

발 간 등 록 번 호

11-1620000-000916-01



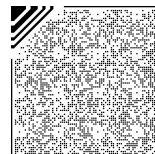
인 권 지능

지키는

인공지능
개발과 활용에 관한
인권 가이드라인
해설서



국가인권위원회



인공지능

인공지능
개발과 활용에 관한
인권 가이드라인
해설서

목차

개요	1. 인권 가이드라인 제정의 배경 / 04
	2. 인권 가이드라인의 목적과 의미 / 08
	3. 용어의 정의 / 10
내용	4. 인간의 존엄성 존중 / 14
	5. 투명성과 설명 의무 / 16
	6. 자기결정권의 보장 / 22
	7. 차별 금지 / 26
	8. 인공지능 인권영향평가 시행 / 32
	9. 위험도 등급 및 관련 법제도 마련 / 36



인공지능
개발과 활용에 관한
인권 가이드라인
해설서

1 개요

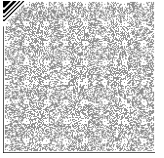
1

인권 가이드라인 제정의 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 바탕으로 다양한 영역에서 사람의 판단을 대신할 수 있는 인공지능은 점차 적용영역을 넓혀 사회 전반과 개인의 삶에 강력한 영향력과 과급력을 행사할 것입니다.

3 인공지능의 발전과 확산은 국가경쟁력과 개인 삶의 질을 높일 것으로 기대되지만 개인정보 및 사생활에 대한 권리, 차별받지 않을 권리 등 인권을 침해하는 문제들도 대두되고 있습니다.



인공지능에 의한 인권침해 사례

- 미국 아마존(社)의 경우, 인공지능 기반 채용 시스템이 개발자·기술직군에 대부분 남성만을 추천하는, 데이터 편향성 문제가 발생하여 해당 시스템의 사용을 중단했습니다.
- 미국의 인공지능 기반 범죄 예측 프로그램인 COMPAS는 재범 예측 과정에서 흑인 범죄자의 재범 가능성을 백인보다 2배 이상 높게 예측하는 편향이 나타났습니다.
- 우리나라에서도 채용 시 인공지능을 활용하는 경우가 증가하고 있습니다. 정필모 의원실에 따르면, 인공지능이 A와 B+로 평가한 합격자에 대해 면접관들은 하위 2등·1등으로 평가하였고, 면접관들이 1·2등으로 평가한 합격자는 인공지능으로부터 B0, B-의 평점을 받는 과리가 있었습니다.
- 금융 분야에서도 다양한 용도로 인공지능을 많이 활용하고 있습니다. 그런데 인공지능 의사결정 과정에 사용된 데이터에 편향성이 있거나 정확성이 떨어질 경우, 시장 리스크와 불확실성을 확대 해석하는 등 금융시장 불안정성을 증폭시킬 가능성이 있습니다. 인공지능 의사결정 과정에 대한 충분한 설명 없이 결과만 제공받는 경우, 알고리즘에 대한 정보 부족으로 인해 대처방안을 마련하는 것이 어렵습니다.



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 상황**입니다.

5 따라서 인공지능을 개인의 삶과 사회적 공익에 기여할 수 있도록 설계하며, 인간의 존엄성과 자기결정권 및 차별을 받지 않을 권리의 보장 등 **기본적 인권에 기반을 두도록 하는 것이** 매우 중요합니다.

인공지능으로 완전히 자동화된 의사결정과 행정처분이 가능해졌습니다

2021. 9. 24. 제정 시행된 「행정기본법」 제20조는 “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.”라고 규정하고 있는데, 이는 인공지능이 단순한 지원을 넘어 의사결정과 행정처분을 할 수 있는 근거가 됩니다.

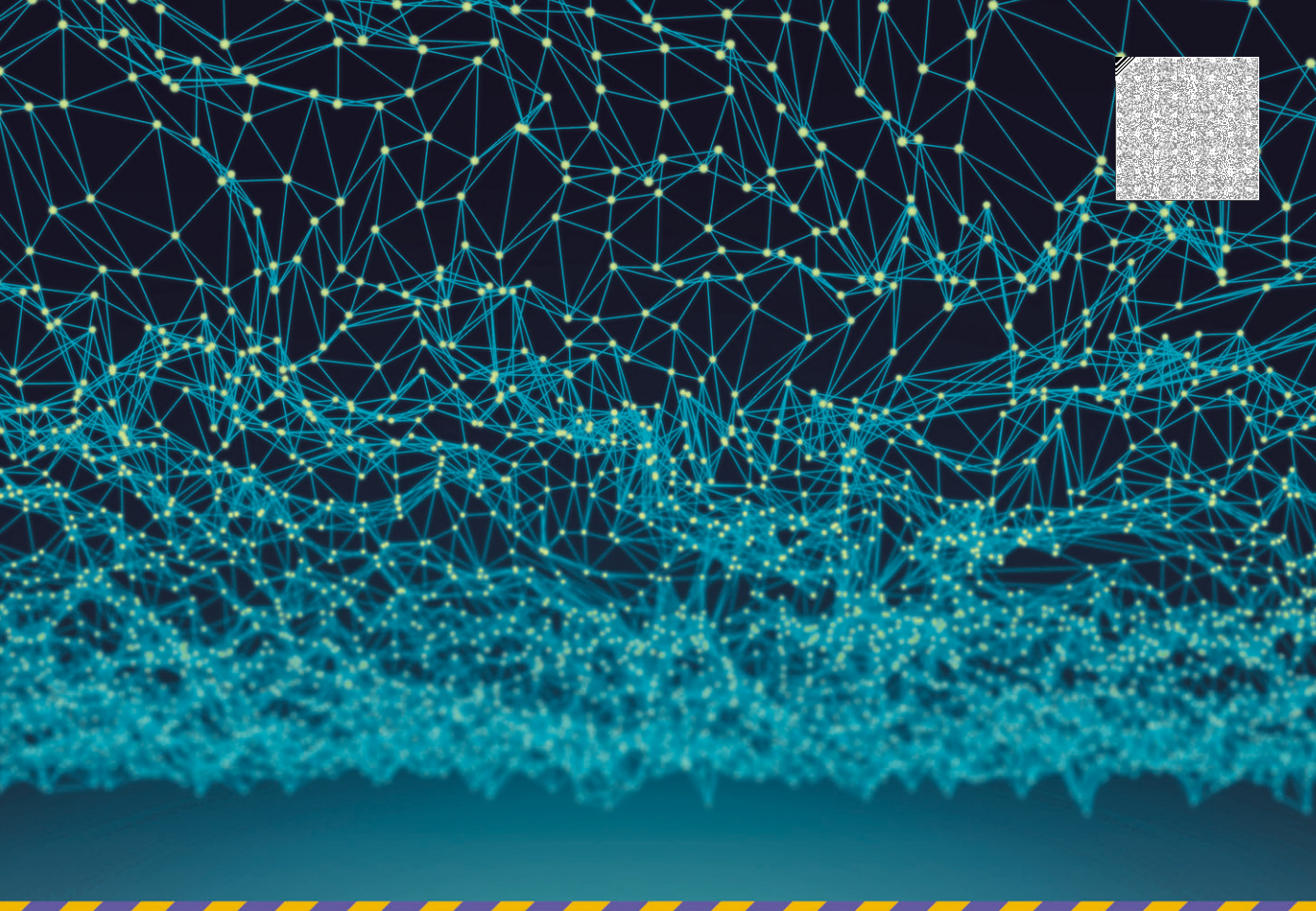
완전히 자동화된 처분에 대하여 사전통지, 이의권 등을 행사하기 어렵습니다

행정처분의 대상이 되는 사람에게는 처분 경위에 대해 설명을 받을 수 있는 권리, 처분으로 인한 피해를 구제받을 권리가 있습니다. 그러나 인공지능으로 완전히 자동화된 처분의 대상자가 되는 사람이 자동화된 처분이 이루어진다는 사실, 인공지능의 논리, 자신에게 미칠 중대성이나 예상 결과를 모른다면 제대로 설명을 요구하거나 이의권을 행사하기 어렵기 때문에 구제받을 권리를 위축될 수 있습니다.

권리구제가 제약될 우려가 있습니다

정부는 인공지능 산업을 정책적으로 적극 지원하고 있는 반면, 권리구제 마련에는 다소 소극적입니다. 유엔 인권이사회는 인공지능으로 인한 인권침해가 발생한 경우 그 피해자가 효과적인 구제 수단에 접근할 수 있도록 법제도를 마련할 것을 각국에 권고하였습니다. 그러나 현재 우리나라 법제도상으로는 실효적 권리구제 절차와 방법이 미흡합니다.

과학기술정보통신부가 2021. 5. 13. 발표한 신뢰할 수 있는 <인공지능 구현 전략 발표>에 따르면 권리구제에 관한 내용이 기획·검토 단계에 있을 뿐입니다.



인공지능 개발과 활용에 관한 인권 가이드라인

6 본 가이드라인은 인권적 관점에서 인공지능의 개발과 활용의 전 과정에서 **인권침해를 예방하기 위하여 준수해야 할 원칙과 내용**을 제시하고자 마련되었습니다.

