

우리나라 정부는 2021년 9월 국회에 발의한 「개인정보보호법」 일부개정법률안에서, 인공지능 등을 이용한 완전히 자동화된 결정이 정보주체의 권리 또는 의무에 중대한 영향을 미치는 경우 정보주체가 이를 거부하거나 해당 결정에 대한 설명 등을 요구할 수 있도록 하였다(안 제37조의2).

184) 유엔문서 A/HRC/48/31, 42문.

2. 사생활의 비밀 및 개인정보자기결정권 보장

35. 인공지능과 관련하여 정보주체의 권리는 처리된 개인정보에 관하여 고지를 받을 권리, 개인정보 접근 및 열람권, 개인정보처리 동의권 및 정정·삭제권, 처리정지권 등을 포함하며, 정보주체는 자신의 데이터가 사용되는 방법을 이해하고 그에 대한 통제권을 가지는 것이 중요합니다. 정보주체는 인공지능 서비스가 언제, 어디서 자신의 데이터를 수집하고, 어떻게 데이터를 처리하여 사용, 보관, 삭제되는지에 대해 알고 참여할 권리가 있습니다.

36. 인공지능의 개발과 활용에서 개인정보는 목적에 필요한 범위에서 최소한의 개인정보만을 처리하여야 하며, 처리목적 달성에 필요한 기간 동안만 보관되어야 합니다. 또한 이러한 개인정보 처리 원칙은 정보주체가 확인할 수 있도록 공개되어야 합니다.

37. 개인정보자기결정권은 정보주체의 자기 정보에 대한 통제력을 보장하기 위해 인정되는 것인데, 그 통제력 보장의 핵심은 정보주체의 동의권입니다. 따라서 개인정보 처리에 대한 정보주체의 동의는 단순한 외형적 의사 표시만이 아니라 정보주체가 개인정보 처리에 대한 제반 상황을 설명·제공받고 스스로의 자유의사에 기하여 결정할 수 있어야 합니다.

38. 인공지능의 개발과 활용에서 민감정보를 처리할 때에는 특별한 주의를 기울여 보호하여야 합니다. 더불어 의사결정의 내용과 관련성이 없거나, 부정확한 데이터에 기반한 의사결정이 이루어지지 않도록 데이터의 정확성, 완전성, 최신성을 보장해야 합니다.

이 절의 내용은 「헌법」과 개인정보보호법 등의 기준에 비추어 당연한 원칙을 열거한 것이다. 즉, 개인정보처리자는 인공지능의 개발과 활용의 전 과정에서 정보주체의 권리를 보장하고 개인정보보호 원칙을 준수하여야 한다. 최근 인공지능이 사용하는 데이터셋에 개인정보를 포함하여 처리하면서 이 사실을 정보주체에게 명확히 알리지 않거나 가명처리했다는 이유로 정보주체의 동의 없이 개인정보를 처리하는 일이 크게 늘었다. 개인정보처리자는 적법하고 공정하며 투명한 개인정보 처리로 정보주체가 신뢰할 수 있는 인공지능 환경을 구축할 필요가 있다.

개인은 인공지능 시스템이 개인정보를 처리하는 목적과 그 결과물 등에 대한 정보를 사전에 제공받고 자유롭게 동의 여부 및 동의 범위를 선택하며 자신에 관한 개인정보를 통제할 수 있어야 한다.[185] 인공지능을 개발하거나 활용하는 개인정보처리자는 개인정보를 사용, 보관, 삭제하는 모든 처리에 대하여 정보주체에게 알리고, 개인정보의 열람권, 정정·삭제권, 처리정지권 등 정보주체의 참여권 행사를 보장하여야 한다.[186] 정보주체가 인공지능의 학습용 데이터에 관한 정보를 얻고 해당 정보가 가명처리된 이후의 과정에도 합리적으로 가능하고 가명처리의 목적 달성에 저해되지 않는 한 참여할 수 있도록 보장하는 것이 바람직하다.

더불어 개인정보보호법은 개인정보 처리의 목적 제한 원칙, 최소성·적법성·정당성 원칙, 정확성·완전성·최신성 원칙, 안전성 보장 원칙, 공개의 원칙 등을 규정하고 있으며, 개인정보처리자는 이를 준수하여야 한다. 정보주체의 민감정보는 특히 주의하여 처리하여야 하며, 별도의 동의 등 보다 제한적인 준수 기준을 충족하여야 한다.

챗봇 이루다에 대한 개인정보보호법 위반 조치

'이루다'는 2020년 12월 출시된 인공지능 챗봇 서비스로, 출시 2주 만에 약 75만 명에 달하는 이용자를 모으며 큰 인기를 끌었다. 그러나 이루다 발화 내용에서 여성·성소수자·장애인·흑인 등을 혐오하는 내용이 발견되면서 인공지능 윤리와 차별 논란을 빚었다.

무엇보다 이 서비스는 개발 과정에서 개인정보보호법을 중대하게 위반한 사실이

185) Council of Europe (2020), B.2.1문.
186) 유엔문서 A/HRC/48/31, 55문.

드러나 개발회사에 총 1억 330만 원의 과징금과 과태료가 부과되었다. 개인정보보호위원회는 개발회사가 자사 과거 서비스 이용자 60만 명의 카카오톡 대화문장 94억 건을 수집하고 이루다 인공지능 학습과 서비스에 이용하면서 그 사실을 이용자에게 명확히 알리고 동의를 받지 않았다고 밝혔다. 이렇게 불법으로 처리된 개인정보 데이터에는 만 14세 미만 아동 20여만 명의 개인정보가 포함되어 있었다.

【참고】 개인정보위, '이루다' 개발사 (주)스캐터랩에 과징금·과태료 등 제재 처분. 개인정보보호위원회 보도자료(2021. 4. 29).

3. 차별 금지

39. 인공지능을 개발하고 활용할 때는 인공지능으로 인해 영향받는 사람의 다양성과 대표성을 반영하기 위해 노력해야 하고, 성별, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향 등 개인과 집단의 특성에 따라 편향적이고 차별적인 결과가 나오지 않도록 해야 합니다.
40. 데이터의 수집·선정 및 시스템 설계, 활용 등 인공지능 개발 전반에 걸쳐 편향이나 차별을 배제해야 하고, 이는 데이터 요소를 검사하고 차별적인 데이터를 조정하는 등의 조치를 포함합니다.
41. 특히 학습용 데이터가 인공지능의 판단에 직접적인 영향을 미치는 상황을 고려할 때, 학습용 데이터의 수집 단계부터 차별적 요소를 통제하고 데이터 편향성을 최소화하여 인공지능을 통한 의사결정이 특정 집단에 부정적 영향을 미치지 않도록 해야 합니다.
42. 개발한 인공지능에 대해 주기적인 모니터링을 거쳐 데이터 품질과 위험을 관리하고, 차별적 결과나 의도치 않은 결과에 대해 개선의 조치를 주기적으로 수행해야 합니다.
43. 인공지능 기술 및 서비스에 대한 접근성과 인공지능이 주는 혜택은

사회적 약자와 취약계층을 포함하여 모든 사회구성원에게 평등하게 제공되어야 합니다.

차별적인 영향을 받을 위험이 있는 당사자를 대표하는 다양한 이들이 인공지능 수명주기의 개발 단계부터 인권 실사에 참여하여 협의하는 절차가 있다면 부정적인 인권 영향을 방지하고 완화할 수 있다. 더불어 국가와 기업은 인공지능이 사용하는 데이터셋과 모델 및 그 결과물의 오류, 편향 및 차별 가능성과 관련된 위험을 평가하고 이를 완화하는 조치를 취해야 한다.

캐나다 〈자동화된 의사결정 정부 훈령〉은 공공기관 인공지능 시스템이 사용하는 데이터의 편향성을 방지하기 위한 규정을 두고 있다. 생산에 착수하기 전에는 시스템이 사용하는 데이터에 대하여 의도하지 않은 데이터 편향이나 결과에 부당하게 영향을 미칠 수 있는 요소에 대해 검사하는 절차를 마련하고, 시스템이 의도하지 않은 결과를 방지하고 관련 법률을 준수하는지 확인하기 위하여 정기적으로 모니터링하는 절차를 두도록 하였다. 유럽평의회는 공공부문 뿐 아니라 민간부문에 대해서도 인공지능 시스템을 설계, 개발, 구현하는 모든 수명 주기에서 그 영향을 받는 개인과 집단에 대한 차별을 유발하거나 고착화하는 결과를 방지할 것을 요구하였다.

특히 아직 평등법(포괄적 차별금지법)이 존재하지 않는 우리나라의 경우 국가인권위원회법에 따라 인공지능 개발과 활용 환경에서 차별을 방지하기 위한 노력이 필요하다. 국가나 기업이 개발하거나 활용하는 인공지능이 차별을 금지하는 「헌법」과 「국가인권위원회법」, 장애·연령 등 분야별 차별금지법을 위반하였을 경우에는 그 차별 행위에 대하여 국가인권위원회에 진정할 수 있다.

국가인권위원회법의 차별 정의 (제2조)

"평등권 침해의 차별행위" 란 합리적인 이유 없이 성별, 종교, 장애, 나이, 사회적 신분, 출신 지역(출생지, 등록기준지, 성년이 되기 전의 주된 거주지 등을 말한다), 출신 국가, 출신 민족, 용모 등 신체 조건, 기혼 · 미혼 · 별거 · 이혼 · 사별 · 재혼 · 사실혼 등 혼인 여부, 임신 또는 출산, 가족 형태 또는 가족 상황, 인종, 피부색, 사상 또는 정치적 의견, 형의 효력이 실효된 전과(前科), 성적(性的) 지향, 학력, 병력

(病歷) 등을 이유로 한 다음 각 목의 어느 하나에 해당하는 행위를 말한다. 다만, 현존하는 차별을 없애기 위하여 특정한 사람(특정한 사람들의 집단을 포함한다. 이하 이 조에서 같다)을 잠정적으로 우대하는 행위와 이를 내용으로 하는 법령의 제정·개정 및 정책의 수립·집행은 평등권 침해의 차별행위(이하 "차별행위"라 한다)로 보지 아니한다.

가. 고용(모집, 채용, 교육, 배치, 승진, 임금 및 임금 외의 금품 지급, 자금의 융자, 정년, 퇴직, 해고 등을 포함한다)과 관련하여 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위

나. 재화·용역·교통수단·상업시설·토지·주거시설의 공급이나 이용과 관련하여 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위

다. 교육시설이나 직업훈련기관에서의 교육·훈련이나 그 이용과 관련하여 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위

라. 성희롱[업무, 고용, 그 밖의 관계에서 공공기관(국가기관, 지방자치단체, 「초·중등교육법」 제2조, 「고등교육법」 제2조와 그 밖의 다른 법률에 따라 설치된 각급 학교, 「공직자윤리법」 제3조의2제1항에 따른 공직유관단체를 말한다)의 종사자, 사용자 또는 근로자가 그 직위를 이용하여 또는 업무 등과 관련하여 성적 언동 등으로 성적 굴욕감 또는 혐오감을 느끼게 하거나 성적 언동 또는 그 밖의 요구 등에 따르지 아니한다는 이유로 고용상의 불이익을 주는 것을 말한다] 행위

국제 인권 규범은 모든 사회구성원이 인공지능 기술 및 서비스에 평등하게 접근하고 그 발전의 혜택 또한 평등하게 누려야 한다고 강조한다. 유엔 사무총장은 각국에 정보격차 및 기술 격차를 해소하고 신기술의 접근성, 가용성, 경제성, 적응성 및 품질을 개선하기 위하여 노력할 것을 촉구하였다. 특히 공공부문은 인공지능의 이용에 관한 정보와 지식을 대중에게 전파하기 위한 노력을 강화해야 한다. 경제·사회·문화적 권리의 향유에 필수적인 제품 및 서비스를 비롯하여 신기술에 대한 접근에 있어서 차별과 편견을 해소해야 함은 물론이다.[187]

187) 유엔문서 A/HRC/43/29, 62(c)문; 같은 문서, 62(e)문; 같은 문서, 62(i)문.

4. 인공지능 인권영향평가 시행

44. 국가는 인공지능의 개발과 활용에 있어 인권적 가치가 우선시 되도록 하여야 하며, 인공지능으로 인해 발생할 수 있는 인권침해와 차별에 대하여 사전적 또는 사후적으로 관리 감독을 할 의무가 있습니다.

45. 국가는 인공지능의 개발과 활용에 있어서 인권 침해와 차별의 가능성 및 정도, 영향을 받은 당사자의 수, 사용된 데이터의 양 등을 고려하여 인권영향평가를 실시해야 합니다. 특히 인공지능 기술이 적용되어 기존 제도로 관리되거나 감독될 수 없는 새로운 분야는 인권영향평가 제도를 도입해야 합니다.

46. 인권영향평가 내용에는 인공지능의 특성, 상황, 범위 및 목적을 감안하여 본 인권 가이드라인이 제시한 기본 원칙 및 주요 내용, 국제 인권 기준, 관련 법률에서 정한 의무 등이 포함되어야 하며, 인권 침해 위험요인의 분석, 개선 사항 등을 도출해야 합니다.

47. 인권영향평가는 개발 및 출시 전에 실시하고 인공지능의 기능 또는 범위 변경 시 평가를 갱신하여야 합니다.

48. 인권영향평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하기 위한 조치사항을 수립하여 적용하여야 하며, 원칙적으로 그 내용이 공개되어야 합니다. 또한, 이를 방지하거나 완화하는 조치를 취하기 전에는 그 개발과 활용을 중단해야 합니다.

49. 국가는 인권영향평가를 인권전문성과 독립성을 확보한 기관이 담당하도록 하고, 해당 기관은 인권영향평가의 활성화를 위하여 관계 전

문가의 육성, 영향평가 기준의 개발 및 보급 등 필요한 조치를 마련해야 합니다.

인공지능이 인권에 미치는 부정적인 영향이나 편향성 및 위험성을 방지하고 완화하기 위해서는 이를 평가하고 관리 감독하는 체계가 마련되어야 한다.

인권영향평가는 공공기관이나 기업의 사업 관계의 결과나 활동으로 인권에 미칠 수 있는 실제적 또는 잠재적으로 부정적인 영향을 평가하는 절차이다.[188] 인권영향평가는 유엔 「기업과 인권 이행지침」에서 권고하는 인권 실사(human rights due diligence) 의무의 핵심 도구이다. 인권영향평가는 기관(기업) 활동이 이해당사자의 인권에 미치는 부정적 영향을 식별하고 평가하며 이에 대응하는 방지 및 완화 조치를 제시함으로써 그 활동이 인권 친화적으로 수행될 수 있도록 돕는다. 우리나라에서도 여러 지방자치단체와 공공기관 들이 다양한 수준으로 인권영향평가를 시행해 왔다.

최근 인공지능이 인권에 미치는 영향, 특히 차별에 미치는 부정적 영향에 대한 우려가 증가함에 따라 국제 인권 규범은 인공지능에 대하여 인권영향평가를 실시할 것을 요구해 왔다. 유엔 인권최고대표는 인공지능 시스템의 설계, 개발, 배치, 판매, 구입, 운영의 수명주기 전반에 걸쳐 체계적인 인권 실사가 수행되어야 하고, 정례적이고 포괄적인 인권영향평가는 그 핵심이라고 강조하였다.[189]

인권영향평가를 자율적 또는 의무적으로 실시하는 대상은 인권 침해와 차별의 가능성 및 정도, 영향을 받는 당사자의 수, 사용된 데이터의 특성 등을 고려하여 결정될 필요가 있다. 유엔 사무총장은 인권에 중대한 영향을 미칠 수 있는 인공지능 시스템의 경우 그 전체 수명 주기에 걸쳐 인권영향평가가 포함된 체계적인 인권 실사의 대상이 된다고 보았다.[190] 특히 공공기관이 직접 개발하였거나 조달하는 인공지능에 대해서는, 여러 인공지능 인권 규범들에서 인권영향평가를 의무적으로 실시할 것을 권고하고 있다. 기존 제도로 관리되거나 감독될 수 없는 새로운 분야는 인권에 미칠 위험을 방지할 수 있도록

188) 인권 위험을 측정하기 위해 기업은 사업 관계의 결과로 또는 기업의 활동으로 인해 인권에 미칠 수 있는 실제적 그리고/또는 잠재적 부정적 영향을 식별하고 평가해야 한다(중략). 유엔 <기업과 인권 이행 지침>; 국가인권위원회(2018). 공공기관 인권경영 매뉴얼, p14.
189) 유엔문서 A/HRC/48/31, 60(a)문.
190) 유엔문서 A/HRC/43/29, 62(g)문.

인권영향평가가 반드시 실시될 필요가 있다.

인공지능 인권영향평가는 인공지능의 특성, 상황, 범위 및 목적을 감안하여 실시하고, 그 내용에 있어서 인공지능 알고리즘 뿐 아니라 그 입력 정보와 출력 결과물이 본 가이드라인이 제시한 기본 원칙 및 주요 내용, 국제 인권 기준, 관련 법률에서 정한 의무를 준수하는지 평가하여야 한다. 인공지능 인권영향평가는 인권 침해 위험 요인을 식별하고 인권 침해 위험요인을 분석하며 이를 개선하기 위한 조치 사항을 도출해야 한다,[191] 특히 인권영향평가는 소수자 및 취약 집단에 미치는 부정적인 영향에 대하여 각별한 주의를 기울여야 하고,[192] 그 영향을 받을 수 있는 당사자 및 집단의 의견을 반영할 필요가 있다.[193]

인공지능 인권영향평가의 실시 시기는 그 위험을 식별하고 조치를 취할 수 있도록 인공지능을 개발하거나 출시하기 전에 이루어져야 한다. 인공지능의 기능 또는 범위 변경 시에는 평가를 갱신하여야 할 것이다. 유엔 인권최고대표는 인공지능 시스템에 대한 인권 실사 의무는 그 구입, 개발, 배치 및 운영할 때 뿐 아니라 개인에 대한 빅데이터를 공유하거나 사용하기 전의 시점에도 적용된다고 강조하였다.[194] 영국 <인공지능 조달 지침>은 조달 공고 전에 데이터 편향성 등에 대하여 평가하도록 하고 이후 인공지능 설계 및 조달 절차에서 식별된 위험을 완화할 것을 요구하였으며, 개발 사양을 모르는 상태에서 이루어진 평가는 완전할 수 없기 때문에 이후 다시 평가할 것 또한 요구하였다.

인공지능 인권영향평가의 결과 식별된 위험에 대해서는 적절한 방치 및 완화 조치가 취해져야 한다. 인권 위험이 식별되었음에도 이를 방지하거나 완화하는 조치를 취하기 전이거나 어려운 경우에는 그 개발과 활용을 중단해야 한다. 인권영향평가의 결과와 조치는 그 영향을 받는 당사자들이 확인할 수 있도록 원칙적으로 공개되어야 한다.[195]

특히 국가는 이러한 인공지능 인권영향평가의 절차를 지원하고 주도할 뿐 아니라 기업의 인권영향평가 실시를 요구하고 장려할 수 있다.[196] 유럽평의회 인권위원장은 인공지능 인권영향평가는 인권기구를 비롯한 독립된 감독 기관이나 외부 전문가가 실시하여

191) Council of Europe Commissioner for Human Rights (2019), p7.
192) 유엔문서 A/HRC/48/31, 49문.
193) 같은 문서, 50문.
194) 같은 문서, 48문.
195) Council of Europe Commissioner for Human Rights (2019), p8; 같은 문서, 50문.
196) 같은 문서, 48문.

야 한다고 지적하였다.[197] 또한 공공기관은 영업비밀로 인해 인권영향평가를 실시할 수 없는 기관(기업)에서 인공지능을 조달해서는 안 된다.[198] 국가는 관계 전문가의 육성, 영향평가 기준의 개발 및 보급 등 필요한 조치를 마련하여 인권영향평가를 활성화할 수 있어야 할 것이다.

5. 위험도 등급 및 관련 법제도 마련

50. 국가는 공공기관과 민간이 개발하고 활용하는 인공지능에서 인권과 책임성을 보장하도록 관련 법률과 감독 체계를 보완해야 하며, 특히 인권을 보장하기 위한 구체적인 지침과 정책을 마련하여야 합니다.

51. 국가는 개인의 인권과 안전에 미치는 위험성이 매우 높아 인공지능이 금지되는 영역, 상당한 제한이 필요한 인공지능 고위험 영역, 위험성이 거의 없는 영역 등 적절하게 위험성 단계를 구분하고, 그에 맞는 규제 수준과 인적 개입이 이루어지도록 법과 제도를 마련하여야 합니다.

52. 특히, 당사자의 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능은 투명성과 설명가능성, 개인 정보 보호, 차별 금지 등 그 규제에 있어서 공공과 민간의 구분 없이 엄격하게 적용되어야 합니다.

53. 감독 기관은 공공기관과 민간의 위법한 인공지능 개발과 활용 여부를 조사하고 피해 구제 및 조치를 취하기 위하여 상세 정보에 접근할 수 있어야 합니다. 이를 위하여 공공기관 인공지능 및 민간 고위험 인

197) Council of Europe Commissioner for Human Rights (2019), p7.
198) 같은 문서, p8.

공지능 개발자 및 운영자는 사용된 데이터와 알고리즘의 주요 요소 등을 기록하고 문서화하여 일정 기간 보관하여야 합니다.

54. 국가는 인공지능을 독립적이고 효과적으로 감독할 수 있는 체계를 수립하여 개인의 인권과 안전을 보장하고 피해를 구제하여야 합니다. 인공지능 국가 감독 체계는 독립적이고 효과적이어야 하며, 진정 또는 인지로 접수한 사건을 조사하기에 충분한 자원, 권한 및 전문지식을 구비해야 합니다.

55. 국가는 인공지능으로 인하여 인권을 침해당하거나 차별을 받은 사람이 진정을 접수하여 권리를 구제받을 수 있는 기회를 보장하는 등 국가기관의 구제수단에 대한 접근을 보장해야 합니다. 인공지능을 개발하고 활용하는 공공기관과 민간은 언제든지 구제가 가능하도록 그 책임자에 대한 정보는 물론, 이의를 제기할 수 있는 기관과 방법에 대한 정보를 일반에 공개하여야 합니다.

국가는 인공지능이 인권에 미치는 부정적 영향을 방지하는 입법 조치를 취해야 할 의무가 있다. 국제 인권 규범은 인공지능의 개발과 활용에 있어 공공부문과 민간부문을 가리지 않고 책무성을 보장하는 구체적인 제도 마련을 각국에 촉구해 왔다.[199] 유럽평의회 인권위원장은 인공지능 시스템의 개발, 배치 또는 사용에서 발생하는 인권 침해에 대한 책임성과 책무성은 항상 자연인 또는 법인에게 있어야 한다고 지적하였다.[200] 특히 인공지능으로 인한 인권 침해 행위를 감독하고 그 피해자를 구제하는 국가적 체계가 마련되

199) 국가는 민간 부문 활동에 관한 조치를 포함하여 입법 조치를 취해야 할 의무를 재확인하고 준수하여야 하고, 이로써 신기술은 경제·사회·문화적 권리를 포함한 모든 사람들의 인권에 대한 완전한 향유에 기여하고 인권에 미치는 부작용이 방지되어야 한다. 유엔문서 A/HRC/43/29. 62(b)문; 신기술이 사용되는 상황에 대해 책임을 완전하게 보장하는 적절한 법률체계와 절차방법을 마련해야 한다. 이를 위해 국내 법제도의 공백을 검토 및 평가하고, 필요한 경우 감독 체제를 수립하고, 신기술로 인한 피해에 대해 국민이 이용할 수 있는 구제 수단이 구비되어야 한다. 같은 문서, 62(h)문.

200) Council of Europe Commissioner for Human Rights (2019), p13.

어야 한다. 입법적인 조치가 도입되기 전이거나 미치지 않는 경우, 인공지능의 개발과 활용으로부터 인권을 보호하기 위한 기관, 기업, 개인들의 자율적인 조치들이 인권 침해를 완화하는데 도움이 될 수 있다.

유엔 인권최고대표는 인권에 부정적인 영향이 발생할 가능성이 있을 때에는 인간이 감독하고 의사결정에 의무적으로 개입하는 규정이 있어야 한다고 지적하였다.[201] 특히 국제 인권 규범을 준수할 수 없는 인공지능을 금지하고, 고위험 인공지능의 경우 그 판매와 사용을 규제하는 규정을 마련하기 전까지 사용을 유예할 것을 각국에 요구하였다.[202] 최근 여러 나라에서 인공지능이 개인의 인권과 안전에 미치는 위험 정도에 따라 규제하는 제도를 도입하고 있다. 인권에 미치는 위험이 큰 분야로는 생명과 안전, 생체인식, 사회기반시설, 고용과 입점, 교육, 사회복지와 신용평가 등 필수사회서비스, 수사, 재판, 출입국관리 등이 꼽힌다. 고위험 인공지능에 대해서는 인적 개입은 물론, 투명성과 설명가능성, 개인정보보호, 차별 금지에 대한 의무가 공공과 민간을 구분하지 않고 반드시 보장되어야 한다.

각국 고위험 인공지능 규제 추진 사례

○ 캐나다는 2019년 <자동화된 의사결정 정부 훈령>에서 공공기관 의사결정에 사용되는 인공지능의 위험 수준별로 전문가 검토, 고지, 인적 개입, 설명, 검사, 모니터링, 교육훈련, 비상 계획, 승인의 의무를 부과하였다. 위험 수준은 의사결정이 개인·공동체의 권리, 개인·공동체의 건강과 복리, 개인·단체·공동체의 경제적 이익, 생태계의 지속가능성에 미치는 영향에 대한 평가 결과에 따라 4개 등급으로 나뉜다.

【참고】 The Government of Canada (2019). Directive on Automated Decision-Making.

○ 뉴질랜드는 2020년 <알고리즘 헌장>을 발표하고 위험 수준별로 투명성, 공익성, 의견 수렴, 데이터 편향 방지, 전문가 검토, 인적 감독 의무를 부과하였다. 위험 수준은 위험의 발생 빈도 및 그 영향의 범위와 심각성에 따라 3개 등급으로 나뉜다.

【참고】 New Zealand Government (2020). Algorithm charter for Aotearoa New Zealand.

201) 유엔문서 A/HRC/48/31, 45문.
202) 같은 문서, 59(c)문.

○ 미국은 연방거래위원회(FTC)가 2020년과 2021년 인공지능 사용 지침을 연달아 발표하고 기업이 개발하고 판매하는 인공지능이 투명성, 설명 가능성, 공정성, 견고성과 실증적 타당성, 책임성을 갖추도록 하였다. 또한 인종적으로 편향된 알고리즘의 판매나 사용, 고용・주택・신용평가・보험・복지급여에 대한 거부 결정시 불공정한 행위, 인종・피부색・종교・출신국가・성별・혼인여부・연령・공공부조 수령여부에 따른 차별을 금지하였다. 미국 정부와 의회는 고위험(highly sensitive) 자동화된 의사결정 시스템에 영향평가 의무를 부여하는 연방 입법을 추진 중이다.

【참고】 FTC (2020). Using Artificial Intelligence and Algorithms; FTC (2021). Aiming for truth, fairness, and equity in your company' s use of AI.

○ 유럽연합은 2021년 4월 <인공지능법(안)>을 발의하고 입법 절차를 밟고 있다. 인공지능법(안)은 ① 국민의 안전과 생계, 권리에 명백한 위협이 되는 인공지능을 금지하고 ② 고위험 인공지능에 대해서는 위험 평가 및 완화, 데이터셋의 고품질 보장, 기록 및 문서화, 정보 공개, 인적 통제, 견고성・보안성・정확성 등의 의무를 부과하고 ③ 제한적인 위험의 인공지능에 대해서는 투명성 의무를 부과하고 ④ 최소 위험의 인공지능에 대해서는 규제를 완화하였다. 이때 고위험 인공지능은 △교통 등 주요 인프라 △입시 등 교육훈련 △장난감 등 안전성 △고용 및 입점 △신용평가・사회복지 등 필수서비스 △법 집행 △출입국관리 △재판 등에 사용되는 인공지능이다.

【참고】 European Commission (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence.

유엔 인권 규범은 인권에 중대한 부정적인 영향을 미칠 수 있는데도 검사할 수 없는 인공지능 시스템을 사용하지 않아야 한다고 강조해 왔다.[203] 인권최고대표는 인공지능 기술에 대한 민관 협업 역시 투명해야 하고 인권 감독의 대상이 되어야 하며, 국가는 인권에 대한 책무를 포기해서는 안 된다고 지적하였다.[204]

특히 감독 기관은 위법한 인공지능의 개발과 활용 여부를 조사하고 피해 구제 및 조치를 취하기 위하여 상세 정보에 접근할 수 있어야 한다. 인권에 중대한 영향을 미칠 수

203) 유엔문서 A/HRC/43/29, 54문; 유엔문서 A/HRC/48/31, 56문.
204) 같은 문서, 59(j)문.

있는 공공기관 인공지능 및 민간의 고위험 인공지능을 개발하거나 운영하는 기관, 기업, 개인 등은 사용된 데이터와 알고리즘의 주요 요소 등을 기록하고 문서화하여 일정 기간 보관하고, 감독 기관이 인공지능 감독에 필요한 정보를 요청할 경우 이를 제공해야 한다. 유엔 인권최고대표와 유럽평의회는 지적재산권 또는 영업 비밀 보호 제도가 인권에 부정적 영향을 미치는 인공지능 시스템에 대한 조사에 지장을 주어서는 안 된다고 지적하였다.[205]

인공지능 시스템의 인권 준수에 대한 감독은 독립적이고 효과적으로 이루어져야 한다. 감독 기관은 인공지능을 개발 및 활용하는 공공기관과 기업들로부터 독립적이어야 하며, 감독과 구제 기능을 효과적으로 수행할 수 있는 권한과 자원을 구비해야 한다.[206]

공공 또는 민간 인공지능으로 인하여 인권 침해를 당하거나 차별을 받은 사람은 감독기관 등 이를 소관하는 국가기관에 진정을 제기하고 침해 행위나 차별에 대한 시정 등 그 권리를 효과적으로 구제받을 수 있어야 한다.[207] 이때 피해자가 이의를 제기하거나 권리 구제를 요구할 수 있도록 인공지능을 개발하거나 활용한 책임자에 대한 정보는 물론, 이의를 제기할 수 있는 기관과 방법에 대한 정보가 일반적으로 공개되어 누구든지 쉽게 접근할 수 있어야 한다. 기업들 역시 인공지능 처리에 대한 이용자들의 불만사항 및 이의제기에 적시에 대응하고 사람에 의한 구제 수단을 마련하여야 한다.[208]

56. 특히 국가는 대량 감시와 차별로 이어질 위험이 높은 얼굴인식 등 원격 생체인식 기술의 사용을 공공장소에서 금지하고, 특별한 경우에 한하여 사용을 허용하되, 인권 침해나 차별의 위험성이 드러난 경우 이를 방지하거나 완화하는 조치를 취하기 전에는 사용을 중단해야 합니다. 또한 국가는 생명의 존엄성 및 윤리를 훼손할 가능성이 높은 자율살상무기에 대하여 인도주의적으로 접근하고 그 연구, 개발, 생산 및 활용을 금지하는 국제 규범을 준수하고, 이에 대한 논의에 적극적으로 참여해야 합니다.

205) Council of Europe (2020), B.4.1문; 유엔문서 A/HRC/48/31, 56문.
206) Council of Europe Commissioner for Human Rights (2019), p10.
207) 같은 문서, p13.

한편, 생체인식기술의 발전으로 온라인과 오프라인의 공개된 장소에서 생체인식정보를 처리하는 경우가 증가하고 있다. 생체인식기술은 지문, 얼굴, 홍채, 정맥, 음성, 필적 등 개인의 신체적, 생리적, 행동적 특징에 관한 정보에서 추출한 특징점 등을 이용(비교・대조)하여 특정 개인임을 인증・식별하는 기술적 수단으로서 이 과정에서 생성된 생체인식정보는 개인정보보호법에서 특별히 보호하는 민감정보에 해당한다(개인정보보호법 제23조 및 동법 시행령 제18조). 한 개인의 생체인식정보는 다른 사람과 구별되는 독특한 특성을 나타내기 때문에 그 사람의 핵심적인 인격 특성을 구성한다. 특히 군중 속에서 원격으로 특정 개인을 식별하는 원격 생체인식은 공공장소에서 개인을 추적할 수 있는 국가와 기업의 능력을 증가시켜 대량 감시와 차별에 대한 우려를 낳았다. 실시간 원격 생체인식은 이동의 자유 뿐 아니라 표현의 자유, 집회 및 결사의 자유에 직접적으로 부정적인 영향을 미치기 때문에 고도로 위험한 인공지능에 해당한다[209]. 원격 생체인식은 잘못된 식별로 개인에게 불이익한 조치를 초래할 수 있고, 성별・출신 국가・출신 민족・인종 면에서 특정 집단에 편향적인 영향을 미치거나 이들 민감한 특성에 기반하여 개인을 프로파일링하는 등 추가적인 인권 침해가 이어질 수 있다. 따라서 국가는 공공장소에서 얼굴인식 등 원격 생체인식 기술을 사용하는 것을 원칙적으로 금지하고 예외적인 사용에 대해서는 법원의 통제 등 인권을 보호하기 위한 제도를 마련해야 한다. 이러한 인권 보호 제도가 마련되기 전까지는 원격 생체인식의 사용을 유예해야 한다. 유엔 인권최고대표는 국가가 원격 생체인식기술의 정확성과 차별적 영향에 대한 문제를 해결하고 국제 인권 규범 준수를 보장할 수 있을 때까지, 공공장소에서 원격 생체인식기술의 사용을 유예할 것을 각국에 권고하였다[210].

자율살상무기의 경우 생명의 존엄성 및 윤리를 중대하게 훼손할 것이라는 비판을 받아 왔으며, 유엔 사무총장은 치명적인 자율 무기 시스템에 대한 세계적인 금지를 촉구해 왔다.[211] 국가는 자율살상무기에 대하여 인도주의적으로 접근하고 그 생산 및 활용을 금지하는 국제 규범에 대한 논의에 적극적으로 참여하고 준수하여야 한다.

---

208) Council of Europe (2020). C.4.2문
209) 유엔문서 A/HRC/48/31, 26~27문.
210) 같은 문서, 59(d)문.
211) Autonomous weapons that kill must be banned, insists UN chief. UN 보도자료 (2019. 3. 25.) <https://news.un.org/en/story/2019/03/1035381 (검색일: 2021. 11. 1.)>; 유엔문서 A/HRC/48/31. 4문.

제8장 결론 및 정책권고

최근 몇 년간 공공기관과 민간 기업에서 인공지능을 기반으로 하는 신기술의 도입과 사용이 크게 증가하였다. 이미 인공지능은 인터넷 서비스의 추천 알고리즘 등으로 일반 시민의 일상생활은 물론 소상공인의 생계에 큰 영향을 미치고 있으며, 채용, 노동, 금융, 행정, 복지, 치안 등 사회 전반에서 완전히 또는 부분적으로 인공지능에 의해 내려진 조치나 결정이 인간의 기본적 권리와 삶에 직접적인 영향을 미치고 있다. 특히 행정처분 등 법적 또는 그에 준하는 효력을 갖는 공공부문 의사결정에서 인공지능의 관여와 역할이 커질 것으로 예상되고 있다.