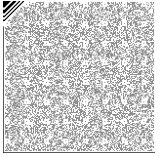
2

인권 가이드라인의 목적과 의미

7 국가인권위원회는 「국가인권위원회법」 제19조 제6호에 따라 인권침해의 유형, 판단기준 및 그 예방조치 등에 관한 지침을 제시할 권한을 가지고 있으므로, 인공지능에 의한 **인권침해와 차별을 판단하고, 개선을 권고하거나 구제절차를 마련하는 데 필요한 기준을 제시하고자** 합니다.

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

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



인공지능 전반에 대해 인권적 기준을 제시합니다

이미 인공지능에 관련한 유사한 가이드라인들이 다수 있지만 대체로 특정 영역에 한정되어 있습니다. 반면, 인권 가이드라인은 인공지능 전반에 걸쳐 구체적인 인권 기준을 제시하고 있습니다. 유엔 인권이사회 <신기술과 인권 결의(2021)>는 인공지능 등 신기술에 대응하는 인권의 증진 및 보호는 전체적이고, 포용적이며, 포괄적인 접근 방식을 따를 필요가 있다고 지적해 왔습니다.

개인정보보호위원회의 <인공지능(AI) 자율점검표(2021.5.)>는 개인정보보호법의 보호 기준을 인공지능에 어떻게 적용해야 하는지에 관한 체크리스트로 의미가 있으며, <금융 분야 AI 개발·활용 안내서(2022. 8.)>는 금융기관의 이용자에 대한 설명의무라는 법적 의무를 전제로 한다는 점에서 해당 분야에 한정되고 있습니다. 과학기술정보통신부의 <인공지능(AI) 윤리기준(2020.12.)>은 산업·경제 분야의 자율 규제 환경을 조성함으로써 인공지능 연구개발과 산업 성장을 제약하는 것을 목표로 내세운다는 점에서 구속력 있는 인권 준수를 요구하는 본 가이드라인과 차이가 있습니다. 그 외 방송통신위원회의 <이용자 중심의 지능정보사회를 실현하기 위한 원칙(2019.11.)>등이 있습니다.

10 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.

강제성은 없으나, 실행력을 갖고 있습니다

본 가이드라인은 인권 기준을 제시하고 그에 대한 자발적인 준수를 권장하는 바, 강제성을 갖고 있지는 않습니다. 하지만 본 가이드라인이 국가인권위원회의 정책권고를 기반으로 하고 있고, 위원회 구제 절차 시 해석의 기준이 되기에 간접적인 실행력을 갖추고 있다고 할 수 있습니다.

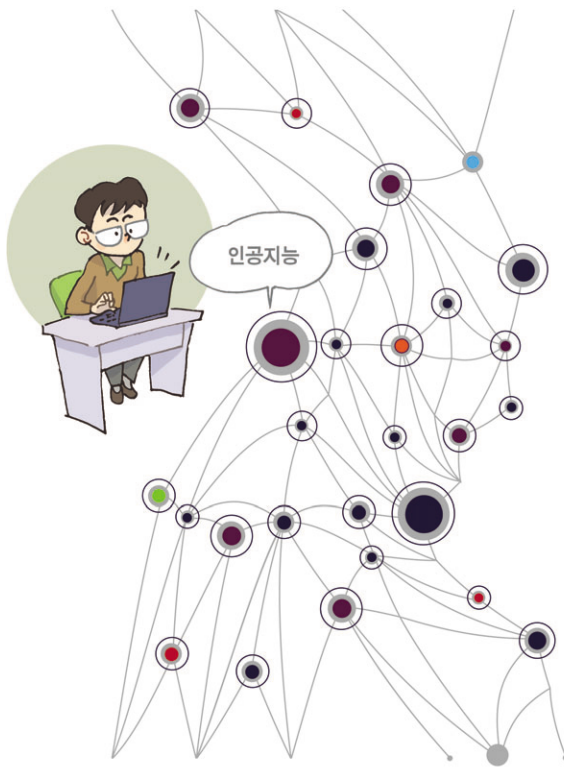
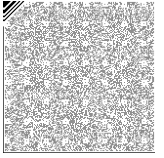
3

용어의 정의

11 인공지능의 개발과 활용을 위한 가이드라인에서 의미하는 **인권**은 「대한민국헌법」과 법률에서 보장하거나 대한민국이 가입·비준한 국제인권조약 및 국제관습법에서 인정하는 **인간으로서의 존엄과 가치 및 자유와 권리**를 의미합니다.

12 본 가이드라인에서 말하는 **인공지능**은 일차적으로는 학습과 추론, 판단을 전자적으로 구현하는 알고리즘과 해당 프로세스를 지칭하나, 이차적으로는 빅데이터 등 **인공지능을 기능하게 하는 일련의 기술들**을 포함하고 있습니다.

13 ‘**영향을 받는 당사자(affected individuals)**’는 국가나 기업의 규정 또는 행위 등으로 인하여 인공지능의 적용대상이 되고, 직·간접적으로 인권에 **영향을 받는 개인 또는 집단**을 의미합니다.



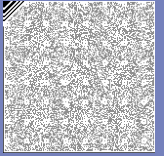
인공지능에 대한 정의는 매우 다양합니다

인공지능의 스펙트럼이 넓고 지금도 계속 발전하고 있는 기술이기에, 현시점에서 하나의 단어로 정의하는 것은 쉽지 않습니다.

- 우리나라 「지능정보화기본법」제2조에 따르면 ‘지능정보기술’이란 전자적 방법으로 학습·추론·판단 등을 구현하는 기술, 데이터를 전자적 방법으로 수집·분석·가공 등 처리하는 기술, 물건 상호간 또는 사람과 물건 사이에 데이터를 처리하거나 물건을 이용·제어 또는 관리할 수 있도록 하는 기술, 클라우드 컴퓨팅 기술, 무선 또는 유·무선이 결합된 초연결 지능정보통신 기반 기술 등을 뜻합니다.
- 유럽평의회의 〈알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고(2020)〉에 따르면 ‘알고리즘 시스템’이란 수학적 최적화 기술을 사용하여 데이터의 수집, 결합, 정리, 정렬, 분류 및 추론뿐 아니라 선택, 순위 지정, 권장 사항 도출 및 의사결정과 같은 업무를 하나 이상 수행하는 응용 프로그램으로 이해됩니다.
- 유럽연합 〈인공지능 백서(2020년)〉는 인공지능을 데이터, 알고리즘, 컴퓨팅 연산 능력 등을 결합한 기술의 집합으로 정의하고 있습니다.
- 유럽연합 「인공지능법(안)(2021년 제안)」은 인공지능 시스템을 기계 학습, 논리·지식기반 또는 통계적 접근 방식으로 개발된 소프트웨어로서, 인간이 정의한 목표에 따라 상호작용하는 환경에 영향을 미치는 콘텐츠 생산, 예측, 추천 또는 의사결정을 할 수 있는 시스템으로 정의하고 있습니다.

본 가이드라인에서의 인공지능에 대한 정의

인공지능에 대한 여러 정의들을 종합하여 볼 때, 인공지능이란 학습·추론·판단을 전자적 방법으로 구현하는 알고리즘 시스템뿐만 아니라 빅데이터, 사물인터넷, 플랫폼 등 인공지능을 가능하게 하는 일련의 기술까지 포함하는 것으로 볼 수 있습니다.



인공지능
인권

인공지능
개발과 활용에 관한
인권 가이드라인
해설서

2 내용

4

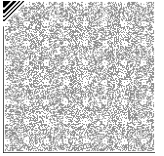
인간의 존엄성 존중

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

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

16 인공지능은 인간으로서의 존엄과 가치 및 행복을 추구할 권리에 부합하는 방향으로 개발 및 활용되어야 하며, 개인의 선택과 판단 및 결정을 강요하거나 자율성을 침해해서는 안 됩니다.

17 인공지능의 개발과 활용은 개인의 행복과 사회적 공공성의 증진에 위배되어서는 안 되고, 표현의 자유, 집회 및 결사의 자유, 노동권 등 인권을 인공지능의 부정적 영향으로부터 보호해야 합니다.



인공지능과 인권

국가인권위원회는 이미 2020. 4. 2. 「인공지능산업 진흥에 관한 법률(안)」(김경진 의원 대표발의)에 대해 인공지능에서 인권 및 인간 존엄성 존중 원칙, 인공지능으로 인한 차별 방지 원칙, 기본계획 수립 시 인권 보호에 관한 사항이 추가로 규정될 필요가 있다고 의견 표명한 바 있습니다.

- 유엔 인권이사회는 <디지털 시대 프라이버시권(2021)> 결의에서 인공지능 수명주기에 걸쳐 인권 실사 및 모니터링과 점검을 시행하는 등, 인공지능 시스템이 인권에 미치는 영향에 대응하는 노력을 증진할 것을 요구하였습니다.
- 유엔 인권최고대표는 2021년 <디지털시대 프라이버시권> 보고서에서 “인공지능은 어떻게 개발되고 활용되는지에 따라 개인정보 및 사생활의 권리, 차별받지 않을 권리뿐 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.”라고 하였습니다.
- 2018년 유엔 의사표현의자유 특별보고관은 “개인이 의견을 형성하고 보유할 수 있는 역량, 정보 환경에서 접근하고 표현할 수 있는 역량을 인공지능이 대체 또는 조작하거나 방해해서는 안 된다.”라고 지적하면서 “인공지능 관련 입법은 이용자가 알고 선택하고 통제할 수 있도록 개인의 자율성을 존중해야 한다.”라고 강조하였습니다. “특히 인공지능의 추론과 예측은 사람들이 자신의 자율성과 정체성을 수립하는 데 깊은 영향을 미친다.”라고 하였습니다.
- 유럽평의회는 <알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고(2020)>에서 “알고리즘 시스템의 설계, 개발, 판매, 배치, 구현 및 서비스에 관여하는 민간 부문의 행위자는 공적·사적 모든 영역에서 인권 실사를 실시해야 한다. 그들은 고객은 물론 자신의 활동에 의해 영향을 받는 여타 당사자들의 기본적 자유와 인권을 존중할 책임이 있다.”라고 하여, 민간 행위자에게도 인권 존중의 책임이 있음을 분명히 하였습니다.

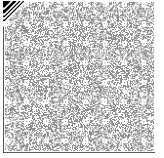
5

투명성과 설명 의무

18 알 권리란 개인이 의사형성에 필요한 정보를 수집하고, 수집된 정보를 취사·선택할 수 있는 자유를 의미하며, 알 권리를 통해 개인은 충분한 정보를 획득할 수 있고, 지식과 이해의 폭을 넓혀 합리적인 판단과 자신의 인격을 발현시킬 수 있습니다.

19 이러한 알 권리의 보장과 인공지능이 미치는 영향력과 중요성을 감안할 때, **인공지능의 판단과정과 그 결과에 대한 적절하고 합리적인 설명**이 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.



불투명한 인공지능 시스템

유엔 인권최고대표는 <디지털 시대 프라이버시권(2021)>에서 “많은 인공지능 시스템의 의사 결정 과정은 불투명하다. 인공지능 시스템의 개발 및 운영을 뒷받침하는 정보환경, 알고리즘, 모델의 복잡성은 물론 정부와 민간 행위자들의 의도적인 비밀주의는 일반 대중으로 하여금 인공지능 시스템이 인권과 사회에 미치는 영향을 이해하지 못하도록 방해하는 요인이다.”라고 지적하였습니다.

완전 투명해...



알 권리와 투명성

인공지능의 학습, 추론, 판단 과정, 결과에 이르는 이유 등은 설명하기 어려운 특성이 있고, 설명의 범위 역시 논란이 있을 수 있습니다. 그럼에도 불구하고 영향을 받는 당사자가 충분한 설명을 듣고, 참여하고 발언할 수 있는 ‘알 권리’는 반드시 보장되어야 합니다.

이를 위해 국가와 기업은 인공지능의 판단 과정과 결과에 대한 적절하고 합리적인 설명을 가능케 하는 기술적·제도적 장치를 마련하고, 투명성 제고를 위해 노력해야 합니다. 더불어 인공지능의 사용이 사회와 인권에 미치는 영향이 심대할수록 투명성에 가중치를 부여하는 등 선제적이고 적극적인 자세의 건치가 필요합니다.

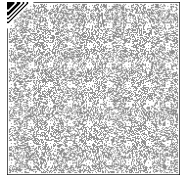
- 금융위원회 <금융 분야 AI 개발·활용 안내서(2022. 8.)>에서는, 금융회사가 「신용 정보의 이용 및 보호에 관한 법률」 제36조의2(자동화 평가 결과에 대한 설명 및 이의 제기 등)를 준수하여 ①자동화 평가 실시 여부, 평가 결과·기준·기초정보 등에 대한 설명 요구권, ②자동화 평가 결과 산출에 유리하다고 판단되는 정보를 제출할 수 있는 정보 제출권, ③기초정보의 정정·삭제, 결과 재산출을 요구할 수 있는 이의 제기권을 포함한 권리구제 방안을 마련하도록 규정하고 있습니다. 더불어 관련 내용을 소비자에게 안내하고 있는지 확인하도록 하고 있습니다.

- 유엔 사무총장은 <사회권의 실현에 있어 신기술의 역할에 대한 보고서(2020)>에서 신기술의 개발 및 도입에 관한 의사결정에 모든 관련 이해당사자의 참여를 보장하고, 특히 공공부문에서 인공지능이 지원하는 의사결정에 대하여 적절한 설명의 보장이 필요하다고 하였습니다.

- 유럽연합은 2020년 7월 12일 발효된 「P2B 온라인 플랫폼 시장 공정성 및 투명성 강화 규칙」에서 플랫폼이 입점 사업자들에게 검색·배열순위 결정과 관련된 매개변수 및 중요도 등을 약관에 공개하도록 하였습니다. 또한 2022년 11월 발효한 유럽연합 「디지털서비스법」은 모든 온라인 플랫폼이 광고가 이용자를 맞춤하기 위해 사용한 주요 매개변수를 공개하고 이용자가 변경할 수 있도록 하였고, 콘텐츠 추천 시스템 또한 주요 매개변수를 공개하고 이용자가 변경할 수 있도록 하였습니다.

- 유럽평의회 인권위원장은 “각국은 인공지능 시스템에 대한 공개 협의를 다양한 단계에서는 물론 최소한 조달 및 인권영향평가 단계에서 도입해야 한다. 공개 협의가 의미가 있으려면 인공지능 시스템의 작동, 기능, 잠재적 또는 식별된 영향을 잘 이해할 수 있는 모든 정보를 시의적절하게 사전 공개해야 한다. 공개 협의는 정부 관계자, 민간 부문 대표자들, 학계, 비영리 부문, 언론 및 영향을 받는 소외 집단이나 단체 대표자를 포함한 모든 이해관계자가 의견을 제시할 기회가 제공되어야 한다. 국가인권기구는 시민사회와 국가 사이의 가교 역할을 하며 유의미한 협의 수행에 도움을 줄 수 있다.”고 밝힌 바 있습니다.

- 미국 연방거래위원회(FTC)는 「인공지능과 알고리즘 사용에 대한 지침」을 통해, 기업의 알고리즘 의사결정에 대하여 소비자에게 설명할 것을 요구했습니다. 즉, 인공지능 알고리즘을 사용하여 소비자에 대한 의사결정을 하는 기업은 제품이나 서비스의 거부에 이르는 이유, 소비자에게 부과된 위험도 점수 및 중요도에 영향을 미치는 요소, 자동화된 방식에 기반하여 거래 조건을 변경하는지 여부 등을 소비자에게 설명할 것을 요구했습니다.



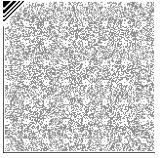
- 호주 국가인권위원회는 〈인권과 기술〉 권고에서 설명을 제공할 수 없는 경우 자동화 또는 인공지능의 사용을 통한 행정적인 결정을 내려서는 안 된다고 권고하였습니다.
- 프랑스 앙티브 시는 알고리즘 주민 공개 제도를 실시하여, 각 알고리즘의 관계 행정부서 연락처 등 행정 정보, 법적 근거나 자동화 수준 등 알고리즘 관련 사업 정보, 영향을 받는 사람 등 의사결정 영향에 대한 정보, 데이터 출처 및 알고리즘 유형 등 알고리즘의 작동 원리에 대한 정보를 공개하였습니다.
- 네덜란드 암스테르담과 핀란드 헬싱키 시는 시민에게 알고리즘 등록부를 시범적으로 공개했습니다. 이 알고리즘 등록부는 시가 도입·운영하는 인공지능 시스템의 학습용 데이터에 대한 정보, 데이터 처리에 대한 정보, 차별 방지에 대한 정보, 인적 감독에 대한 정보, 위험성에 대한 정보 등을 읽기 쉬운 평문으로 공개하고, 책임 부서와 연락처도 함께 공개하여 시민이 의견 제시할 수 있도록 하였습니다.
- 캐나다의 「자동화된 의사결정 지침」에서는 의사결정 전에 의사결정의 전체 또는 부분이 자동화된 의사결정 시스템에 의해 수행된다는 내용을 웹사이트 등을 통해 알리고, 의사결정 후에도 결정이 내려진 방법과 이유에 대해 영향을 받는 개인들이 이해할 수 있는 설명을 하도록 규정하고 있습니다.
- 스페인은 이른바 「라이더 법(디지털 플랫폼 유통에 종사하는 개인의 고용 상태에 관한 법률)」을 제정하고, 플랫폼 노동자의 근로조건, 고용과 해고 결정에 영향을 미칠 수 있는 알고리즘의 논리 등에 관한 정보를 노동자 대표에게 공개하도록 의무화하고 일부 협상할 수 있도록 하였습니다.
- 호주는 「뉴스 미디어 협상법(News Media and Digital Platforms Mandatory Bargaining Code)」을 제정하고 플랫폼이 뉴스 노출 알고리즘을 변경할 때 해당 사항을 언론사에 알리고 협상할 수 있도록 하였습니다.