2016년 한국 사회에 알파고가 던진 충격 이후 인공지능의 놀라운 발전은 언론, 산업계, 학계, 정부와 국회에서 큰 관심을 받았지만, 이러한 관심 대부분은 대체로 산업적 가치에 초점을 맞추어 왔다. 인공지능의 발전과 확산은 생산성·편의성을 높여 국가경쟁력과 삶의 질을 높일 것으로 기대되지만, 개인정보 및 사생활 침해와 차별 등 기본적 인권을 침해하는 문제들도 제기되고 있다.

최근 유엔은 인권이사회와 인권최고대표를 비롯한 인권 기구들 뿐 아니라 사무총장과 조약 기구 등이 인공지능 등 신기술의 개발과 활용에서 국제인권규범을 준수할 것을 촉구해 왔고, 인권 책무를 실현하는 기준 및 제도 또한 제안해 왔다. 특히 유엔 사무총장은 공공기관 및 민간 기업의 인공지능 개발과 활용에 있어 책임성을 보장하는 법률체계와 절차방법을 마련하여 감독 체제를 수립하고 구제 수단을 구비할 것을 요구하였다. 캐나다 및 유럽연합 등 국가 수준에서도 공공기관 인공지능이나 위험도가 높은 인공지능의 개발과 활용으로부터 시민의 안전과 기본권을 보장하기 위한 법제도를 도입해 왔다.

우리나라에서 인공지능의 문제에 대한 접근은 각 부처의 소관업무에 따라 개별적으로 이루어져 왔으며, 과학기술정보통신부나 방송통신위원회가 마련한 인공지능 윤리기준들은 산업에 미치는 영향을 고려하여 자율적 준수를 강조해 왔다. 그러나 이 경우 이해관계자의 입장에 따라 자의적 해석과 충돌이 발생할 수 있으며, 개발자나 사업자가 편의에 따라 보호 장치를 취사선택한다면 인공지능으로 인한 피해에 대한 구제가 이루어지지 못하거나 부분적으로만 이루어지는 등 그 규범적 효력이 의문시된다.

유엔인권이사회는 신기술이 갖는 영향력, 기회, 도전 과제에 대응하는 인권의 증진 및

보호에 있어 전체적, 포용적, 포괄적 접근방식이 필요하다고 강조하였고, 전세계 인권기구들 또한 국제적으로 인정되고 국내 법률 등에 명시된 인권 규범을 준수하는 인공지능의 개발과 활용을 요구해 왔다.

이에 인공지능의 개발과 활용에 있어 우리사회의 인권적 가치를 훼손하지 않고, 인간의 존엄성과 기본권 보장을 위해 다음과 같은 인권 가이드라인(안)을 제시한다. 이 가이드라인(안)은 인공지능 윤리 등 관련 기준 및 관련 법률에 적용가능한 구체적인 인권 기준으로 고안되었다.

인공지능 개발과 활용에서의 인권 가이드라인(안)

제1장 의의

제1절 제정 배경

1. 사회 전반에 걸쳐 인공지능(AI: Artificial Intelligence)의 활용이 증가함에 따라 고용, 금융, 행정, 복지 등 거의 모든 분야에서 인간의 기본적인 삶과 인권에 영향을 미치는 사례가 증가하고 있습니다.
2. 앞으로도 막대한 양의 빅데이터를 분석하고 학습하는 과정을 통해 다양한 영역에서 사람의 판단을 대신할 수 있는 인공지능은 점차 적용영역을 넓혀 사회 전반과 개인의 삶에 강력한 영향력과 파급력을 행사할 것입니다.
3. 인공지능의 발전과 확산은 국가경쟁력과 개인 삶의 질을 높일 것으로 기대되지만 개인정보 및 사생활 침해, 차별 등과 같은 인권을 침해하는 문제들도 대두되고 있습니다.
4. 반면 인공지능으로 영향을 받는 당사자들은 인공지능의 도입, 운영, 결정에 대하여 참여의 기회를 보장받고 있지 못하며, 인공지능으로 인한 인권침해가 발생한 경우에도 적절하고 효과적인 권리구제를 받을 수 있는 절차와 방법이 미흡한 상황입니다.

5. 따라서 인공지능을 개인의 삶과 사회적 공익에 기여할 수 있도록 설계하며, 인간의 존엄성과 차별금지, 자기결정권 보장 등 기본적 인권에 기반을 두도록 하는 것이 매우 중요합니다.
6. 본 가이드라인은 인권적 관점에서 인공지능의 개발과 활용의 전 과정에서 인권침해를 예방하기 위하여 준수해야 할 기본 원칙과 주요 내용을 제시하고자 마련되었습니다.

제2절 목적 및 의미

7. 국가인권위원회는「국가인권위원회법」제19조 제6호에 따라 인권침해의 유형, 판단기준 및 그 예방조치 등에 관한 지침을 제시할 권한을 가지고 있습니다.
8. 이에 국가인권위원회는 인공지능으로 인한 인권침해와 차별을 판단하고, 개선 등의 권고나 구제절차를 마련하는 데 필요한 기준을 제시하고자 합니다.
9. 본 가이드라인은 인공지능의 개발과 활용에 있어서 우리 사회의 인권적 가치가 훼손되지 않고 인간의 존엄성을 보장하며 기본적 인권을 실현하는 방향으로 나아가도록 하는 데 목적이 있습니다.
10. 또한 인공지능의 개발과 활용 과정에 적용할 인권원칙을 제시하고, 인공지능 서비스 이용자, 영향을 받는 당사자 등에게 주어진 권리 및 피해구제수단을 제공하며, 정부에게 적절한 법령과 정책을 수립할 수 있는 가이드라인을 마련하고자 합니다.

제2장 적용범위 및 정의

제1절 적용 범위

11. 인공지능의 개발과 활용을 위한 가이드라인에서 의미하는 인권은「대한민국 헌법」및 법률에서 보장하거나 대한민국이 가입·비준한 국제인권조약 및 국제관습법에서 인정하는 인간으로서의 존엄과 가치 및 자유와 권리를 의미합니다.
12. 본 가이드라인은 인공지능의 개발부터 활용에 이르는 모든 과정에 참여하는 사회

구성원들을 대상으로 하며, 여기에는 인공지능 개발자, 이용자, 영향을 받는 당사자, 정부 및 공공기관, 기업 등이 포함됩니다.

13. 인공지능 개발과 활용에 관한 법률을 제정하거나 개정할 때에는 본 가이드라인의 목적과 기본 원칙 및 주요 내용을 참고할 수 있습니다.

제2절 정의

14. 본 가이드라인에서 말하는 인공지능은 일차적으로는 학습과 추론, 판단을 전자적으로 구현하는 알고리즘과 해당 프로세스를 지칭하나, 이차적으로는 빅데이터 등 인공지능을 기능하게 하는 일련의 기술들을 포함하고 있습니다.
15. 본 가이드라인에서 기본 원칙은 인공지능의 개발과 활용에 있어서 기본이 되는 인권적 가치를 도출한 것으로 본 가이드라인이 추구하는 방향을 제시한 원칙입니다.
16. 본 가이드라인에서 주요 내용은 기본 원칙을 바탕으로 구체적으로 적용해야 할 과제들을 제시하였으며, 향후 입법 및 제도화 과정에서 주요 내용을 보다 구체화할 필요가 있습니다.

제3장 기본 원칙

제1절 인간의 존엄성

17. 「헌법」 제10조에서 보장하고 있는 인간의 존엄과 가치는 누구나 누려야 할 불가침의 기본적 인권으로, 모든 권리의 출발점인 동시에 종국적으로 보장되어야 할 인권적 가치입니다.
18. 어떠한 활동도 인간의 존엄에서 유래하는 다양한 권리들의 희생을 강요해서는 안 되며, 궁극적으로 모든 활동은 인간의 존엄과 가치를 증대시키는 방향으로 수행되어야 합니다.
19. 따라서 인공지능은 인간으로서의 존엄과 가치 및 행복을 추구할 권리에 부합하는 방식으로 개발 및 활용되어야 하며, 개인의 선택과 판단 및 행동을 강요하거나 자

율성을 침해해서는 안 됩니다.

제2절 알 권리

20. 알 권리란 개인의 의사형성에 필요한 정보를 수집하고, 수집된 정보를 취사·선택할 수 있는 자유를 의미합니다.
21. 개인은 알 권리를 통해 충분한 정보를 획득할 수 있고, 지식과 이해의 폭을 넓힐 수 있으며, 이를 바탕으로 합리적인 판단과 자신의 인격을 발현시킬 수 있습니다.
22. 개인의 의사 판단과 인격 발현의 중요한 요소인 알 권리의 보장을 위해 인공지능의 판단 과정 및 결과에 대한 합리적인 설명 등을 보장하도록 하고, 이를 뒷받침하는 기술적·제도적 장치를 마련해야 합니다.

제3절 자기결정권

23. 「헌법」 제10조가 규정하고 있는 인간의 존엄과 행복추구권에서 도출되는 자기결정권은 결정의 주체인 개인이 중대한 사안에 대해 외부의 강요 없이 누구나 스스로의 판단에 따라 타인의 간섭 없이 결정하고 행동할 수 있으며, 동시에 선택하지 않을 자유를 의미합니다.
24. 인공지능은 인간을 보조하여 서비스를 제공하는 것뿐만 아니라 주어진 데이터를 바탕으로 스스로 학습하여 의사결정을 내릴 수 있는 수준에 이르렀습니다.
25. 따라서 자기결정권은 인공지능의 개발과 활용의 전 과정에서 우선적으로 보장되어야 합니다.

제4절 평등과 차별 금지

26. 「헌법」 제11조는 모든 국민에게 평등권을 보장하고 있으며, 「국가인권위원회법」 제2조는 합리적인 이유 없이 성별, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향 등 개인 특성에 따라 특정한 사람을 우대·배제·구별하거나 불리하게 대우

하는 행위를 평등권 침해의 차별행위로 정의하고 있습니다.

27. 인공지능의 개발과 활용은 개인의 행복과 사회적 공공성의 증진에 위배되어서는 안 되고, 인공지능을 통한 경제·사회·문화적 권리의 향유에 있어서 다양한 계층을 포용하고 참여의 기회를 동등하게 보장해야 합니다.
28. 또한 인공지능의 결정이 특정 집단이나 일부 계층에게 차별적이거나 부정적 영향을 초래하지 않기 위해 개발 단계부터 다양한 계층의 의견을 수렴하고, 차별적 결과가 발생하지 않도록 필요한 조치를 취해야 합니다.

제4장 주요 내용

제1절 투명성과 설명 의무

29. 인공지능 및 이를 이용한 의사결정이 개인에게 미치는 영향력과 중요성이 갈수록 증가하고 있으므로, 인공지능의 판단과정과 그 결과에 대한 적절하고 합리적인 설명이 보장되어야 합니다. 학습 및 추론, 판단의 과정과 결과에 이른 이유를 설명하기 어려운 인공지능은 이에 대한 대응의 불확실성과 영향을 받는 당사자의 불안감을 유발하고, 인권 및 안전에 관한 법령과 정책의 집행효과를 불분명하게 할 수 있습니다.
30. 인공지능과 상호작용하는 사람에게는 언제나 그 상대방이 인공지능이라는 사실을 알려야 합니다.
31. 공공기관은 인공지능의 개발과 활용 계획 등을 사전 공개하여야 하고, 관련 당사자들의 의견을 공청회 등으로 수렴하여야 합니다. 공공기관이 인공지능을 통한 의사결정을 할 때 설명할 수 없는 인공지능을 활용해서는 안 되며, 특히 조달의 경우 입찰 단계에서부터 설명가능성이 보장되어야 합니다.
32. 공공기관이 개발하고 활용하는 모든 인공지능과 민간이 개발하고 활용하는 인공지능 중 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능(이하 ‘고위험 인공지능’이라 한다.)은 사용된 데이터와 인공지능 알고리즘의 주요 요인을 일반에게 공개하고 설명하여야 합니다.

33. 또한, 인공지능에 의한 자동화된 의사결정이 예정되어 있는 경우, 영향을 받는 당사자들은 사전에 그 사실을 알아야 합니다. 자동화된 의사결정에 의하여 영향을 받는 당사자는 그 결정의 이유에 대하여 설명을 듣고, 당사자 진술을 할 수 있으며, 이의를 제기할 수 있어야 합니다.
34. 특히, 완전히 자동화된 의사결정으로만 개인에게 법적 효력 또는 생명·신체·정신·재산에 중대한 영향을 미치는 일은 제한되어야 하고, 이러한 의사결정이 이루어진 경우에는 당사자가 해당 방식을 거부하거나 인적 개입을 요구할 수 있는 권리를 보장받아야 합니다.

제2절 사생활의 비밀 및 개인정보자기결정권 보장

35. 인공지능과 관련하여 정보주체의 권리는 처리된 개인정보에 관하여 고지를 받을 권리, 개인정보 접근 및 열람권, 개인정보처리 동의권 및 정정·삭제권, 처리정지권 등을 포함하며, 정보주체는 자신의 데이터가 사용되는 방법을 이해하고 그에 대한 통제권을 가지는 것이 중요합니다. 정보주체는 인공지능 서비스가 언제, 어디서 자신의 데이터를 수집하고, 어떻게 데이터를 처리하여 사용, 보관, 삭제되는지에 대해 알고 참여할 권리가 있습니다.
36. 인공지능의 개발과 활용에서 개인정보는 목적에 필요한 범위에서 최소한의 개인정보만을 처리하여야 하며, 처리목적 달성에 필요한 기간 동안만 보관되어야 합니다. 또한 이러한 개인정보 처리 원칙은 정보주체가 확인할 수 있도록 공개되어야 합니다.
37. 개인정보자기결정권은 정보주체의 자기 정보에 대한 통제력을 보장하기 위해 인정되는 것인데, 그 통제력 보장의 핵심은 정보주체의 동의권입니다. 따라서 개인정보 처리에 대한 정보주체의 동의는 단순한 외형적 의사 표시만이 아니라 정보주체가 개인정보 처리에 대한 제반 상황을 설명·제공받고 스스로의 자유의사에 기하여 결정할 수 있어야 합니다.
38. 인공지능의 개발과 활용에서 민감정보를 처리할 때에는 특별한 주의를 기울여 보호하여야 합니다. 더불어 의사결정의 내용과 관련성이 없거나, 부정확한 데이터에

기반한 의사결정이 이루어지지 않도록 데이터의 정확성, 완전성, 최신성을 보장해야 합니다.

제3절 차별 금지

39. 인공지능을 개발하고 활용할 때는 인공지능으로 인해 영향받는 사람의 다양성과 대표성을 반영하기 위해 노력해야 하고, 성별, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향 등 개인과 집단의 특성에 따라 편향적이고 차별적인 결과가 나오지 않도록 해야 합니다.
40. 데이터의 수집·선정 및 시스템 설계, 활용 등 인공지능 개발 전반에 걸쳐 편향이나 차별을 배제해야 하고, 이는 데이터 요소를 검사하고 차별적인 데이터를 조정하는 등의 조치를 포함합니다.
41. 특히 학습용 데이터가 인공지능의 판단에 직접적인 영향을 미치는 상황을 고려할 때, 학습용 데이터의 수집 단계부터 차별적 요소를 통제하고 데이터 편향성을 최소화하여 인공지능을 통한 의사결정이 특정 집단에 부정적 영향을 미치지 않도록 해야 합니다.
42. 개발한 인공지능에 대해 주기적인 모니터링을 거쳐 데이터 품질과 위험을 관리하고, 차별적 결과나 의도치 않은 결과에 대해 개선의 조치를 주기적으로 수행해야 합니다.
43. 인공지능 기술 및 서비스에 대한 접근성과 인공지능이 주는 혜택은 사회적 약자와 취약계층을 포함하여 모든 사회구성원에게 평등하게 제공되어야 합니다.

제4절 인공지능 인권영향평가 시행

44. 국가는 인공지능의 개발과 활용에 있어 인권적 가치가 우선시 되도록 하여야 하며, 인공지능으로 인해 발생할 수 있는 인권침해와 차별에 대하여 사전적 또는 사후적으로 관리 감독을 할 의무가 있습니다.
45. 국가는 인공지능의 개발과 활용에 있어서 인권 침해와 차별의 가능성 및 정도, 영

향을 받은 당사자의 수, 사용된 데이터의 양 등을 고려하여 인권영향평가를 실시해야 합니다. 특히 인공지능 기술이 적용되어 기존 제도로 관리되거나 감독될 수 없는 새로운 분야는 인권영향평가 제도를 도입해야 합니다.

46. 인권영향평가 내용에는 인공지능의 특성, 상황, 범위 및 목적을 감안하여 본 인권 가이드라인이 제시한 기본 원칙 및 주요 내용, 국제 인권 기준, 관련 법률에서 정한 의무 등이 포함되어야 하며, 인권 침해 위험요인의 분석, 개선 사항 등을 도출해야 합니다.
47. 인권영향평가는 개발 및 출시 전에 실시하고 인공지능의 기능 또는 범위 변경 시 평가를 갱신하여야 합니다.
48. 인권영향평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하기 위한 조치사항을 수립하여 적용하여야 하며, 원칙적으로 그 내용이 공개되어야 합니다. 또한, 이를 방지하거나 완화하는 조치를 취하기 전에는 그 개발과 활용을 중단해야 합니다.
49. 국가는 인권영향평가를 인권전문성과 독립성을 확보한 기관이 담당하도록 하고, 해당 기관은 인권영향평가의 활성화를 위하여 관계 전문가의 육성, 영향평가 기준의 개발 및 보급 등 필요한 조치를 마련해야 합니다.

제5절 위험도 등급 및 관련 법제도 마련

50. 국가는 공공기관과 민간이 개발하고 활용하는 인공지능에서 인권과 책임성을 보장하도록 관련 법률과 감독 체계를 보완해야 하며, 특히 인권을 보장하기 위한 구체적인 지침과 정책을 마련하여야 합니다.
51. 국가는 개인의 인권과 안전에 미치는 위험성이 매우 높아 인공지능이 금지되는 영역, 상당한 제한이 필요한 인공지능 고위험 영역, 위험성이 거의 없는 영역 등 적절하게 위험성 단계를 구분하고, 그에 맞는 규제 수준과 인적 개입이 이루어지도록 법과 제도를 마련하여야 합니다.
52. 특히, 당사자의 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능은 투명성과 설명가능성, 개인 정보 보호, 차별 금지 등 그 규제에 있어서

공공과 민간의 구분 없이 엄격하게 적용되어야 합니다.

53. 감독 기관은 공공기관과 민간의 위법한 인공지능 개발과 활용 여부를 조사하고 피해 구제 및 조치를 취하기 위하여 상세 정보에 접근할 수 있어야 합니다. 이를 위하여 공공기관 인공지능 및 민간 고위험 인공지능 개발자 및 운영자는 사용된 데이터와 알고리즘의 주요 요소 등을 기록하고 문서화하여 일정 기간 보관하여야 합니다.

54. 국가는 인공지능을 독립적이고 효과적으로 감독할 수 있는 체계를 수립하여 개인의 인권과 안전을 보장하고 피해를 구제하여야 합니다. 인공지능 국가 감독 체계는 독립적이고 효과적이어야 하며, 진정 또는 인지로 접수한 사건을 조사하기에 충분한 자원, 권한 및 전문지식을 구비해야 합니다.

55. 국가는 인공지능으로 인하여 인권을 침해당하거나 차별을 받은 사람이 진정을 접수하여 권리를 구제받을 수 있는 기회를 보장하는 등 국가기관의 구제수단에 대한 접근을 보장해야 합니다. 인공지능을 개발하고 활용하는 공공기관과 민간은 언제든지 구제가 가능하도록 그 책임자에 대한 정보는 물론, 이의를 제기할 수 있는 기관과 방법에 대한 정보를 일반에 공개하여야 합니다.

56. 특히 국가는 대량 감시와 차별로 이어질 위험이 높은 얼굴인식 등 원격 생체인식 기술의 사용을 공공장소에서 금지하고, 특별한 경우에 한하여 사용을 허용하되, 인권 침해나 차별의 위험성이 드러난 경우 이를 방지하거나 완화하는 조치를 취하기 전에는 사용을 중단해야 합니다. 또한 국가는 생명의 존엄성 및 윤리를 훼손할 가능성이 높은 자율살상무기에 대하여 인도주의적으로 접근하고 그 연구, 개발, 생산 및 활용을 금지하는 국제 규범을 준수하고, 이에 대한 논의에 적극적으로 참여해야 합니다.

참 고 문 헌

[1] 오요한・홍성욱 (2018). 인공지능 알고리즘은 사람을 차별하는가?. 과학기술학연구 제18권 제3호 153-215.

[2] Australian Human Rights Commission (2021). Human Rights and Technology. <https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf (검색일: 2021. 9. 1.)>.

[3] Council of Europe (2020). Guidelines on addressing the human rights impacts of algorithmic systems. Appendix to Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems. <https://rm.coe.int/09000016809e1154 (검색일: 2021. 9. 1.)>.

[4] Council of Europe Commissioner for Human Rights (2019). 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 (검색일: 2021. 9. 1.)>.

[5] European Commission (2020a). WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust.

[6] European Commission (2020b). White Paper on Data Ethics in Public Procurement of AI-based Services and Solutions.

[7] European Commission (2021). Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN (검색일: 2021. 11. 1.)>.

[8] European Union Agency for Fundamental Rights (2020a), “Strong and Effective national Human Rights Institutions: Challenges, Promising Practices and Opportunities”. 3.4.1.절 참조.

[9] European Union Agency for Fundamental Rights (2020b). Getting the Future Rights : Artificial Intelligence and Fundamental Rights.

[10] Government of Canada (2019). Directive on Automated Decision-Making. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 (검색일: 2021. 11. 1.)>.

[11] High-Level Expert Group on Artificial Intelligence (2019). Ethics guidelines for trustworthy AI. European Commission.

[12] UK government (2020). Guidelines for AI procurement. <https://www.gov.uk/government/publications/guidelines-for-ai-procurement (검색일: 2021. 9. 1.)>

[13] 유엔문서 A/73/348 (2018. 8. 29). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.

[14] 유엔문서 A/HRC/43/29 (2020. 3. 4). Question of the realization of economic, social and cultural rights in all countries: the role of new technologies for the realization of economic, social and cultural rights.

[15] 유엔문서 A/HRC/RES/47/23 (2021. 7. 16). Resolution adopted by the Human Rights Council on 13 July 2021, 47. 23. New and emerging digital technologies and human rights.

[16] 유엔문서 A/HRC/48/31 (2021. 9. 13). The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights.

부 록

Ⅰ. 유엔 인권최고대표 <최종 사용에서 인권 위험 식별 및 평가>

Ⅱ. 유엔 인권최고대표 <디지털 시대 프라이버시권 (2021)>

Ⅲ. 유럽평의회 인권위원장 <인공지능 블랙박스 개봉: 인권 보호를 위한 10단계>

Ⅳ. 유럽평의회 <알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고>

Ⅴ. 캐나다 정부 <알고리즘 영향평가 도구>

Ⅵ. 신기술 관련 유엔 인권이사회 특별절차 보고서 목록

부록 I.

유엔 인권최고대표
<최종 사용에서 인권 위험 식별 및 평가>[1]

- B-테크 기초 자료[2]

개요

유엔 <기업과 인권 이행 지침>은 기술 회사들이 자신이 제공하는 제품과 서비스와 관련된 사람들에 미치는 위험을 해결하고자 할 때 주의를 집중할 만한 강력한 접근법을 제시한다. 이는 특히 거의 모든 장소에서 여러 민간, 공공 또는 개인 사용자 다수가 대규모로 사용하는 제품 및 서비스에 대해 정기적으로 검토하고 결정을 내려야 하는 기업들에게 유용하다.

인권 위험을 식별하고 평가하는 시점(인권 실사의 첫 단계)에 <기업과 인권 이행 지침>이 기업들에 요구하는 바는 다음과 같다.

발생할 수 있는 영향에 대하여 폭넓은 관점의 유지 : 이는 기업의 사업 활동 및 관계 전반에서 관련된 인권 모두에 미치는 위험을 식별하는 것을 의미한다. 여기에는 제품 및 서비스의 설계, 개발, 판촉, 판매/라이선스, 계약 및 사용이 인권에 부정적인 영향을 미칠 수 있는지 여부 및 그 방법에 대한 식별이 포함된다.

가장 심각한 해악에 초점 : 이는 사업 수행이 사람들에게 가장 심각하거나 광범위하거나 지속적인 피해를 야기하거나 할 수 있는 지점에 우선순위 초점을 맞춰야 함을 의

---

1) Office of the High Commissioner for Human Rights (2020). Identifying and Assessing Human Rights Risks related to End-Use. https://www.ohchr.org/Documents/Issues/Business/B-Tech/identifying-human-rights-risks.pdf.

2) B-테크(B-Tech) 프로젝트는 기술 분야에서 <기업과 인권 이행지침>을 이행하기 위한 지침과 자료를 제공하는 프로젝트이다. https://www.ohchr.org/EN/Issues/Business/Pages/B-TechProject.aspx 참조.

미한다.

이해관계자의 참여와 의미 있는 의사소통 : 이는 기업이 관련 외부 이해관계자와 관계를 맺고 그 인권 위험 평가 및 우선순위를 처음에 고지하고 나중에 설명해야 함을 의미한다.

이 자료에 대한 소개

이 자료는 제품과 서비스에서 인권 위험을 식별하고 평가할 때 <기업과 인권 이행지침>의 기본적인 요구를 이해하고자 하는 기술 기업 경영진을 위하여 주로 작성되었다. 이 자료는 <기업과 인권 이행지침>이 기술 기업 및 국가에 갖는 의미를 재서술하고 설명하며 명확히 하는 유엔 인권 B-테크 프로젝트 기초 자료 시리즈의 일부이다.

이 자료는 인권 실사의 모든 측면에 대한 세부 사항을 제공하지 않으며, 유엔 <기업과 인권 이행 지침> 해석가이드 또는 OECD 책임 경영 실사 지침 등 기존 산업 간 지침을 대체하거나 반복하려는 것이 아니다.

이 자료 시리즈는 정책 입안자/규제자, 시민 사회, 투자자 및 기업을 비롯하여 기술 경영에서 인권 존중을 구현하기 위해 노력하는 모든 이해관계자에게 공통된 출발점을 제시하는 것을 목표로 한다. 이 자료들은 또한 B-테크 프로젝트의 핵심 영역 전반에 걸쳐 프로젝트 활동, 지침 및 권장 사항의 기반이 되는 공통의 이해 체계를 수립한다. 이는 프로젝트 작업의 끝이 아니라 시작점이다.

기초 자료 시리즈의 링크는 다음과 같다.

B-체크 프로젝트 포털

https://www.ohchr.org/EN/Issues/Business/Pages/B-TechProject.aspx

유엔 <기업과 인권 이행 지침>

https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

인권 실사 절차

1. 인권 위험의 성격과 정도를 측정하기 위하여 영향을 식별하고 평가함
2. 내부 기능 및 절차에 대한 반영을 비롯하여 사람에 대한 위험을 방지하고 완화하기 위하여 조치함
3. 위험 완화 대응의 시간 경과에 따른 효과성을 추적함
4. 인권 영향을 해결하기 위한 활동에 대하여 적절하게 소통함

주요 사항

1. 인권 위험을 식별하고 평가하는 분석 범위에는 자사 제품과 서비스의 설계, 개발, 판촉, 배포, 판매/라이선스 및 사용이 인권에 부정적인 영향을 미칠 수 있는지 여부 및 그 방법을 식별하는 활동이 포함된다.

2. 기술 기업이 대량의 제품, 서비스 및 사용자를 보유하고 있는 경우, 인권 관점에서 어떤 제품, 서비스, 솔루션, 사업 관계 또는 사용 환경이 더 높은 위험인지 알기 위해 첫 평가를 실시해야 할 수 있으며, 따라서 보다 상세한 인권 실사를 위해 우선순위를 정해야 한다.

3. 기업이 보다 심층적인 인권 실사를 위해 특정 제품/서비스, 사용자 유형 또는 사용 환경에 우선순위를 부여한 경우, 기업은 이와 관련하여 인권에 미칠 수 있는 실제적 또는 잠재적 부정적인 영향에 대한 분석을 실시하여야 한다.

4. 기업이 초점을 맞출 지점의 우선순위를 부여해야 할 때마다, 기업은 사람에 대한 위험의 '심각성'에 초점을 맞추는 원칙적인 접근법을 사용해야 한다. <기업과 인권 이행지침>에 따르면 영향의 심각도는 다음과 같은 규모, 범위 및 구제불가능성에 의해 판단된다.

규모는 그 영향이 얼마나 중대하거나 심각한지와 관련이 있다.

범위는 영향을 얼마나 광범위하게 미치는지 또는 영향을 받는 사람의 수와 관련이 있다.

구제가능성은 상황에 영향을 받은 사람들을 적어도 영향 이전 상황과 같거나 동등한 수준으로 회복시킬 수 있는 능력을 의미한다.

첫째, 인권 위험을 식별하고 평가하는 분석 범위에는 기업이 자사 제품과 서비스의 설계, 개발, 판촉, 배포 및 사용이 인권에 부정적인 영향을 미칠 수 있는지 여부 및 그 방법을 식별하는 것이 포함된다.

인권 위험을 식별하고 평가하는 것은 인권 실사 절차의 첫 번째 단계이다. <기업과 인권 이행지침>에 따르면 "인권 위험을 측정하기 위해 기업은 사업 관계의 결과로 또는 기업의 활동으로 인해 인권에 미칠 수 있는 실제적 또는 잠재적 부정적인 영향을 식별

하고 평가해야 한다." (<기업과 인권 이행지침> 18문).

실제적 또는 잠재적 인권 영향에 대한 평가는 기술 기업이 야기할 수 있는 모든 영향을 포괄하여야 하며, 이는 자체 활동을 통해 유발되는 영향 뿐 아니라, 그 영향의 원인이 되지 않은 경우라 하더라도 사업 관계에 의해 업무, 제품 및 서비스에 직접 연결될 수 있는 영향을 포함해야 한다. 이 상황에서 기술 기업의 '자체 활동'에는 제품, 서비스 및 솔루션의 설계, 개발, 마케팅, 판매/라이센스 및 배포가 포함된다.

평가는 국제적으로 인정된 모든 인권과 관련하여 이루어져야 한다. "기업은 사실상 국제적으로 인정된 인권 전체 범위에 영향을 미칠 수 있기 때문에 기업의 인권 존중 책임은 모든 인권에 적용된다. 실제로, 특정 산업이나 환경에서는 일부 인권이 다른 인권보다 더 큰 침해의 위험에 처해질 수 있기 때문에 더 큰 관심을 기울여야 한다. 그러나 상황은 변할 수도 있으므로, 모든 인권은 정기적 검토의 대상이 되어야 한다." (<기업과 인권 이행지침> 주석12).

기술 분야의 많은 사람들은 기업이 개인정보보호 및 표현의 자유에 미치는 영향에 대해 의문을 제기할 것이다. 그러나 기술의 사용과 오용이 온라인과 오프라인에서 다양한 인권에 영향을 미칠 수 있다는 증거가 이미 있다. 예를 들어, 법집행 기관과 형사 사법 시스템에서 인공지능 도구를 사용하면 자의적 체포에 대한 개인의 자유 또는 법 앞의 평등에 대한 권리에 영향을 미칠 수 있다. 감시 기술은 평화적 집회의 권리에 영향을 미칠 수 있다. 소셜 미디어 플랫폼의 사용은 정신 건강에 대한 권리에 영향을 미칠 수 있다. 부동산 임대 플랫폼은 주택 시장을 변화시켜 적절한 생활 수준에 대한 권리에 영향을 미칠 수 있다.

또한 <기업과 인권 이행지침> 주석18에 따르면 "기업은 소외되고 취약한 높은 위험에 처해 있는 집단이나 인구집단에 속한 개인에게 미치는 특정한 인권 영향에 특별한 주의를 기울여야 한다. 여기에는 아동, 소수 민족, LGBTI 커뮤니티 구성원 및 인권 활동가가 포함된다. (유엔 기업과 인권 실무단에서 젠더 차원 보고서에서 자세히 설명한 대로) <기업과 인권 이행지침>은 젠더 기반 위험 및 영향을 염두에 둘 필요성이 있음을 강조한다.

외부 자문단 또는 기타 외부 이해관계자 및 예정 사용자 참여 방식의 수립은 기업이 제품 및 서비스의 실제적 또는 잠재적 위험을 식별하고 이해하고자 할 때 또한 도움이

될 수 있으며 지속적이고 누적적인 참여를 가능케 할 수 있다.

둘째, 기술 기업이 대량의 제품, 서비스 및 사용자를 보유하고 있는 경우, 인권 관점에서 어떤 제품, 서비스, 솔루션, 사업 관계 또는 사용 환경이 더 높은 위험인지 알기 위해 첫 평가를 실시해야 할 수 있으며, 따라서 보다 상세한 인권 실사를 위해 우선순위를 정해야 한다.

기술 기업의 제품, 최종 사용자, 사용 사례 및 사용 환경(예: 대륙적, 국가적 또는 지역적)의 수와 다양성은 모두 기술 기업의 사업 및 관계와 관련된 잠재적 인권 위험의 심각성과 위험 상황의 복잡성에 영향을 미칠 것이다. 제품, 서비스와 솔루션 및 최종 사용 시나리오의 포트폴리오가 복잡할수록, 그 시스템은 모든 관련 위험을 식별하고 해결할 수 있도록 더 포괄적이고 정교해질 필요가 있을 것이다. 결과적으로 하나의 제품이나 솔루션을 가지고 있거나 소수의 타겟 고객 또는 최종 사용자를 가진 일부 소규모 기업 및 스타트업 기업에게는 관련 인권 위험을 식별하고 평가하는 것이 상당히 복잡하지 않을 수 있다.

예를 들면,

제품, 서비스, 솔루션 및 최종 사용자 유형(기존 및 잠재 고객)에 있어 크고 복잡한 포트폴리오를 보유한 기술 기업의 경우, 인권 위험이 가장 중대할 수 있는 일반 영역을 식별하고 인권 실사를 위해 그 우선순위를 정하는 것을 목표로 회사의 인권 정책과 시스템을 개발해야 한다. 이는 모든 제품과 서비스, 관련 최종 사용자 및 최종 사용 시나리오에서 최소 수준의 인권 위험에 대하여 식별하고, 인권 위험이 높은 시나리오에 대한 보다 상세한 분석을 우선시하는 것이 필요할 수 있다. 이를 위해서는 회사의 전략 및 제품/서비스 개발 활동을 파악할 수 있는 상급 또는 다기능 거버넌스 기구가 이 검토를 수행해야 할 수 있다. 그러나 직원이 자체 연구 및 제품 아이디어를 추진할 수 있고 한정된 감독 하에 혁신에 도달할 수 있는 경우, 일반적인 인권 위험을 식별하는 이러한 예비 절차는 모종의 상향식 절차(예: 신속한 인권 감각 검사 도구 또는 정기적인 부서 워크숍)에 의해 뒷받침되어야 할 수 있다.