20 인공지능과 상호작용하는 사람에게는 언제나 **그 상대방이 인공지능이라는 사실을 알려야 합니다.**

23 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.**

사전 통지와 설명

- 유럽연합은 「일반개인정보보호규정(GDPR)」에서 완전히 자동화된 의사결정의 대상이 되는 정보주체에게 자동화된 개인정보 처리를 바탕으로 한 결정에 사용되는 개인정보는 물론, 자동화된 의사결정 여부, 관련 로직(logic), 처리의 중대성 및 예상 결과를 통지하도록 하였고, 이후 정보주체가 인간의 개입 요구, 의견 제시, 이의 제기권의 행사가 가능하도록 하였습니다.
- 유럽연합 집행위원회가 2021년 제안한 「인공지능법(안)」은 원칙적으로 인공지능과 상호작용하는 사람에게 그 상대방이 인공지능이라는 점을 알리도록 하였습니다. 고위험 인공지능은 시스템의 공급자, 시스템의 특성, 기능 및 한계, 시스템 결과 해석을 위한 기술적 조치 및 인적 감독 조치 등에 대하여 사용자에게 명확한 정보를 제공해야 할 뿐 아니라, 일부 고위험 인공지능의 경우 유럽연합 데이터베이스에 그 운용을 추적할 수 있는 추가적인 정보를 등록하여 일반에게 공개하도록 하였습니다.
- 영국의 2020년 <인공지능 조달 지침>은 공공에 사용되는 인공지능의 조달에 있어 설명 가능성과 해석 가능성을 설계 기준으로 삼아 알고리즘의 대상이 되는 사람에 대한 책무성을 요구하였고, 이로써 특정 공급업체에 고착되지 않을 수 있다는 점도 강조하였습니다.



21 공공기관은 인공지능의 개발과 활용 계획 등을 **사전 공개**하여야 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.

22 공공기관이 개발하고 활용하는 모든 인공지능과 민간이 개발하고 활용하는 인공지능 중 개인의 생명이나 안전 등 기본적 인권에 중대한 영향을 미치는 인공지능은 **사용된 데이터와 인공지능 알고리즘의 주요 요소를 일반에게 공개하고 설명**하여야 합니다.

24 특히, 완전히 자동화된 의사결정으로만 개인에게 법적 효력 또는 생명·신체·정신·재산에 **중대한 영향을 미치는 일은 제한**되어야 하고, 이러한 의사결정이 이루어진 경우에는 당사자가 해당 방식을 거부하거나 인적 개입을 요구할 수 있는 권리를 보장받아야 합니다.

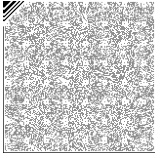
6

자기결정권의 보장

25 「대한민국헌법」 제10조가 규정하고 있는 인간의 존엄과 행복추구권에서 도출되는 **자기결정권**은 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.

26 **개인정보 자기결정권**은 정보주체의 자기 정보에 대한 통제력을 보장하기 위해 인정되는 것인데, 그 통제력 보장의 핵심은 **정보주체의 동의권**입니다.

따라서 개인정보 처리에 대한 정보주체의 동의는 단순한 외형적 의사 표시만이 아니라 정보주체가 개인정보 처리에 대한 제반 상황을 설명받고 자유의사에 따라서 결정될 수 있어야 합니다.



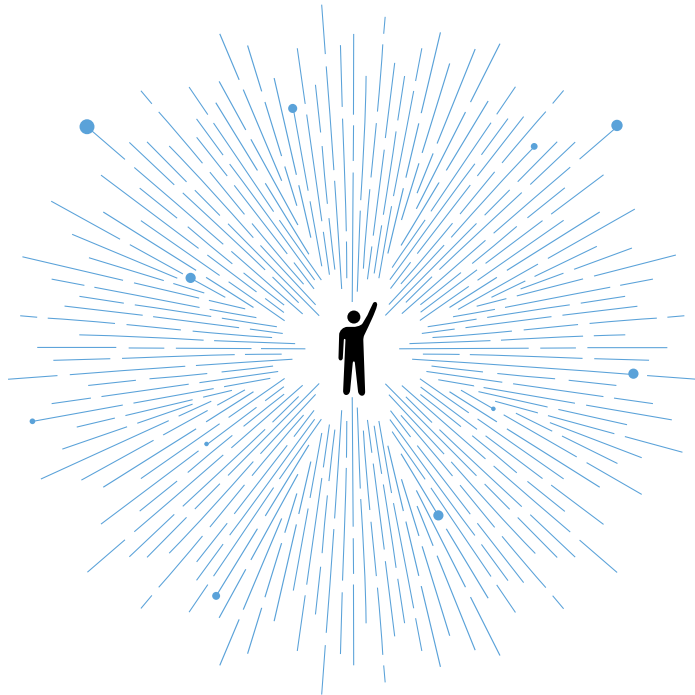
개인정보 자기결정권

개인정보 자기결정권은 자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리를 말합니다(헌재 2005. 5. 26. 결정 99헌마513). 이때 정보주체는 처리되는 정보에 의하여 알아볼 수 있는 사람으로서 그 정보의 주체가 되는 사람을 의미합니다(개인정보보호법 제2조 제3호).

국가와 기업이 인공지능을 개발하거나 활용하기 위하여 개인정보가 포함된 데이터를 처리할 때에는 정보주체의 개인정보 자기결정권을 보장하고 개인정보 보호법을 준수하여야 합니다.

- 유럽평의회의 <알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고(2020)>는 '정보적 자기결정권'을 규정하면서, 국가는 개인정보 처리에 대한 내용을 사전에 통지하고 상호 운용성을 비롯하여 자신의 개인정보를 통제할 수 있는 방안을 보장해야 한다고 하였습니다.

또한 민간 부문 행위자는 인공지능 시스템의 영향을 받는 개인에게 학습 데이터를 비롯하여 개인정보의 모든 사용에 관하여 동의하거나 철회할 수 있는 선택권을 쉽게 행사할 수 있도록 보장하여야 한다고 하였습니다.

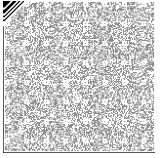


27 인공지능과 관련하여 정보주체의 권리는 처리된 개인정보에 관하여 고지를 받을 권리, 개인정보 접근 및 열람권, 개인정보처리 동의권 및 정정·삭제권, 처리정지요구권 등을 포함하며, 정보주체는 자신의 개인정보가 사용되는 방법을 이해하고 그에 대한 통제권을 가지는 것이 중요합니다. 정보주체는 인공지능 서비스가 언제, 어디서 자신의 개인정보를 수집하고, 어떻게 개인정보를 처리하여 사용, 보관, 삭제하는지에 대해 알고 참여할 권리가 있습니다.

‘이루다’ 사례

‘이루다’는 2020년 12월 출시된 인공지능 챗봇 서비스로, 출시 2주 만에 약 75만 명에 달하는 이용자를 모으며 큰 인기를 끌었습니다. 그러나 이루다 대화 내용에서 여성, 성소수자, 장애인, 흑인 등을 혐오하는 내용이 발견되면서 인공지능 윤리와 차별 논란을 빚었습니다.

더불어 이 서비스는 개발 과정에서 「개인정보보호법」을 중대하게 위반한 사실이 드러나 개발회사인 스캐터랩에 총 1억 330만 원의 과징금과 과태료가 부과되었습니다. 개인정보보호위원회는 개발회사가 과거 서비스 이용자 60만 명의 카카오톡 대화 문장 94억 건을 수집하고 이루다 인공지능 학습과 서비스에 이용하면서 그 사실을 이용자에게 명확히 알리고 동의를 받지 않았다고 밝혔습니다. 이렇게 불법으로 처리된 개인정보 데이터에는 만 14세 미만 아동 20여만 명의 개인정보가 포함되어 있었습니다.



28 인공지능의 개발과 활용에서 개인정보는 목적에 필요한 최소한의 범위 내에서 처리되어야 하며, 처리목적 달성에 필요한 기간 동안만 보관되어야 합니다. 또한 이러한 개인정보 처리 방침은 정보주체가 확인할 수 있도록 공개되어야 합니다.

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



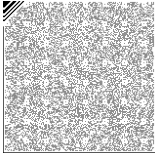
7

차별 금지

30 「대한민국헌법」 제11조는 모든 국민에게 **평등권**을 보장하고 있으며, 「국가인권위원회법」 제2조는 합리적인 이유 없이 성별, 종교, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향, 사회적 신분 등 개인 특성에 따라 특정한 사람을 우대·배제·구별하거나 불리하게 대우하는 행위를 **평등권 침해의 차별행위**로 정의하고 있습니다.

31 따라서 모든 사람은 평등하고, 어떠한 이유로도 생활의 모든 영역에서 구별, 배척, 제한을 받지 아니하며, 인권의 인정과 향유를 보장받을 권리가 있습니다.

32 인공지능을 개발하고 활용할 때는 인공지능으로 인해 영향을 받는 사람의 **다양성과 대표성을 반영**하기 위해 노력해야 하고, 성별, 종교, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향, 사회적 신분 등 개인과 집단의 특성에 따라 **편향적이고 차별적인 결과가 나오지 않도록 하여야 합니다.**



33 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.