다양한 운영 환경 및 규모에서 사용되는 제품, 서비스 및 솔루션의 수가 적거나 비교적 단순한 포트폴리오를 보유한 기술 기업의 경우, 해당 회사의 인권 정책과 시스템 또한 모든 관련 현장 상황에서 회사의 기술이 사용되는 데 대한 일반적인 위험 영역

을 식별하려는 목표로 개발되어야 한다. 이를 위해서는 회사의 전략, 신규 시장 확대 및 사업 개발 업무를 파악할 수 있는 상급 또는 다기능 거버넌스 기구가 이 검토를 수행해야 할 수 있다.

일부 기술 기업의 경우, 제품/서비스/솔루션 및 최종 사용자 유형에 대한 크고 복잡한 포트폴리오 뿐 아니라 다양한 운영 환경에서 대규모로 사용되는 경우를 모두 탐색할 필요가 있을 것이다.

그러나 어떤 경우에는 특히 심각한 위험이 분명히 존재하는 경우, 전체적인 분석을 먼저 수행하지 않고 명백한 고위험 영역부터 시작하는 것이 필요하고 현명할 수 있다. 어느 쪽이든, 기업은 최소 수준의 평가가 지속적이고 동적일 수 있도록 보장해야 하며, 이는 시간이 지남에 따라 회사의 실사 초점을 변경하고 넓히는 결과로 이어질 수 있다.

셋째, 기업이 보다 심층적인 인권 실사를 위해 특정 제품/서비스, 사용자 유형 또는 사용 환경에 우선순위를 부여한 경우, 기업은 이와 관련하여 인권에 미칠 수 있는 실제적 또는 잠재적 부정적인 영향에 대한 분석을 실시하여야 한다.

대부분의 경우 보다 세분화된 수준에서 실제적 및 잠재적 인권 영향을 식별하는 활동은 몇 단계 계층적 분석이 필요하며 분석 시작점에 따라 다양할 수 있다.

고급 수준에서는:

분석의 시작점이 특정 제품 또는 서비스인 경우, 회사는 예상 사용자의 영역, 해당 사용자의 다양한 사용 사례(의도되지 않았거나 의도된 오용 포함) 및 해당 이용 사례와 관련된 실제적 또는 잠재적 인권 영향을 검토할 수 있는 체계적인 방법을 보유해야 할 필요가 있을 것이다.

분석의 시작점이 특정 사용자 또는 사용자 범주인 경우(예: 민간 부문 잠재 고객, 타겟 산업, 특정 국가 국가 기관 또는 기술 부문의 동료) 기업은 해당 사용자에 대해 가능한 사용 사례(의도되지 않았거나 의도된 오용 포함) 및 해당 이용 사례와 관련된 실제적 또는 잠재적 인권 영향을 매핑할 때와 동일한 논리를 따라야 할 필요가 있을 것이다. 여기에는 '고객 실사 파악' (Know Your Customer due diligence)과 같은 기존 절차 위에 실시되는 것도 포함될 수 있다. 이러한 기존 절차를 사용하는 경우, 기업은 인권 위험을 처리하는 데 있어 고객의 실적 또는 고객의 영업 모델이 인권에 내재된 위험을 수반할 수 있는지 여부 등 사람에 대한 위험의 관점에서 관련 고객의 특성에 초점

을 맞추도록 주의해야 한다. 또한 기술 기업과 고객이 공동으로 인권 실사를 실시하여 인권에 미치는 부정적 영향을 보다 용이하게 방지하고 완화하기 위하여 협력하는 것도 가능할 수 있다.

분석의 시작점이 그 사용이 부정적인 영향을 미칠 수 있는 지리적 지역, 국가 또는 지역적 환경에 대해서라면, 기업은 사용자 영역, 잠재적 사용 사례 및 관련 영향에 대한 기존의 이해를 심화시키고자 할 가능성이 높다. 그러나 회사는 한편 지역적 사회경제적, 정치적, 인권적 현실이 인권 해악을 어떻게 악화시키거나 보호하는지 이해하는 데 초점을 맞출 필요가 있을 것이다.

이러한 유형의 분석에서 두 가지 주요 특성에 특히 주의할 가치가 있다. 첫째, 분석을 수행하는 데 있어 유일한 '올바른' 방법은 없다. 회사는 작업을 수행하는 방법과 회사 내에서 분석을 주도하거나 참여하는 사람에 대해 선택해야 한다. 그리고 회사는 이해관계자에게 절차를 설명할 준비가 되어 있어야 한다.

둘째, 〈기업과 인권 이행지침〉은 회사의 분석이 100% 완전하고 절대적일 것을 요구하지 않는다. 그러한 분석들에는 불확실성이 있고 주관적인 판단이 있다. 관련 내부 및 외부 이해관계자를 관여시켜 오판과 사각 지대를 최소화해야 한다. 지침이 요구하는 것은 회사가 접근가능한 사실, 지적인 예측 및 건전한 판단에 기반하여 분석을 수행하기 위한 모든 합리적인 조치를 취해야 한다는 것이다.

넷째, 기업이 초점을 맞출 지점의 우선순위를 부여해야 할 때마다, 기업은 사람에 대한 위험의 '심각성'에 초점을 맞추는 원칙적인 접근법을 사용해야 한다.

기술 기업이 특정 기술, 사용자 또는 사용 환경에 초점을 맞춘 경우에도 인권 위험 평가는 때때로 더 많은 수의 사용자, 사용 사례 및 인권 위험을 식별하는 결과를 낳는다. 이로 인해 회사는 관심을 집중할 부분의 우선순위를 정해야 할 수 있다. 그렇다면 회사는 먼저 인권 위험의 심각성을 기반으로 우선순위를 모색하고 부정적인 영향이 발생할 가능성에도 주의를 기울여야 한다. 그런 다음 계속해서 다른 영역을 다루어야 한다.

〈기업과 인권 이행지침〉이 언급한 것처럼 "실제적, 잠재적으로 인권에 미치는 부정적 영향을 다루는 활동에서 우선순위를 정해야 할 때, 기업은 가장 심각하거나, 대응이 지체되었을 때 구제가 불가능 할 수도 있는 영향을 제일 먼저 방지하고 완화시키도록 노

력해야 한다." (<기업과 인권 이행지침> 24문). 주석24는 "영향의 심각성은 규모, 범위 및 구제불가능성에 따라 판단될 것"이라고 설명한다.

규모는 그 영향이 얼마나 중대하거나 심각한지와 관련이 있다.

범위는 영향을 얼마나 광범위하게 미치는지 또는 영향을 받는 사람의 수와 관련이 있다.

구제가능성은 상황에 영향을 받은 사람들을 적어도 영향 이전 상황과 같거나 동등한 수준으로 회복시킬 수 있는 능력을 의미한다.

취약하거나 소외될 위험이 높은 개인이나 집단에 따라 영향의 규모, 범위 및 구제가능성이 어떻게 다를 수 있는지, 그리고 남성과 여성 등 서로 다른 집단이 직면하는 위험이 다를수 있다는 것을 기업이 고려하는 것이 중요하다.

<기업과 인권 이행지침>에 대한 유엔 인권 해석 지침에 설명된 바와 같이 "운영 환경에 따라 아동, 여성, 원주민 또는 소수 민족 등에 속하는 사람들과 같이 취약하거나 소외될 위험이 높은 집단에 속한 사람들이 가장 심각한 인권 영향에 직면할 수 있다. 기업이 인권 영향에 대한 대응의 우선순위를 정할 필요가 있다고 판단하면 그러한 집단의 취약성과 특정 영향에 대한 지체된 대응이 이들에게 불균형적으로 영향을 미칠 수 있는 위험을 고려해야 한다." 기술 제품 및 서비스의 사용이 특정 집단에 차등적인 영향을 미칠 수 있는지 여부에 대한 이러한 추가적인 고려사항은 영향의 '범위'가 특히 광범위한 경우(예: 수십만, 수백만 또는 수십억) 특별한 유용성을 가질 수 있다.

기업이 가능한 모든 사용자와 사용 사례에 걸쳐 인권 영향의 심각성을 고려하더라도 여전히 주의를 요하는 광범위한 문제의 영역이 남아 있을 수 있다. 기업이 우선순위를 더 부여할 필요가 있는 경우, '가능성'의 필터를 사용해야 한다. 여기서 회사는 다음과 같은 사항을 고려하기를 원할 수 있다.

사용자 이익, 동기 및 인센티브: 제품, 서비스 또는 솔루션을 위험을 초래할 수 있는 방식으로 사용하거나 오용하는 것이 사용자의 이익 범위 안에 있는가?

사용자의 기술적 노하우 및 능력: 사용자의 노하우(또는 노하우 부족)가 확인된 사용 사례 및 부정적인 영향이 발생할 가능성을 변화시키는가? 기존 기술적 장벽(예: 컴퓨팅 성능에 대한 접근)이 사용 사례를 실제 불가능하게 만드는가?

현지 정책 및 법률: 사용 사례가 실제 발생할 가능성을 증가시키거나 낮추는 정부 정책 및 법률이 있는가?

기업의 실사 접근 방식에서 우선순위 결정이 이루어지는 방법과 이 결정이 신뢰를 구축하는 데 중요한 이유를 공개적으로 소통한다.

부록 Ⅱ.

유엔 인권최고대표 <디지털 시대 프라이버시권 (2021)>[3)]

요약

최고대표는 인권이사회가 결의안 42/15에서 위임한 본 보고서에서 프로파일링, 자동화된 의사결정 및 머신러닝 기술을 비롯한 인공지능을 국가와 기업이 광범위하게 사용하는 것이 프라이버시권 및 관련 권리의 향유에 어떻게 영향을 미치는지 분석하였다. 최고대표는 국제법 체계를 개괄한 데 이어 인공지능의 프라이버시권 침해 측면을 강조하면서 4대 주요 분야[역주: 수사/국가안보/형사사법/출입국관리, 공공서비스 제공, 고용, 온라인 정보 관리]에서 프라이버시권 및 관련 권리에 영향을 미친 사례를 소개하였다. 이후 최고대표는 문제를 해결하기 위한 접근방식을 논의하면서 안전장치의 설계 및 구현에 관하여 국가와 기업에 일련의 권고를 제시하였고, 이를 통하여 인공지능이 제공할 수 있는 편익을 최대한 누리면서 유해한 결과물은 방지하고 완화할 수 있도록 하였다.

Ⅰ. 도입

1. 본 보고서는 인권이사회 결의 42/15에 따라 제출되며, 이 결의에서 인권이사회는 적절한 안전장치가 없으면 프로파일링, 자동화된 의사결정 및 머신러닝 기술을 포함한 인공지능이 어떻게 프라이버시권의 향유에 영향을 미칠 수 있는지를 논의하는 전문가 세미나를 개최하고, 이 문제에 대한 주제 보고서를 마련하여 인권이사회 제45차 회의에 제출할 것을 유엔인권최고대표에게 요청한 바 있다.

2. 최근 몇 년간 인공지능(AI), 특히 머신러닝 기술만큼 대중의 상상력을 사로잡은 기

3) The right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights. 유엔문서 A/HRC/48/31.

술 발전이 없었다. 이러한 기술들은 실제로 큰 현대 문제 중 일부를 사회가 극복하도록 도우면서 선을 위한 엄청난 힘이 될 수 있다. 그러나 이러한 기술이 인권에 미치는 영향을 충분히 고려하지 않고 배치될 경우 부정적인, 심지어 치명적인 영향을 미칠 수도 있다.

3. 본 보고서는 코로나19 감염병(COVID-19) 대유행에 초점을 맞추지 않았지만, 현재의 지구적 건강 위기는 세계 각지 여러 생활 영역에서 인공지능의 속도, 규모 및 영향이 강력하다는 사실을 매우 잘 보여주는 사례들을 낳았다. 여러 유형의 데이터(위치정보, 신용카드정보, 교통정보, 건강정보 및 인구통계학적 정보)와 인적 관계에 대한 정보를 이용한 접촉 추적 시스템이 질병의 확산을 추적하는 데 사용되어 왔다. 개인을 감염된 사람이나 감염되었을 수 있는 사람으로 지목하고, 격리나 자가격리를 요구하는데 인공지능 시스템이 사용되어 왔다. 성적의 예측적 할당에 사용된 인공지능 시스템은 공립학교와 가난한 지역의 학생들을 차별하는 결과를 낳았다. 이러한 발전 양태는 인공지능 시스템이 사람들의 일상생활에 광범위한 영향을 미치고 있음을 보여준다. 이러한 경우들에서 무엇보다 프라이버시권이 영향을 받는데, 인공지능이 개인정보를 사용하고 종종 사람들의 삶에 실제적인 영향을 미치는 결정을 내리게 되었기 때문이다. 또한 프라이버시 문제와 깊이 얽혀 있는 건강권, 교육권, 이동의 자유, 평화로운 집회의 자유, 결사의 자유, 표현의 자유 등 다른 권리의 향유도 다양한 영향을 받는다.

4. 2019년 유엔 사무총장은 “최고의 열망: 인권을 위한 행동 촉구”에서 디지털 시대가 인류 복지, 지식 및 탐험의 새로운 경계를 열었다고 인정했다. 그는 디지털 기술이 인권을 옹호하고 방어하며 행사할 수 있는 새로운 수단을 제공한다고 강조했다. 그럼에도 불구하고, 신기술은 예를 들어 인권 활동가 등에 대한 감시, 억압, 검열 및 온라인 괴롭힘을 통해, 특히 이미 취약하거나 뒤쳐진 사람들의 권리를 침해하기 위해 너무 자주 사용된다. 복지제도의 디지털화는 효율성을 향상시킬 수 있는 그 잠재력에도 불구하고 가장 필요로 하는 사람들을 제외시킬 위험이 있다. 사무총장은 신기술의 발전이 인권을 침해하고, 불평등을 심화시키고, 기존의 차별을 악화시키는 데 사용되어서는 안 된다고 강조했다. 그는 인공지능의 거버넌스는 공정성, 책무성, 설명성 및 투명성을 보장할 필요가 있다고 강조했다. 안보 분야에 있어서 사무총장은 치명적인 자율

무기 시스템에 대한 세계적인 금지를 재차 촉구했다.

5. 본 보고서는 디지털 시대의 프라이버시권에 관한 최고대표의 이전 두 보고서를 토대로 작성되었다. 또한 이사회 결의안 42/15에 따라 2020년 5월 27일, 28일에 개최된 온라인 전문가 세미나에서 얻은 통찰과 최고대표에게 본 보고서 발간을 요청한 데 대한 응답이 담겼다.

II. 법적 체계

6. 세계인권선언 제12조, 시민적 및 정치적 권리에 관한 국제규약 제17조 및 기타 여러 국제적 및 지역적 인권기구는 프라이버시권을 기본 인권으로 인정한다. 프라이버시권은 국가와 개인의 힘의 균형에서 중추적인 역할을 하며 민주주의 사회의 기본적 권리이다. 데이터 중심 세계에서 온라인과 오프라인에서 다른 인권을 향유하고 행사하는 것의 중요성도 증가하고 있다.

7. 프라이버시권은 인간의 존엄성의 표현이며 인간의 자율성과 개인 정체성의 보호와 관련이 있다. 인공지능 사용 환경에서 특히 중요한 프라이버시 측면으로 정보 프라이버시가 포함되며, 이러한 정보 프라이버시에는 개인과 개인의 삶에 대해 존재하거나 도출할 수 있는 정보와 해당 정보에 기초한 결정을 포함하며, 개인의 정체성을 결정할 자유도 포함된다.

8. 프라이버시권에 대한 어떠한 간섭도 자의적이거나 불법적이어서는 안 된다. '불법'이란 국가가 법률에 근거를 두고 해당 법률을 준수하는 한도에서만 프라이버시권을 간섭할 수 있음을 의미한다. 법률 자체는 시민적 및 정치적 권리에 관한 국제규약의 조항, 목적 및 목표를 준수해야 하며 그러한 간섭이 허용되는 정확한 상황을 상세히 명시해야 한다. 자의성 개념은 법률에 의해 규정된 간섭조차도 규약의 조항, 목적 및 목표에 따라야 하며, 어떤 경우에도 특정 상황에서 합리적이어야 한다는 것을 보장하기 위해 도입된 것이다. 따라서 프라이버시권에 대한 모든 간섭은 정당한 목적에 부합해야 하며, 정당한 목적을 달성하기 위해 필요하며, 비례적이어야 한다. 어떠한 제한도 최소 침해적인 선택이 가능해야 하며 프라이버시권의 본질을 침해해서는 안 된다.

9. 프라이버시권은 모두에게 적용된다. 인종, 피부색, 성, 언어, 종교, 정치적 또는 기타의 의견, 민족적 또는 사회적 출신, 재산, 출생 또는 기타의 신분에 따라 그 보호에 차이를 두는 것은 시민적 및 정치적 권리에 관한 국제규약 제2조 (1) 및 제3조에 명시된 차별금지 원칙과 일치하지 않는다. 이러한 이유로 차별을 하는 것은 또한 규약 제26조에 담긴 법 앞의 평등권을 침해한다.

10. 시민적 및 정치적 권리에 관한 국제규약 제2조 제1항은 국가가 자국 영토 내의 모든 개인과 그 관할권하에 있는 모든 개인에 대해 차별 없이 규약에서 인정된 권리를 존중하고 보장하도록 요구하고 있다. 즉, 국가는 규약에서 인정된 권리를 침해하는 것을 삼가야 할 뿐만 아니라 그러한 권리의 향유를 보호하기 위해 적극적인 조치를 취해야 할 의무가 있다. 이는 국가 당국, 자연인, 법인 누구로부터 유발된 것이건 간에 개인의 사생활에 대한 간섭으로부터 개인을 보호하기 위한 적절한 입법 및 기타 조치를 채택할 의무를 의미한다. 이러한 의무는 기업과 인권 이행지침의 기둥 I에도 반영되어 있으며, 여기서는 기업이 관련된 부정적인 인권 영향으로부터 보호해야 할 국가의 의무를 개괄하였다.

11. 기업은 국제적으로 인정된 모든 인권을 존중할 책임이 있다. 이는 타인의 인권 침해를 방지하고 자신이 관련된 부정적인 인권 영향을 해소해야 한다는 의미다. 기업과 인권 지침의 기둥 II는 이러한 책임을 이행하는 방법에 관하여 모든 기업에게 권위 있는 청사진을 제공한다. 존중 책임은 기업의 활동 및 사업 관계 전반에 걸쳐 적용된다.

III. 인공지능이 프라이버시권 및 기타 인권에 미치는 영향

A. 인공지능 시스템의 관련 기능

12. 인공지능 시스템의 운영은 다양한 방식으로 프라이버시 침해 및 기타 권리에 대한 간섭을 초래하고 심화시킬 수 있다. 여기에는 완전히 새로운 애플리케이션뿐만 아니라 개인정보의 수집 및 사용을 증가시켜 프라이버시권리 간섭을 확대, 강화 또는 장려하는 인공지능 시스템의 기능이 포함된다.

13. 인공지능 시스템은 일반적으로 개인정보를 포함한 대용량 데이터셋에 의존한다. 이는 광범위한 데이터 수집, 저장 및 처리를 촉진한다. 많은 기업이 가능한 한 많은 데이터를 수집하기 위해 서비스를 최적화한다. 예를 들어, 소셜 미디어 회사와 같은 온라인 기업은 인터넷 이용자에 대한 방대한 양의 데이터를 수집하고 수익화하는데 의존한다. 이른바 사물 인터넷(Internet of Things)은 데이터 급증의 원천으로 기업과 국가 모두 사용한다. 데이터 수집은 친밀한 공간, 사적인 공간 및 공공 장소를 가리지 않고 이루어진다. 데이터 브로커는 개인정보를 수집, 결합, 분석하고 수많은 수령인과 공유한다. 이러한 데이터 거래는 대부분 공공 조사를 받지 않으며 기존 법적 체계 내에서는 극히 부분적으로만 제한될 뿐이다. 결과 데이터셋이나 수집된 정보의 방대함은 전례 없는 비율이다.

14. 이러한 데이터셋은 기업과 국가에 사람들의 사생활을 노출하는 것 외에도, 여러 가지 방법으로 개인들을 취약하게 만든다. 개인정보 유출은 수백만 명의 민감 정보를 여러 번 노출시켰다. 대용량 데이터셋은 셀 수 없이 많은 유형의 분석과 제3자 공유를 가능하게 하며, 이는 종종 추가적인 프라이버시 침해로 이어지고 인권에 기타 부정적인 영향을 끼친다. 예를 들어, 정부 기관이 기업이 보유한 이러한 데이터셋에 직접 접근할 수 있도록 하는 방식은 관련된 개인의 프라이버시권에 자의적이거나 불법적인 간섭의 가능성을 증가시킨다. 한 가지 특별한 우려사항은 다양한 출처의 데이터를 결합함으로써 탈익명화가 촉진될 가능성이다. 동시에 데이터셋 설계가 개인의 정체성에 영향을 미칠 수 있다. 예를 들어 성별을 2개항으로 기록하는 데이터셋은 남성 또는 여성으로 식별되지 않는 사람들을 오인하게 한다. 또한 개인정보의 장기간 보관은 개인정보 수집 당시 예상하지 않았던 장래의 악용 가능성이 있기 때문에 특별한 위험이 수반된다. 시간이 지남에 따라 데이터가 부정확해지거나 부적절해지거나 역사적인 편견에 따른 오인을 초래하여 향후 개인정보 처리의 편향 또는 오도된 결과를 초래할 수 있다.

15. 인공지능 시스템은 개인정보 처리에만 전적으로 의존하지 않는다는 점에 유의해야 한다. 그러나 개인정보가 관련되지 않더라도 프라이버시권을 비롯한 인권은 아래와 같이 개인정보 이용에 의해 부정적인 영향을 받을 수 있다.

16. 인공지능 도구는 인간 행동의 패턴에 대한 통찰력을 얻기 위해 널리 사용된다. 해당 데이터셋에 접근하면 특정 지역 예배 장소에 얼마나 많은 사람들이 참석할 가능성이 있는지, 어떤 TV 프로그램을 선호하는지, 심지어 대략 몇 시에 일어나고 잠을 자는 경향이 있는지에 대한 결론을 도출할 수 있다. 인공지능 도구는 개인의 정신적, 신체적 상태를 비롯하여 개인에 대해 광범위한 추론을 할 수 있으며 특정 정치적 또는 개인적 성향을 가진 사람들과 같은 집단을 식별할 수 있다. 인공지능은 또한 미래의 행동이나 사건이 일어날 가능성을 평가하는 데 사용된다. 인공지능이 만든 추론과 예측은 그 확률적 특성에도 불구하고 때때로 완전히 자동화된 방식으로 사람들 권리에 영향을 미치는 의사결정의 기반이 될 수 있다.

17. 많은 추론과 예측은, 사람들의 자율성과 자신의 정체성에 대한 세부사항을 확립할 권리를 포함하여, 프라이버시권의 향유에 깊은 영향을 미친다. 이는 또한 사상과 의견의 자유에 대한 권리, 표현의 자유, 공정한 재판 관련 권리 등 다른 권리에도 많은 문제를 야기한다.

18. 인공지능 기반 결정은 오류에서 자유롭지 않다. 사실, 인공지능 솔루션의 확장성은 작아 보였던 오류율의 부정적인 영향을 극적으로 증가시킬 수 있다. 인공지능 시스템의 출력 결함에는 다양한 원천이 있다. 우선, 인공지능 알고리즘의 출력에는 확률적 요소가 있으며, 이는 그 결과물에도 불확실성이 포함되어 있음을 의미한다. 더불어 사용된 데이터의 관련성과 정확성은 종종 의문스럽다. 또한 비현실적인 기대는 원하는 목표를 달성할 채비를 갖추지 않은 인공지능 도구의 배치로 이어질 수 있다. 예를 들어, COVID-19 위험을 진단하고 예측한다는 수백 개의 의료 인공지능 도구를 분석한 결과, 높은 기대 속에 개발되었음에도 이들 중 어느 것도 임상 용도에 적합하지 않은 것으로 나타났다.

19. 결함이 있는 데이터에 의존하는 인공지능 시스템의 출력물은 예를 들어, 한 개인을 테러범으로나 부정 수급을 저지른 것으로 오지목함으로써 여러 가지 방법으로 인권 침해의 원인이 될 수 있다. 편향된 데이터셋은 인공지능 시스템에 기반한 차별적 의사결정으로 이어지는 바 특히 우려된다.

20. 많은 인공지능 시스템의 의사 결정 과정은 불투명하다. 인공지능 시스템의 개발 및 운영을 뒷받침하는 정보 환경, 알고리즘, 모델의 복잡성은 물론 정부와 민간 행위자들의 의도적인 비밀주의는, 인공지능 시스템이 인권과 사회에 미치는 영향을 일반 대중이 이해할 수 있는 뜻깊은 여정을 방해하는 요인이다. 머신러닝 시스템은 불투명성의 핵심적인 요인이다. 이 시스템은 설명하기 어렵거나 불가능한 방식으로 패턴을 식별하고 설명하기 어렵거나 불가능한 처방을 내릴 수 있다. 이를 흔히 "블랙박스" 문제라고 한다. 불투명성으로 인해 인공지능 시스템을 유의미하게 조사하는 것이 어려워지고, 인공지능 시스템이 위해를 야기하는 경우 불투명성이 효과적인 책무성 확보에 장벽이 될 수 있다. 그럼에도 불구하고 이 시스템들이 전적으로 조사될 수 없는 것은 아니라는 점에 주목할 필요가 있다.

B. 주요 분야에서 인공지능 시스템에 대한 우려

21. 본 절에서는 인공지능 도구의 적용으로 인해 문제가 발생한 4가지 주요 영역을 검토하여 이러한 우려 사항이 실제로 어떻게 경험되고 있는지 설명한다.

수사/국가안보/형사사법/출입국관리 분야의 인공지능

22. 국가는 점점 더 많은 인공지능 시스템을 법 집행(수사), 국가안보, 형사사법 및 국경 출입국관리 시스템에 통합하고 있다. 이러한 응용의 많은 부분에 대하여 실제로 우려가 제기되고 있지만, 본 절에서는 새롭게 부상하는 인권 문제 일부를 대표하는 선별된 사례 몇 가지에 초점을 맞출 것이다.

23. 인공지능 시스템은 종종 예측 도구로 사용된다. 이들은 알고리즘을 사용하여 과거 데이터를 비롯한 대량의 데이터를 분석하여 위험을 평가하고 미래 추세를 예측한다. 학습 데이터와 분석된 데이터는 목적별로 예를 들어 범죄 기록, 체포 기록, 범죄 통계, 특정 지역의 경찰 개입 기록, 소셜 미디어 게시물, 통신 데이터 및 여행 기록을 포함할 수 있다. 이 기술은 사람들의 프로필을 생성하고, 범죄 또는 테러 활동이 증가할 가능성이 있는 장소를 식별하며, 심지어 개인을 용의자나 미래의 재범자로 지목하는 데 사용될 수 있다.

24. 이러한 활동이 프라이버시권 및 광범위한 인권에 미치는 영향은 막대하다. 첫째, 사용된 데이터셋에는 다수의 개인에 대한 정보가 포함되어 있으므로 프라이버시권이 관련된다. 둘째, 인공지능 평가는 그 예측에서 확률적 특성을 보유하고 있으므로 합리적인 의심의 근거로 생각해서는 안 되는데도 불구하고, 수색, 검문, 체포 및 기소 등 국가의 개입을 촉발할 수 있다. [이때] 영향을 받는 권리에는 프라이버시권, 공정한 재판에 대한 권리, 자의적인 체포와 구금으로부터의 자유, 그리고 생명에 대한 권리가 포함된다. 셋째, AI 기반 결정에 내재된 불투명성은, 특히 인공지능이 강제적인 조치의 기반이 될 때 국가의 책무성에 대한 중요한 문제를 제기하며, 대테러기관 활동처럼 일반적으로 투명성의 결여를 지적받아온 분야에서는 더욱 그러하다. 넷째, 예측 도구는, 특정 소수자 집단에 편중된 치안 집중의 사례에서처럼, 사용된 데이터셋에 내재된 역사적인 인종적 및 민족적 편견을 반영하여 차별을 영구화하거나 강화할 위험을 내재하고 있다.

25. 생체 인식 기술 분야의 발전으로 법 집행 기관 및 국가 보안 기관에서 생체 인식 기술의 사용이 증가하고 있다. 생체 인식은 얼굴, 지문, 홍채, 음성 또는 보행과 같은 개인의 특정 특징의 디지털 표상을 데이터베이스 내 이런 다른 표상과 비교하는 방식이다. 비교 결과, 더 높거나 낮은 확률로 그 사람이 실제로 그 식별된 사람임을 추론한다. 원격에서 실시간으로 이러한 프로세스가 수행되는 일이 점점 더 많아지고 있다. 특히, 점점 더 많은 국가 당국이 원격 실시간 얼굴 인식을 도입하고 있다.

26. 원격 실시간 생체 인식은 국제 인권법 상으로 심각한 우려를 낳고 있으며, 최고대표는 이전에도 이에 대해 조망했던 바 있다. 이러한 우려사항 중 일부는 예측 도구와 관련된 문제로 나타나며, 이는 개인에 대한 오식별 가능성이나 특정 집단 구성원에 편중된 영향을 미치는 문제 등이다. 또한 얼굴 인식 기술은 개인의 민족, 인종, 출신지, 성별 및 기타 특성에 기반하여 개인을 프로파일링하는 데 사용될 수 있다.

27. 원격 생체 인식은 프라이버시권에 대한 깊은 간섭으로 이어진다. 한 개인의 생체 인식 정보는 다른 사람과 구별되는 독특한 특성을 나타내기 때문에 그 사람의 인격의 핵심 특성 중 하나를 구성한다. 게다가, 원격 생체 인식은 공공 장소에서 체계적으로 개인의 신원을 확인하고 추적할 수 있는 국가의 능력을 극적으로 증가시켜, 사람들이

관찰되지 않고 자신의 삶을 영위할 수 있는 능력을 약화시키고, 이동의 자유 뿐 아니라 표현의 자유, 평화로운 집회 및 결사의 자유에 대한 권리 행사에 직접적으로 부정적인 영향을 미친다. 따라서 이러한 상황에서 최고대표는 실시간 생체 인식 기술의 사용을 제한하거나 금지하려는 최근의 노력을 환영한다.

28. 인공지능 도구는 또한 사람들의 얼굴 표정 및 기타 '예측적 생체 인식'에서 감정 및 정신 상태를 추론하여 보안 위협 여부를 판단하도록 개발되었다. 안면감정인식 시스템은 인간의 감정상태를 표정에서 자동적이고 체계적으로 추론할 수 있다는 전제하에 운영되며, 이는 과학적 근거가 부족하다. 연구자들은 얼굴 표정과 감정의 약한 연관성만을 발견했고 얼굴 표정은 문화와 맥락에 따라 다양하여 감정 인식이 편견과 오해에 취약하다는 점을 강조했다. 이러한 우려를 고려하였을 때, 예를 들어 경찰이 검문이나 체포를 위해 개인을 선별하거나 심문 중 진술의 진실성을 평가하기 위해 공공 기관이 감정 인식 시스템을 사용하는 것은 프라이버시권, 자유로울 권리 및 공정한 재판에 대한 권리 등 인권 침해 위험을 초래한다.

인공지능 시스템과 공공서비스 제공

29. 공공 서비스 전달 지원에 인공지능 시스템이 점점 더 많이 사용되고 있으며, 이는 대개 적시에 정확한 서비스를 전달하는 보다 효율적인 시스템을 개발한다는 명시적 목표를 가지고 있다. 이는 또한 물품과 서비스의 전달이 인공지능 시스템과 연계될 수 있는 인도주의적 상황에서 점점 더 많이 나타나고 있다. 이것은 정당하고 심지어 칭송받을 만한 목표이긴 하지만, 적절한 안전장치가 마련되어 있지 않다면 공공적이고 인도적인 서비스의 전달에 인공지능 도구를 배치하는 것이 인권에 부정적인 영향을 미칠 수 있다.

30. 인공지능은 복지 수급에 대한 의사 결정에서 보육 서비스 방문 가정을 지정하는 것에 이르기까지 다양한 공공 서비스에 사용된다. 이러한 의사 결정은 대규모 데이터셋을 사용하여 이루어지는데, 이러한 데이터셋은 국가 보유 데이터를 포함할 뿐 아니라 소셜 미디어 회사나 데이터 브로커와 같은 민간 기관에서 취득한 정보도 포함하며, 일부는 보호적 법체계 바깥에서 수집된다. 게다가, 인공지능 시스템에 대한 컴퓨팅 지

식과 권한을 민간 기업이 보유하는 경향이 있기 때문에, 이러한 협업은 종종 민간 기업이 인구집단의 많은 부분에 대한 정보를 담은 데이터셋에 접근할 수 있다는 것을 의미한다. 이는 프라이버시권에 대한 우려 뿐 아니라 데이터에 내재된 역사적 편견이 공공 기관의 의사결정에 어떤 영향을 미칠 것인가에 대한 우려를 낳는다.

31. 공공 서비스에 인공지능을 사용하는 것에 대한 가장 큰 우려는 그것이 차별적일 수 있다는 것이며, 특히 소외된 집단과 관련한 것이다. 극빈 및 인권에 관한 특별보고관은 '디지털 복지 디스토피아'를 경고한 바 있는데, 이는 복지 수급자를 발견, 조사, 처벌하기 위해 규제되지 않은 데이터 매칭을 사용하며, 수급자에게 개인의 자율성과 선택을 저해하는 조건을 부과한다. 이러한 우려에 대한 사례가 최근 네덜란드에서 나타났는데, 디지털 복지 부정 탐지 시스템이 프라이버시권을 침해하는 것으로 밝혀져 법원이 이를 금지한 판결이 널리 알려졌다. 해당 시스템은 중앙 및 지방 정부에 대하여 고용, 주택, 교육, 복지 급여 및 건강보험 및 기타 식별 가능한 형태의 데이터를 비롯하여 이전에 분리하여 보관했던 데이터를 공유하고 분석할 수 있도록 광범위한 권한을 제공했다. 또한 이 도구는 저소득 및 소수자 집단 주민을 타겟으로 하여 사실상 사회경제적 배경에 기반한 차별을 초래했다.

고용 환경에서 인공지능 사용

32. 모든 유형의 비즈니스 규모에 걸쳐 다양한 고용주들이 인공지능 시스템을 포함한 데이터 기반 기술을 사용하여 노동자를 모니터링하고 관리하고자 하는 수요 증가가 나타나고 있다. 소위 인력 분석은 직원들에 대한 보다 효율적이고 객관적인 정보 제공을 요구한다. 여기에는 채용, 승진 계획 또는 해고에 대한 자동화된 의사 결정이 포함될 수 있다.

33. 이러한 기술의 대부분은 직무 관련 행동 및 성과 모니터링에 초점을 맞추고 있지만, 인공지능 시스템의 적용 범위는 직무와 관련이 없는 행동 및 데이터로도 확대된다. COVID-19 대유행은 두 가지 방법으로 이러한 추세를 가속화시켰다. 첫째, 노동자에게 예방적 건강 체계를 제공하는 일부 기업은 건강 관련 데이터를 점점 더 많이 수집한다. 둘째, 사람들이 재택으로 일하는 동안 많은 업무가 디지털로 실행됨에 따라,

인공지능 시스템에 의한 직장 감시가 사람들의 가정으로 옮겨왔다. 두 추세 모두 직장 모니터링 데이터를 비업무 데이터 입력과 결합시킬 위험을 증가시킨다. 이러한 인공지능 기반 모니터링 관행은 전체 데이터 수명 주기 동안 방대한 프라이버시 위험을 구성한다. 여기에 더해, 데이터가 직원에게 처음 공지된 목적외 다른 목적으로 사용될 수 있으며, 이는 소위 기능 확대(function creep)를 초래할 수 있다. 동시에, 사람 관리에 사용되는 많은 인공지능 시스템의 기반이 되는 정량적 사회과학은 견고하지 않고 편견에 취약하다. 예를 들어, 회사가 남성, 백인, 중년 남성을 선호하는 과거 데이터셋으로 학습된 인공지능 채용 알고리즘을 사용하는 경우, 결과 알고리즘은 구인에 적합한 자격을 동등하게 갖춘 여성, 유색인종 및 젊은이나 노년층을 선호하지 않을 것이다. 동시에, 근로자를 보호하기 위한 책무성 구조와 투명성이 결여된 경우가 많으며, 노동자들은 인공지능 기반 모니터링 실시에 대한 설명을 거의 또는 전혀 듣지 못하고 있다. 어떤 상황에서는 기업이 순수하게 직장 내 부정행위를 예방하려는 관심을 가지고 있지만, 그러한 관심을 위해 취하는 조치들이 직장 내 상호작용의 사회적 모드 및 연결된 성과 목표를 과도하게 침입적으로 정량화하는 것을 정당화할 수 없다. 직장 환경 및 사용자와 노동자 사이 권력 관계를 고려해 볼 때, 우리는 노동자들이 노동에 대한 대가로 프라이버시권을 포기하도록 강요받고 있을 가능성의 시나리오를 떠올릴 수 있다.

온라인 정보 관리를 위한 인공지능

34. 소셜 미디어 플랫폼은 인공지능 시스템을 사용하여 콘텐츠 관리에 대한 의사결정을 지원한다. 기업은 이러한 시스템을 사용하여 콘텐츠의 순위를 매기고 무엇을 증폭시킬지, 무엇을 격하시킬지 결정하며, 각기 다른 개별 이용자 프로필별로 이러한 의사결정을 개인화하는 등의 방식을 사용한다. 자동화는 또한 관할권 내에서 또는 관할권 간에 서로 다른 법적 요건에 대응하는 등 콘텐츠에 대한 제한을 실시할 때도 사용된다. 인식된 온라인 위해성과 관련하여 인터넷 매개사업자에 대해 필터 의무를 도입하는 것은 이러한 시스템이 지역적 및 지구적 수준의 표현의 자유 및 프라이버시권에 미치는 심각한 영향을 고려하지 않고 인공지능에 광범위하게 의존하는 경향을 확대할 위험이 있다.

35. 내용 추천, 증폭 및 조정 시스템이 의존하는 방대한 데이터셋은 플랫폼 이용자와 개인 관계망에 대한 광범위한 온라인 모니터링 및 프로파일링을 통해 생성되고 지속적으로 확장된다. 정보를 수집하고 그로부터 추론하는 이러한 영구적 프로세스는 극심한 시장 집중과 결합되어 전세계적으로 소수의 기업들이 수십억 명의 개인에 대한 프로파일과 네트워크화된 일반 공공 영역을 쥐고 통제하는 상황을 초래했다.

36. 막대한 시장 지배력을 가진 기업들이 수행한 인공지능 지원 콘텐츠 큐레이션은 연이은 두 명의 의사 표현의 자유 특별보고관 수임자들이 지적했듯이 개인의 의견 형성 및 개발 능력에 미치는 영향에 대한 우려를 불러 일으키고 있다. 또한 플랫폼 추천 시스템은 사람들의 선호도, 인구통계학적이고 행태적인 패턴에 대한 통찰력에 의존하면서 사용자 참여를 극대화하는 데 주력하는 경향이 있으며, 이는 종종 선정적인 콘텐츠를 촉진하여 양극화 경향을 잠재적으로 강화하는 것으로 나타났다. 더욱이, 정보의 타겟팅은 달갑지 않고 심지어 위험한 사생활 침해로 이어질 수 있다. 예를 들어 추천 시스템은 소셜 미디어 플랫폼에서 폭력 생존자의 잠재적인 친구로 가해자를 추천하는 결과를 낳았고, 그 반대의 추천으로 생존자를 위험에 빠뜨리는 결과도 낳았다. 또한, 검색 결과의 데이터에 반영된 다수 또는 지배적 집단의 편향이 소수자 집단 또는 취약 집단에 대해 또는 그 안에서 공유되는 정보에 대해 영향을 미치는 것으로 나타났다. 예를 들어, 연구 결과 구글의 검색 결과에서 성별과 인종적 편향은 충격적인 정도인 것으로 나타났다.

Ⅳ. 문제에 대한 해결

37. 일반적으로 새로운 기술, 특히 인공지능에 대한 인권 기반 접근의 필요성이 점점 더 많은 전문가, 이해관계자 및 국제사회에서 인식되고 있다. 인권 기반 접근법은 사회가 기술적 발전의 혜택을 극대화하는 동시에 위해를 방지하고 제한할 수 있는 방법을 식별하는 데 도움이 되는 도구를 제공한다.

A. 기본 원칙

38. 인공지능에 대한 인권 기반 접근법은 평등과 차별금지, 참여와 책무성, 지속 가능

한 개발 목표와 기업과 인권 이행지침의 중요 원칙을 포함한 여러 가지 핵심 원칙의 적용을 요구한다. 또한, 합법성, 정당성, 필요성 및 비례성의 요건이 AI 기술에도 일관되게 적용되어야 한다. 또한, 인공지능은 가용성, 경제성, 접근성 및 품질이라는 핵심 요소를 달성함으로써 경제적, 사회적 및 문화적 권리의 실현을 촉진하는 방식으로 배치되어야 한다. 인공지능 사용과 관련된 인권 침해와 남용 피해를 입은 사람들은 효과적인 사법적 및 비사법적 구제수단을 이용할 수 있어야 한다.

39. 위에서 지적한 바와 같이, 프라이버시권에 대한 제한은 법률에 규정되어야 하며, 정당한 목적을 달성하기 위해 필요한 것이어야 하며, 그 목적에 비례해야 한다. 실제에서 이는 어떠한 조치가 정해진 목표를 달성할 수 있는지, 그 목표가 얼마나 중요한지, 그리고 조치의 영향이 어떤지를 각국이 신중하게 판단해야 함을 의미한다. 각국은 또한 덜 침해적인 접근법이 동일한 효과로 동일한 결과를 달성할 수 있는지 판단해야 하며, 만약 그렇다면 해당 조치를 채택해야 한다. 최고대표는 이미 정보 기관과 법 집행 기관 감시에 필요한 제한과 안전장치들을 개략적으로 설명한 바 있다. 필요성 및 비례성 검사 또한 특정 조치를 취해서는 안 된다는 판단으로 이어질 수 있다는 점에 유의해야 한다. 예를 들어, 통신 회사 등에 부과된 포괄적이고 무차별적인 통신 데이터 보존 요구는 비례성 검사를 통과하지 못할 수 있다. 마찬가지로 복지급여 수급자에게 생체인식 요구를 부과하는 것은 대체수단이 제공되지 않는다면 불균형적이다. 또한 조치를 단독으로 평가하는 것이 아니라 별개의 조치이지만 상호 작용하는 조치들과의 누적 효과를 적절히 고려하는 것이 중요하다. 예를 들어, 국가는 새로운 인공지능 기반 감시 도구를 배치하기로 결정하기 전에 기존 역량과 기능이 프라이버시권 및 기타 권리의 향유에 미치는 영향을 조사해야 한다.

B. 입법 및 규제

40. 프라이버시권 및 상호 연결된 권리에 대한 효과적인 보호는 국가가 수립한 법률, 규제 및 제도적 체계에 달려 있다.

41. 데이터 기반 인공지능 시스템이 출현함에 따라 개인정보보호법으로 효과적이고 법적으로 보호해야 할 중요성이 커지고 있다. 이러한 보호는 최고대표의 프라이버시권

관련 이전 보고서에서 확인된 최소 기준을 충족해야 한다.

42. 개인정보보호 체계는 인공지능 사용과 관련된 새로운 위협을 처리해야 한다. 예를 들어 법적으로 추론되거나 나아가 사용 및 공유될 수 있는 데이터의 유형에 대하여 법률로써 제한할 수 있다. 또한 입법자들은 개인의 권리 강화를 고려하고, 자신의 권리에 영향을 미치는 완전히 자동화된 결정에 대하여 유의미한 설명에 대한 권리와 거부할 권리 등을 부여해야 한다. 인공지능 기술의 계속적인 발전에 맞추어, 개인정보보호 체계 내에서 더 많은 안전장치를 지속적으로 개발할 필요가 있다.

43. 방대한 정보 비대칭성을 비롯한 글로벌 데이터 환경의 복잡성과 불투명성 증가에 대응하기 위한 핵심 요소 중 하나는 독립적인 개인정보보호 감독 기관이다. 이 기관들은 효과적인 집행력을 가지고 있어야 하고 적정한 자원을 확보해야 한다. 강력한 진정 메커니즘 구축 등을 통하여 시민 사회가 개인정보보호법의 집행을 지원할 수 있도록 장려하여야 한다.

44. 개인 정보 보호법을 넘어, 인공지능의 문제를 권리 존중 방식으로 해결하기 위해 광범위한 법안들을 검토하고 장래 채택해야 한다.

45. 인공지능 애플리케이션, 시스템 및 용도의 다양성을 고려하여, 규제는 부문별 문제를 해결하고 관련 위험에 대한 대응을 맞춤화할 수 있을 만큼 충분히 구체적이어야 한다. 인권에 대한 위험이 높을수록 인공지능 기술의 사용에 대한 법적 요구사항이 엄격해져야 한다. 따라서 법 집행, 국가안보, 형사사법, 사회보장, 고용, 보건의료, 교육 및 금융 등 개인의 이해관계가 특히 높은 분야가 우선되어야 한다. 법률과 규제에 있어 위험기반 접근방식은 특정 인공지능 기술, 애플리케이션 또는 사용 사례에 대한 금지를 요구해야 하며, 필요성과 비례성 검사를 통과하지 못하는 등 국제 인권법 하에서 정당화되지 않는 잠재적 또는 실제적 영향을 초래하는 경우가 이에 해당한다. 게다가, 차별 금지와 본질적으로 충돌하는 인공지능의 사용은 허용되어서는 안 된다. 예를 들어, 정부 또는 인공지능 시스템이 금지된 차별 기준에 의해 개인을 집단으로 분류하는 개인 사회신용점수는 이러한 원칙에 따라 금지되어야 한다. 특정 상황에 배치될 때 그 사용이 인권에 위험을 초래하는 시스템의 경우, 국가는 관할권 내외에서 부정적인 인

권 영향을 방지하고 완화하기 위해 해당 시스템의 사용과 판매를 규제할 필요가 있다. 인권에 부정적인 영향이 발생할 가능성이 있을 때 인간의 감독 및 의사결정의 의무적 개입이 규정되어야 한다. 위험을 평가하고 해결할 수 있기까지 시간이 걸릴 수 있다는 점을 감안할 때, 국가는 또한 원격 실시간 얼굴 인식과 같은 잠재적인 고위험 기술의 사용에 대하여, 그 사용이 인권을 침해할 수 없음이 보장될 때까지 유예(모라토리엄)하여야 한다.

46. 또한 국가는 인권 활동가 또는 언론인을 타겟으로 하는 등 인권 침해에 그러한 기술이 사용될 위험이 있을 때 그 판매를 방지하기 위해 국경간 감시 기술의 무역에 대해 강력한 수출 통제 체제를 도입하여야 한다.

47. 인공지능 시스템에서 발생하는 위험의 스펙트럼은 인공지능 시스템의 개발, 배치 및 사용에 대한 적정하고 독립적이고 공정한 감독이 필요하다는 사실을 시사한다. 이러한 감독은 행정적, 사법적, 준사법적 및 의회 감독 기관의 조합으로 수행될 수 있다. 예를 들어, 개인 정보보호 기관, 소비자 보호 기관, 부문별 규제 기관, 차별 방지 기구 및 국가 인권 기구는 감독 시스템의 일부를 구성해야 한다. 또한 인공지능 사용을 감독하는 부문 간 규제 기관은 기본 표준을 수립하고 정책 및 집행의 일관성을 보장하는 데 도움이 될 수 있다.

C. 인권 실사

48. 국가와 기업은 인공지능 시스템의 구입, 개발, 배치 및 운영 시 뿐 아니라 개인에 대한 빅데이터를 공유하거나 사용하기 전에 포괄적인 인권실사를 실시해야 한다. 또한 국가는 이러한 프로세스의 자원을 조달하고 주도할 뿐 아니라 기업에 포괄적인 인권 실사를 실시하도록 요구하거나 장려할 수 있다.

49. 인권실사절차의 목적은 기업이 원인이 되었거나 초래하였거나 직접 관련된 부정적인 인권 영향을 식별, 평가, 방지 및 완화하는 것이다. 실사 절차에서 인공지능의 사용이 인권과 양립할 수 없는 것으로 드러나는 경우, 위해를 완화할 수 있는 유의미한 방법이 없기 때문에, 이러한 형태의 사용은 더 이상 추진되어서는 안 된다. 인권영향평

가는 인권 실사 과정의 필수적인 요소이다. 인권 실사는 인공지능 시스템의 수명주기 전반에 걸쳐 실시하여야 한다. 여성과 십대 여성, 레즈비언, 게이, 양성애자, 트랜스젠더 및 성소수자, 장애인, 소수자 집단에 속하는 사람, 노인, 빈곤층 및 기타 취약한 상황에 처해 있는 사람들에게 불균형한 영향을 미치는 것에 각별한 주의를 기울여야 한다.

50. [인공지능이 인권에 미치는 위험의] 완화 방법 개발 및 평가를 비롯한 인권실사는 잠재적으로 영향을 받는 권리 주체 및 시민사회와 유의미한 협의로 이루어져야 하며, 여러 학제간 전문성을 갖춘 전문가도 이에 참여해야 한다. 국가와 기업은 그들이 사용하는 인공지능 시스템이 인권에 부정적인 영향을 미치는지 여부를 확인하기 위해 지속적으로 모니터링해야 한다. 인권영향평가의 결과, 인권 위험을 해결하기 위해 취해진 조치 및 공공 협의 내용은 그 자체로 공개되어야 한다.

D. 국가-기업 연합

51. 국가와 기술 기업 사이에 긴밀한 연합 관계가 있는 상황에 대해서는 집중적인 주의가 요구된다. 국가는 법적, 정책적 조치에 대한 국가적 역할을 넘어 인공지능이 어떻게 개발되고 사용되는지를 형성할 수 있는 중요한 경제 주체이다. 국가가 민간 부문의 인공지능 개발자 및 서비스 제공자와 협력하는 경우, 국가는 인권과 양립할 수 없는 목적을 위해 인공지능이 사용되지 않도록 추가적인 조치를 취해야 한다. 이러한 조치는 국영 기업 경영, 연구개발 자금 지원, 기타 국가가 인공지능 기술 기업에 제공하는 금융 등 지원, 민영화 시도 및 공공 조달 관행 전반에 적용되어야 한다.

52. 국가가 경제 행위자로 활동하는 경우, 그들은 국제인권법 상 주요 의무의 보유자이며 그 의무를 적극적으로 이행하여야 한다. 동시에, 기업들은 국가와 협력할 때 인권을 존중할 책임이 있으며 인권법과 상충되는 국가의 요구에 직면했을 때 인권을 존중할 방법을 모색해야 한다. 예를 들어, 인권 기준을 충족하지 못하는 개인 정보 제출 요구를 받았을 때, 기업은 초래될 위해를 방지하거나 완화하기 위한 수단을 사용해야 한다.

53. 국가는 책임 있는 사업 행위를 지속적으로 요구함으로써 인권 보호를 강화할 수 있다. 예를 들어, 수출 신용 기관이 인공지능 기술 회사에 지원을 제공할 때, 그들은 이러한 기업들이 권리 존중 행위 부문에서 견실한 실적을 보유하고 강력한 실사 과정을 통해 이를 입증할 수 있도록 보장해야 한다.

54. 국가가 공공재 또는 서비스를 제공하기 위해 인공지능 기업에 의존하는 경우, 국가는 인공지능 시스템의 개발과 배치를 감독할 수 있도록 보장해야 한다. 이는 인공지능 애플리케이션의 정확성과 위험에 대한 정보를 요구하고 평가함으로써 이루어질 수 있다. 위험을 효과적으로 완화할 수 없는 경우, 국가는 공공재 또는 서비스 전달에 인공지능을 사용하지 말아야 한다.

E. 투명성

55. 인공지능 시스템의 개발자, 판매자, 운영자, 사용자는 인공지능 사용의 투명성에 대한 노력을 대폭 강화해야 한다. 첫 번째 단계로, 국가, 기업 및 기타 인공지능 사용자는 그들이 사용하는 시스템의 종류, 사용 목적, 시스템 개발자와 운영자의 신원에 대한 정보를 제공해야 한다. 의사결정이 자동으로 이루어지고 있거나 이루어졌을 때 또는 자동화 도구의 도움을 받아 이루어졌을 때 영향을 받는 개인에게 이를 체계적으로 알려야 한다. 개인이 제공하는 개인 정보가 인공지능 시스템에서 사용하는 데이터 셋의 일부가 될 경우 이를 통지해야 한다. 또한 인권에 중대한 애플리케이션에 대하여 국가는 그 인공지능 도구와 사용에 대한 주요 정보를 포함하는 등록제도를 도입해야 한다. 개인 정보 보호 체계에 포함된 투명성 의무와 개인정보 열람, 삭제 및 정정권의 효과적인 시행이 보장되어야 한다. 개인이 자신에 대해 작성된 프로파일을 더 잘 이해하고 통제할 수 있도록 특별히 주의를 기울여야 한다.

56. 투명성 증진은 위에서 설명한 ‘블랙박스’ 문제를 극복하기 위한 지속적인 노력을 포함함으로써 더 전진해야 한다. 인공지능 시스템을 보다 설명 가능하게 하는 (종종 알고리즘 투명성이라 언급되는) 방법론의 개발 및 체계적 배치는 적절한 권리 보호를 보장하는 데 있어 가장 중요하다. 이것은 인공지능이 사법 절차 내 또는 경제적, 사회적, 문화적 권리 실현에 필수적인 사회서비스와 관련된 중요한 문제를 결정하는

데 사용될 때 가장 필수적이다. 연구자들은 이미 그 목표를 달성하기 위한 다양한 접근 방식을 개발했으며, 이 분야에 대한 투자 증대가 중요하다. 국가는 또한 지적재산권 보호가 인권에 영향을 미치는 인공지능 시스템에 대한 유의미한 조사를 방해하지 않도록 조치를 취해야 한다. 조달규칙은 인공지능 시스템에 대한 감사가능성 등 투명성의 필요성을 반영하여 갱신되어야 한다. 특히 국가는 인권에 중대한 부정적인 영향을 미칠 수 있음에도 유의미한 감사 대상이 될 수 없는 인공지능 시스템의 사용을 피해야 한다.

V. 결론 및 권고

A. 결론

57. 본 보고서는 인공지능 기술이 프라이버시권 및 기타 인권 행사에 미치는 영향이 부정할 수 없고 지속적으로 증가하는 실태의 긍정적이고 부정적인 면 모두를 조망하였다. 보고서는 널리 사용되는 인공지능 시스템의 부분을 이루는 대체로 투명하지 않은 개인정보 수집 및 교환의 광범위한 생태계를 비롯하여 우려스러운 발전상에 대하여 지적하였다. 이러한 시스템은 치안 유지와 사법 행정에 대한 정부 접근법에 영향을 미치고, 공공 서비스의 접근성을 판단하며, 누가 채용될 기회를 가질 것인지 결정하고, 사람들이 온라인에서 어떤 정보를 보고 공유할 수 있는지에 영향을 미친다. 게다가, 인공지능 기반 결정과 관련된 차별의 위험은 너무 현실적이다. 보고서는 포괄적인 인권 기반 접근만이 모두에게 이익이 되는 지속 가능한 해결책을 보장할 수 있음을 강조하면서 인공지능과 관련된 근본적인 문제를 해결하는 다양한 방법을 개괄하였다.

58. 그럼에도 불구하고, 인공지능 주변에서 발생하고 있는 다양하고 새로운 문제들을 고려해 보았을 때, 이 보고서는 끊임없이 진화하는 인공지능 환경에 대한 한순간의 스냅샷이다. 추가적으로 분석되어야 할 분야에는 보건의료, 교육, 주택 및 금융 서비스가 포함된다. 국가, 국제기구 및 기술 회사들에게 점점 더 믿음직한 해결책이 되고 있는 생체 인식 기술 분야에는 더 많은 인권 지침이 시급하다. 또한 인권 관점에서 향후 초점을 맞추어야 하는 할 작업 분야는 글로벌 데이터 환경에서 엄청난 책무성 공백을

메울 수 있는 방법을 찾는 것이어야 한다. 마지막으로, 인공지능발 차별을 극복하기 위한 솔루션을 시급히 확인하고 구현하여야 한다.

B. 권고

59. 최고대표는 각국에 다음을 권고한다.

(a) 인공지능의 개발, 사용 및 거버넌스에 있어 모든 인권을 보호하고 강화해야 할 필요성을 인정하고, 온라인과 오프라인에서 모든 인권을 동등하게 존중하고 시행할 수 있도록 보장한다.

(b) 인공지능의 사용이 모든 인권에 부합하도록 보장하고, 인공지능의 사용에서 프라이버시권 및 기타 인권에 대한 간섭이 법률에 의해 규정되고, 정당한 목적을 추구하며, 필요성과 비례성의 원칙을 준수하며, 문제의 권리의 본질을 손상시키지 않도록 보장한다.

(c) 국제 인권법을 준수하며 운영할 수 없는 인공지능 애플리케이션을 명시적으로 금지하고 인권 향유에 고위험을 수반하는 인공지능 시스템의 판매 및 사용에 대해 인권 보호를 위한 적정한 안전장치가 마련될 때까지 모라토리엄을 부과한다.

(d) 최소한 책임 당국이 프라이버시 및 개인정보보호 기준을 준수하고 중대한 정확성 문제와 차별적 영향이 없음을 입증할 수 있을 때까지, 그리고 A/HRC/44/24 제53항 (j)(i-v)에 명시된 모든 권고안이 시행될 때까지, 공공장소에서 원격 생체 인식 기술의 사용에 대하여 모라토리엄을 부과한다.

(e) 인공지능 환경에서 프라이버시권을 보호하기 위한 필수 전제조건으로 공공 및 민간 부문에 대한 개인정보보호법을 독립적이고 공정한 권한을 통해 채택하고 효과적으로 시행한다.

(f) 공공 및 민간부문의 인공지능 사용과 관련된 다층적이고 부정적인 인권 영향을 적절하게 방지하고 완화하는 입법 및 규제 체계를 도입한다.

(g) 인공지능 시스템의 사용과 관련된 인권 침해 및 남용의 피해자가 효과적인 구제수단을 이용할 수 있도록 보장한다.

(h) 특히 공공 부문에서 인권에 중대한 영향을 미칠 수 있는 모든 인공지능 지원 의사결정에 대하여 적절한 설명 가능성을 요구한다.

(i) 국가 및 기업의 인공지능 시스템 사용과 관련된 차별을 해소하기 위한 노력을 강화하며, 여기에는 인공지능 시스템의 결과물 및 그 배치로 인한 영향에 대하여 체계적인 평가와 모니터링을 실시하고, 요구하고, 지원하는 것이 포함된다.

(j) 인공지능 기술의 제공 및 사용에 대한 민관 파트너십은 투명해야 하고 독립적인 인권 감독의 대상이 되어야 하며, 인권에 대한 정부 책무를 포기하는 결과로 이어지지 않도록 보장해야 한다.

60. 최고대표는 국가와 기업에 대하여 다음을 권고한다.

(a) 인공지능 시스템의 설계, 개발, 배치, 판매, 구입, 운영의 수명주기 전반에 걸쳐 체계적으로 인권실사를 수행한다. 그 인권 실사의 핵심 요소는 정례적이고 포괄적인 인권영향평가여야 한다.

(b) 일반인과 영향을 받는 개인들에게 적절히 알리고 자동화 시스템에 대한 독립적이고 외부적인 감사를 가능하게 하는 등 인공지능 사용의 투명성을 크게 증가시킨다. 인공지능 사용과 관련된 잠재적 또는 실제적 인권 영향이 더 많고 심각할수록, 더 높은 투명성이 필요하다.

(c) 인공지능 개발, 배치 및 사용에 대한 결정에 모든 관련 이해당사자, 특히 영향을 받는 개인 및 집단의 참여를 보장한다.

(d) 자금 지원 및 해당 목적 연구 수행 등을 통하여 인공지능 기반 의사 결정의 설명 가능성을 증진한다.

61. 최고대표는 기업에 대하여 다음을 권고한다.

(a) 기업과 인권 이행지침을 완전히 운용하는 등 모든 인권을 존중하는 책임을 이행하기 위한 노력을 다한다.

(b) 인공지능 시스템의 출력물 및 그 배치의 영향에 대한 체계적인 평가 및 모니터링 수행을 비롯하여 인공지능 시스템의 개발, 판매 또는 운영과 관련된 차별을 해소하기 위한 노력을 강화한다.

(c) 인공지능 개발을 책임지는 인력의 다양성을 보장하기 위해 과감한 조치를 취한다.

(d) 인권에 부정적인 영향을 미치거나 이를 초래하였을 때 효과적인 운영 차원의 고충처리 메커니즘 등 정당한 절차를 이용하는 구제 수단을 제공하거나 협조한다.

부록 Ⅲ.

유럽평의회 인권위원장
<인공지능 블랙박스 개봉: 인권 보호를 위한 10단계>[4)]

1. 인권영향평가

회원국은 공공기관이 구입, 개발 또는 배치하는 인공지능 시스템에 대해 인권영향평가 실시 절차를 도입하는 법적 체계를 수립해야 한다. 인권영향평가는 규제 영향평가 및 개인정보보호 영향평가 등 공공기관이 수행하는 다른 영향평가와 유사한 방식으로 구현 및 운영되어야 한다.

회원국은 법률에서 인권영향평가의 적용을 받는 인공지능 시스템의 유형을 기술할 수 있지만, 그러한 설명은 인공지능 시스템의 모든 수명주기 단계에서 개인의 인권을 간섭할 가능성이 있는 모든 인공지능 시스템을 포괄할 수 있을 만큼 충분히 포괄적이어야 한다.

인권영향평가의 법적 체제의 일부로서, 공공기관에 대하여 계획 중이거나 기존에 사용 중인 인공지능 시스템에 대하여 자체적인 평가를 실시하도록 해야 한다. 이 자체 평가는 인공지능 시스템의 특성, 맥락, 범위 및 목적을 고려하여 시스템이 인권에 미칠 수 있는 잠재적 영향을 평가해야 한다. 공공기관이 제안된 인공지능 시스템을 아직 조달하거나 개발하지 않은 경우, 시스템의 구입 또는 개발에 앞서 이 평가를 수행해야 한다.

인권 영향 평가는 인공지능 시스템에 대하여 독립된 감독 기관이나 관련 전문지식을 갖춘 외부 연구자/감사관이 실시하는 유의미한 외부 검토를 포함하여, 인권 영향과 시간의 흐름에 따른 위험을 발견하고 측정하고 전체적으로 파악하는 데 도움이 되어야 한다. 공공기관은 이러한 유의미한 외부 검토를 수행할 때 국가인권기구를 참여시킬 것을 고려

4) Council of Europe Commissioner for Human Rights (2019). 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.

해야 한다.

자체 평가와 외부 검토는 인공지능 시스템 뒤의 모델이나 알고리즘의 평가에 국한되지 않고 의사결정자가 어떻게 입력 정보를 수집하거나 영향을 미치는지, 그러한 시스템의 출력물을 해석할 수 있는지에 대한 평가를 포함해야 한다. 또한 인공지능 시스템이 시스템의 수명주기 전체에 걸쳐 유의미한 인간의 통제를 유지하는지 여부를 평가하는 것도 포함해야 한다.

자체 평가나 외부 검토 결과 인공지능 시스템이 인권을 침해할 실질적인 위험이 있음이 드러난 상황이면 인권영향평가는 그러한 위험을 방지하거나 완화하기 위해 고안된 수단, 안전장치 및 방법을 수립해야 한다. 공공기관이 이미 구축한 인공지능 시스템에서 그러한 위험이 식별된 상황이면 위에서 언급한 수단, 안전장치 및 방법이 적용될 때까지 그 사용을 즉시 중단해야 한다. 식별된 위험을 유의미하게 완화할 수 없는 경우, 공공기관은 인공지능 시스템을 배치하거나 기타 방법으로 사용해서는 안 된다. 자체 평가나 외부 검토에서 인권 침해 사실이 드러난 경우 공공기관은 즉시 조치를 취해 인권 침해를 해소 및 구제하고 그러한 침해 위험이 재발될 위험을 방지 또는 완화하는 조치를 취해야 한다.

외부 검토 절차를 거친 연구 결과 및 결론을 비롯한 인권영향평가는 쉽게 접근할 수 있고 기계로 판독할 수 있는 형식으로 일반에 공개되어야 한다.

공공기관은 제3자가 정보(기밀사항 혹은 영업 비밀)에 대한 제한을 포기하지 않는 상황에서 그 제한이 (1) (외부 연구/검토의 수행을 포함하여) 인권영향평가를 수행하거나 (2) 인권영향평가를 공개하는 것을 지연시키거나 방해하는 경우에는 해당인으로부터 인공지능을 구입하지 말아야 한다.

공공기관들에 대하여 인공지능 시스템을 구입·개발하는 시점뿐 아니라 정기적으로 인권영향평가를 실시하도록 요구하여야 한다. 인권영향평가는 최소한 인공지능 시스템의 각 수명주기의 새로운 단계 또는 그와 유사하게 중요한 시점에 수행되어야 한다.

2. 공개적인 의견 수렴

국가의 인공지능 시스템 사용은 공개 조달 기준에 의해 통제되고 투명하게 운영되는

절차를 적용하여 모든 이해관계자가 투입할 수 있도록 개방되어야 한다. 회원국은 인공지능에 특화된 요구사항을 반영하기 위해 정보 접근, 정부 공개 및 공공 조달 법률 및 정책의 개정을 검토하여야 한다.

회원국은 인공지능 시스템과 관련된 다양한 단계는 물론 최소한 조달 및 인권영향평가 단계에서 공개적인 의견 수렴을 실시해야 한다. 유의미한 공개 의견 수렴 과정은 인공지능 시스템 관련 모든 정보의 시기적절하고 사전적인 공개를 수반하여 인공지능 시스템의 작동, 기능 및 잠재적 또는 측정된 영향에 대한 적절한 이해를 촉진한다. 공개적인 의견 수렴은 정부 관계자, 민간 부문 대표자들, 학계, 비영리 부문, 언론 및 소수자 집단이나 영향을 받는 집단 및 공동체의 대표자를 비롯하여 모든 이해관계자가 의견을 제시할 수 있는 기회를 제공해야 한다. 국가인권기구는 시민사회와 국가 기관 사이의 가교 역할을 맡아 유의미한 협의 수행에 도움을 줄 수 있다.

3. 민간부문의 인권기준 이행을 촉진해야 하는 회원국 의무

회원국은 유엔 기업 및 인권 이행지침과 인권과 기업 관련 각료위원회 권고 CM/Rec(2016)3을 효과적으로 이행해야 한다. 기업은 그 이행에 있어 젠더 관련 위험과 관련하여 차별적이지 않은 방법을 사용하여야 한다. 또한 기업은 소관 사항 내에 정주하거나 운영되는 모든 인공지능 행위자(예: 인공지능 개발자, 소유자, 제조업체, 관리자, 서비스 제공자 및 기타 인공지능 기업)도 마찬가지로 운영 전반에 걸쳐 이러한 원칙을 구현해야 한다는 요구를 분명히 제시하여야 한다.

유럽인권협약에 따른 능동적이고 절차적인 의무를 준수하기 위해 회원국은 인공지능 시스템의 수명주기 전체에 걸쳐 인공지능 행위자의 인권 침해로부터 개인의 인권을 보호하는 데 필요한 조치를 적용해야 한다. 회원국은 자국 법률이 인공지능 행위자의 인권 존중에 도움이 되는 여건을 조성하고 인공지능 관련 인권 침해에 대하여 효과적인 책무와 구제 수단에 장벽을 만들지 않도록 구체적으로 보장해야 한다.

회원국은 적절한 경우 인권 실사를 비롯하여 인공지능 행위자가 인권을 존중하도록 요구하는 추가적인 조치를 적용해야 한다. 회원국은 인공지능 행위자가 인공지능 시스템에 의해 야기되는 위해를 방지하거나 완화하기 위한 효과적인 조치를 취하도록 요구해야

하며, 인공지능 행위자는 인공지능 시스템에 의해 야기되는 위해를 식별, 방지 및 완화하려는 노력에 대해 투명해야 한다. 회원국은 인권에 부정적인 영향을 미치는 것으로 식별된 위험이 적절히 완화되지 않고 해결되지 않을 경우 적절한 조치 결과를 공개해야 한다.

4. 정보 및 투명성

개인의 인권에 유의미한 영향을 미치는 의사 결정 과정에서 인공지능 시스템이 사용되는 사실은 식별 가능해야 한다. 인공지능 시스템의 사용은 명확하고 접근하기 쉬운 용어로 공개되어야 할 뿐만 아니라 개인은 의사결정에 도달하는 방법과 그러한 결정이 어떻게 검증되었는지 이해할 수 있어야 한다.

인공지능 시스템이 공공 서비스, 특히 사법, 복지 및 보건의료 서비스에서 개인과 상호작용하는 데 사용되는 경우, 그 사실을 이용자에게 고지하여야 하며 요청시 전문가의 자문을 지체없이 구할 수 있음을 전달하여야 한다. 인공지능 시스템의 출력물이 오로지 또는 중대하게 지원하는 공공기관의 의사결정이 누군가에 대해 이루어진 경우 그 사람에게 위 정보를 고지하고 즉시 제공해야 한다.

전체 인공지능 시스템에 대한 감독 또한 투명성 요구에 따라야 한다. 이는 해당 시스템의 프로세스, 인권에 대한 직간접적 영향, 인권에 부정적인 시스템의 영향을 식별하고 완화하기 위해 취한 조치 등 시스템에 대한 정보를 공개하는 형태로 이루어지거나, 독립적이고 포괄적이며 효과적인 감사의 형태로 이루어질 수 있다. 모든 경우에, 공개되는 정보는 인공지능 시스템에 대한 유의미한 평가를 가능하게 해야 한다. 어떤 인공지능 시스템도 인간의 검토와 정밀 조사를 허용하지 못할 정도로 복잡해서는 안 된다. 투명성과 책무성의 적절한 기준을 따를 수 없는 시스템은 사용해서는 안 된다.

5. 독립적 감독

회원국은 공공기관과 민간 기업이 개발, 배치, 사용하는 인공지능 시스템의 인권 준수에 대한 독립적이고 효과적인 감독을 위해 입법 체계를 수립해야 한다. 이 입법 체계는

효과적으로 상호협력하는 행정적, 사법적, 준사법적 또는 의회 감독기구들의 조합으로 구성된 구조를 포함할 수 있다. 회원국은 적절한 경우 기존 국가인권기구에 인공지능 시스템의 인권 준수에 대해 독립적이고 효과적인 감독 역할을 수행할 수 있도록 권한 부여를 검토해야 한다.