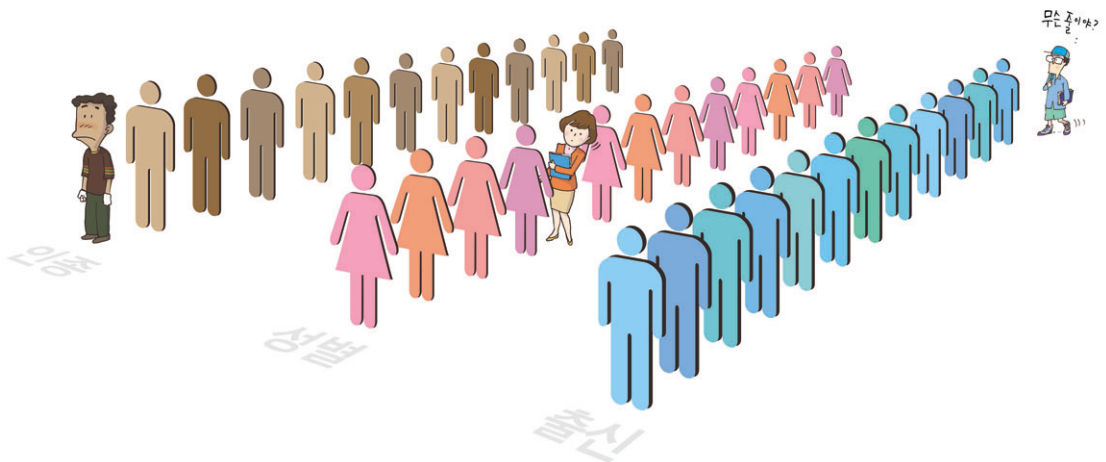
인공지능에 의한 차별

인공지능의 개발과 활용 결과에서 국내·외 인권 규범이 금지하는 직·간접 차별의 발생 우려가 커지고 있습니다. 인공지능이 편견이 내포된 데이터를 사용하거나 차별적인 결과로 이어질 수 있는 데이터를 사용하는 경우, 또는 데이터를 구성하는 인구집단의 다양성과 대표성이 부족할 경우, 이를 학습한 알고리즘과 결과물도 편향성을 나타내어 차별을 영구화하거나 강화할 위험이 있다는 것입니다.

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 있습니다.

특히 고용, 교육, 주택, 금융, 보건의료, 사회복지, 법 집행, 재판 등 중요 분야에서 인공지능으로 부분적 또는 완전히 자동화된 의사결정을 내리는 국가나 기업은, 편향이나 차별이 발생하지 않도록 예방하고 조치하여야 합니다.

국가나 기업이 개발하거나 활용하는 인공지능이 차별을 금지하는 「헌법」과 「국가인권위원회법」, 장애·연령 등 분야별 차별 금지법을 위반하였을 경우에는 그 차별 행위에 대하여 국가인권위원회에 진정할 수 있습니다.



인공지능에 의한 차별 사례

● 거주 지역에 따른 차별

영국 시험 감독청은 2020년 코로나19로 대학수학능력시험에 해당하는 A 레벨 시험을 취소하는 대신 인공지능 알고리즘을 통해 학생 성적을 부여하기로 결정하였습니다. 이 알고리즘은 학생의 A 레벨 예비 시험 점수와 교사의 예상 점수 등을 바탕으로 성적을 산출하고 소속 학교의 역대 학업능력을 고려한 가중치를 부과하였습니다.

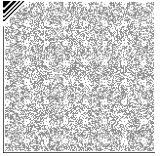
그러나 평가 결과 부유한 지역 학생에 비하여 가난한 지역 학생이 상대적으로 차별을 받은 것으로 나타났습니다. 학교별 학업 성취도가 개별 학생들의 성적에 중대한 영향을 미쳤기 때문입니다. 그러자 인공지능이 불평등을 강화한다며 전국에서 항의 시위가 일었고, 교육부 담당 공무원과 시험감독청장이 사임하였습니다. 2020년 8월 영국 교육부 장관과 시험감독청장은 A 레벨 알고리즘 성적을 철회한다고 밝혔습니다.

● 인종과 성별 편향

2017년 미국 언론은 페이스북 맞춤형광고가 여성이나 노인에게는 구인 광고를 노출하지 않고 유색인종, 십 대 부모, 장애인, 이주민 가정 등에 주택 광고를 노출시키지 않는 등, 이용자를 차별하는 옵션을 제공했다고 밝혔습니다. 이에 미국 법원에 시민단체를 비롯한 여러 원고의 집단소송이 잇따르자, 페이스북은 2019년 일부 광고에서 맞춤 옵션을 제한하기로 원고들과 합의하였습니다. 그러나 옵션 제한 이후로도 광고 표시 알고리즘이 자동적으로 광고 내용에 따라 인종과 성별 편향적인 표시를 계속하여 논란이 멈추지 않고 있습니다.

● 생체 인식에서의 편향성

2017년 한 연구에서 유튜브 음성인식 자동 자막 서비스가 여성 음성과 사투리 인식에서 그 정확도가 현저히 낮은 것으로 나타났습니다. 2018년 또 다른 연구에서는 상용 얼굴인식 시스템들이 어두운 피부색의 여성 얼굴에 대한 오인식률이 밝은 피부색의 남성 얼굴에서보다 높은 것으로 나타났습니다. 연구자들은 이러한 문제의 원인으로 학습 데이터의 편향성을 들었습니다.



34 특히 **학습용 데이터**가 인공지능의 판단에 직접적인 영향을 미치는 상황을 고려할 때, 학습용 데이터의 수집 단계부터 **차별적 요소를 통제**하고 데이터 **편향성을 최소화**하여 인공지능을 통한 의사결정이 특정 집단에 부정적 영향을 미치지 않도록 해야 합니다.

35 학습 데이터의 수집과 선정, 알고리즘의 설계와 활용방향 설정 등 인공지능 개발 전 과정에 걸쳐 편향이나 차별의 요소가 배제될 수 있도록 **점검하는 절차를 마련**해야 합니다. 여기에는 학습 데이터의 개별 요소를 검사하고 차별적 영향을 초래할 수 있는 **데이터를 조정하는 등의 조치**가 포함되어야 합니다.

편향적이지 않은 데이터

인공지능의 차별적 결과물은 데이터와 알고리즘 모두에서 유래할 수 있습니다. 데이터의 경우 그 수집과 활용의 전 과정에서 편향을 방지하거나 최소화할 수 있도록 데이터의 품질과 위험을 관리해야 합니다. 편향과 차별 위험은 기술적인 수준의 조치뿐 아니라 차별적인 영향을 받을 위험이 있는 당사자 집단 및 이들을 대표하는 단체들이 인공지능을 개발하고 활용하는 과정에서 참여하고 협의하는 절차를 통하여 방지하고 완화할 수 있습니다.

- 유엔 인권최고대표는 2021년 〈디지털 시대 프라이버시권〉 보고서에서 인공지능이 사용한 데이터 셋에 내포된 역사적인 편견이 그 결과물에 반영되어 차별을 영구화하거나 강화할 위험이 있다고 지적하였습니다.

차별 금지와 본질적으로 충돌하는 인공지능의 사용은 허용되어서는 안 되며, 예를 들어, 정부 또는 인공지능 시스템이 금지된 차별 기준에 의해 개인을 집단으로 분류하는 개인 사회 신용점수는 이러한 원칙에 따라 금지되어야 한다고 하였습니다.

36 개발한 인공지능에 대해 **주기적인 모니터링**을 거쳐 데이터 품질과 위험을 관리하고, 차별적 결과나 의도치 않은 결과에 대해 **개선**의 조치를 **주기적으로** 수행해야 합니다.

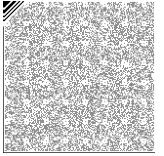
모니터링과 주기적인 관리

데이터 편향이 없어도 알고리즘에서 차별적이거나 의도치 않은 결과가 나타날 수 있습니다.

따라서 알고리즘의 경우 배치하기 전에 결과물의 편향성을 테스트하고, 배치 후에는 주기적인 모니터링을 통해 추적하고 개선하는 조치를 취해야 합니다.

- 유엔 인권이사회는 〈디지털 시대 프라이버시권(2021)〉 결의에서 인공지능 수명주기에 걸쳐 인권 실사 및 모니터링과 점검을 시행하는 등, 인공지능 시스템이 인권에 미치는 영향에 대응하는 노력을 증진할 것을 기업에 권고하였습니다.
- 미국 연방거래위원회(FTC)는 「AI와 알고리즘 사용에 대한 지침」을 통해 기업이 인공지능을 개발하고 활용할 때 차별이 발생하지 않도록 알고리즘의 사용 전과 후에 엄격하게 테스트하고, 알고리즘을 통한 의사결정이 특정 집단에 부당한 영향을 미치지 않도록 관리할 것을 요구하였습니다.
- 캐나다는 자동화된 의사결정 시스템이 사용하는 데이터와 정보에 대하여 의도하지 않은 데이터 편향 및 결과에 부당하게 영향을 미칠 수 있는 기타 요소에 대해 검사할 수 있는 절차를 사전에 개발하고 그 결과를 정기적으로 모니터링하도록 하고 있습니다.