감독 기관은 인공지능 시스템을 개발, 배치 및 기타 사용하는 공공기관과 민간 기업들로부터 독립적이어야 하며 감독기능을 수행하기에 적절하고 충분한 분야간 전문지식, 역량 및 자원을 구비해야 한다.

독립적인 감독 기관은 인공지능 시스템의 인권 준수를 능동적으로 조사 및 모니터링하고 영향을 받은 개인들로부터 진정을 접수, 처리하며 인공지능 시스템의 성능과 기술 개발에 대한 정기적인 검토를 보다 일반적으로 실시해야 한다. 감독 기관은 인권 침해 발생의 위험을 포착한 상황에 개입할 수 있는 권한을 보유해야 한다. 감독 기관은 또한 정기적으로 의회에 보고하고 활동에 대한 보고서를 발표해야 한다.

공공기관과 민간 기업들은 감독 기관이 인공지능 시스템의 효과적인 감독에 필요한 모든 정보를 요청할 때 이를 제공해야 하고 감독 기관에 정기적으로 보고해야 한다. 또한 인공지능 시스템이 인권에 미친 영향과 관련해 감독 기관의 권고를 이행해야 한다. 더불어 감독 절차는 투명해야 하며 적절한 공공 조사의 대상이 될 수 있어야 하고 감독 기관의 결정은 이의제기 또는 독립 심사의 대상이 될 수 있어야 한다.

6. 차별금지 및 평등

모든 상황에서 인공지능으로 불균형한 영향을 받을 위험이 증가하는 집단에 대하여 각별한 주의를 기울여 차별 위험을 방지 및 완화하여야 한다. 여기에는 여성, 아동, 노인, 경제적으로 취약한 사람들, LGBTI 집단의 구성원, 장애인 및 ‘인종적’, 민족적 또는 종교적 집단이 포함된다. 회원국은 차별적이거나 차별적인 결과를 초래하는 인공지능 시스템의 사용을 억지하여야 하며, 관할권 내에서 제3국의 인공지능 시스템의 사용 결과로부터 개인을 보호하여야 한다.

인공지능 수명주기의 모든 단계에 이들 집단을 효과적으로 대표하는 다양한 공동체의 적극적인 참여와 유의미한 협의 과정을 포함하는 것은 부정적인 인권 영향을 방지하고

완화하는 데 중요한 구성요소이다. 또한, 인공지능 시스템 개발에 사용되는 학습 데이터에 대한 투명성과 정보 접근성에 각별한 주의를 기울여야 한다. 인권영향평가 및 기타 인권 실사 방법은 정기적으로 반복되어야 하며, 책무성 실현과 구제 조치를 위해 적절하고 접근 가능한 방법이 제공되어야 한다.

회원국은 특히 예측적 또는 예방적 치안과 같은 방법론이 관련되는 법집행 상황에서 인공지능 시스템을 사용할 때 최고 수준의 정밀 조사를 실시하여야 한다. 그러한 시스템은 특정 집단에 대하여 사실상의 프로파일링을 수행할 수 있는 차별적 효과에 대하여 배치 전에 독립적으로 감사가 이루어질 필요가 있다. 이러한 효과가 탐지되면 해당 시스템을 사용할 수 없다.

7. 개인정보보호 및 프라이버시

개인정보 처리에 의존하는 인공지능 시스템의 개발, 학습, 검사 및 사용은 유럽인권협약 제8조에 따라 개인정보와 관련한 '정보적 자기결정권'을 비롯하여 개인의 사생활 및 가정생활을 존중할 권리를 완전히 보장해야 한다.

인공지능 시스템 환경에서 이루어지는 개인정보 처리는 그러한 처리가 추구하는 정당한 목적에 비례해야 하며, 처리의 모든 단계에서 인공지능 시스템의 개발 및 배치로 추구하는 이익과 문제되는 권리와 자유 간에 공정한 균형을 반영해야 한다.

회원국은 회원국에 구속력이 있는 개인정보보호 및 프라이버시권에 관한 기타 국제법뿐만 아니라 개인정보의 자동화된 처리에 관한 유럽평의회 현대화 협약("108+호 협약")을 효과적으로 이행하여야 한다. 인공지능 시스템 수명주기의 모든 단계에서 이루어지는 개인정보의 처리는 108+호 협약에 명시된 원칙에 근거하여야 하며, 특히 (i) 관련 인공지능 시스템 수명주기 단계에서 개인정보를 처리하기 위해 법률에 의해 규정된 합법적인 근거가 있어야 하고, (ii) 개인정보가 합법적으로 처리되어야 하며, (iii) 개인정보는 명시적이고 구체적이며 합법적인 목적으로 수집되어야 하며 이러한 목적과 양립할 수 없는 방식으로 처리되어서는 안 되고, (iv) 개인정보는 처리 목적에 적절하고 관련이 있으며 과도하지 않아야 하며, (v) 개인정보는 정확해야 하고, 필요한 경우 최신 상태로 유지해야 하며, (vi) 개인정보는 해당 정보가 처리되는 목적에 필요한 기간을 넘어서지 않는

동안만 정보주체를 식별할 수 있는 형태로 보존되어야 한다.

회원국은 인공지능 시스템이 유전정보, 범죄·형사 절차·유죄 판결 및 관련 보안조치와 관련된 개인정보, 생체인식 정보, '인종적'·민족적 기원, 정치적 의견, 노동조합 가입여부, 종교적 또는 기타 신념, 건강 또는 성생활에 대한 개인정보의 처리에 의존하는 경우 적절한 안전장치를 규정하는 입법 체계를 도입하여야 한다. 그러한 안전장치는 또한 이들 개인정보가 차별적이거나 편향된 방식으로 처리되지 않도록 보호하여야 한다.

8. 표현의 자유, 집회 및 결사의 자유, 노동권

모든 개인의 인권과 기본적 자유를 존중, 보호 및 이행해야 할 책임의 맥락에서 회원국은 인공지능의 사용에 의해 관여될 수 있는 국제 인권 기준의 모든 범위를 고려해야 한다.

표현의 자유: 회원국은 다양하고 다원적인 정보 환경을 조성해야 할 의무와 인공지능의 콘텐츠 관리와 큐레이션이 표현의 자유, 정보에 대한 접근 및 의견의 자유 행사에 미칠 수 있는 부정적인 영향을 염두에 두어야 한다. 회원국은 또한 인공지능 전문지식과 권력의 집중이 정보의 자유로운 흐름에 부정적인 효과를 미치는 것을 방지하기 위하여 기술 독점을 규제하는 적절한 조치를 검토하는 것이 권장된다.

집회 및 결사의 자유: 특히 집회 및 결사의 자유를 오프라인에서 행사하기 어려운 상황에서 콘텐츠 관리를 위해 인공지능 시스템을 사용하는 것이 이들 권리에 미칠 수 있는 영향에 각별한 주의를 기울여야 한다. 얼굴인식 기술의 사용은 회원국이 엄격하게 규제하여야 하며, 그 사용을 명확히 한정하고, 효과적인 집회의 자유 행사를 보호하기 위하여 공공 투명성을 규정하는 법률을 입법하여야 한다.

노동권: 자동화를 가속화하여 노동 가용성에 부정적인 영향을 미칠 수 있는 인공지능의 잠재력을 주의 깊게 관찰하여야 한다. 인공지능 개발로 인해 창출되고 손실된 일자리의 수와 유형을 추적하기 위한 정기적인 평가가 이루어져야 한다. 인적 노동 수요 감소로 인해 분명히 영향을 받는 노동자들을 재교육하고 일자리를 재할당하기 위해 적절한 계획이 개발되어야 한다. 또한 회원국은 교육 커리큘럼을 조정하여 인공지능 시스템과 관련된 역량이 필요한 일자리에 대한 접근을 보장하여야 한다.

9. 구제 수단

인공지능 시스템은 기계 학습 또는 유사한 기술을 통해 인공지능 시스템이 인간의 특정한 개입과 독립적으로 의사결정을 내릴 수 있는 상황에서도 항상 인간의 통제 하에 있어야 한다. 회원국은 인공지능 시스템 수명주기의 다양한 단계에서 발생할 수 있는 인권 침해에 대하여 명확한 책임선을 수립해야 한다. 인공지능 시스템의 개발, 배치 또는 사용에서 발생하는 인권 침해에 대한 책임성과 책무성은 항상 자연인 또는 법률인에게 있어야 하며, 이는 심지어 인권을 침해한 해당 조치를 인간 책임자 또는 운영자가 직접 지시하지 않은 경우에도 해당한다.

공공 또는 민간 기관의 인공지능 시스템의 개발, 배치, 사용으로 인하여 인권 침해를 당했다고 주장하는 사람은 소관 국가기관으로부터 효과적인 구제 수단을 제공받아야 한다. 나아가 회원국은 투명하지 않은 방식 또는 인지하지 못한 상태에서 인공지능 시스템의 출력물이 오로지 또는 중대하게 지원하는 조치의 대상이 되었다고 의구심을 갖는 사람들에게 효과적인 구제 수단을 제공해야 한다.

효과적인 구제 수단이란 인공지능 시스템의 개발, 배치 또는 사용으로 인한 위해에 대하여 신속하고 적절한 배보상과 시정을 포함하여야 하며, 민사·행정적 또는 형사적 조치를 포함할 수 있다. 국가인권기구는 각자의 소관에 따라 자체적인 결정을 수행함으로써 그러한 시정조치의 근거가 될 수 있다.

회원국은 유의미한 인적 개입 없이 오로지 자동화된 의사결정에 기반하여 자신에게 중대한 영향을 미치는 의사결정의 대상이 되지 않을 권리를 개인에게 보장하여야 한다. 최소한 개인은 그러한 자동화된 의사결정에서 인적 개입을 구할 수 있어야 하며 그러한 의사결정이 실행되기 전에 자신의 견해에 대한 검토를 받을수 있어야 한다.

회원국은 인공지능 시스템으로 인해 인권 침해를 당했다는 주장의 입증과 관련이 있을 경우, 개인이 상대방 또는 제3자가 소유한 정보에 접근할 수 있는 권리를 보장하여야 한다. 이는 적절한 경우 관련 학습 또는 검사 데이터, 인공지능 시스템이 사용된 방법에 대한 정보, 인공지능 시스템이 특정 권고, 의사결정, 예측에 도달한 방법에 대한 유의미하고 이해가능한 정보, 인공지능 시스템의 출력물이 해석되고 작동하는 방식에 대한 세부사항을 포함한다.

국가 기관이 인공지능 시스템의 개발, 배치 또는 사용으로 인한 인권 침해 문제를 검토할 때에는 인공지능 시스템이 제시하는 '객관성의 유혹'에 대하여 적절한 거리를 두어야 하고 인권 침해에 도전하는 개인이 문제가 된 조치에 대하여 책임이 있는 측보다 더 높은 입증 기준을 요구받지 않도록 하여야 한다.

10. 인공지능 리터러시 증진

정부 기관, 독립적인 감독 기관, 국가인권기구, 사법부와 법 집행 기관은 물론 일반 대중에게 인공지능에 대한 지식과 이해를 촉진하여야 한다. 회원국은 정부 내에 모든 인공지능 관련 문제에 대해 협의하는 자문 기구 설립을 고려해야 한다.

인공지능 시스템의 개발이나 적용에 직간접적으로 관련된 사람들은 시스템이 어떻게 기능하는지에 대해 필요한 지식과 이해를 갖추어야 하고 그것이 인권에 미치는 영향에 대하여 알아야 한다. 그러한 행위자들이 그들의 시스템이 인권에 미치는 영향에 대하여 알기 위해서는, 이들 또한 그들의 업무에 관여될 수 있는 인권 기준의 범위를 인지할 수 있어야 한다.

회원국은 (특히) 학교를 비롯한 공간에서 강력한 인식 제고, 훈련 및 교육에 대한 노력으로 일반 대중의 인공지능 리터러시 수준에 투자하여야 한다. 이는 인공지능 작동에 대한 교육에만 한정되어서는 안 되며, 인공지능이 인권에 미치는(긍정적이고 부정적인) 잠재적 영향에도 투여되어야 한다. 소외된 집단과 일반적으로 IT 리터러시에서 취약한 사람들에게 접근하기 위해 특별한 노력을 기울여야 한다.

체크리스트

	해야 할 일
인권영향평가	- 공공기관이 구입, 개발 또는 배치하였거나 예정인 인공지능 시스템에 대하여 인권영향평가 실시를 요구하는 법률과 규제를 도입하는 조치를 취할 것 - 인권영향평가 관련 법률 체계가 도입되는 시점에 공공기관은 이미 배치하였거나 사용하고 있는 인공지능 시스템에 대하여 인권영향평가를

	즉시 수행할 것. 또는, 공공기관은 인공지능 시스템의 구입 및 개발 전에 인권영향평가를 먼저 수행해야 함 - 전체 수명주기에 걸쳐 인공지능 시스템이 인권에 미치는 영향을 지속적으로 모니터링하고, 수명주기의 새로운 단계 및 시스템의 환경, 범위, 특성 및 목적에 변화가 있을 때 정기적으로 인권영향평가를 수행할 것
	하지 말아야 할 일
	- 인권영향평가에 관한 법적 체계를 도입할 때 시민사회단체와 인공지능 및 인권에 관련된 전문가를 비롯한 관련 이해관계자와 유의미하게 협의하고 의견을 수렴하는 데 실패하지 말 것 - 인권영향평가를 투명하지 않은 방법으로 시행하지 말고, 인권영향평가의 수행이나 공개를 방해하기 위한 목적으로 기밀, 사생활, 영업 비밀, 국가 기밀 또는 지적 재산에 관한 법률을 사용하거나 사용을 유도하지 말 것 - (i) 인권영향평가의 적용을 받지 않았거나 (ii) 인권영향평가가 인공지능 시스템의 실제적인 인권 침해 위험을 나타냈음에도 식별된 위험을 방지하거나 완화하기 위한 조치, 안전장치 또는 방법을 채택하지 않은 상황에서 인권을 간섭할 잠재성이 있는 인공지능 시스템을 구입, 개발, 배치 및 사용하지 말 것
공개적인 의견 수렴	**해야 할 일**
	- 인공지능 시스템의 사용에 공개 조달 기준과 투명한 절차를 적용할 것 - 최소한 조달 및 인권영향평가 단계에서 공개적인 의견 수렴에 영향을 받는 집단 또는 공동체를 비롯한 모든 이해관계자를 포함할 것
	하지 말아야 할 일
	- 인공지능 시스템과 관련된 모든 관련 정보를 적시에 사전에 공개하고 모든 관련 이해관계자의 참여를 적극적으로 추진하는 등 유의미한 적절한 조치를 취하지 않고 의견 수렴을 실시하지 말 것
민간부문의 인권기준 이행을 촉진해야 하는 회원국 의무	**해야 할 일**
	- 인공지능 관련 인권 침해에 대해 인공지능 행위자가 책무를 지는 데 있어 간극 또는 장벽을 파악하기 위하여 기존 형법 및 민법은 물론 기타 동등한 법적 책임 체제에 대한 감사를 수행할 것 - 인공지능 행위자의 침해로부터 개인의 인권을 보호하기 위해 국가의 의무를 이행할 필요가 있는 경우 기존 법률을 시행할 것 - 인공지능 행위자가 인권 존중에 대한 책임을 다하고 있음을 '알고 보여' 주도록 조치를 취할 것. 여기에는 인공지능 시스템과 관련된 인권 위험을 식별하고 그러한 시스템이 초래하는 피해를 방지 및 완화하기

	위하여 효과적인 조치를 취하는 투명한 인권 실사 절차가 포함됨
	하지 말아야 할 일
	- 인공지능 부문에 적용되는 법률, 정책 및 규정을 회원국에 대한 인권 의무로부터 분리하거나 정보가 없는 것으로 취급하지 말 것 - 인공지능 부문에서 인권 기준의 시행과 집행을 차별적인 방식으로 실시하지 말 것
정보 및 투명성	**해야 할 일**
	- 특히 공공 서비스에서 인공지능 시스템이 언제 어떻게 사용되고 있는지 개인이 이해할 수 있도록 필요한 모든 정보를 제공할 것
	하지 말아야 할 일
	- 적절한 투명성 및 책무성 기준에 따라 사람이 검토하고 조사할 수 없을 정도로 복잡한 인공지능 시스템을 사용하지 말 것
독립적 감독	**해야 할 일**
	- 가능한 경우 국가인권기구를 포함한 기존 감독 기관을 활용하여 인공지능 시스템의 인권 준수에 대한 독립적이고 효과적인 감독을 위한 체계의 수립을 입법화할 것 - 모든 관련 감독 기관이 충분한 전문 지식에 접근하고, 인공지능 시스템 및 그 인권 영향에 대한 적절한 교육을 받으며, 기능을 효과적으로 수행하는 데 적정한 재정 및 기타 자원을 제공받도록 보장하기 위해 조치를 취할 것 - 인공지능 시스템의 인권 침해(개발, 검사 및 사용 중에 발생하는 행위 포함)에 책임이 있을 수 있는 모든 행위자를 공공기관 또는 민간기관 여부를 불문하고 조사 및 모니터링하려는 목적에 적정하도록 관련 감독 기관의 기능을 보장할 것
	하지 말아야 할 일
	- 인권 침해 발생(위험)을 식별하는 상황에서 유의미하게 개입할 수 없을 정도로 감독 기관의 기능과 권한을 제한하지 말 것 - 인공지능 시스템의 인권 준수에 대한 조사 및 모니터링을 담당하는 감독 기관의 제도적, 운영적, 재정적 및 개인적 독립성을 손상시키지 말 것 - (학습 및 검사용) 데이터셋, 인공지능 입력/출력물, 모델/알고리즘, 운영 지침 및 인권 실사에 대한 접근권 박탈을 비롯하여, 감독 기관이 효과적으로 기능을 수행하는 데 필요한 정보를 박탈하거나 제3자가 박탈하도록 허용하지 말 것
차별금지	**해야 할 일**

및 평등	- 인공지능 시스템으로 인해 그 권리가 불균형적으로 영향을 받을 위험이 높은 집단에 대하여 인공지능 시스템 사용의 차별 위험을 방지하고 완화할 것 - 특히 법 집행 상황에서 인공지능 시스템을 사용할 때 특정 집단에 속한 개인의 프로파일링을 방지하기 위하여 가장 높은 수준의 정밀 조사를 적용할 것
	하지 말아야 할 일
	- 차별적이거나 차별적인 결과를 초래하는 인공지능 시스템을 사용하거나 제3자가 사용하도록 허용하지 말 것
개인정보 보호 및 프라이버시	**해야 할 일**
	- 기존 개인정보보호법에 대한 검토 및 평가를 수행하여 인공지능 시스템 환경에서 사생활권과 개인정보보호권을 충분히 보호하는지 판단하고, 그렇지 않은 경우 법적 개혁을 실시할 것 - 인공지능 시스템의 개발, 배치 및 사용과 관련된 민간 및 공공 기관이 정보주체의 권리를 존중하고 해당 개인정보보호법에 따른 의무를 이행하도록 사전에 조치를 취할 것
	하지 말아야 할 일
	- 인공지능 시스템을 개발, 배치 또는 사용하는 사람들에 대하여 개인정보 처리에 관한 의무의 광범위하고 불균형적인 예외 또는 면제를 규정하지 말 것 - 그 학습 또는 검사에서 사생활권 및 개인정보보호권을 위반하여 수집되거나 처리된 데이터셋에 의존하는 인공지능 시스템의 개발 또는 사용을 허용하지 할 것 - 그 입력 또는 출력 데이터로 사생활권 및 개인정보보호권을 위반하는 개인정보를 처리하는 인공지능 시스템의 개발 또는 사용을 허용하지 말 것
표현의 자유, 집회 및 결사의 자유, 노동권	**해야 할 일**
	- 인공지능의 사용에 의해 잠재적으로 관련될 수 있는 국제 인권 기준의 모든 범위를 고려할 것 - 인공지능 중심의 콘텐츠 관리 및 큐레이션이 표현의 자유, 정보에 대한 접근 및 의견의 자유에 미칠 수 있는 영향을 유념할 것 - 집회의 자유를 효과적으로 행사할 수 있도록 얼굴 인식 기술의 사용을 엄격하게 규제할 것 - 인공지능이 노동권에 미칠 수 있는 부정적인 영향을 모니터링하고 학교 교육을 비롯하여 이를 완화할 계획을 수립할 것
	하지 말아야 할 일

	- 국제 인권 기준에 따라 보호되는 인권을 침해하는 인공지능 시스템의 개발, 배치 또는 사용을 허용하거나 촉진하지 말 것
구제 수단	**해야 할 일**
	- 민사, 형사 및 행정법을 포함한 기존 법률에 대한 평가를 수행하고 해당 법률이 인공지능 시스템의 개발, 배치 또는 사용으로 인해 발생한 인권 침해의 피해자임을 주장하는 사람들에게 효과적인 구제수단을 제공하지 않는 경우 개정을 추진할 것 - 인공지능 시스템의 수명 주기 각 단계에서 발생할 수 있는 모든 범위의 인권 침해에 대해 법적 책임이 있는 사람을 명확하게 규정하도록 책임 체제를 보장할 것 - 사법부 및 기타 관련 국가 기관이 인공지능 시스템의 가정/인식된 정확성 또는 객관성에 부적절한 비중을 부여하지 않고, 인공지능 시스템으로 인한 인권 침해에 문제를 제기하는 피해자와 대상자 간에 동등한 힘의 제공을 보장할 것
	하지 말아야 할 일
	- 전적으로 인간의 통제 범위 밖에서 작동하는 인공지능 시스템의 개발, 배포 또는 사용을 허용하지 말 것 - 유의미한 인적 개입을 구할 기회가 제공되지 않고 의사결정이 실행되기 전에 자신의 의견이 검토되지 않은 상황에서 개인이 자신에게 중대한 영향을 미치는 자동화된 의사결정의 대상이 되도록 허용하지 말 것
인공지능 리터러시 증진	**해야 할 일**
	- 인공지능 관련 문제에 대해 협의하는 정부 자문 기구를 설치할 것 - 인공지능 시스템 개발 관계자부터 일반 대중에 이르기까지 모든 사람의 인공지능 및 인권에 대한 지식과 이해를 증진할 것
	하지 말아야 할 일
	- 인공지능 리터러시 활동에서 인권에 대한 잠재적 영향을 포함하지 않고 기술적 측면으로 한정하지 말 것

부록 Ⅳ.

유럽평의회
<알고리즘 시스템의 인권 영향에 대한 대응 지침>[5)]

제1373차 각료 대표단 회기 2020년 4월 8일 각료위원회 채택

권고 CM/Rec(2020)1 부록

A. 범위 및 맥락

1. 이 지침은 알고리즘 시스템의 설계, 개발 및 진행 중인 배치에 대한 모든 조치에서 국가, 공공 및 민간 부문 행위자에 대해 조언하기 위해 고안되었다. 유럽인권협약(“협약”) 및 기타 관련 조약에 명시된 바와 같이 모든 개인의 인권과 기본적 자유가 기술 발전 전반에 걸쳐 효과적으로 보호되도록 보장하기 위하여, 유럽 평의회 회원국은 알고리즘 시스템 사용에서 인권 침해를 억지하고, 모든 행위자로 하여금 인권을 존중 및 증진하고 침해 가능성을 방지하게끔 노력하도록 환경을 조성하는 입법 및 규제 체제를 개발해야 한다. 국가의 의무 및 관할권과 별개로, 공공 및 민간 부문 행위자는 국제적으로 인정된 인권을 존중할 책임이 있다.

2. 이 권고의 목적 하에서, ‘알고리즘 시스템’이란 종종 수학적 최적화 기술을 사용하여 데이터 수집, 결합, 정리, 정렬, 분류 및 추론 뿐 아니라 선택, 우선 순위, 권장 사항 및 의사 결정과 같은 업무를 하나 이상 수행하는 응용프로그램으로 이해된다. 적용되는 설정에서 요구 사항을 충족하기 위해 하나 이상의 알고리즘에 의존하는 알고리즘 시스템은 대규모 및 실시간으로 적응형 서비스를 생성하는 방식으로 업무를 자동화한다.

---

5) Guidelines on addressing the human rights impacts of algorithmic systems. Appendix to Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems. https://rm.coe.int/09000016809e1154.

3. 일반적으로 대규모 데이터셋에서 패턴을 감지하여 작동하는 알고리즘 시스템은 (특히 정밀도, 타겟팅 및 일관성 향상을 통해) 서비스 성능을 개선하고 새로운 솔루션을 제공하며 업무 및 시스템 성능의 효율성 및 효과성 측면에서 성과를 낼 가능성을 제공한다. 이는 디지털 정보의 분류 및 검색 가능성을 크게 향상시켰고 의료 진단, 운송 및 물류와 같은 분야에서 중요한 발전을 촉진하여 전 세계적으로 더 광범위하고 빠른 정보 공유를 가능하게 하고 새로운 형태의 협업 및 협력을 가능하게 했다. 그 결과 알고리즘 시스템은 현대 인간 생활의 많은 측면에 스며들었다.

4. 그러나 일상 생활에서 알고리즘 시스템에 대한 의존도가 높아짐에 따라 공정한 재판을 받을 권리, 프라이버시권 및 개인정보에 대한 권리, 사상·양심 및 종교의 자유, 의사 표현의 자유, 집회의 자유, 평등권, 경제적·사회적 권리 등에서 중대한 인권 문제도 제기되고 있다. 알고리즘 시스템의 기능은 종종 온라인 및 오프라인에서 개인 및 집단의 신원 및 행동에 대한 대규모 디지털 추적으로 수집된 데이터의 체계적인 집계 및 분석을 기반으로 한다. 대규모 추적은 개인의 사생활 침해와 고도로 개인화된 조작 가능성의 증가 외에도 알고리즘 시스템의 제안 단계부터 이후 전체 수명 주기에 걸쳐 고려되어야 하는 인권의 행사에 심각한 악영향을 미칠 수 있다.

5. 이러한 손실이 합리화와 정확성의 향상으로 상쇄된다는 주장이 자주 제기되지만 대부분의 알고리즘 시스템은 오류가 불가피한 통계 모델을 기반으로 하며 때로는 기존의 편향, 오류 및 가정을 유지, 복제 및 강화하는 피드백 순환구조를 가지고 있다는 사실에 주목해야 한다. 더 큰 데이터셋이 반복적인 패턴과 상관 관계를 찾을 수 있는 더 나은 기회를 제공하는 것처럼 보일 수 있지만 정확도는 데이터셋의 크기에 따라 자동으로 증가하지 않는다. 알고리즘 시스템이 많은 사람들에게 영향을 미친 결과, 위양성 및 위음성 형태의 오류와 이들 오류 및 내재된 편향의 영향을 받는 사람들의 수도 확대되어 인권 행사에 다양한 방식의 추가 간섭이 유발될 것이다.

6. 알고리즘 시스템은 개인정보에만 기반하여 결과물을 처리 및 생성하지 않는다. 알고리즘 시스템은 시뮬레이션, 합성 데이터 또는 일반화된 규칙이나 절차 등 비관찰 및 비개인정보를 기반으로도 작동할 수 있다. 그러나 그러한 사용 또한 인권에 부정적인 영향을 미칠 수 있다. 특히 알고리즘 시스템을 사용하여 의사 결정을 지원하고 권장 사항을 적용하거나 물리적 환경을 형성할 때 해당 데이터가 처리되지 않거나 달리 고려되지

않은 개인 및 집단은 직접적으로 관련 되고 상당한 영향을 받을 수 있다.

7. 많은 알고리즘 시스템은 개발 및 구현 단계가 밀접하게 얽혀 있는 최적화 기술을 사용한다. 알고리즘 시스템은 미리 정의된 결과물의 좁은 범위에 기반한 결과를 달성하기 위해 사용시마다 그 기능이 조정될 수 있다. 이러한 프로세스는 특히 대규모로 운영될 때 환경을 형성하거나 방해할 수 있다. 프로세스는 특정 가치를 다른 것보다 우선시하며, 예를 들어 일반적인 이익을 특정 손실보다 우선시할 수 있다. 이러한 작동은 일반적으로 명시적이고, 투명하며, 책임성 있고, 영향을 받는 개인이 통제할 수 있는 방식으로 이루어지지 않으며, 특히 소수자 및 소외집단 또는 취약집단에 부정적인 영향을 미칠 수 있다.

8. 일상 생활에서 알고리즘 시스템의 광범위한 유형과 적용을 감안할 때 알고리즘 시스템이 인권에 미치는 긍정적 또는 부정적 영향의 수준은 항상 알고리즘이 사용되는 특정 목적, 기능, 정확성, 복잡성, 효과 및 배치 규모에 따른다. 알고리즘 시스템의 영향은 또한 그것들이 사용되는 보다 광범위한 조직적, 주제별, 사회적, 법적 맥락에 따라 달라지며, 그 각각은 특정 공공 및 윤리적 가치와 연관된다. 이메일 스팸 필터, 건강 관련 데이터 분석 또는 교통 흐름 합리화 등 응용 프로그램의 쓰임새는 매우 다양하다. 또한 알고리즘 시스템은 경찰 및 국경 통제의 맥락에서 예측 목적으로, 자금 세탁 및 사기를 방지하기 위한 목적으로, 또는 노동, 고용, 교육 시설의 공적·사적 채용과 선발 절차 등에서 사용된다.

9. 알고리즘 시스템의 설계, 개발 및 진행 중인 배치로 인해 인권에 부정적일 수 있는 영향을 평가할 때 지속적으로 평가하고 시스템의 사용 맥락, 법적 근거, 목적, 정확성, 부작용 및 규모를 문서화해야 한다. 처리되는 데이터의 규모, 특성 및 미래 가치로 인해 이러한 시스템이 적대적 머신 러닝 또는 기타 수단(사이버 공격 등)을 통해 공격을 받거나 혼동을 일으킬 수 있는 내재적 위험도 고려해야 한다. 알고리즘 시스템이 인권에 미칠 수 있는 영향의 정도를 평가할 때는 인권 침해를 야기할 수 있는 심각성, 규모 및 가능성을 고려해야 한다.

10. 알고리즘 시스템이 민주적 절차나 법치에 대한 영향을 비롯하여 개인, 특정 집단 또는 인구집단 일반에 대해 부정적인 인권 영향을 생성할 가능성이 있는 경우 이러한 영향은 인권과 관련한 국가의 의무 및 민간 부문의 책임을 불러온다.

11. 국가에서 공공 서비스 또는 공공 정책 전달을 위해 알고리즘 시스템을 사용하면 개인이 여기서 탈퇴(opt out)할 가능성이 없거나 탈퇴 결정의 결과로 부정적인 결과를 겪게 되기 때문에, 경우에 따라 알고리즘 시스템의 적용은 인권에 더 특별하고 높은 위험을 유발할 수 있다. 공공 기관이나 민간 당사자가 특별한 비중이나 법적 결과를 수반하는 의사 결정 과정에서 알고리즘 시스템을 사용하면 유사하게 높은 위험이 발생한다. 예를 들어, 법적 분석, 예측 또는 개인별 위험 평가 목적으로 사법 분야에서 알고리즘 시스템을 사용하는 것은 협약 제6조에 명시된 공정한 재판을 받을 권리에 따라 세심한 주의를 기울여 도입되어야 한다. 이 권고에서 '고위험'이라는 용어는 개인에게 심각한 결과를 초래할 수 있는 절차 또는 결정에서 알고리즘 시스템의 사용이 언급되거나 혹은 대안이 없을 경우 분배 불공정을 유발하거나 확대하는 등 인권 침해 가능성이 특히 높은 상황에 적용된다.

12. 인권에 부정적일 수 있는 영향을 평가할 때 명확하게 공적이거나 명확하게 민간에 속하지 않은 알고리즘 시스템이 광범위하게 사용되는 (책임 할당의 문제로 귀결되는) 상황에 특히 주의를 기울여야 한다. 이는 공공 서비스의 일부가 민간 부문 제공자를 아웃소싱하여 다른 서비스 제공자에게 의존하게 되는 경우이거나, 공공 기관이 민간 부문에서 알고리즘 시스템 및 서비스를 조달하는 경우, 또는 회사가 국가에서 정의한 공공 정책 목표를 달성하기 위해 알고리즘 시스템을 배치하는 경우일 수 있다.

13. 교통이나 통신에서처럼 전통적으로 공공 기관이 수행하는 기능이 민간 당사자의 알고리즘 시스템 제공에 전체 또는 부분적으로 의존하게 되는 경우도 역시 복잡하다. 이들 시스템이 그후 상업적인 이유로 철수될 때 그 결과는 품질 및 효율성의 감소로부터 개인 및 공동체가 필수적으로 간주하는 서비스의 손실에 이르기까지 다양할 수 있다. 특히 민간 부문 행위자가 영향력을 행사하거나 통제할 수 있는 지위를 유지하는 방식으로 시장을 지배하는 상황에서 국가는 필수 서비스가 상업적 생존 가능성에 관계없이 계속 이용 가능하도록 보장하기 위해 비상 사태에 대비해야 한다.

14. 알고리즘 시스템의 설계, 개발 및 진행 중인 배치에는 소프트웨어 설계자, 프로그래머, 데이터 소스, 데이터 작업자, 소유자, 판매자, 사용자 또는 고객, 인프라 제공자, 공공 및 민간 행위자와 기관을 비롯한 많은 행위자가 관여한다. 또한 많은 알고리즘 시스템은 학습 방식이든 비학습 방식이든 상당한 수준으로 때로는 고의적으로 불투명하게 작

동한다. 입력 데이터, 최적화 대상 및 모델을 비롯하여 시스템의 가장 중요한 목표와 매개변수를 통상적으로 설정하는 설계자 또는 운영자조차도 시스템이 결정을 내리기 위해 어떤 정보에 의존하는지 알지 못할 가능성이 높으며, 시스템이 작동하도록 의도된 것보다 더 넓은 환경에서 시스템이 사용자에게 미치는 직간접적 영향에 대해 확실히 알지 못할 수 있다.

15. 이러한 복잡성을 감안할 때 회원국은 이러한 프로세스의 구체적인 인권 영향을 인식하고 이들 시스템에 대한 모든 투자에 있어 효과적인 모니터링, 평가, 검토 절차 및 부작용에 대한 시정 조치는 물론, 필요한 경우 인권 기준을 충족하지 못하는 프로세스를 포기할 수 있는 적절한 비상 대책을 포함하여야 한다. 위험 관리 절차는 알고리즘 시스템의 유해한 사용과 그 부정적 영향을 탐지하고 방지해야 한다. 회원국은 사전 예방적 접근법을 취해야 하고, 특정 시스템의 배치가 돌이킬 수 없는 손상의 높은 위험을 초래하거나 그 불투명성으로 인해 인간의 통제 및 감독이 불가능할 때는 이 시스템을 거부하도록 요구해야 한다.