37 인공지능 기술 및 서비스에 대한 접근성과 인공지능이 주는 혜택은 사회적 약자를 포함하여 모든 사회구성원에게 평등하게 제공되어야 합니다. 또한 모든 사람의 인공지능에 대한 지식과 이해를 촉진시키고, 인공지능의 활용이 어려운 계층에 대한 교육 및 지원이 이루어져야 합니다.

평등한 접근권

경제·사회·문화적 권리의 향유에 필수적인 제품 및 서비스 등 인공지능을 활용한 서비스에 대한 접근에 있어서 차별과 편견을 방지해야 합니다. 또한 장애인, 여성, 아동을 포함한 사회적 소수계층을 비롯한 이러한 시스템의 영향을 받는 개인 및 집단에 대하여 직·간접적인 차별이 발생하지 않도록 해야 합니다.

교육을 통한 인식의 제고

알고리즘의 차별을 극복하는 중요한 대책 중의 하나로 교육이 있습니다. 알고리즘의 개발자들에게 헌법상 원칙인 차별 금지 및 평등의 원칙에 대한 교육을 강화하고, 규제 기관과 회사 담당자, 공무원, 더 나아가 사회 전반에 걸쳐서 인공지능 작동에 대한 지식과 인권에 미치는 영향에 대한 이해를 촉진하는 것이 필요합니다.

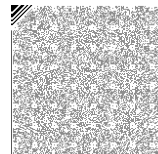
또한 인공지능의 활용이 어려운 계층이 인공지능을 이해하고 사용할 수 있도록 교육할 뿐 아니라, 이들이 인공지능 의사결정에 부당하게 중속되거나 배제되지 않도록 지원하여야 합니다.

38 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.**

39 국가는 인공지능의 개발과 활용에 있어서 인권침해와 차별의 가능성 및 정도, 영향을 받는 당사자의 수, 사용된 데이터의 양 등을 고려하여 공공기관 및 민간기업을 대상으로 **인권영향평가를 실시하여야 합니다.** 특히 인공지능 기술이 적용되어 기존 제도로 관리되거나 감독될 수 없는 새로운 분야는 **인권영향평가 제도를 도입해야 합니다.**

인권영향평가 제도

인공지능이 기본권을 침해하지 않도록 국제기구 및 세계 각국 정부가 다양한 방식의 인공지능 영향 평가를 도입하였거나 도입을 검토 중이며, 인권영향평가는 가장 효과적인 영향평가 방법으로 권장되고 있습니다.



- 유엔인권최고대표는 2021년 <디지털시대 프라이버시권>을 통해 국가와 기업은 인공지능 시스템의 구입, 개발, 배치 및 운영 시 포괄적인 인권실사를 실시해야 하며, 인권영향평가는 인권실사의 필수적인 요소라고 밝혔습니다. 만약 실사 절차에서 인공지능의 사용이 인권과 양립할 수 없는 것으로 드러나는 경우, 이 인공지능은 사용되어서는 안 됩니다.
- 유엔 의사표현의 자유 특별보고관 역시 2018년 보고서(A/73/348)에서 공공부문에서 인공지능 시스템의 조달 및 도입 전에 공개적으로 의견을 수렴하고 인권영향평가 또는 공공기관 알고리즘 영향평가를 실시할 것을 권고한 바 있습니다.
- 유엔 사무총장은 <사회권의 실현에 있어 신기술의 역할에 대한 보고서(2020)>에서 인권의 향유에 대한 중대한 영향을 미칠 수 있는 신기술 시스템, 특히 인공지능 시스템의 전체 수명주기 동안 체계적으로 인권 실사를 실시해야 한다고 밝히고 있습니다.
- 캐나다는 2019년 「자동화된 의사결정 지침」을 제정하여 공공기관 인공지능 요건을 법규화하면서 알고리즘 영향평가의 사전 실시 및 공개를 의무화하였습니다. 영향평가 결과에 따라 전문가 자문, 고지, 인적 개입, 설명 요건, 검사, 모니터링, 교육훈련, 비상계획, 시스템 승인 등 부대 의무를 차등화하고 있습니다.

인권에 미치는 구체적인 영향을 평가하는 절차

유사한 영향 평가로는 「지능정보화기본법」의 사회적 영향 평가와 「과학기술기본법」의 기술 영향 평가가 있습니다.

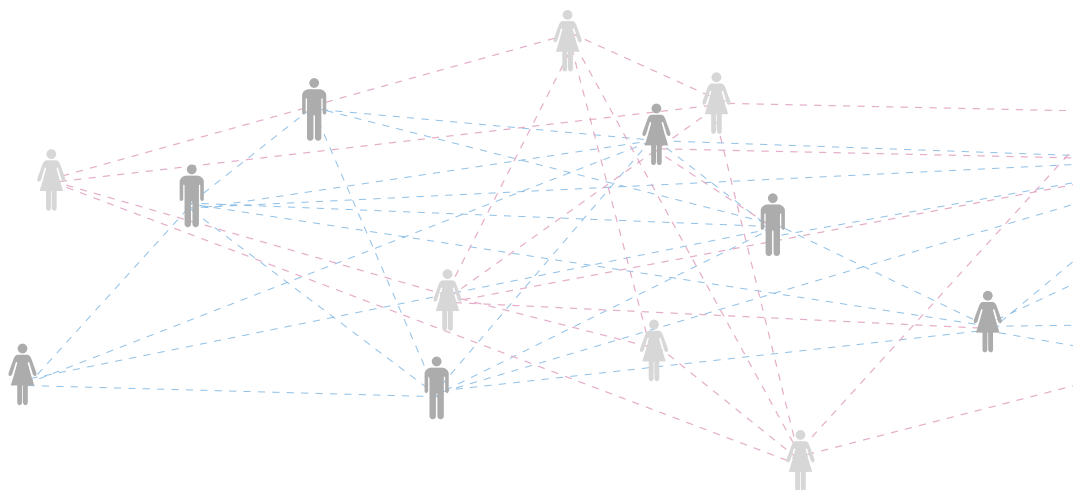
사회적 영향 평가는 ‘국민의 생활에 파급력이 큰 지능정보 서비스 등’에 대해 지능정보 서비스의 안전성, 사생활 보호 등에 미치는 영향, 사회경제에 미치는 영향 등을 평가합니다. 기술 영향 평가는 ‘미래의 신기술 및 기술적·경제적·사회적 영향과 파급효과 등이 큰 기술’에 대해 국민 생활의 편익 증진에 미치는 영향, 새로운 과학기술이 여러 제도적 환경에 미치는 영향, 초래할 부작용 등을 평가하는 절차입니다.

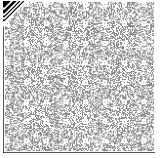
두 평가는 대규모의 사회적 위험을 평가하는 절차인 반면, 인공지능에 대한 인권 영향 평가는 인공지능의 영향을 받는 개인이나 집단의 인권에 미치는 부정적 영향을 평가하고 이를 방지하는 절차라는 점에서 차이가 있습니다.

40 **인권영향평가 내용**에는 인공지능의 특성, 상황, 범위 및 목적을 감안하여 인권 가이드라인이 제시한 원칙 및 내용, 국제 인권 기준, 관련 법률에서 정한 의무 등이 포함되어야 하며, 인권침해 위험요인의 분석, 개선 사항 등을 도출해야 합니다.

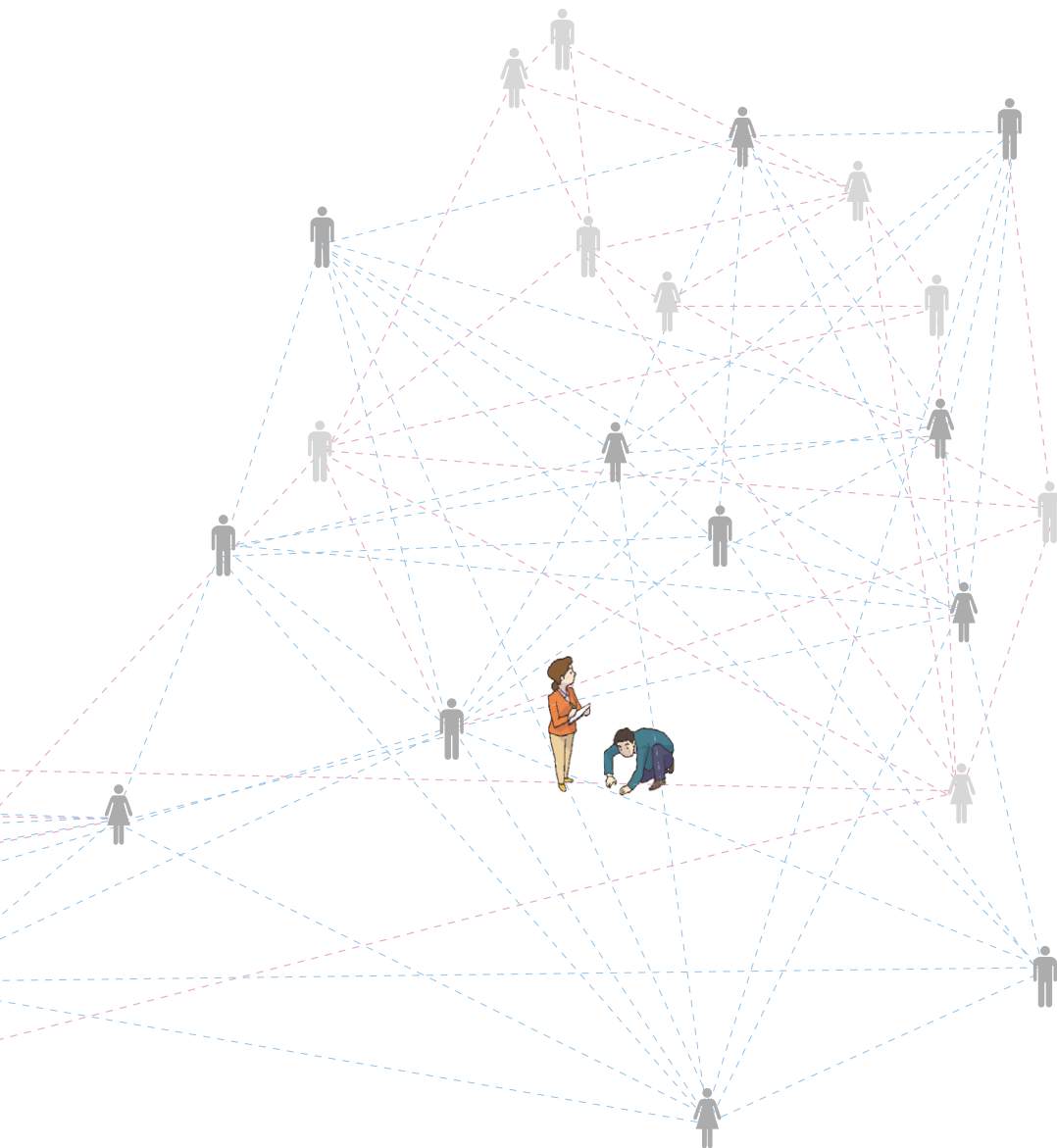
41 **인권영향평가는 개발 및 출시 전에 실시**하고 인공지능의 기능 또는 범위 **변경 시 평가를 갱신**하여야 합니다.

42 **인공지능을 개발 또는 활용하는 국가와 기업은 인권영향평가 결과에서 인권에 미치는 부정적인 영향이나 편향성 및 위험성이 드러난 경우 이를 방지하거나 완화하기 위한 조치사항을 수립하여 적용**하여야 하며, 원칙적으로 그 내용이 공개되어야 합니다. 또한, 이를 방지하거나 완화하는 조치를 취하기 전에는 그 개발과 활용을 중단해야 합니다.





43 국가는 인권영향평가를 **인권전문성과 독립성을 확보한 기관**이 담당하도록 하고, 인권영향평가의 활성화를 위하여 관계 전문가의 육성, 영향평가 기준의 개발 및 보급 등 필요한 조치를 마련해야 합니다.



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

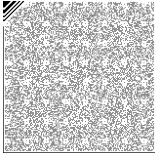
국가의 역할

인공지능이 사용되는 상황에서 책임을 완전하게 보장하는 적절한 법률체계와 절차 방법이 마련되어야 합니다. 이를 위해 국내 법 제도의 공백을 검토 및 평가하고, 필요한 감독 체계를 수립하여 인공지능으로 인한 피해에 대한 구제 수단을 구비해야 합니다.

인공지능으로부터 영향을 받는 사람들을 기본권에 대한 자의적인 간섭으로부터 효과적으로 보호하고, 법적 확실성을 갖추기 위해서 관련 안전조치가 법으로 규정될 필요가 있습니다.

- 국가인권위원회는 김경진 의원이 대표발의 한 「인공지능산업법안」에 대해 ‘인공지능과 같은 미래 산업 지원을 위한 법률의 제정이 필요하다는 데에는 이견은 없다. 다만, 법안은 인공지능 산업의 육성과 지원 관련 규정만을 두고 있고 인권 보호에 대한 내용은 반영하고 있지 않으므로 ①인공지능에서의 인권 보호 기준의 마련, ②인권 보호 원칙의 최우선적 적용, ③인공지능 활용에서 인권 존중 원칙의 반영 등에 대한 필요성이 있다.’ 라고 판단하였습니다.

한편, 인공지능 산업에서의 인권 보호 원칙의 실효성을 담보하기 위한 수단으로, 인공지능 기본계획에 ‘인공지능 산업에서의 인권 보호에 관한 사항’을 추가할 것을 밝힌 바 있습니다.



국가는 인공지능이 인권에 미치는 부정적 영향을 방지하는 입법 조치를 취해야 할 의무가 있습니다.

- 유엔 사무총장은 2020년 보고서에서 인공지능의 개발과 활용에 있어 공공부문과 민간부문을 가리지 않고 책무성을 보장하는 구체적인 제도 마련을 각국에 촉구했습니다.

- 유엔 인권이사회는 <디지털 시대 프라이버시권 결의(2021)>를 통해서 모든 국가에 대하여 국제 인권법에 따른 모든 의무를 완전하고 효과적으로 이행함으로써 인권을 보장하려는 취지에 따라 프로파일링, 자동화된 의사 결정, 머신 러닝 및 생체 인식 기술의 사용에 관한 절차, 관행 및 법률을 정례적으로 검토할 것을 권고하였습니다.

또한 인권이 침해된 개인에게 효과적인 구제 수단에 접근할 수 있도록 규정하는 법률, 규제 및 정책의 도입이나 검토를 고려할 것을 권고하였습니다.

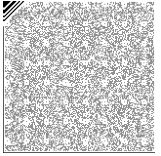


45 국가는 개인의 인권과 안전에 미치는 위험성이 매우 높아 인공지능이 금지되는 영역, 상당한 제한이 필요한 인공지능 고위험 영역, 일정 정도의 제한이 필요한 제한된 위험을 가진 영역, 위험성이 거의 없는 영역 등 적절하게 **인공지능의 위험성 단계를 구분**하고, 그에 맞는 규제 수준과 인적 개입이 이루어지도록 법과 제도를 마련하여야 합니다.

46 특히, 당사자 개인의 생명이나 안전 등 기본적 인권에 **중대한 영향을 미치는 인공지능**은 투명성과 설명가능성, 개인 정보 보호, 차별 금지 등 그 규제에 있어서 공공과 민간의 구분 없이 **엄격하게 적용**되어야 합니다.

위험성 단계에 따른 대응

- 유럽연합 집행위원회가 2021년 제안한 「인공지능법(안)」은 인공지능의 분류를 ○수용 불가 한 위험, ○고위험, ○제한된 위험, ○최소한의 위험 등으로 나누고, 각 수준별로 규제수준을 정하는 위험 기반 접근 방식을 제안하고 있습니다.
- 유엔 인권최고대표는 2021년 〈디지털 시대 프라이버시권〉 보고서에서 인권에 부정적인 영향이 발생할 가능성이 있을 때에는 인간이 감독하고 의사결정에 의무적으로 개입하는 규정이 있어야 한다고 지적하였습니다. 특히 국제 인권 규범을 준수할 수 없는 인공지능을 금지하고, 고위험 인공지능의 경우 그 판매와 사용을 규제하는 규정을 마련하기 전까지 사용을 유예할 것을 각국에 요구하였습니다.
- 캐나다는 2019년 〈자동화된 의사결정 지침〉에서 공공기관 의사결정에 사용되는 인공지능에 알고리즘 영향 평가를 실시하고 4개 등급으로 진단된 위험 수준 별로 전문가 검토, 통지, 인적 개입, 설명, 검사, 모니터링, 교육훈련, 비상 계획, 승인의 의무를 부과하였습니다.



47 **감독 기관**은 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 보관**하여야 합니다.

48 국가는 인공지능을 독립적이고 효과적으로 감독할 수 있는 체계를 수립하여 개인의 인권과 안전을 보장하고 피해를 구제하여야 합니다. 이러한 인공지능 **국가 감독 체계**는 인지 또는 진정으로 접수한 사건을 조사하기에 **충분한 자원, 권한 및 전문지식**을 구비해야 합니다.

독립적인 감독 기관

국가는 인공지능 시스템의 부정적인 영향을 감독하고, 문제 발생 시 실질적으로 해결할 수 있는 책임 시스템을 구축해야 합니다. 그리고 이 역할을 적절한 대응력을 갖춘 독립 기관이 수행하도록 해야 합니다.

- 유엔 인권최고대표는 인공지능 시스템에 대한 적정하고 독립적이고 공정한 감독이 필요하고, 이는 행정적 · 사법적 · 준사법적 기관 및 의회 감독 기관의 조합으로 수행될 수 있다고 했습니다. 예를 들어 개인정보 보호 기관, 소비자 보호 기관, 부문별 규제 기관, 차별방지 기구 및 국가 인권 기구에 인공지능 감독 기능을 구성하는 것입니다.