B. 알고리즘 시스템의 맥락에서 인권과 기본적 자유의 보호 및 증진에 관한 국가의 의무

1. 일반 적용 원칙

1.1 입법: 알고리즘 시스템의 설계, 개발 및 진행 중인 배치에 적용할 수 있는 정책과 법률 또는 규정의 초안, 시행 및 평가 절차는 투명하고 책임성 있고 포용적이어야 한다. 국가는 적절한 경우 부문별 수준을 비롯하여 모든 관련 이해관계자 및 영향을 받는 당사자와 정기적으로 협의해야 한다. 국가는 관련 행위자의 법적 준수 여부를 확인하기 위해 적절한 문서 제출을 요구하는 등 법의 집행 가능성 및 집행을 보장해야 한다. 공공 및 민간 부문 행위자가 그 법적 의무를 이행하지 않는 경우 책임을 져야 한다.

1.2. 지속적인 검토: 제안 단계로부터 영향평가에 이르기까지 알고리즘 시스템의 전체 수명 주기 동안 개별 시스템의 인권 영향 및 다른 기술과의 상호 작용을 정기적으로 평가해야 한다. 이러한 평가는 이들 시스템이 작동하는 속도와 규모, 그리고 시스템이 작동하는 기술 환경의 빠른 진화 때문에 필요하다. 평가는 영향을 받거나 영향을 받을 가

능성이 있는 사람들과의 광범위하고 효과적인 협의를 기반으로 수행되어야 한다.

1.3 민주적 참여 및 인식 제고: 인권과 민주적 자유의 완전한 행사를 보장하기 위해 국가는 알고리즘 시스템의 기능, 권한 및 결과적 영향에 대한 일반 대중의 인식을 촉진해야 한다. 이는 자원을 조작, 착취, 기만 또는 배포할 가능성이 있는 사용을 비롯하여 모든 개인과 집단이 자신의 권리를 인식하고 이를 실행하는 방법과 자신의 이익을 위해 디지털 기술을 사용하는 방법을 알 수 있도록 해야 한다. 또한, 알고리즘 시스템을 고려중이거나 사용 중인 공공, 민간 및 시민 사회 부문을 비롯한 모든 관련 행위자가 알고리즘 시스템을 유능하고 비판적으로 고려하고 사용할 수 있도록 미디어, 디지털 및 정보 리터러시의 수준을 (예를 들어 연령, 성별, 인종, 민족, 문화 또는 사회경제적 배경과 관련된 다양성을 고려하여) 알맞고 포용적인 방식으로 촉진, 장려 및 지원해야 한다.

1.4 제도적 체제: 알고리즘 시스템의 설계, 개발 및 진행 중인 배치를 인권과 양립시키기 위하여, 국가는 일반 또는 부문별 기준(benchmarks)과 안전장치를 수립하는 적절한 제도 및 규제 체제와 표준을 식별하고 개발해야 한다. 이러한 노력들은 개별 시스템의 누적적 효과성을 비롯하여 인권에 대한 직간접적 위험을 즉시 식별하고 적절한 시정 조치를 개시할 수 있도록 해야 한다. 국가는 적절한 자원을 갖춘 규제 및 감독 당국에서 사용할 수 있도록 관련 전문 지식에 투자해야 한다. 국가는 독립적 감독 기관, 평등 기구, 국가 인권 기구, 대학, 표준 수립 기구, 서비스 운영자, 알고리즘 시스템 개발자 및 특히 인권 옹호 등 다양한 분야의 관련 비정부 기구와 긴밀하게 협력해야 한다.

2. 데이터 관리

2.1 정보적 자기결정권: 국가는 알고리즘 시스템의 모든 설계, 개발 및 진행 중인 배치에 있어 개인이 자기 관련 개인정보 처리에 대한 정보를 (그 목적과 가능한 결과를 비롯하여) 사전에 얻고 상호 운용성을 비롯하여 자신의 개인정보를 통제할 수 있는 방안을 제공받을 수 있도록 보장하여야 한다. 자동화 및 기타 기계적 인식 형태나 조작에 대해 난독화를 비롯하여 자기 자신, 그 물리적 환경 또는 자신의 활동을 판독하기 어렵게 만들기 위한 개인 또는 집단의 의도적인 노력은 유효한 정보적 자기결정권의 행사로 인식되어야 하며, 이는 민주사회에서 필요하고 법률에 의해 규정될 수 있는 제한을 따라야 한다.

2.2 데이터셋: 국가는 직접 또는 자신을 위해 설계, 개발, 진행 중인 배치 및 조달되는 알고리즘 시스템에서 입력되거나 추출되는 데이터 품질의 결과로 어떤 인권 및 차별금지 원칙이 영향을 받을 수 있는지 신중하게 평가해야 한다. 데이터가 종종 편향을 포함하고 성별, 인종, 종교, 정치적 의견 또는 사회적 출신과 같은 분류 기준의 대리변수 역할을 할 수 있기 때문이다. 데이터셋의 출처와 결함 가능성, 부적절하거나 탈맥락적 사용 가능성, 이러한 결함과 부적절한 사용으로 인한 부정적인 외부 효과, 데이터셋이 사용되거나 사용될 수 있는 환경도 신중하게 평가되어야 한다. 이전에 익명 또는 가명 기반으로 처리된 데이터를 사용하여 개인을 식별할 수 있는 가능성, 자동화된 수단으로 새롭고, 추론적이고 민감할 수 있는 데이터 및 분류 형식의 생성과 같은 내재적 위험에 특별한 주의를 기울여야 한다. 이러한 평가를 바탕으로 국가는 부작용을 방지하고 효과적으로 최소화하는 적절한 조치를 취해야 한다.

2.3 인프라 시설: 데이터 집중화와 데이터 처리 용량의 증가(클라우드 처리 포함) 및 인프라 시설에 대한 선택성 부족은 협약에 따른 인권 의무를 이행하는 국가의 역량에 부정적인 영향을 미칠 수 있다. 따라서 국가는 공공 및 민간 행위자 모두가 고품질 개인정보 처리와 컴퓨팅 기능을 사용할 수 있도록 대안적이고 안전하고 보안이 보장되는 인프라 시설을 개발해야 한다.

3. 분석 및 모델링

3.1 컴퓨팅 실험: 국가는 중대한 인권 영향을 유발할 가능성이 있는 컴퓨팅 실험은 인권영향평가 후에만 수행하도록 보장해야 한다. 참여 개인의 자유롭고 구체적이며 설명에 기반한 명확한 동의가 사전에 획득되어야 하며, 동의를 철회할 수 있는 접근 가능한 수단도 함께 제시되어야 한다. 기만적이거나 착취적인 효과를 위해 고안된 실험은 명시적으로 금지되어야 한다.

3.2. 안전장치 포함: 개인과 사회에 대한 인권 침해와 기타 부정적인 영향의 위험을 방지 및 완화하기 위하여, 국가는 알고리즘 설계, 개발 및 계속 중인 배치 절차가 설계에서부터 안전, 프라이버시, 개인정보보호, 보안 안전장치를 포함하도록 보장해야 한다. 데이터셋 및 모델의 출처와 품질을 보장하기 위해 지역과 국제 표준에 기반한 인증 체계를 설계하고 적용해야 한다. 그러한 안전장치는 조달 절차의 일부를 구성해야 하며 알고

리즘 시스템의 특정 사용을 금지하는 규제 체제에 의해 알려지고 준수되어야 한다.

3.3 검사: 완전성, 관련성, 프라이버시, 개인정보보호, 기타 인권, 부당한 차별적 영향, 보안 유출과 관련된 최신 표준에 입각하여 알고리즘 시스템의 생산 및 배치 전과 후에 정기적으로 검사, 평가, 보고, 감사하는 것이 검사 활동의 필수적인 부분을 구성해야 하며, 특히 자동화 시스템이 실시간 환경에서 검사되고 실시간 효과를 생성하는 경우 그러해야 한다. 국가는 시스템이 달성하거나 최적화하려는 목표의 합법성과 정당성, 그리고 인권과 관련하여 미칠 수 있는 영향에 대하여 공개적이고 협의적이며 독립적으로 평가를 수행하기 위하여 노력하여야 한다. 또한 그러한 평가가 조달 절차의 일부를 구성해야 한다. 그러한 시스템 검사 중 인권에 대한 중대한 제한이 식별될 경우 즉각적인 시정이 이루어져야 하며, 그렇지 않은 경우 시정이 이루어질 때까지 시스템이 정지되어야 한다.

3.4 데이터셋 및 시스템의 외부 효과 평가: 편향적이고 차별적인 결과물을 비롯한 알고리즘 시스템의 결과물은 그 배치의 특정한 맥락과 시스템을 학습시키는데 사용되는 데이터셋의 크기 및 특성에 따라 다양할 수 있다. 국가는 이 같은 사실을 충분히 고려하여 알고리즘 시스템이 구현하는 기능이 테스트되고 평가되도록 보장해야 한다. 알고리즘 시스템이 인권에 미칠 수 있는 영향에 따라 검사는 가능한 한 실제 개인정보를 사용하지 않고 수행되어야 하며, 제안된 시스템이 인구집단 및 그 환경에 미치는 외부 효과를 충분히 고려하여 배치 전과 후에 다양하고 대표적인 이해관계자 절차를 포함하는 검사가 이루어져야 한다. 국가는 또한 시스템이 본래 개발된 것과 다른 맥락에서 다른 알고리즘 시스템의 개발에 사용되는 등 검사 표본 또는 결과물이 재사용될 가능성과 위험을 인식해야 한다. 이러한 사용은 새로이 검사되어야 하고 그러한 사용의 적절성에 대한 평가 없이 허용되어서는 안 된다.

3.5 개인정보에 대한 검사: 국가는 개인정보에 대한 알고리즘 시스템의 평가 및 검사가 다양하고 충분히 대표적인 표본 집단으로 수행되도록 보장해야 한다. 관련 인구통계학적 집단은 과대 대표되거나 과소 대표되어서는 안 된다. 또한 고의적이거나 의도하지 않은 편향을 방지하기 위하여 국가는 그러한 활동에 관련된 직원이 충분히 다양한 배경을 갖도록 보장해야 한다. 또한 검사와 배치에 특정 개인, 집단, 인구 및 그 환경에 대한 위험 또는 손실이 표면화하는 경우 알고리즘 시스템의 개발 중단을 보장해야 한다. 관련 입법 체제는 그러한 표출을 억지해야 한다. 실제 환경 하의 검사와 관련해서는 특별한

주의를 기울여야 한다.

3.6 대안적 및 병렬적 접근법: 공공 서비스 전달에서 알고리즘 시스템을 사용하거나 기타 고위험 맥락에서 국가가 이러한 기술을 사용하는 경우, 다른 옵션과 비교하여 알고리즘 시스템을 평가하고 그 수행과 결과물이 적절한지 검사하기 위해 대안적 및 병렬적 모델링과 같은 방법론이 수행되어야 한다.

4. 투명성, 책무성 및 효과적인 구제 수단

4.1 투명성 수준: 국가가 직접 구현하거나 민간 부문 행위자들이 국가를 위해 구현하는 알고리즘 시스템의 공공 조달, 사용, 설계에 관련하여 적절한 수준의 투명성과 기본적인 처리의 기준 및 방법이 설정되어야 한다. 지적 재산 또는 영업 비밀에 대한 입법 체제는 이러한 투명성을 배제해서는 안 되며, 국가 또는 민간 당사자가 그러한 목적을 위해 법률을 악용해서는 안 된다. 투명성 수준은 가능한 한 높아야 하며 부정적인 인권 영향의 심각성에 비례해야 하고, 사용자가 시스템 간에 탐색할 수 있도록 알고리즘 시스템에 대한 윤리 인증 또는 인장을 포함해야 한다. 의사 결정 절차에서 알고리즘 시스템을 사용하여 인권에 대한 높은 위험을 수반하는 경우 과정과 결과의 설명 가능성과 관련하여 특히 높은 기준을 따라야 한다.

4.2. 알고리즘 의사결정의 식별 가능성: 공적 영역이든 사적 영역이든 관계없이 알고리즘 시스템에 의해 수행되거나 지원되는 모든 선택 절차 또는 결정은 인권 행사에 중대한 영향을 미칠 수 있기 때문에, 국가는 알고리즘 선택 또는 결정에 대하여 초기 상호작용에서 명확하고 접근 가능한 방식으로 식별 가능하고 추적 가능하도록 보장해야 한다.

4.3 이의제기 가능성: 알고리즘 관련 결정과 판단으로 영향을 받는 개인 및 집단은 이의를 제기할 수 있는 효과적인 수단을 제공받아야 한다. 알고리즘 시스템의 존재, 프로세스, 근거, 추론 및 개인 및 집단 수준에 대해 미치는 결과가 관련 공공 기관은 물론 그 권리 또는 정당한 이익이 침해될 수 있는 개인에게 시의적절하고 공정하며 쉽게 읽을 수 있고 접근 가능한 방식으로 설명되고 명시되어야 한다는 것이 필수적인 전제 조건이다. 이의제기에는 청문 기회, 결정에 대한 철저한 검토와 자동화되지 않은 결정을 구할 수 있는 방안이 포함되어야 한다. 이 권리는 포기될 수 없으며 쉽게 접근할 수 있는 연락처와 핫라인을 제공받는 등 알고리즘 시스템의 배치 전, 배치 중 및 배치 후에

저렴하고 손쉬운 실현이 보장되어야 한다.

4.4 협의 및 적절한 감독: 국가는 직접 또는 자신을 위해 직간접적으로 구현하는 특정 알고리즘 시스템에 대하여 영향을 받는 개인 또는 집단이 제기하는 이의제기의 수량과 유형에 대응하는 적절한 자원을 갖춘 독립 기관이 적정한 감독을 수행하도록 보장해야 한다. 국가는 특정한 경우에 결과물에 대한 시정 조치가 이루어지도록 보장해야 할 뿐 아니라 문제적 결과가 반복되지 않도록 시스템 자체에 시정 조치를 반영하여야 하며, 이를 개선하거나, 인권에 부정적 영향을 미칠 가능성이 있는 경우 특정 시스템의 도입 또는 진행 중인 배치를 중단할 수도 있어야 한다. 이러한 이의제기와 그 후속 조치에 대한 정보는 정기적으로 문서화되고 공개적으로 접근할 수 있어야 한다.

4.5 효과적인 구제 수단: 협약 제6조, 제13조 및 제14조에 따라, 국가는 공공 또는 민간 부문 행위자의 알고리즘 시스템의 사용으로 협약상 권리가 침해당했다고 주장하는 모든 진정에 대하여 공정하게 검토하기 위하여 평등하고 접근 가능하며 저렴하고 독립적이며 효과적인 사법 및 비사법 절차를 보장해야 한다. 국가는 입법 체제를 통해 개인과 집단이 자신의 고충과 관련하여 효과적이고 신속하며 투명하고 기능적이며 효과적인 구제 수단에 접근할 수 있도록 보장해야 한다. 내부적 및 대안적 분쟁 해결 메커니즘이 불충분하거나 영향을 받는 당사자 중 하나가 사법적 심사 또는 항고를 선택하는 경우 사법적 구제가 사용 가능하고 접근 가능한 상태로 남아 있어야 한다.

4.6 장벽: 국가는 직간접적으로 영향을 받는 개인 및 집단이 효과적인 고충 구제 수단을 거부당하는 결과를 낳는 모든 법적, 실무적 또는 기타 관련 장벽을 감소시키기 위해 적극적으로 노력하여야 한다. 여기에는 적절하게 훈련된 직원이 사례를 유능하게 검토하고 적절한 조치를 효과적으로 취할 수 있도록 보장해야 할 필요성이 포함된다.

5. 예방적 조치

5.1 표준: 국가는 인권영향평가의 최신 절차에 관한 적절한 지침(예: 표준, 체계, 지표 및 방법)을 개발하고 구현하기 위하여, 시민 사회를 비롯한 모든 관련 이해관계자와 서로 협력해야 한다. 이러한 절차는 중대한 인권 영향을 미칠 수 있는 모든 알고리즘 시스템에 대하여 그 수명 주기의 모든 단계에서 잠재적 위험을 평가하고 그러한 위험을 예방하거나 완화하기 위한 조치, 보호장치 및 메커니즘을 수립하기 위해 수행되어야 한다.

특히 그러한 시스템이 표적이 아닌 탐색 목적으로 적용될 때 실제 피해를 추적해야 한다. 인권영향평가는 이러한 권리에 높은 위험을 수반하는 모든 알고리즘 시스템에 대해 의무화되어야 한다.

5.2 인권영향평가: 국가는 국가뿐 아니라 국가와 협력하거나 국가를 대신하여 일하는 모든 민간 행위자가 공공 조달 이전, 개발 기간 중, 정규 일정 및 상황별 배치 전반에 걸쳐, 반인권적 결과물의 위험을 식별하기 위하여 정기적이고 협의적으로 인권영향평가를 수행하도록 해야 한다. 기밀 유지 고려 사항 또는 영업 비밀이 효과적인 인권영향평가의 실행을 방해해서는 안 된다. 민간 부문 행위자가 알고리즘 시스템에 의존하는 서비스를 제공하고 그 서비스가 인권의 효과적인 향유를 위해 현대 사회에서 필수적인 것으로 간주되는 경우, 회원국은 대안 솔루션의 장래 실행 가능성을 보존하고 영향을 받는 개인 및 집단이 이러한 서비스에 지속적으로 접근할 수 있도록 보장해야 한다. 인권에 대하여 고위험을 수반하는 알고리즘 시스템의 경우, 영향평가에는 이러한 시스템이 기존의 사회적, 제도적 또는 거버넌스 구조를 변경할 수 있는 가능성에 대한 평가가 포함되어야 하며, 인권에 대한 고위험을 방지하거나 완화하는 방법에 대한 명확한 권장 사항을 포함해야 한다.

5.3 전문성 및 감독: 국가는 고위험 알고리즘 시스템과 관련된 모든 인권영향평가가 독립적인 전문가 검토 및 검사를 위해 제출될 수 있도록 보장해야 한다. 독립적인 감독에 필요한 경우 계층화된 프로세스를 식별하거나 생성해야 한다. 국가에 의해 또는 국가를 위해 수행되는 인권영향평가는 공개적으로 접근할 수 있어야 하고, 적절한 전문가 투여가 있어야 하며, 효과적으로 후속 조치가 취해져야 한다. 후속 조치로는 동적 검사 방법 및 출시 전 시험을 수행하고, 영향을 받을 수 있는 개인과 집단 및 관련 현장 전문가와 협의하며, 적절한 경우 설계, 검사와 검토 단계에서 이들을 실질적인 의사결정권을 가진 행위자로 포함할 수 있다.

5.4. 후속 조치: 인권영향평가가 완화할 수 없는 중대한 인권 위험을 식별하는 상황에서는 공공 기관이 해당 알고리즘 시스템을 구현하거나 사용해서는 안 된다. 이미 배치된 알고리즘 시스템과 관련하여 위험이 식별되면 최소한 위험 완화를 위한 적절한 조치가 취해질 때까지 시스템 구현을 중단해야 한다. 식별된 인권 침해는 즉시 해결 및 구제되어야 하며 추가 침해를 방지하기 위한 조치가 취해져야 한다.

5.5 인간의 관리: 국가는 인권에 중대한 영향을 미치는 알고리즘 시스템의 조달, 개발, 구현, 평가, 검토에 관여하는 모든 관련 직원이 해당되는 인권 및 차별금지 원칙에 대하여 적절하게 훈련되고 철저한 기술적 검토뿐 아니라 인권을 준수해야 하는 자신의 의무를 인식할 수 있도록 보장해야 한다. 시스템 검토 과정에서 다양한 관점을 고려하는 능력을 향상시키기 위해서는 고용 관행이 양성평등과 다양한 인력 확충을 목표로 해야 한다. 공공 부문을 넘어 이러한 접근법을 촉진하기 위해 문서화해야 한다. 또한 국가들은 경험을 공유하고 모범 사례를 발굴하기 위해 협력해야 한다.

5.6 시스템의 상호 작용: 국가는 부정적인 외부 효과를 식별하고 방지하기 위해 여러 알고리즘 시스템이 동일한 환경에서 작동하는 설정을 주의 깊게 모니터링해야 한다. 시스템의 상호 의존과 상호 작용 가능성으로 인해 예방적 접근법이 필요한 경우에 특히 그렇다. 공공 서비스 전달에서 민간 서비스의 조달 또는 참여 메커니즘을 활용할 때, 국가는 알고리즘 시스템의 사용과 상호 작용에 대한 감독, 노하우, 소유권, 통제를 유지해야 할 필요성을 충분히 고려해야 한다.

5.7 공적 토론: 인권 행사에 영향을 미치는 공공 서비스의 어떤 영역을 알고리즘 시스템으로 결정, 판단 또는 최적화하는 대상에서 제외할 것인지 정의하기 위하여, 국가는 지속적이고 포용적이며 간학제적이고 설명에 입각한 공적 토론을 지원하고 참여해야 한다.

6. 연구, 혁신 및 대중적 인식 제고

6.1 권리 촉진 기술: 세금, 조달 또는 기타 인센티브의 사용을 통하여 국가는 인권과 기본적 자유에 대한 평등한 접근과 향유를 향상시키는 알고리즘 시스템 및 기술의 개발을 촉진해야 한다. 여기에는 알고리즘 시스템의 영향을 평가하기 위한 메커니즘의 개발, 취약하고 과소 대표되는 인구집단의 요구를 해결하기 위한 시스템의 개발 뿐 아니라, 비상 대책 또는 개인이 탈퇴할 수 있는 효과적인 기회로서 아날로그 수단을 통하여 기본 서비스의 지속 가능성을 보장하는 단계가 포함될 수 있다.

6.2 공익의 증진: 알고리즘 시스템이 소외되고 취약한 개인과 집단의 이익이 충분히 고려되고 대표되도록 보장하는 등의 조치로 긍정적인 인권 효과를 창출하고 공익을 증진할 수 있는 가능성을 평가, 검사 및 발전시킬 수 있도록 국가는 독립적인 연구를 지원하

고 참여해야 한다. 적절한 경우, 상업적으로 가장 실행 가능한 최적화 프로세스를 절대적으로 선호하려는 영향력에 대한 억제가 필요할 수 있다. 국가는 알고리즘 시스템 개발 또는 진행 중인 배치에 종사하는 직원이 규제 기관 및 대중에게 자신이 구축을 담당했던 시스템이 현재 또는 미래의 인권 기준을 준수하지 못한 사실을 알릴 필요가 있다고 인식하고 내부 고발 또는 기타 행동을 취하는 것을 적절히 보호해야 한다.

6.3 인간 중심적이고 지속 가능한 혁신: 국가는 사회적 권리와 국제적으로 인정된 노동 및 고용 기준을 비롯하여 기존 인권과 일치하는 기술의 혁신을 장려해야 한다. 천연자원의 추출과 개발 관련 문제 등 국제적으로 합의된 지속 가능한 개발 목표에 부응하려는 노력 및 기존 환경과 기후 문제를 해결하기 위한 노력들이 민간 부문 행위자들의 경쟁력을 견인해야 한다.

6.4 독립적인 연구: 국가는 알고리즘 시스템의 진행 중인 배치가 사회와 인권에 미치는 영향을 모니터링하기 위해 독립적인 연구를 개시, 장려 및 발표해야 한다. 또한 이러한 독립적인 연구는 알고리즘 시스템의 불투명성, 설명 불가능성 및 이의제기 불가능성과 관련하여 현존하는 책임성 격차에 대응하는 효과적인 책무성 메커니즘 및 솔루션의 개발을 연구해야 한다. 이러한 독립적인 연구에 참여하는 연구자, 언론인, 학자의 공정성, 세계적 대표성 및 보호를 보장하기 위해 적절한 메커니즘이 마련되어야 한다.

C. 알고리즘 시스템의 맥락에서 인권과 기본적 자유에 관한 민간 부문 행위자의 책임

1. 일반 적용 원칙

1.1 인권 존중에 대한 책임: 알고리즘 시스템의 설계, 개발, 판매, 배치, 구현 및 서비스에 관여하는 민간 부문 행위자는 공적 영역에서건 사적 영역에서건 인권을 존중하는 실사를 실시하여야 한다. 그들은 고객은 물론 자신의 활동에 의해 영향을 받는 여타 당사자들의 기본적 자유와 국제적으로 인정된 인권을 존중할 책임이 있다. 이러한 책임은 자신의 인권 의무를 이행하는 국가의 역량이나 의지와 독자적으로 존재한다. 이러한 책임을 충족하기 위한 일환으로 민간 부문 행위자는 인권 침해의 원인이 되거나 유발하지 않도록 하고 혁신적인 프로세스를 포함한 그들의 활동이 인권을 존중하도록 지속적이고

사전적이며 사후적인 조치를 취해야 한다. 그들은 또한 사회에 대한 책임과 민주주의 사회의 가치를 염두에 두어야 한다. 인권 준수를 보장하기 위한 노력들은 문서화되어야 한다.

1.2 조치 규모: 인권을 존중하고 적절한 조치를 취해야 하는 민간 부문 행위자의 책임은 그 규모, 부문, 운영 상황, 소유권 구조 또는 특성에 관계없이 적용된다. 그러나 이들이 책임을 완수하는 수단은 그 자원 및 서비스와 시스템이 인권에 미칠 수 있는 영향의 심각성에 따라 그 규모와 복잡성이 다양할 수 있다. 다양한 민간 부문 행위자가 협력하고 잠재적인 인권 침해에 투여하는 상황에서는 모든 파트너의 노력이 필요하며 이 노력은 각자의 영향력과 역량에 비례해야 한다.

1.3 추가적인 핵심 기준: 인권의 수평적 효과로 인해, 그리고 알고리즘 시스템의 설계, 개발 및 진행 중인 배치로 인하여 민간 부문 행위자가 공공 행위자와 매우 긴밀한 협력 상태로 관여하고 있다는 점을 감안할 때, B장에서 국가의 의무로 설명되고 있는 일부 주요 조항은 국내적 수준에서 법적 및 규제적 요구사항으로 해석되며 민간 부문 행위자로서 기업의 책임으로 해석된다. 국가가 해당 규제 조치를 취했는지 여부와 관계없이 민간 부문 행위자는 지속적인 검토, 민주적 참여 및 인식 제고, 정보적 자기 결정권, 컴퓨팅 실험, 검사 및 알고리즘 의사 결정의 식별 가능성과 관련된 위의 B장의 1.2, 1.3, 2.1, 3.1, 3.3, 4.2.문에 포함된 관련 기준을 준수해야 한다.

1.4 차별: 알고리즘 시스템을 설계, 개발 또는 구현하는 민간 부문 행위자는 시스템의 모든 수명 주기 동안 차별을 조장하거나 고착화하는 것을 방지하기 위해 인권 실사 표준 체계를 따라야 한다. 민간 행위자들은 알고리즘 시스템의 설계, 개발 및 진행 중인 배치가 특별한 필요가 있거나 장애가 있는 사람들, 인권에 대한 접근에서 구조적인 불평등에 직면할 수 있는 사람들을 비롯하여 이러한 시스템의 영향을 받는 개인 및 집단에 대하여 직간접적인 차별적 효과를 미치지 않도록 보장해야 한다.

2. 데이터 관리

2.1 동의 원칙: 민간 부문 행위자는 알고리즘 시스템의 영향을 받는 개인에게 알고리즘 데이터셋에 포함된 개인정보를 비롯하여 개인정보의 모든 사용에 관하여 동의하거나 철회할 수 있는 선택권을 쉽게 접근할 수 있는 형태로 동등하게 행사할수 있음을 알려야 한다. 또한 사용자는 자신의 개인정보가 어떻게 사용되는지, 문제의 알고리즘 시스템의 실제적 및 잠재적 영향이 무엇인지, 개인정보 처리에 어떻게 반대하는지, 특정 결과물에 대해 어떻게 이의를 제기하는지 알 수 있어야 한다. 알고리즘 시스템의 추적, 저장 및 성능 측정 도구 사용에 대한 동의 원칙은 명확하고 간결하며 완전하게 표현되어야 하고 서비스 약관에 숨겨져서는 안 된다.

2.2 개인정보보호 설정: 민간 부문 행위자는 모든 관련 개인정보보호 기준에 따라 서비스에 대한 접근을 계속하면서 프라이버시를 효과적으로 보호받을 수 있는 정보주체의 권리를 보장해야 한다. 일련의 프라이버시 설정 옵션 중에서 선택할 수 있는 옵션은 쉽게 볼 수 있고 중립적이며 이해하기 쉬운 방식으로 제시되어야 하며 프라이버시 증진 기술을 사용해야 한다. 기본 설정은 개인정보 처리의 구체적이고 합법적인 목적에 필요하고 비례적인 개인정보를 수집해야 하며 추적 설정은 옵트아웃 모드가 기본값으로 설정되어야 한다. 예를 들어 보안 목적으로 사용자 개인정보를 차단, 삭제 또는 격리하는 메커니즘을 적용할 때에는 개인정보가 잘못 사용되거나 과도하게 사용되는 경우에 요구되는 적법 절차를 보장하고 신속한 구제조치를 동반하여야 한다.

3. 분석 및 모델링

3.1 데이터 및 모델 품질: 민간 부문 행위자는 데이터셋과 모델의 오류, 편향 및 차별 가능성을 특정 맥락 안에서 적절하게 대응하기 위한 목적으로 알고리즘 시스템 학습에 사용하는 데이터의 품질, 특성 및 출처와 관련된 위험을 알고 있어야 한다.

3.2. 표본 집단: 알고리즘 시스템의 개인정보에 대한 평가 및 검사는 충분히 다양하고 대표적인 표본 집단으로 수행되어야 하며 특정 인구통계학적 집단을 의존하거나 차별하지 않아야 한다. 알고리즘 시스템의 개발, 검사 또는 배치가 특정 개인, 집단, 인구집단 및 환경에 대한 위험 또는 손실의 외부 효과를 수반하는 경우 그 시스템의 개발을 중단

하거나 조정해야 한다.

3.3. 시스템 및 데이터 보안: 민간 부문 행위자는 해당되는 표준에 따라 자체 직원 또는 제3자의 불법적인 접근, 제3자의 시스템 간섭 및 장치・데이터・모델 오용을 방지하는 방법으로 알고리즘 시스템을 구성해야 한다.

4. 투명성, 책무성 및 효과적인 구제 수단

4.1 이용 약관: 민간 부문 행위자는 자신이 제공하는 제품 및 서비스에서 중대한 인권 영향을 유발할 수 있는 알고리즘 시스템의 사용에 대하여, 일반 대중은 물론 영향을 받는 당사자는 개인이든 법인이든 관계없이 모든 이들에게 명확하고 간결한 언어 및 접근 가능한 형식으로 알려야 한다. 알고리즘 시스템에 대한 이의와 반대를 허용하기 위하여, 그 특성과 기능에 대한 적절한 정보가 제공되어야 한다. 이용 약관은 합리적으로 간결하고 쉽게 이해할 수 있어야 하며 사용자가 설정을 관리할 수 있음을 명확하고 간결한 언어로 설명해야 한다. 여기에는 시스템 기능, 해당되는 불만 제기 절차, 절차의 다양한 단계, 연락처의 정확한 기능, 표시 시간대 및 예상 결과물을 변경하기 위해 사용 가능한 옵션에 대한 정보가 포함되어야 한다. 영향을 받는 모든 당사자, 신규 고객 또는 적용 규칙이 변경된 제품 및 서비스의 사용자는 관련 변경 사항을 사용자 친화적인 형식으로 통지받고 해당되는 경우 변경 사항에 대한 동의를 요청받아야 한다. 동의하지 않는다고 해서 기본 서비스를 사용할 수 없게 되어서는 안 된다.

4.2 이의제기 가능성: 이의제기가 가능하도록 민간 부문 행위자는 인간 검토자가 접근 가능한 상태를 유지하고 쉽게 접근할 수 있는 연락처 및 핫라인을 제공하는 등 직접적인 접촉이 가능하도록 보장해야 한다. 개인과 집단은 이의를 제기할 수 있을 뿐만 아니라 개선을 위한 제안을 하고 사람의 검토가 체계적으로 요구되는 분야 등에서 기타 유용한 피드백을 제공할 수 있어야 한다. 고객 불만 처리에 관여하는 모든 관련 직원은 해당 인권 기준에 대해 적절하게 숙지하고 있어야 하며 정기적인 교육 기회를 보장받아야 한다.

4.3. 투명성: 민간 부문 행위자는 그들이 제공하는 제품 및 서비스와 관련하여 영향을 받는 개인 및 집단이 제기한 불만의 수와 유형, 불만의 결과에 대한 정보를 공개해야 한다. 이는 그 결과가 특정 사례에 대한 구제 조치로 이어지고 대규모 피해가 발생하기 전

에 불만 사항으로부터 교훈을 얻어 오류의 시정을 시스템에 반영하기 위함이다.

4.4 효과적인 구제 수단: 민간 부문 행위자는 인권 침해 가능성이 있는 시스템의 도입 또는 진행 중인 사용에 이의를 제기하거나 권리 침해를 구제하려는 개인, 집단 및 법인이 온라인과 오프라인에서 집단 시정 메커니즘을 비롯하여 효과적인 구제 수단 및 분쟁 해결 시스템을 이용할 수 있도록 보장해야 한다. 사용 가능한 구제 수단의 범위는 한정되지 않을 수 있다. 우선순위 지정이 필요하고 대응 지연이 복구 가능성에 영향을 미칠 수 있는 경우 가장 심각한 인권 영향을 먼저 해결해야 한다. 모든 불만 사항은 공정하고 독립적인 검토를 허용해야 하며 부당한 지연 없이 처리되어야 하고 적법 절차를 보장하며 성실하게 수행되어야 한다. 관련 메커니즘은 불만 제기자가 독립적인 사법 기관 또는 규제 기관의 국가적 심사 메커니즘을 통해 청구할 수 있는 기회에 부정적인 영향을 미쳐서는 안 된다. 구제 수단에 효과적으로 접근할 수 있는 권리를 포기하거나 방해하는 내용이 이용 약관에 포함되어서는 안 된다. 기업 협회는 무역 협회와 협력하여 모범적인 불만 처리 메커니즘을 수립하는 데 더 많은 투자를 해야 한다.

4.5 협의: 민간 부문 행위자는 개인 및 영향을 받는 당사자의 이익을 대표하는 소비자 단체, 인권 단체 및 기타 단체는 물론, 개인정보보호 기관 및 기타 독립적인 행정청 또는 규제 당국과 함께 알고리즘 시스템의 설계, 개발, 진행 중인 배치 및 평가 뿐 아니라 그 불만 처리 메카니즘에 대한 참여 절차에 적극적으로 참여해야 한다.

5. 예방적 조치

5.1 지속적인 평가: 민간 부문 행위자는 알고리즘 시스템의 설계, 개발 및 진행 중인 배치가 가능한 기술적 오류뿐 아니라 시스템에 야기할 수 있는 잠재적인 법적, 사회적, 윤리적 영향을 탐지하기 위해 지속적으로 평가 및 검사가 이루어지도록 내부 절차를 개발하고 문서화해야 한다. 알고리즘 시스템의 적용이 스스로 회피하거나 완화할 수 있는 마이크로 타겟팅 프로세스를 비롯하여 인권에 고위험을 수반하는 경우, 민간 부문 행위자는 문제되는 서비스의 재설계를 포함하여 이러한 위험을 관리하는 방법에 대한 조언과 지침을 구하기 위해 모든 관련 소관 감독 기관에 통지하고 협의할 수 있어야 한다. 민간 부문 행위자는 정기적이고 독립적인 전문가 검토 및 감독을 위해 이러한 알고리즘 시스템을 제출해야 한다.

5.2 직원 교육: 인권영향평가 및 알고리즘 시스템 검토에 관여하는 모든 관련 직원은 적절한 교육을 받아야 하며 해당 개인정보보호 및 프라이버시 표준 및 그 이상으로 인권을 존중해야 할 책임을 인식해야 한다.

5.3 인권영향평가: 인권영향평가는 영향을 받는 개인 및 집단의 적극적인 참여와 함께 가능한 한 공개적으로 수행되어야 한다. 고위험 알고리즘 시스템을 배치하는 경우, 진행되는 인권영향평가의 결과, 위험 완화를 위해 식별된 기술, 관련 모니터링 및 검토 절차에 대한 사항이 법으로 보호되는 비밀을 침해하지 않고 공개적으로 접근 가능해야 한다. 비밀 규칙을 시행해야 하는 경우 모든 기밀 정보는 평가 보고서의 별도 부록으로 제공되어야 한다. 이 부록은 관련 감독 당국이 접근할 수 있어야 한다.

5.4 후속 조치: 민간 부문 행위자는 인권 행사에 대한 부정적인 영향과 위험을 방지하고 완화하려는 목적으로, 알고리즘 시스템의 전체 수명 주기 동안 기록된 발견 사항에 기반하여 적절한 조치를 취하고 대응 효과를 모니터링함으로써 인권영향평가에 대한 적절한 후속 조치를 보장해야 한다. 식별된 오류는 가능한한 신속하게 해결되어야 하고, 적절한 경우 관련 활동이 일시 중단되어야 한다. 이를 위해서는 설계, 검사 및 배치 단계 전반에 걸쳐 정기적이고 지속적인 품질 보증 검사와 실시간 감사가 필요하다. 나아가 상황별 현장별 알고리즘 시스템의 인권 영향을 모니터링하고 적절하고 시기적절한 방식으로 오류와 피해를 시정하기 위하여 영향을 받는 개인과의 정기적인 협의가 추가적으로 필요하다. 이는 부정적인 인권 영향을 악화시키고 고착화할 수 있는 피드백 순환구조의 위험을 고려할 때 특히 중요하다.

6. 연구, 혁신 및 대중적 인식 제고

6.1 연구: 민간 부문 행위자는 긍정적인 인권 영향을 창출하고 공익을 증진할 수 있는 알고리즘 시스템을 평가, 검사 및 발전시키는 것을 목표로, 연구 윤리에 따라 수행되는 연구에 참여하고 자금을 지원하고 이를 출판해야 한다. 그들은 또한 이러한 목표를 가진 독립적인 연구를 지원하고 연구자와 연구 기관의 진실성을 존중하여야 한다. 이러한 연구는 알고리즘 시스템의 영향을 평가하기 위한 메커니즘의 개발과 취약하고 소외된 인구집단의 요구를 해결하기 위한 알고리즘 시스템의 개발과 관련될 수 있다. 민간 부문 행위자는 알고리즘 시스템의 배치와 관련된 위험을 식별하고 대응하기 위해 특히 인권 우

려가 높은 지역에서 지역 시민 사회 단체와 효과적인 의사 소통 방안을 찾아야 한다.

6.2 데이터 접근: 알고리즘 시스템과 디지털화된 서비스가 권리 행사, 커뮤니케이션 네트워크 및 민주주의 체제에 미치는 영향을 분석하기 위해 민간 부문 행위자는 삭제를 위해 분류된 데이터에 적절한 당사자, 특히 독립 연구자, 언론 및 시민 사회 단체가 접근하도록 하는 등 관련 개인정보 및 메타 데이터셋에 대한 접근을 확대해야 한다. 이러한 접근 확대는 법적으로 보호되는 이익과 모든 해당 프라이버시 및 개인정보보호 원칙을 완전히 존중하여 이루어져야 한다.

부록 V.

캐나다 정부
<알고리즘 영향평가 도구>[6)]

2021. 4. 1.

1. 소개

알고리즘 영향평가(AIA)는 재무부의 <자동화된 의사결정 훈령>(이하 '훈령')을 지원하는 의무적 위험성 평가 도구입니다. 이 도구는 자동화 된 의사결정 시스템의 영향 정도를 판단하는 질문지 형식입니다. 질문지는 48개의 위험성과 33개의 완화 조치에 대한 질문으로 구성되어 있습니다. 시스템 설계, 알고리즘, 의사결정 유형, 영향 및 데이터를 비롯한 여러 요소를 기반으로 평가 점수가 내려집니다.

AIA는 내부 및 외부 이해관계자의 자문과 모범 사례를 바탕으로 개발되었습니다. 이 도구는 공개적으로 개발되었으며 일반인도 공개 라이선스에 따라 공유 및 재사용이 가능합니다.

2. 평가의 사용 및 채점 방식

이 평가는 재무부 사무처가 학계, 시민 사회 및 기타 공공기관과 협의하여 수립한 것으로서 자동화된 의사결정 시스템의 위험성 영역에 대하여 캐나다 정부의 정책, 윤리, 행정법적 사항들을 고려하여 구성되었습니다. AIA는 각 부처 및 기관이 자동화 된 의사결정 시스템과 관련된 위험성을 더 잘 이해하고 관리 할 수 있도록 설계되었습니다. AIA는 표 1에 정의된 위험성 영역을 평가하기 위해 다양한 형식의 문항으로 구성됩니다.

---

6) Algorithmic Impact Assessment Tool.
https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.

<Table 1> 정의된 위험성 영역

위험성 영역	정의
1. 프로젝트	
프로젝트 단계	프로젝트 소유자, 설명 및 단계 (설계 또는 구현)
사업 동기 / 긍정적 영향	의사결정 절차에 자동화를 도입하려는 동기
위험성 개요	높은 수준의 프로젝트 위험성 지표
프로젝트 권한	프로젝트에 대해 새로운 정책 권한을 추구할 필요가 있음
2. 시스템	
시스템 관련	시스템의 사양 (예 : 이미지 인식, 위험성 평가 등)
3. 알고리즘	
알고리즘 관련	알고리즘의 투명성, 쉽게 설명되는지 여부 등
4. 의사결정	
의사결정 관련	자동화된 의사결정의 분류 (예 : 보건의료서비스, 사회 지원, 면허 등)
5. 영향	
영향평가	지속성, 가역성 및 영향을 받는 분야 (예 : 권리, 건강, 경제 또는 환경)
6. 데이터	
소스	의사결정을 자동화하는 데 사용된 데이터의 출처 및 보안 등급
유형	정형 또는 비정형으로 사용되는 데이터의 종류 (오디오, 텍스트, 이미지 또는 비디오)

AIA는 또한 식별된 위험성을 관리하기 위해 마련된 완화 조치를 평가합니다. 이 완화성에 대한 문항들은 표 2에 정의된 범주로 구성됩니다.

<Table 2> 완화성 영역 정의

완화성 영역	정의
7. 자문	
내부 및 외부 이해관계자	개인정보보호 및 법률 전문가를 비롯해 자문받은 내부 및 외부의 이해관계자
8. 위험성 제거 및 완화 조치	
데이터 품질	데이터가 대표성을 가지고 편향적이지 않도록 보장하는 절차 및 이 절차들과 관련된 투명성 조치
절차적 공정성	시스템과 그 의사결정을 감사하는 절차 및 회복 조치
개인정보보호	개인정보를 보호하는 조치

2.1 채점

각 영역에는 1 개 이상의 문항이 포함되어 있으며, 문항들에 대한 답변들이 해당 영역의 최대 점수를 구성합니다. 각 문항의 값은 본래의 영향과 완화성에 기초하여 가중치가 부여되며, 이 가중치가 점수에 반영됩니다. 본래 영향에 대한 점수는 자동화의 위험성을 측정하는 반면, 완화성 점수는 자동화의 위험성을 관리하는 방법을 측정합니다.

위험성 영역 1~6 문항은 본래 영향 점수를 높여가고, 완화성 영역 7과 8 문항은 완화성 점수를 높여갑니다. 이 문항들은 다음과 같은 다양한 요소에 걸쳐 의사결정이 미칠 영향을 측정하기 위해 고안되었습니다.

개인 또는 공동체의 권리
개인 또는 공동체의 건강 또는 복리
개인, 단체 또는 공동체의 경제적 이익
생태계의 지속가능성
영향의 지속성과 가역성

<Table 3> 위험성 영역에서 본래 영향 점수

위험성 영역	문항수	최대 점수
1 - 프로젝트	15	15
2 - 시스템	1	0
3 - 알고리즘	2	6
4 - 의사결정	1	6
5 - 영향	16	36
6 - 데이터	13	44
본래 영향 점수	**48개 문항**	**최대 107 점**

<Table 4> 완화성 영역에서 완화성 점수

완화성 영역	문항수	최대 점수
7 - 자문	2	2
8 - 위험성 제거 및 완화 조치	31	43
완화성 점수	**33개 문항**	**최대 45점**

현재 점수는 다음과 같이 결정됩니다.

a. 완화성 점수가 달성 가능한 최대 점수의 80% 미만일 경우, 현재 점수는 본래 영향 점수와 같음

b. 완화성 점수가 달성 가능한 최대 점수의 80% 이상일 경우, 본래 영향 점수에서 15%를 차감함

이 현재 점수는 백분율 범위로 표시되며, <Table 5>에 제시된 영향 수준에서 해당하는 범위를 맞춰 봅니다.

2.2 영향 수준

AIA를 완료하면 해당 시스템에 대한 영향 수준이 제시됩니다. 영향 수준은 수준 I (작은 영향)으로부터 수준 IV (매우 높은 영향)에 이르기까지 다양합니다. 영향 수준은 달성 가능한 최대 점수 대비 현재 점수의 백분율로 결정됩니다.

<Table 5> 영향 수준 정의

영향 수준	정의	점수 백분율 범위
수준 I	영향이 거의 없거나 전혀 없음	0 % ~ 25 %
수준 II	중간 영향	26 % ~ 50 %
수준 III	높은 영향	51 % ~ 75 %
수준 IV	매우 높은 영향	76 % ~ 100 %

3. 지침

AIA는 정보공개 포털에서 온라인 질문지로 사용할 수 있습니다. 모든 문항을 마친 후 영향 수준 및 훈령상 요구사항 링크가 보여집니다. 또한 상세한 결과 페이지에서 해당 시스템이 그 수준으로 평가된 이유에 대한 설명을 볼 수 있습니다. 이 결과 및 설명은 PDF 포맷으로 인쇄 및 저장할 수 있습니다.

AIA는 서비스 대상자, 업무 절차, 데이터 및 시스템 설계 결정을 비롯해 광범위한 주제에서 자동화된 결정을 평가합니다. 이러한 영역들의 전문지식을 제공하는 다분야 팀과 함께 AIA를 실시하는 것이 가장 좋습니다.

AIA의 각 질문에 모두 답변해야 합니다. 질문에 대한 답변을 알 수없는 경우 이 질문에서 가장 낮은 점수를 선택하십시오. 해당되는 경우 요청을 받으면 문서 증거를 제공할 준비를 하십시오.

훈령 제6장은 기관이 수행해야 할 요구사항의 일반적인 목록을 제공합니다. 전문가검토 유형 및 결정에 인적으로 개입하는 정도를 포함하여 더 높은 영향 수준에서 일부 요구사항이 증가할 수 있습니다. 영향 수준별로 다양한 요구사항들을 준수하려면, 훈령 부

록의 영향 수준별 요구사항을 참고하십시오. 그외 요구사항은 영향 수준별로 달라지지 않는 기본 요구사항으로서, 시스템 개발에 앞서 기관의 법률부서와 협의하고, 직원을 교육하며, 의사결정에 이의를 제기할 수 있는 법적인 소구 방안을 제공해야 합니다.

3.1 AIA 실시 시기

AIA는 프로젝트 설계 단계 초기에 실시되어야합니다. AIA의 결과에 따라 자동화된 의사결정 시스템을 구현할 때 준수해야 하는 훈령상의 완화성 및 자문 요구사항을 따르게 됩니다.

시스템의 생산 전에 두 번째로 AIA를 실시하여 해당 결과가 구축된 시스템에 정확하게 반영되었는지 확인해야합니다. 이 수정된 AIA 결과는 최종 결과로서 정보공개 포털에 공개되어야합니다.

3.2 결과 공개

각 기관은 AIA의 최종 결과를 정부 정보공개 포털에서 일반에 접근가능한 형식과 공용어 두 가지로 공개할 책임이 있습니다. AIA의 결과 페이지는 AIA에 입력 된 텍스트에 대한 번역을 제공하는 옵션을 제공합니다. 결과 페이지는 이 요구사항을 만족하기 위해 일반에 접근가능한 PDF 포맷으로 결과를 다운로드하는 옵션도 제공합니다.

알고리즘 영향평가 질문지

AIA의 정보는 사용자의 로컬 컴퓨터에만 저장되며, 캐나다 정부는 사용자가 도구에 입력한 정보에 접근할 수 없습니다. 작업을 보관하려면 나중에 사용할 수 있도록 ‘저장’ 버튼을 사용하여 데이터를 로컬에 저장하십시오. ‘JSON 파일 업로드’ 버튼을 사용하여 이전에 저장된 AIA 양식을 다시 로드할 수 있습니다.

제1장 : 프로젝트 세부정보

답변자 이름

직책

기관명

부서명

프로젝트명

IT 계획상 프로젝트 ID

부서 프로그램명(부서 성과 체계상)

프로젝트 단계(필수)

☐ 설계

☐ 구현

프로젝트 설명을 입력하십시오.

제2장 : 사업 동기 / 긍정적 영향

귀 기관의 팀이 이 의사결정 과정에 자동화를 도입하는 동기는 무엇입니까? (해당 항목 모두 체크하시오)

☐ 기존 업무 또는 사건의 적체
☐ 전반적인 의사 결정의 품질 향상
☐ 기존 프로그램의 처리 비용 절감
☐ 시스템이 인간이 합리적인 시간 내에 수행할 수 없었던 업무를 수행함
☐ 혁신적인 접근 방식의 사용
☐ 기타 (구체적으로 작성하십시오)

제3장 : 위험성 개요

프로젝트가 (예를 들어 개인정보보호 문제로 인해) 집중적인 공공 조사 영역에 속하거나 소송이 빈번한 영역에 속합니까?

☐ 예
☐ 아니오

이 업계의 고객이 특히 취약합니까?

☐ 예
☐ 아니오

의사결정의 이해관계가 첨예합니까?

☐ 예
☐ 아니오

프로젝트가 직원의 수나 역할 측면에 큰 영향을 미칩니까?

☐ 예
☐ 아니오

제4장 : 프로젝트 권한

이 프로젝트에 새로운 정책 권한이 필요합니까?

☐ 예
☐ 아니오

제5장 : 시스템 관련

다음 중 시스템에 적용되는 기능을 체크하십시오.

☐ 이미지 및 개체 인식: 이미지 또는 개체와 관련된 인식, 분류 및 맥락화를 자동화하기 위해 대단히 큰 규모의 데이터셋을 분석함
☐ 텍스트 및 음성 분석: 텍스트, 발화, 음성을 식별, 처리, 태그하고 태그에 기반하여 권고하기 위하여 큰 규모의 데이터셋을 분석함
☐ 위험성 평가: 패턴을 식별하고 및 행동 방침을 권고하거나 경우에 따라 특정 조치를 개시하기 위하여 대단히 큰 규모의 데이터셋을 분석함
☐ 콘텐츠 생성: 특정 상황에서 특정 컨텐츠를 분류, 처리, 분배, 개인화 및 서비스하기 위하여 큰 규모의 데이터셋을 분석함
☐ 프로세스 최적화 및 업무절차 자동화: 이상 징후를 식별하고, 패턴을 군집화하고, 결과 또는 최적화 방법을 예측하고, 특정 업무절차를 자동화하기 위하여 큰 규모의 데이터셋을 분석함
☐ 기타 (특정하시오)

제6장 : 알고리즘 관련

알고리즘에 다음 특성이 있습니까?

사용되는 알고리즘이 (영업) 비밀이 될 것이다

☐ 예
☐ 아니오

알고리즘 처리가 해석하거나 설명하기 어려울 것이다

☐ 예
☐ 아니오

제7장 : 의사결정 관련

의사결정이 아래 범주 중 하나에 해당합니까(해당 항목 모두 체크하십시오)?

☐ 보건의료 관련 서비스
☐ 경제적 이익(보조금 및 기금, 세금 혜택, 채권 추심)
☐ 사회적 지원(예 : 장애인 신청 등)
☐ 접근성 및 이동성(보안 허가서, 국경 통과)
☐ 면허 및 허가서 발급
☐ 기타 (특정하시오)

제8장 : 영향평가 관련

이 시스템은 의사결정자를 보조하는 데만 사용됩니까?
- ☐ 예
- ☐ 아니오

이 시스템이 인간에 의해 이루어질 수 있는 의사결정을 대체합니까?
- ☐ 예
- ☐ 아니오

이 시스템이 판단이나 재량권이 필요한 인간의 의사결정을 대체합니까?
- ☐ 예
- ☐ 아니오

시스템을 개발한 부서가 아닌 다른 부서에서 이 시스템을 사용합니까?
- ☐ 예
- ☐ 아니오

의사결정에 따른 영향을 되돌릴 수 있습니까?
- ☐ 되돌릴 수 있음
- ☐ 비교적 되돌릴 수 있음
- ☐ 되돌리는 것이 어려움
- ☐ 되돌리는 것이 불가능함

의사결정으로 인한 영향이 얼마나 지속됩니까?
- ☐ 잠시 영향을 미치는 편임
- ☐ 어떤 영향은 몇 개월 동안 지속되지만, 어떤 영향은 더 오래 지속될 수 있음
- ☐ 영향이 수년간 지속됨
- ☐ 대부분의 영향이 영구적임

위에서 선택한 답변에 따라 의사결정으로 인한 영향이 발생한 이유를 설명하십시오.

의사결정이 개인의 권리나 자유에 미치는 영향이 다음과 같습니다.
- ☐ 영향이 거의 없거나 전혀 없음
- ☐ 온건한 영향을 미침
- ☐ 높은 영향을 미침
- ☐ 매우 높은 영향을 미침

(위에서 선택한 답변에 따라) 의사결정으로 인한 영향이 발생한 이유를 설명하십시오.

의사결정이 개인의 건강과 복리에 미칠 영향이 다음과 같습니다.
☐ 영향이 거의 없거나 전혀 없음
☐ 온건한 영향을 미침
☐ 높은 영향을 미침
☐ 매우 높은 영향을 미침

(위에서 선택한 답변에 따라) 의사결정으로 인한 영향이 발생한 이유를 설명하십시오.

의사결정이 개인의 경제적 이익에 미치는 영향이 다음과 같습니다.
☐ 영향이 거의 없거나 전혀 없음
☐ 온건한 영향을 미침
☐ 높은 영향을 미침
☐ 매우 높은 영향을 미침

(위에서 선택한 답변에 따라) 의사결정으로 인한 영향이 발생한 이유를 설명하십시오.

의사결정이 환경 생태계의 지속가능성에 미칠 영향이 다음과 같습니다.
☐ 영향이 거의 없거나 전혀 없음
☐ 온건한 영향을 미침
☐ 높은 영향을 미침
☐ 매우 높은 영향을 미침

(위에서 선택한 답변에 따라) 의사결정으로 인한 영향이 발생한 이유를 설명하십시오.

제9장 : 데이터 관련

A. 데이터 소스

자동화된 의사결정 시스템이 개인 정보를 입력 데이터로 사용합니까?
☐ 예
☐ 아니오

이 시스템에서 사용하는 입력 데이터의 최고 보안 등급은 무엇입니까? (하나를 선택하십시오)
☐ 없음
☐ A급 보호
☐ 기밀 / 대외비 (Classified / Confidential)
☐ B급 보호 / C급 보호
☐ 비밀 / 최고비밀 (Secret / Top Secret)

누가 데이터를 관리합니까?
☐ 공개 데이터
☐ 연방 정부
☐ 지방자치단체
☐ 민간기관 / 시민단체

이 시스템이 여러 다른 소스의 데이터를 사용합니까?
☐ 예
☐ 아니오

이 시스템이 인터넷 또는 원격 통신 장치로 데이터를 입력합니까? (예: 사물 인터넷, 센서)
☐ 예
☐ 아니오

이 시스템이 다른 IT 시스템과 상호 작용할 것인가?
☐ 예
☐ 아니오

이 시스템을 학습시키는데 사용되는 데이터를 누가 수집했습니까?
☐ 우리 기관 자체
☐ 다른 연방 기관
☐ 다른 정부 기관
☐ 외국 정부 / 제3의 비정부기관

이 시스템에서 사용하는 입력 데이터를 누가 수집했습니까?
☐ 우리 기관 자체
☐ 다른 연방 기관
☐ 다른 정부 기관
☐ 외국 정부 / 제3의 비정부기관

B. 데이터 유형

이 시스템이 권고사항이나 의사결정을 내리기 위해 비정형 데이터의 분석을 필요로 합니까?
☐ 예
☐ 아니오

('예' 인 경우) 비정형 데이터의 유형은 무엇입니까?
☐ 오디오 및 텍스트 파일
☐ 이미지 및 비디오 파일

제10장 : 자문

다음 집단이 관여합니까?

내부 이해관계자(전략 정책 및 계획 부서, 데이터 거버넌스 부서, 프로그램 정책 부서 등)
☐ 예
☐ 아니오

외부 이해관계자(시민사회, 학계, 산업계 등)
☐ 예
☐ 아니오

제11장 : 위험성 제거 및 완화 조치 - 데이터 품질

편향성 및 기타 예상치 못한 결과물에 대해 데이터셋을 검사하기 위해 문서화된 절차를 두고 있습니까? 이런 절차들은 체계, 방법론, 지침 또는 기타 평가 도구를 적용한 사례 등을 말합니다.
☐ 예
☐ 아니오

설계 단계에서 데이터 품질 문제를 해결하는 방식에 대해 문서화하는 절차를 두고 있습니까?
☐ 예
☐ 아니오

위 정보가 공개됩니까?
☐ 예
☐ 아니오

이 데이터에 대하여 '젠더 기반 분석 플러스'를 적용할 예정입니까?
☐ 예
☐ 아니오

위 정보가 공개됩니까?
☐ 예
☐ 아니오

이 시스템의 설계, 개발, 유지보수 및 개선에 대한 책임을 소속 기관 내에서 할당했습니까?
☐ 예
☐ 아니오

오래된 데이터나 신뢰할 수 없는 데이터가 자동화된 의사 결정에 사용되는 위험을 관리하기 위하여 문서화된 절차를 두고 있습니까?
☐ 예
☐ 아니오

위 정보가 공개됩니까?
☐ 예
☐ 아니오

이 시스템에 사용된 데이터가 정보공개포털에 공개됩니까?
☐ 예
☐ 아니오

제12장 : 위험성 제거 및 완화 조치 - 절차적 공정성

감사추적(audit trail) 기능이 법률에서 명시된 권한 또는 위임된 권한을 확인합니까?
☐ 예
☐ 아니오

이 시스템은 시스템에 의해 수행된 모든 권고사항 또는 의사결정을 기록한 감사추적 정보를 제공합니까?
☐ 예
☐ 아니오

모든 주요 의사결정 사항이 감사추적 정보에서 확인됩니까?
☐ 예
☐ 아니오

자동화된 시스템 로직 내의 모든 주요 의사결정 사항들이 관련 법률, 정책 또는 절차와 관련되어 있습니까?
☐ 예
☐ 아니오

모델과 시스템의 모든 변경 사항을 상세히 기록한 로그기록을 유지합니까?
☐ 예
☐ 아니오

감사추적 기능이 시스템에 의해 결정된 모든 사항을 명확하게 규정합니까?
☐ 예
☐ 아니오

시스템이 생성한 감사추적 기능을 사용하여 필요한 경우 결정 통지(이유서 또는 기타 통지 포함)를 생성할 수 있습니까?
☐ 예
☐ 아니오

감사추적 기능이 시스템의 각 결정에 대해 어떤 버전의 시스템이 사용되었는지 정확하게 식별할 수 있습니까?
☐ 예
☐ 아니오

감사추적 기능이 권한이 있는 의사결정자가 누구인지 식별할 수 있습니까?
☐ 예
☐ 아니오

시스템이 필요할 경우 의사결정이나 권고사항에 대한 이유를 제시할 수 있습니까?
☐ 예
☐ 아니오

시스템에 대한 접근 권한을 부여, 모니터링 및 취소하는 절차가 있습니까?
☐ 예
☐ 아니오

시스템 사용자가 피드백을 포착할 수 있는 메커니즘이 있습니까?
☐ 예
☐ 아니오

의사결정에 대해 이의를 제기하고자 하는 고객을 위해 계획 또는 수립된 소구 절차가 있습니까?
☐ 예
☐ 아니오

시스템의 결정에 대해 인간의 기각이 가능합니까?
☐ 예
☐ 아니오

기각이 이루어졌을 때 그 사실을 기록하는 절차가 있습니까?
☐ 예
☐ 아니오

감사추적 기능에 시스템의 작동 또는 성능에 대한 수정사항을 기록하는 변경사항 관리 절차가 포함됩니까?
☐ 예
☐ 아니오

캐나다 정부 <정보기술아키텍처 검토위원회>에 컨셉 케이스를 제출할 예정입니까?
☐ 예
☐ 아니오

제13장 : 위험성 제거 및 완화 조치 - 개인정보보호

시스템에 개인정보 사용이 포함된 경우, 개인정보 영향평가를 수행하거나 수행한 적이 있거나, 기존 영향평가를 갱신할 예정입니까?
☐ 예
☐ 아니오

프로젝트의 개념 수립 단계에서부터 시스템에 보안과 개인정보보호조치를 설계하고 구축합니까?
☐ 예
☐ 아니오

개인정보가 폐쇄형 시스템(예 : 인터넷, 인트라넷 또는 다른 시스템에 대한 연결이 없음) 내에서 사용됩니까?
☐ 예
☐ 아니오

개인정보 공유와 관련된 경우, 적절한 보호조치가 수반된 동의서 또는 약정이 수립되어 있습니까?
☐ 예
☐ 아니오

알고리즘 영향 수준별 요구사항[7)]

요구사항	수준 Ⅰ	수준 Ⅱ	수준 Ⅲ	수준 Ⅳ
전문가 검토 (peer review)	비해당	다음중 1개 이상 수행 - 연방, 주, 준주 또는 시 정부기관에서 자격이 인증된 전문가의 검토 - 고등교육기관 학부 유자격 구성원의 검토 - 관련 비정부기구 소속 유자격 연구자의 검토 - 관련 전문성을 갖춘 서드파티 공급자와 계약 - 자동화된 의사결정 시스템의 사양을 전문가가 검토하는 저널에 게재 - 재정위원회 사무처에서 지정한 데이터 및 자동화 자문기구의 검토		다음중 2개 이상 수행하거나 - 캐나다 국립연구위원회, 캐나다 통계청, 또는 캐나다 통신보안기구에서 자격이 인증된 전문가의 검토 - 고등교육기관 학부 유자격 구성원의 검토 - 관련 비정부기구 소속 유자격 연구자의 검토 - 관련 전문성을 갖춘 서드파티 공급자와 계약 - 재정위원회 사무처에서 지정한 데이터 및 자동화 자문기구의 검토, 또는 - 자동화된 의사결정 시스템의 사양을 전문가가 검토하는 저널에 게재

7) (Appendix C) Impact Level Requirements

공지	비해당	프로그램이나 서비스 웹사이트에 쉬운 용어로 된 공지 게시	관련 웹사이트에 자동화된 의사결정 시스템에 대한 쉬운 용어로 된 문서 발간하며 다음 사항을 포함할 것 - 구성 요소의 작동 방식 - 행정 결정을 지원하는 방식 - 모든 검토 또는 감사의 결과 - 학습 데이터에 대한 설명, 또는 이 데이터를 공개적으로 사용할 수 있는 경우 익명화된 학습 데이터에 대한 링크
의사결정에 대한 인간의 개입 (Human-in-the- loop for decisions)	의사결정이 인간의 직접적인 개입 없이 내려질 수 있음		의사결정 절차에서 특정 시점에 인적 개입이 없으면 의사결정이 내려질 수 없음. 더불어 최종 의사결정은 사람에 의해 이루어져야 함
설명 요구사항	해당 법정 요구사항에 추가적으로 공통적인 의사결정 결과에 대하여 유의미한 설명이 제공되도록 보장할 것. 여기에는 웹사이트 자주 묻는 질문 코너(FAQ)를 통해 설명을 제공하는 것이 포함될 수 있음	해당 법정 요구사항에 추가적으로 수혜, 서비스, 기타 규제 조치를 거부하는 의사결정 결과에 대하여 요청이 있을 경우 유의미한 설명이 제공되도록 보장할 것	해당 법정 요건에 추가적으로 수혜, 서비스, 기타 규제 조치를 거부하는 모든 의사결정 결과에 대하여 유의미한 설명이 제공되도록 보장할 것
검사 (testing)	- 생산에 착수하기 전, 학습 데이터가 의도하지 않은 데이터 편향 및 결과에 부당하게 영향을 미칠 수 있는 기타 요소에 대해 검사할 수 있는 적절한 절차를 개발할 것 - 자동화된 의사결정 시스템에서 사용 중인 데이터가 여전히 관련성이 있고 정확하며 최신인지 확인하기 위해 정기적으로 검사할 것		

모니터링	자동화된 의사결정 시스템의 결과를 지속적으로 모니터링하여 의도하지 않은 결과로부터 보호하고 본 지침뿐만 아니라 기관 및 프로그램 관련 법률의 준수를 보장할 것			
교육훈련	비해당	시스템의 설계 및 기능에 대한 문서화	시스템의 설계 및 기능에 대한 문서화. 교육 과정 이수 필수.	시스템의 설계 및 기능에 대한 문서화. 교육 과정 반복적 이수. 교육 이수 확인 수단 마련.
비상 계획 (Contingency Planning)	비해당		자동화된 의사결정 시스템을 사용할 수 없는 경우 비상 계획 및 백업 시스템을 사용할 수 있도록 보장할 것	
시스템 구동 승인	비해당	비해당	부처 실장 승인	재정위원회 승인

부록 Ⅵ. 신기술 관련 유엔 인권이사회 특별절차 보고서 목록[8)]

특별절차	연도 (문서)	보고서 내용
아프리카 전문가 실무그룹	2019 (A/HRC/42/59)	- '인종 정의를 위한 데이터'를 주제로 스위스 제네바에서 개최된 실무그룹 23, 24차 세션에 대하여 2019년 인권이사회 42차 세션에 제출된 보고서 - 형사사법 분야에서 인공지능 및 알고리즘의 인종편향 가능성에 대하여 논의함
초국적기업 및 기타 사업체의 인권에 대한 실무그룹	2021 (A/HRC/47/39/Add.2)	- 기업과 인권 이행지침에 대하여 2021년 인권이사회 47차 세션에 제출된 보고서 - 기술과 소셜미디어 회사에 대한 장에 인권 옹호자 존중을 위한 지침이 포함됨
	2020 (A/75/212)	- 기업, 인권 및 갈등에 영향을 받는 지역에 대한 2020년 총회 75차 세션에 제출된 보고서 - 사이버 시대의 과제에 대한 장이 포함됨
	2019 (A/HRC/41/43)	- 기업과 인권 이행지침 이행에 있어 젠더 관점 통합에 관하여 2019년 인권이사회 41차 세션에 제출된 보고서 - 기업은 원칙 13에 따라 인공지능 및 자동화 등 신기술이 여성의 인권에 부당한 역효과를 일으키지 않도록 보장해야 함
	2019 (A/HRC/41/49)	- 기업과 인권 7차 연례포럼에 대하여 2019년 인권이사회 41차 세션에 제출된 보고서 - 기술 및 기업의 인권 보호에 대한 장이 포함됨
	2018 (A/73/163)	- 기업과 인권 이행지침에 명시된 기업 인권실사 이행을 증진시키기 위한 기업과 정부의 조치에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서 - 신기술이 공급망에서 인권 영향의 확보 방식을 개선하는 혁신적 해결 가능성을 제공한다는 점에 주목함 - 중요한 혁신에는 노동자의 목소리가 반영되는 기술이 포함되며, 이는 공급망에서 인권 실사와 복원적 접근법 모두를 향상시킬 수 있음 - 공급망의 모든 노드에서 영향을 모니터링하기 위해 블록체인 기술을 사용하려는 협력적

8) Office of the High Commissioner for Human Rights (2021). Non-exhaustive list of Special Procedures reports relevant to new technologies. <https://www.ohchr.org/Documents/HRBodies/SP/List_SP_Reports_NewTech.pdf (검색일: 2021. 9. 12)>.