- 유럽평의회 인권위원장은 “인공지능 감독 기구는 관련 기관 및 기업들로부터 독립적이어야 하며, 감독 기능을 수행하기에 충분한 전문지식과 역량 및 자원을 갖추어야 한다. 감독 기구는 인공지능 시스템의 인권 준수를 능동적으로 조사 및 모니터링하고, 영향을 받은 개인들로부터 진정을 접수 · 처리하며 인공지능 시스템에 대한 주기적인 검토를 실시해야 한다. 감독 기구는 인권 침해 발생의 위험성을 포착한 상황에 개입할 수 있는 권한을 보유해야 한다. 감독 기구는 정기적으로 의회 보고 및 활동 보고서 발표를 해야 한다. 관련 기관 및 기업들은 감독 기구가 감독에 필요한 정보를 요청할 때 이를 제공해야 하고, 정기적인 보고 및 감독 기구의 권고를 이행해야 한다. 감독 절차는 투명해야 하며 공공 조사의 대상이 될 수 있어야 하고, 감독기구의 결정 역시 이의 제기 또는 심사의 대상이 될 수 있어야 한다”라고 하며 독립적인 감독을 강조했습니다.

국가는 인공지능으로 인하여 인권을 침해당하거나 차별을 받은 사람이 진정을 제기하여 권리를 구제받을 수 있는 기회를 보장하는 등 **국가기관의 구제수단**에 대한 접근을 제공해야 합니다.

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

실효적인 구제 수단

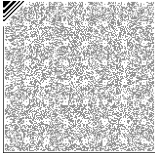
국가는 인공지능으로 인해 피해를 입거나 인권을 침해당한 사람의 진정을 접수하는 등 국가기관의 구제수단에 대한 접근을 보장해야 하며, 이용자들은 구제 수단에 쉽게 접근할 수 있어야 합니다.

기업도 인공지능 시스템에 노출되는 이용자들의 불만사항 및 이의제기에 대해, 사람에 의한 검사 및 구제 시스템을 갖추어야 합니다. 불만 접수 및 구제 신청 건수, 구제수단의 종류와 효과 등도 주기적으로 공개해야 합니다.

- 유엔 인권이사회는 <디지털 시대 프라이버시권(2021)>를 통해 기업에 대해서도 “프라이버시권 및 여타 인권에 대한 존중이 자동화된 의사결정 및 머신러닝 기술의 설계, 운영, 평가 및 규율에 반영되도록 보장하고, 이 기술이 야기하였거나 기여한 인권 침해에 대하여 보상 등 효과적인 구제수단을 제공할 것”을 요구하였습니다.

- 유럽평의회 인권위원장은 2020년 <인공지능 비밀해제 인권보호를 위한 10단계> 보고서에서 ‘효과적인 구제수단’이란, 피해(손해)에 대한 신속하고 적절한 배상 또는 보상, 시정을 포함해야 한다고 했습니다. 이는 민·형사적 조치를 포함할 수 있으며 국가인권기구의 자체 결정 수행은 그러한 시정 조치의 근거가 될 수 있으며, 또한 피해를 주장하는 개인이 자신의 주장을 입증하기 위해서 상대방 또는 제3자 정보에 접근할 수 있는 권리를 보장해야 한다고 했습니다.

- 유럽평의회는 <알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고(2020)>에서 ‘이의 제기 가능성’을 규정하며, “이의 제기에는 청문 기회, 결정에 대한 재검토, 자동화되지 않은 방법으로 결정을 구할 수 있는 방안이 포함되어야 한다. 바로 연락할 수 있는 연락처와 핫라인을 제공받는 등 알고리즘 시스템의 배치 전후 언제라도 손쉽게 이의 제기가 가능해야 한다.”라고 권고했습니다.



50 국가는 대량 감시와 차별로 이어지고, 집회 및 결사의 자유에 부정적 영향을 행사할 위험이 높은 얼굴인식 등 **원격 생체인식 기술의 사용을 공공장소에서 원칙적으로 금지해야 합니다.** 해당 기술은 특별한 경우에만 사용하되, 인권침해나 차별의 위험성이 드러난 경우 이를 방지하거나 완화하는 조치를 취하기 전까지는 사용을 중단해야 합니다.

생체인식 기술의 위험성

생체인식 기술의 발전으로 법 집행 기관에서나 시설 보안 목적으로 생체인식 기술을 사용하는 경우가 증가하고 있습니다.

생체인식은 개인을 식별하기 위하여 얼굴, 지문, 홍채, 음성 또는 보행과 같은 개인의 디지털 표상을 데이터베이스 내 다른 표상과 자동적으로 비교하는 방식입니다. 한 개인의 생체인식 정보는 다른 사람과 구별되는 독특한 특성을 나타내기 때문에 그 사람의 인격의 핵심 특성 중 하나입니다.

게다가, 원격 생체인식은 공공장소에서 개인의 신원을 확인하고 추적할 수 있는 국가의 능력을 증가시켜, 이동의 자유, 표현의 자유, 집회 및 결사의 자유에 대한 권리 행사에 부정적인 영향을 직접적으로 미칩니다. 또한 실시간 생체인식은 개인을 프로파일링 하는 데 사용될 수 있어 심각한 우려가 제기되고 있습니다.

따라서 국가는 공공장소에서 원격 생체인식 기술의 사용을 원칙적으로 금지하고, 예외적인 사용에 대해서 법원의 통제를 받는 등 인권 보호를 위한 체계를 마련해야 합니다.

- 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 권고하였습니다.

또한 유엔 인권이사회는 <디지털 시대 프라이버시권(2021)> 결의에서 얼굴인식 기술 등 생체인식 기술이 평화적 집회의 권리 행사 등에 대한 자의적이고 불법적 감시로 이어지지 않도록 보장할 것을 요구하였습니다.

- 유럽연합의 「인공지능법(안)」은 공공장소에서 실시간으로 원격 생체인식 기술을 사용하는 것은 주민 대부분의 사생활에 영향을 주고, 끊임없이 감시당하는 느낌을 촉발하고, 집회의 자유를 비롯한 기본권 행사를 특히 침해할 수 있다고 지적하면서, 그 사용을 원칙적으로 금지하고 있습니다.

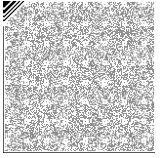
- 호주 국가인권위원회는 얼굴인식 및 기타 생체인식 기술의 사용을 규제하는 법률을 도입할 것과 해당 법이 발효될 때까지 그 사용을 유예할 것을 권고하였습니다.

51 국가는 다양한 정보가 자유롭게 유통되는 정보 환경을 조성하고, 표현의 자유, 정보 접근 및 의견 표명 등에 부정적 영향을 미치는 인공지능에 대해 적절한 조치를 마련해야 합니다.

- 유엔 의사표현의자유 특별보고관은 인공지능이 정보 환경의 중요한 부분이 되었다고 지적하며, 콘텐츠 표시 · 개인화, 콘텐츠 조정 · 삭제, 프로파일링 · 광고 · 타겟팅에 사용되는 인공지능 알고리즘과 그 독점력은 개인이 디지털 의사소통 수단을 통해 정보에 접근하고 이를 배포할 권리를 행사하는 데 큰 영향을 미치고 있다고 분석하였습니다.
따라서 국가는 다양하고 다원적인 정보 환경을 조성하고, 표현의 자유, 정보 접근 및 의견 표명에 부정적인 영향을 미치는 인공지능에 대하여 적절한 조치를 취해야 합니다.

52 국가는 확대되는 인공지능의 개발과 활용에 따라 변화가 예상되는 노동 시장의 일자리 감소, 전자적 노동 감시 등에 대한 대책 마련과 함께 이에 대한 평가와 점검을 해야 합니다.

- 유럽평의회 인권위원장은 2019년 《인공지능 비밀해제 인권보호를 위한 10단계》 보고서에서 인공지능이 개인 및 집단의 자유로운 의견 행사에 미칠 수 있는 부정적 영향, 개인 및 집단에 대한 디지털 추적 이 집회의 자유에 미치는 부정적 영향, 인공지능으로 인한 일자리 감소와 노동 감시 등에 대해 적절한 조치를 취할 것을 강조하였습니다.
또한 대중의 인공지능에 대한 지식과 이해 촉진, 취약계층에 대한 디지털 및 정보 문해력(literacy) 교육 과 지원을 요구하였습니다.



53 또한 국가는 생명의 존엄성 및 윤리를 훼손할 가능성이 높은 자율 살상무기에 대하여 인도주의적으로 접근하고 그 연구, 개발, 생산 및 활용을 금지하는 국제 규범을 준수하고, 이에 대한 논의에 적극적으로 참여해야 합니다.

- 유엔 사무총장은 2019년 치명적인 자율 무기 시스템에 대한 세계적인 금지를 촉구했습니다.
- 유럽의회는 2018년 9월 12일 〈자율 무기체계에 대한 유럽의회 권고(2018/2752(RSP))〉를 통해 인간의 통제 없는 자율 살상 무기체계의 사용, 개발, 생산을 금지하고 이에 대한 효과적인 협의를 시작할 것을 거듭 촉구했습니다.



인공지능
개발과 활용에 관한
인권 가이드라인
해설서

발행일 2022.12. 발행인 송두환 기획 국가인권위원회 인권정책과 제작 젊은기획 02) 2264-2015

주소 서울시 중구 삼일대로 340 나라키움 저동빌딩 13층 전화 02) 2125-9831 팩스 02) 2125-0918

이메일 m740217@humanrights.go.kr 홈페이지 www.humanrights.go.kr

ISBN 978-89-6114-943-3 03350





ISBN 978-89-6114-943-3 03350