		기획이 몇몇 부문에서 추진되고 있음 - 그러나 이와 동시에 기술을 사용할 때는 모범 관행을 훼손하는 위험에 대해 충분한 고려가 필요함
문화권 분야 특별보고관	2019 (A/74/255)	- 문화권을 실현시키는데 있어 공공 공간의 중요성과 모든 사람이 이에 접근하고 향유하기 위하여 해결되어야 할 문제에 대하여 2019년 유엔총회 74차 세션에 제출된 보고서 - 공공 공간으로서 사이버공간에 대하여 논하고, 디지털 시대에 공공 공간이 더이상 엄격하게 물리적 공간에 한정되지 않고 사이버공간을 포함하며 이는 인권이 온라인에서도 계속 보장되어야 함을 의미한다고 결론을 내림 - 이 공간은 특유의 권리에 대하여 국제인권법에서 확인된 제한 체제에 똑같이 구속되며, 인권에 대한 의무를 충족해야 하는 공권력은 모든 사람이 사이버공간에 접근하고 참여하도록 보장하는 조치를 취할 필요가 있음
	2016 (A/71/317)	- 분쟁 및 비분쟁 상황에서 국가/비국가적 행위로 인하여 세계 각지 문화 유산의 파괴에 대한 인권 접근에 대하여 2016년 유엔총회 71차 세션에 제출된 보고서 - 국가가 전시 문화 유산에 대한 모든 위협 가능성에 대하여 평화시 대비할 것을 권고함. 여기에는 자국 관할권내 유무형 문화유산에 대한 문서화는 물론 가능한 디지털기술과 뉴미디어의 적용을 포함함
	2015 (A/70/279)	- 특허 정책 및 과학문화향유권에 대하여 2015년 유엔총회 70차 세션에 제출된 보고서 - 주요 기술에 대한 접근을 보장하는 특허 정책의 영향에 대하여 논함. 예를 들어 에너지 영역에서 정보통신기술, 나노기술, 합성생물학 등 과학 발전이 인권에 중대한 영향을 미칠 수 있는 함의에 주목할 필요성이 있음
	2015 (A/HRC/28/57)	- 과학문화향유권의 관점에서 저작권법과 정책에 대하여 2015년 인권이사회 28차 세션에 제출된 보고서 - 저자의 권리 보호 필요성과 문화생활 참여 기회의 확대 필요성을 모두 강조함
	2012 (A/HRC/20/26)	- 과학 발전과 그 적용의 혜택을 향유할 권리에 대하여 2012년 인권이사회 20차 세션에 제출된 보고서
발전권 특별보고관	2018 (A/73/271)	- 과학기술 협력 측면을 포함한 남남(south-south) 협력에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서

장애인권리 특별보고관	2019 (A/74/186)	- 장애 노인의 상황에 대하여 2019년 유엔총회 74차 세션에 제출된 보고서 - 이들의 상황에서 보조기기와 기술, 정보통신기술을 통한 디지털 서비스와 디지털 거버넌스에 대하여 논함
	2017 (A/HRC/34/58)	- 보조기기 및 기술을 포함한 장애인에 대한 다양한 형태의 권리 기반 지원 및 보조에 대하여 2017년 인권이사회 34차 세션에 제출된 보고서
	2016 (A/71/314)	- 장애인 협약에서 확인되고 지속가능 개발목표 달성에 기여하는 장애인 통합정책 수립 방안에 대하여 2016년 유엔총회 71차 세션에 제출된 보고서 - 보조기기 및 기술에 대하여 논함
여성 차별에 대한 실무그룹	2020 (A/HRC/44/51)	- 노동의 변화와 여성 인권에 대하여 2020년 인권이사회 44차 세션에 제출된 보고서 - 기술 변화에 주목하고 국가가 “기술 공급자가 온라인에서 여성에 대한 모든 형태의 폭력을 방지하고 제거하도록 규제 체제를 강화할 것”을 권고함
교육권 특별보고관	2020 (A/HRC/44/39)	- 코로나 위기가 교육권에 미치는 영향에 대하여 2020년 인권이사회 44차 세션에 제출된 보고서로, 교육의 디지털화에도 주목함 - “정부는 디지털 기술을 통한 민간 부문의 대규모 진입과 그로 인하여 장기적 관점에서 교육권 및 교육 체제에 미치는 주요 위기 측면을 고려해야 하고, 민간 부문의 증대되는 역할이 위기에서 이윤을 추구하는 기업이 제한된 공공 교육 자원을 차지하고 학생 및 교사의 데이터를 수집하며 아동청소년을 대상으로 한 광고로 이어지지 않도록 적절한 규제를 채택하는 등 조치할 것”을 권고함. 교육 및 학습 솔루션은 교육권 향유를 위협하고 불평등을 심화시키는 상업적 라이선스 제약 없이 공익을 위해 개발되어야 함
	2018 (A/73/262)	- 교육권 측면에서 난민 문제에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서 - 특히 지속가능 개발목표 달성의 측면에서 이러닝 등 혁신 솔루션을 제공하는 정보통신기술의 문제 해결 역할에 대하여 논함
	2016 (A/HRC/32/37)	- 고등교육의 측면에서 디지털 시대 교육권의 쟁점과 과제에 대하여 2016년 인권이사회 32차 세션에 제출된 보고서 - 교육권에 내재된 규범과 원칙을 디지털 기술을 수용하며 유지할 방안에 대하여 논함
	2016 (A/71/358)	- 평생교육과 교육권에 대하여 2016년 유엔총회 71차 세션에 제출된 보고서 - 평생교육에서 정보통신기술의 역할에 대하여 논함

안전, 위생적, 건강 및 지속가능한 환경의 향유와 인권 의무 문제 특별보고관	2019 (A/HRC/40/55)	- 건강한 환경권 및 대기 오염에 대하여 2019년 인권이사회 40차 세션에 제출된 보고서 - 고가의 대기질 측정소와 다른 위생 기술을 뛰어넘는 가능성을 보이는 신기술에 대하여 언급함
의사 표현의 자유 특별보고관	2021 (A/HRC/47/25)	- 가짜뉴스와 의사 표현의 자유에 대하여 2021년 인권이사회 47차 세션에 제출된 보고서 - 가짜뉴스 배포에 있어 디지털 기술의 역할을 검토함
	2019 (A/74/48050)	- 온라인 ‘혐오 표현’ 규제에 적용되는 인권법에 관하여 2019년 유엔총회 74차 세션에 제출된 보고서
	2019 (A/HRC/41/35)	- 민간 감시기술과 의사 표현의 자유 등 인권에 관하여 2019년 인권이사회 41차 세션에 제출된 보고서 - 부록(A/HRC/41/35/Add.3)에서 보고서 준비 중 접수된 의견들에 대하여 개괄하고 있음
	2018 (A/73/348)	- 정보환경에서 인공지능 기술이 인권에 미치는 의미에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서 - 의사 표현의 자유, 프라이버시권 및 차별금지 문제에 주목함
	2018 (A/HRC/38/35)	- 이용자 콘텐츠 규제에 대하여 2018년 인권이사회 38차 세션에 제출된 보고서 - 부록(A/HRC/38/35/Add.1)에서 보고서 준비 중 접수된 의견들에 대하여 개괄하고 있음
	2018 (A/HRC/38/35/Add.5)	- 디지털 시대 의사 표현의자유 행사에 있어 암호화 및 익명성에 대한 2015년 보고서 후속으로 2018년 인권이사회 38차 세션에 제출된 보고서 - 이전 보고서 이후 새로운 동향과 문제를 검토함
	2017 (A/HRC/35/22)	- 인터넷과 통신 접속서비스 제공에서 민간 부문이 수행하는 역할 등 디지털 접속서비스 제공자의 역할에 대하여 2017년 인권이사회 35차 세션에 제출된 보고서 - 결론에서 민간 부문에서 인권을 보장하는 지침이 될 수 있는 일련의 원칙을 제시함 - 부록(A/HRC/35/22/Add.4)에서 보고서에 부대되는 자료들을 포함함
	2016 (A/71/373)	- 최신 의사 표현의 자유의 문제에 대하여 2016년 유엔총회 71차 세션에 제출된 보고서 - 대규모 표적 감시와 인터넷 차단 등을 다룸

	2016 (A/HRC/32/38)	- 디지털 시대 표현의 자유, 국가 규제 및 민간부문에 대하여 2016년 인권이사회 32차 세션에 제출된 보고서 - 내용규제, 인터넷 차단, 망중립성, 감시 및 디지털 보안의 문제를 다루었으며, 이 주제에 대하여 각국가 및 시민사회 의견을 제출받음
	2015 (A/70/361)	- 원자료 및 내부고발자 보호에 대하여 2015년 유엔총회 70차 세션에 제출된 보고서 - 관련 감시 문제 및 그로부터 보호를 증진하는 암호화 및 익명화 프로그램 등 디지털 도구 문제를 포함하였으며, 이 주제에 대하여 각국가 및 시민사회 의견을 제출받음
	2015 (A/HRC/29/32)	- 디지털 시대 의사 표현의 자유 행사에서 암호화 및 익명성에 대하여 2015년 인권이사회 29차 세션에 제출된 보고서 - 이 주제에 대하여 각국가 및 시민사회 의견을 제출받고, 추가적인 참고 문헌을 포함함
	2013 (A/HRC/23/40, A/HRC/23/40/Corr.1)	- 프라이버시 및 의사 표현의 자유 행사에 있어 국가의 통신 감시가 가지는 의미에 대하여 2013년 인권이사회 23차 세션에 제출한 보고서 - 통신 감시, 데이터, 인터넷 필터링 및 내용 규제를 포함함
	2012 (A/67/357)	- 혐오 표현 및 증오 선동에 대하여 2012년 유엔총회 67차 세션에 제출된 보고서 - 혐오 표현의 온라인 전파, 내용 삭제, 온라인 익명성을 다룸
	2011 (A/66/290)	- 인터넷에서 의사 표현의 자유 행사에 대하여 2011년 유엔총회 66차 세션에 제출된 보고서 - 온라인 컨텐츠에 대한 접근과 인터넷 접속 문제, 디지털 리터러시의 문제를 다룸
	2011 (A/HRC/17/27)	- 모든 사람이 인터넷에서 모든 유형의 정보와 사상을 추구하고 수신하고 전달할 권리의 주요 동향과 문제에 대하여 2011년 인권이사회 17차 세션에 제출된 보고서 - 컨텐츠 접근, 인터넷 인프라, 온라인 정보 검열 증가 및 사이버공격에 대하여 다룸
	2007 (A/HRC/4/27)	- 2007년 인권이사회 4차 세션에 제출된 보고서 - 인터넷 거버넌스 기구의 추진 및 인터넷 자유에 대한 인권적 접근에서 상업적 압력을 제한하는 역할 등 4가지 주제를 분석함
	1998 (E/CN.4/1998/40)	- 정보에 대한 동등한 접근의 권리 및 의사 표현의 자유 행사에 있어 신정보기술의 영향에 대하여 1998년 인권위원회 54차 세션에 제출된 보고서

평화적 집회 및 결사의 자유 특별보고관	2021 (A/HRC/47/24/Add.2)	- 평화적 시위에서 인터넷 셧다운의 역할에 대하여 2021년 인권이사회 47차 세션에 제출된 보고서
	2020 (A/HRC/44/50)	- 시민공간을 수호하는 전세계의 10년을 돌아보며 2020년 인권이사회 44차 세션에 제출된 보고서 - “얼굴인식, 인공지능, 해킹도구, 디지털 식별과 같은 기술의 발전은 집회결사의 자유에 복잡한 문제를 낳고 있다. 정부가 대규모 집회를 억압하고 선거시기 반대의 목소리를 침묵시키기 위해 인터넷과 모바일 네트워크에 대한 접속을 끊어버리는 일이 증가하고 있다. 시민사회의 많은 이들에게 인터넷은 더이상 안전한 장소가 아니며, 이들이 감시의 표적과 온라인 폭력의 대상이 되는 경우가 늘고 있다. 이러한 문제에 대한 해결이 진전을 보지 못하고 있다는 사실은 실천과 책임에 대한 약속을 넘어서야 할 긴급한 필요성을 드러낸다.” 고 지적함
	2019 (A/HRC/41/41)	- 디지털 시대 평화적 집회 및 결사의 자유가 직면한 기회와 문제에 대하여 2019년 인권이사회 41차 세션에 제출된 보고서 - 감시, 온라인 컨텐츠에 대한 자의적 차단 및 필터링, 네트워크 방해 등 국가적 제한에 있어 국가의 의무, 기업의 역할 및 책임, 기회, 동향을 다룸
	2018 (A/HRC/38/34)	- 디지털 공간에서 만나게 되는 장벽 등 평화적 집회및 결사의 자유 행사의 세계적 동향에 대하여 2018년 인권이사회 38차 세션에 제출된 보고서
	2017 (A/HRC/35/28)	- 최근 몇년간 시민사회의 성취에 대하여 2017년 인권이사회 35차 세션에 제출된 보고서 - 시민사회가 어떻게 “디지털 기술을 이용하여 조직하고, 숙려하고 혁신해 왔는지”를 다루고, 동시에 이 기술이 프라이버시, 검열 및 감시 우려를 야기하는 “양날의 검”이라는 점을 지적함
	2014 (A/HRC/26/29)	- 특히 위기 집단에 있어 평화적 집회 및 결사의 자유에 대한 위협에 대하여 2014년 인권이사회 26차 세션에 제출된 보고서 - “결사의 자유는 온라인과 오프라인에 모두 적용되며” “부당하게 인터넷 표현의 자유를 제한하고 사람들이 이 매체에서 결사할 역량을 제한하는 법은 수용될 수 없다”고 지적함
	2013 (A/HRC/23/39)	- 단체를 재정적으로 후원하고 평화적 집회를 개최할 역량에 대하여 2013년 인권이사회 23차 세션에 제출된 보고서 - 신통신기술이 평화적 집회를 촉진하고 조직하는 데 있어서 중요하다는 사실을 포함함

	2012 (A/HRC/20/27)	- 평화적 집회와 결사의 자유를 증진하고 보호하는 모범 관행에 대하여 2012년 인권이사회 20차 세션에 제출된 보고서 - 개인이 평화적 집회를 조직하는 기본 도구로서 소셜미디어 등 인터넷 기술, 기타 정보통신 기술 이용의 중요성 및 이들 기술에 대한 국가의 제한에 대하여 논함
도달 가능한 최고의 신체적·정신적 건강을 누릴 권리 특별보고관	2020 (A/HRC/44/48)	- 정신건강의 권리를 증진하는 인권 기반 세계 의제에서 필요한 요소에 대하여 2020년 인권이사회 44차 세션에 제출된 보고서 - 디지털 감시에 대하여 논함
	2019 (A/HRC/74/174)	- 보건의료 종사자의 교육에 대하여 2019년 유엔총회 제74차 세션에 제출된 보고서 - 보건의료 종사자들이 적절히 역할을 수행하기 위하여 필요한 필수의약품, 백신 및 의료제품 등과 같은 보급품으로서 기술 및 통신 기술을 언급함 - “필수 보건사업 및 기술은 생명의학 제품에 국한되지 않아야 하며 효과적인 심리사회적, 인구 기반 공공 보건 사업을 포함해야 한다”고 강조함. 나아가 “의사 등 보건의료 종사자를 위한 중요 통신 기술 및 문화적 인식 교육 시행이 장애인의 보건의료 접근권 증진에 효과적인 것으로 나타”났다고 지적함
	2017 (A/HRC/35/21)	- 정신건강의 권리에 대하여 2017년 인권이사회 35차 세션에 제출된 보고서 - 정신건강 접근성과 더불어 지역사회에 거주하고 참여할 권리는 정신건강의 보장을 종합병원, 1차 의료기관, 사회복지서비스 및 인권 준수적 모바일 기술의 사용에 통합함으로써 달성될 수 있음
	2016 (A/HRC/32/32)	- 도달 가능한 최고의 신체적·정신적 건강을 누릴 청소년의 권리에 대하여 2016년 인권이사회 32차 세션에 제출된 보고서 - 신통신기술의 사용은 청소년이 정보 전달, 데이터 공유, 건강 캠페인 설계, 건강 교육, 사람대사람 교육 및 갈등중재/자문을 통해 건강권을 증진하는 관계망을 구축하고 이용하는 것을 지원할 수 있다고 설명함. 디지털 환경 속 폭력에 대해서도 보호받을 권리가 적용되며 국가는 사이버폭력에 대응하는 조치를 취해야 함
인권옹호자 특별보고관	2019 (A/74/159)	- 인권옹호자 인권 침해에 대한 지속적인 불처벌과 그 대책에 관하여 2019년 유엔총회 74차 세션에 제출된 보고서 - 디지털 공격은 복잡한 조사를 필요로 하며 대부분의 국가는 인권 옹호자에 대한 디지털 공격에 동원할 수 있는 기술과 막강한 소프트웨어를 확보할 수 있는 자원을 보유하고 있으며

		공공 정보에 대한 접근을 제한하는 법들이 존재하고 독립적인 책임 체제의 부재가 확보한 기술이 어떻게 사용될지 결정하는 것은 물론 책임체제의 확립도 불가능하게 만든다고 언급함 - 국가는 디지털 기술을 감시에 사용하는 것을 모니터링하고 조사하는 독립 체제를 수립하여 그러한 사용이 합법성, 필요성 및 목적의 정당성에 부합하는지 보장할 것을 권고하고, 감시 기술 판매 기업은 그 기술들이 인권 침해적 방식으로 사용되는 사실이 드러날 경우 이를 자제할 것을 권고함
법관과 변호사의 독립에 관한 특별보고관	2019 (A/74/176)	- 사법 독립 기본 원칙이 사법 독립 보장 방안으로서 수행한 주요 역할과 법관 및 변호사의 독립 및 세계 사법 체계 전반 측면에서 당면한 위협과 과제의 실태에 대하여 2019년 유엔총회 74차 세션에 제출된 보고서 - 신기술과 소셜 미디어의 부적절한 사용은 사법 체계에 대한 대중의 인식에 부정적인 영향을 미칠 수 있고 사법 제도 기능에 대한 대중의 신뢰를 해칠 수 있다고 지적함
인권 침해 및 자기결정권 행사를 저해하는 수단으로서 용병 사용에 대한 실무그룹	2020 (A/HRC/45/9)	- 이민 및 국경 관리에 있어 민간 군사 및 보안 서비스 이용이 이주민의 권리 보호에 미치는 영향에 대하여 2020년 인권이사회 45차 세션에 제출된 보고서 - 국경 보안 기술 및 모니터링 서비스에 대하여 주목함
이주민 인권 특별보고관	2019 (A/HRC/41/38)	- 젠더 관점으로 이민이 이주 여성에 미친 영향에 대하여 2019년 인권이사회 41차 세션에 제출된 보고서 - 이민당국 및 공공서비스 간에 정보 차단으로 이주 여성들이 추방에 대한 공포 없이 자신의 권리를 행사할 필요성을 논함
	2018 (A/73/178/Rev.1)	- 이주민의 효과적인 사법 접근에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서 - 추방에 대한 공포 없이 이주민의 사법 접근권을 보장하는 “차단장벽” 구축 추진의 필요성을 강조함
소수자 문제 특별보고관	2021 (A/HRC/46/57)	- 2021년 인권이사회 46차 세션에 제출된 보고서로 소셜미디어 혐오발언에서 소수자를 표적으로 삼는 현상이 만연한 문제를 다룸 - “경찰 및 정보기관이 인공지능 얼굴인식과 디지털 기술을 사용하는 것은 … 차별 금지, 이동의 자유, 표현과 결사의 자유, 특히 소수자 및 원주민 등 특정 집단의 권리에 대하여

		간섭할 수 있”고 “경찰이 얼굴인식 기술을 이용하여 특정한 소수자를 표적으로 삼는 것은 인종적인 프로파일링을 수행하고 소수자 집단의 구성원을 특별히 골라낼 수 있다.”고 지적함
	2015 (A/HRC/28/64)	- 미디어에서 소수자에 대한 혐오 발언 및 증오 선동에 대하여 2015년 인권이사회 28차 세션에 제출된 보고서 - 디지털 미디어가 소수자에 공적 토론에 참여할 기회를 제공하면서 미디어 지형이 어떻게 변화하고 있으며, 동시에 온라인 혐오 표현 확산과 관련한 문제가 증가하는 문제에 대하여 논함
미얀마 인권 상황 특별보고관	2019 (A/74/342)	- 미얀마 인권 상황에 대하여 2019년 유엔총회 74차 세션에 제출된 보고서 - 인터넷 셧다운 등 인터넷과 인권 문제 및 혐오표현 금지법안 발의에 대한 우려 등 온라인 표현 규제 문제에 대하여 논함. 자동화가 이루어지고 있지만 인간 중재자가 여전히 필요하다고 지적함. 미얀마에서 운영되는 인터넷 기업들이 정책적 기반 및 내용 규제 절차로 국제인권법의 확립된 원칙을 채택할 것을 권고함
노인의 인권 향유에 대한 독립전문가	2020 (A/75/205)	- 코로나 바이러스 질병이 노인의 인권 향유에 미치는 영향에 대하여 2020년 유엔총회 75차 세션에 제출된 보고서 - 노인의 정보의 권리 및 신기술 사용 문제에 주목함
	2019 (A/HRC/42/43)	- 위기 상황에서 노인의 인권에 대하여 2019년 인권이사회 42차 세션에 제출된 보고서 - 디지털 정보 및 통신 기술, 위성 데이터 및 디지털 정보의 컴퓨터 처리, 생체인식 등 디지털 기술이 위기 지원 서비스 및 위기 관리에 미치는 영향에 대하여 검토함
	2017 (A/HRC/36/48)	- 2017년 인권이사회 36차 세션에 제출된 보고서 - 보조 및 로봇 기술, 인공지능 및 자동화가 노인 인권에 미치는 영향에 대하여 검토함
	2017 (A/HRC/36/48/Add.1)	- 싱가폴 국가방문에 대하여 2017년 인권이사회 36차 세션에 제출된 보고서 - 노인 인권 관련 보조 및 로봇 기술에 대하여 살펴봄
	2015 (A/HRC/30/43)	- 노인의 자율성과 돌봄에 대하여 2015년 인권이사회 30차 세션에 제출된 보고서 - 교육, 훈련, 평생교육에 있어 신기술의 중요성 논함. 지속 교육 및 직업 재활 문제에서 신기술 접근성의 중요성을 지적함. 정보통신 기술에 대한 지식 결핍의 결과에 따른 세대 격차를 완화하고 타인 의존을 줄이기 위해 노인들에게 원격 교육 및 디지털 교육훈련을 제공할 것을 권고함

극빈 및 인권에 관한 특별보고관	2019 (A/74/48037)	- 디지털 기술, 사회보장 및 인권에 대하여 2019년 유엔총회 74차 세션에 제출된 보고서 - 빈곤한 사람들의 인권에 신기술이 미치는 영향을 검토함. 사회보장 및 부조 체제에서 자동화, 예측, 식별, 감시, 탐지, 표적, 처벌하기 위해 사용되는 디지털 데이터 및 기술의 주도성이 증가하고 있으며, 취약계층의 생활 수준을 향상시키기 위해 어떻게 복지 예산을 기술을 통해 변화시킬수 있을 것인지 논하고 몇가지를 권고함
	2019 (A/HRC/41/39/Add.1)	- 영국 국가방문에 대하여 2019년 인권이사회 41차 세션에 제출된 보고서 - 빅데이터, 인공지능, 알고리즘 및 자동화된 의사결정 등 신기술이 특히 복지 체제 기능의 측면에서 빈곤한 사람들의 인권에 미치는 영향에 대하여 검토함
	2018 (A/HRC/38/33/Add.1)	- 미국 국가방문에 대하여 2018년 인권이사회 38차 세션에 제출된 보고서 - 소위 홈리스에 대한 '기관간 연계 시스템'의 인권적 함의 및 예측 분석이 미국 형사 사법 제도의 재판전 결정에 미치는 영향에 대하여 다룸
프라이버시 특별보고관	2021 (A/HRC/46/37)	- 인공지능과 프라이버시 및 아동의 프라이버시에 대하여 2021년 인권이사회 47차 세션에 제출된 보고서
	2020 (A/75/147)	- 코로나바이러스 질병 팬데믹과 프라이버시 문제에 대하여 2020년 유엔총회 75차 세션에 제출된 보고서 - 개인정보보호 및 기술적 감시 문제에 주목함
	2020 (A/HRC/43/52)	- 젠더기반 프라이버시 침해로부터 보호에 대하여 2020년 인권이사회 43차 세션에 제출된 보고서 - "프라이버시 침해는 다층적이고 상호관련적이며 반복적인 형태로 발생하며 디지털 기술로 악화된다. - "프라이버시 침해는 물리적 또는 국가적 경계를 넘어서는 민간 및 공공 환경 속에서 다층적이고 상호 관련적이며 반복적인 여러 형태로 발생하며 디지털 기술로 악화된다. 온라인 프라이버시 침해는 오프라인 프라이버시 침해를 반영하고 확대한다. 디지털 기술은 그 침해의 범위및 강도를 증폭시킨다" 고 지적함
	2019 (A/74/277)	- 건강관련 데이터의 보호 및 이용에 대하여 2019년 유엔총회74차 세션에 제출된 보고서 - 인공지능, 알고리즘 투명성, 빅데이터에 대하여 검토하고 보건의료 알고리즘은 투명하고, 공정하며 예측가능하게 규제할 것을 권고함. 모든 알고리즘 및 인공지능은 법률 및 UN 협약으로

		보호받는 권리에 대한 부작용 모니터링을 증진해야 함. 처리 및 시스템은 알고리즘 편향을 식별하고 해결하기 위해 설계되고 실행되어야 함
	2019 (A/HRC/40/63)	- 젠더 관점에서 보안 및 감시 문제, 건강정보 관련 프라이버시에 대하여 2019년 인권이사회 40차 세션에 제출된 보고서 - 건강정보 전담반은 ‘스마트’ 이식형 건강 기기/보형물들이 이면에서 기업등에 지속적으로 실생활 데이터를 전송함으로써 ‘데이터로서 신체’를 구축하고 법적 절차에 사용하는 문제, 인공지능/기계 학습 및 자동 처리 등과 같은 문제를 확인함
	2018 (A/73/45712)	- 빅데이터/개방데이터 전담반의 활동에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서
	2018 (A/HRC/37/62)	- 2018년 인권이사회 37차 세션에 제출된 보고서 - 특히 감시 및 프라이버시에 주력한 첫 3년 임기 활동에 주목함
	2017 (A/72/540)	- 빅데이터/개방데이터 전담반의 활동에 대하여 2017년 유엔총회 72차 세션에 제출된 중간보고서
	2017 (A/HRC/34/60)	- 국가적, 초국적 관점에서 정부 감시 활동에 대하여 2017년 인권이사회 34차 세션에 제출된 보고서 - 정부 감시에 대한 보다 프라이버시 친화적인 감독 체제를 개괄함
	2016 (A/71/368)	- 2016년 유엔총회 71차 세션에 제출된 보고서 - 빅데이터/개방 데이터, 보안 및 감시, 건강 정보, 기업의 개인정보 처리, “프라이버시 이해 증진” 등 일명 ‘주제별 활동’의 임기 우선순위를 개괄함
	2016 (A/HRC/31/64)	- 2016년 인권이사회 31차 세션에 제출된 보고서 - 프라이버시 및 기술, 빅데이터 분석, 감시 등 임기 우선순위를 개괄함
인종주의, 인종차별, 외국인혐오 및 관련 불관용주의의	2020 (A/75/590)	- 신디지털기술이 이주민, 무국정자, 난민 및 기타 비국적자에 미치는 차별적 영향에 대하여 2020년 유엔총회 75차 세션에 제출된 보고서
	2020 (A/HRC/44/57)	- 인종차별 및 신디지털기술에 대하여 2020년 인권이사회 44차 세션에 제출된 보고서

현대적 형태에 대한 특별보고관	2018 (A/73/312)	- 네오나치즘 및 관련 불관용주의 확산에서 디지털 기술의 현대적 사용에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서 - 이 문제를 해결하기 위한 국가 및 기술 기업의 모범 관행도 소개함
	2018 (A/HRC/38/53)	- 최근 우려스러운 나치즘 및 네오나치즘 이념의 이동 및 지지 및 숭배 경향에 대하여 2018년 인권이사회 38차 세션에 제출된 보고서 - 네오나치즘 이념의 전파에서 기술의 역할을 검토함
	2014 (A/HRC/26/49)	- 인터넷 및 소셜미디어의 인종주의 징후에 대하여 2014년 인권이사회 26차 세션에 제출된 보고서 - 인터넷 및 소셜미디어에서 인종주의 상황, 주요 동향, 징후에 대하여 검토하고 국제적, 지역적, 국내적 수준에서 취해지는 법정책 체제 및 수단 뿐 아니라 인터넷 및 소셜미디어 사업자가 채택한 규제 정책들을 개괄함
	2012 (A/67/326)	- 인종주의 사상 및 인종적 증오 및 폭력 선동 전파를 위한 인터넷 사용 증가로 인한 주요 문제 및 과제에 대하여 2012년 유엔총회 67차 세션에 제출된 보고서 - 인종주의, 인종 차별, 외국인 혐오 및 관련 불관용주의에 대응하는 효과적인 도구로서 인터넷의 역할 가능성 및 긍정적인 기여 또한 강조함
종교・신념 특별보고관	2019 (A/HRC/40/58)	- 2019년 인권이사회 40차 세션에 제출한 보고서 - 상호밀접하고 상호보완적인 권리로서 종교 및 사상의 자유와 표현의 자유를 살펴보고, 얼굴인식 기술의 사용 등 온라인 플랫폼과 관련 제한의 영향 문제를 논함
아동 매매, 아동 성매매 및 아동포르노 특별보고관	2020 (A/HRC/43/40)	- 2020년 인권이사회 43차 세션에 제출된 보고서 - 이전 보고서 이후 특별보고관 활동을 개괄하며 ICT와 아동 성착취간의 관계에 주목함
	2019 (A/74/162)	- 대리모 출산 아동권 보호장치에 대하여 2019년 74차 세션에 제출된 보고서 - 자신의 건강권에서 중요할 수 있는 보조생식기술과 출생에 관한 정보에 아동이 접근할 권리를 논함
	2018 (A/HRC/37/60)	- 대리모 및 아동 매매에 대하여 2018년 인권이사회 37차 세션에 제출된 보고서 - 보조생식 신기술과 관련한 문제를 다룸

	2015 (A/HRC/28/56)	- 정보통신기술과 아동 매매 및 성착취 문제에 대하여 2015년 인권이사회 28차 세션에 제출된 보고서
	2009 (A/HRC/12/23)	- 인터넷 아동포르노에 대하여 2009년 인권이사회 12차 세션에 제출된 보고서
	2005 (E/CN.4/2005/78)	- 인터넷 아동포르노에 대하여 2005년 인권위원회 61차 세션에 제출된 보고서
성소수성(SOGI)에 대한 폭력 및 차별 독립전문가	2019 (A/HRC/41/45)	- 성소수성에 대한 폭력 및 차별에 대한 인식을 제고하는 수단으로서 정보 수집과 관리에 대하여 2019년 인권이사회 41차 세션에 제출된 보고서
현대적 노예 형태 및 그 원인과 결과에 대한 특별보고관	2020 (A/75/166)	- 2020년 유엔총회 75차 세션에 제출된 보고서로 새로운 임기의 우선순위를 개괄함 - 기술과 노예의 새로운 형태의 관련성에 대하여 논함
	2019 (A/HRC/42/44)	- 노예반대를 위한 최근의 노력이 확산중인 현대적 노예 형태에 효과적으로 대응할 수 있는지, 이런 노력이 장래의 현대적 노예 형태와 징후를 해결하는데 적절할 것인지에 대하여 2019년 인권이사회 42차 세션에 제출된 보고서 - 현대적 노예 형태의 유형과 요인이 주요 기술 변화에 영향을 받는 방식에 대하여 설명하고, 노예제 반대 활동은 “업무 효과를 높이고 재정 접근의 금융에 대해 새로운 접근방식을 취하는 디지털 기술”의 사용으로 “스마트”해져야 함을 주장함
수단 인권 상황 독립전문가	2019 (A/HRC/42/63)	- 수단 인권 상황에 대하여 2019년 인권이사회 42차 세션에 제출된 보고서 - 인터넷 셧다운 등 언론인에 대한 표현의 자유 제한 및 괴롭힘에 대하여 논함
특별 사법, 즉결 또는 임의 형집행에 대한 특별 보고관	2020 (A/75/384)	- 집단매장지에 대하여 2020년 유엔총회 75차 세션에 제출된 보고서 - 역사적으로 세계적으로 수많은 집단 학살지와 불법적인 사망 사건들에 대하여 강조하고, 집단매장지의 발견와 관리에 기여하는 디지털 기술을 검토함
	2020 (A/HRC/44/38)	- 2020년 인권이사회 44차 세션에 제출된 보고서 - 군사드론을 이용한 표적 사살 문제와 지난 5년간 드론 사용 확산 및 강화된 성능 문제를 살펴봄. 그 사용의 규제와 책임성 증진에 대하여 권고함
	2016 (A/71/372)	- 2016년 유엔총회 71차 세션에 제출된 보고서 - 지난 임기 분야에 대한 업데이트 및 최신 생명권 문제에 대한 논평하며 군사드론, 자율무기

		등 신기술의 영향과 생명권 보호 문제 및 정보통신기술 실태조사를 포함함
	2015 (A/HRC/29/37)	- 생명권 보호를 위한 정보통신기술 사용에 대하여 2015년 인권이사회 29차 세션에 제출된 보고서
	2014 (A/69/265)	- 법집행에서 덜치명적인 무인 무기에 대하여 2014년 유엔총회 69차 세션에 제출된 보고서 - 기술 진보에 대한 구체적인 규제 체제가 필요하다는 관점을 피력함
	2014 (A/HRC/26/36)	- 법집행에서 생명권 보호에 대하여 2014년 인권이사회 26차 세션에 제출된 보고서 - 드론 사용과 관련된 국제 무력 사용 규칙 해석에 있어 법적 불확실성을 논하고, 인권이사회가 원격 항공기 및 군사드론 사용에 대한 법 체제의 기본 개요를 명시할 것을 제안함. 인권이사회에 자율 무기 시스템 문제에 관여할 것을 요청함
	2013 (A/68/382)	- 생명권 보호 관점에서 군사드론 등 치명적 무력 사용에 대하여 2013년 유엔총회 68차 세션에 제출된 보고서 - 국가간 무력 사용을 소관하는 국제법과 국제인권법 하 군사드론의 사용에 대한 상세 사항을 검토함
	2013 (A/HRC/23/47)	- 치명적 자율 로봇과 생명권 보호에 대하여 2013년 인권이사회 23차 세션에 제출된 보고서
	2010 (A/65/321)	- 인권실태조사, 표적 사살 및 책임성과 신기술, 특별사법 집행 및 로봇 기술 간의 관련성에 대하여 2010년 유엔총회 65차 세션에 제출된 보고서
테러리즘 대응에서 인권 및 기본적 권리 증진과 보호에 대한 특별보고관	2021 (A/HRC/46/36)	- 여성, 10대 여성 및 가족에 대한 (폭력적) 극단주의와 테러리즘에 대응하는 정책과 실무에서 인권영향에 대하여 2021년 인권이사회 46차 세션에 제출된 보고서 - “신기술과 개인정보 수집 방법은 특히 소수자에 대해 서로 다른 영향을 미치고 있으며 매우 젠더 편향적이다.” 라고 확인함
	2019 (A/HRC/40/52)	- 테러리즘, 폭력적 극단주의 대응 조치가 시민 공간, 시민사회 활동가 및 인권 옹호자의 권리에 미치는 영향에 대하여 2019년 인권이사회 40차 세션에 제출된 보고서 - 테러 목적의 인터넷 이용에 대하여 검토함
	2017 (A/HRC/34/61)	- 역외 치명적 대테러 작전에서 원격 항공기 사용 및 테러대응 목적의 대량 디지털 감시 등 특별보고관 이전 보고서 상의 원칙 문제와 관련한 상황 발전에 대하여 2017년 인권이사회 34차 세션에 제출된 보고서

	2014 (A/69/397)	- 테러 대응과 디지털 대량 감시에 대하여 2014년 유엔총회 69차 세션에 제출된 보고서
	2014 (A/HRC/25/59)	- 역외 치명적 대테러 작전에서 드론 사용의 시민적 영향에 대하여 2014년 인권이사회 25차 세션에 제출된 보고서
고문 및 기타 잔인하고 비인간적이며 모멸적인 취급 및 처벌에 대한 특별보고관	2020 (A/HRC/43/49)	- 심리적 고문에 대하여 2020년 인권이사회 43차 세션에 제출된 보고서 - 신기술로 인한 현대적 가능성 및 과제에 대응하는 고문 금지 조항의 해석이 요구됨. 예비적 방안으로 “사이버고문”으로 서술될 수 있는 기본 개요에 대하여 검토함
	2018 (A/73/207)	- 고문 및 부당 대우에 대한 완전한 금지의 국제적 이행과 관련한 발달 상황과 우선적인 과제에 대하여 2018년 유엔총회 73차 세션에 제출된 보고서 - 신기술 및 장비가 고문 및 부당 대우를 금지하고 가해자 책임 촉구를 지원할 방법이 무엇인지, 새로운 유형의 무기, 장비 및 기술의 등장이 그 오남용에 대한 심각한 우려를 낳고 있음을 논함. 국가는 신무기, 구속 및 기타 장치 및 기술의 사용이 고문과 부당 대우 금지 원칙 및 국제법 하 여타 의무를 침해하는지 여부, 또는 그러한 침해가 야기한 위험이 심각하게 증가하는지 여부를 판단하는 데 대한 체계적인 법적 검토를 수행하고 권고함
여성 폭력 및 그 원인과 결과에 대한 특별보고관	2018 (A/HRC/38/47)	- 인권 관점에서 여성에 대한 온라인 폭력 및 정보통신기술에 의해 조장된 폭력에 대하여 2018년 인권이사회 38차 세션에 제출된 보고서