

〈인공지능 개발과 활용에 관한 인권 가이드라인〉의 의의와 과제

2023. 2. 24.
정보인권연구소
변호사 김기중

I. 주요 경과와 의의

인공지능 인권가이드라인 주요경과

- 2021.11. 정보인권연구소, 「인공지능 개발과 활용에서의 인권 가이드라인」 연구보고서 제출
- 2022. 4. 11. 인권위, <인공지능 개발과 활용에 관한 인권가이드라인>과 가이드라인에 따른 정책 등을 국무총리 등에게 권고
- 2022. 10. 국무총리, 과기부, 개인정보위, 방통위, 공정위, 금융위, 인권위의 권고를 대체적으로 수용
- 2023. 1. 12. 인권위, <얼굴인식 기술의 도입·활용에 있어서 인권을 보호를 위한 권고> 의결
- 2023. 2. 인권위, <인공지능 인권가이드라인 안내서> 발간

인공지능 가이드라인 주요 내용

제1장 개요

제1절 제정배경, 제2절 목적과 의미, 제3절 정의 규정

제2장 가이드라인 내용

제1절 인간의 존엄성 존중

제2절 투명성과 설명 의무

제3절 자기결정권의 보장

제4절 차별 금지

제5절 인공지능 인권영향평가 시행

제6절 위험도 등급 및 관련 법제도 마련

기타(제6절에 포함) : 원격 생체인식 기술의 사용 기준, 노동 시장 및 전자적 노동 감시에 대한 기준, 자율살상무기 기준

국내 다른 가이드라인 사례(1)

- 개인정보위, <인공지능 자율점검표>(2021.5.)
 - 개보법 준수를 위한 점검표

- 금융위, <금융 분야 AI 가이드라인>(2021) 및 활용 안내서(2022.8.)
 - 시행 전, AI윤리위원회 설치, 위험관리정책 수립 등의 거버넌스 구축
 - 기획/설계 단계에서, 윤리원칙 부합 여부, 피해 가능성 평가
 - 개발 단계에서, 학습 데이터의 품질 등 검증, 설명 가능성을 고려하고, 가용한 설명 가능한 인공지능 기술 등을 확인하여 이를 도입하기 위해 노력
 - 평가/검증 단계에서, 적절한 성능 목표/성능 측정 지표 선정,관리, 적절한 공정성 목표 수준/판단 지표 관리, 안정성·신뢰성 등을 훼손하지 않는 범위 내에서 설명가능성을 합리적인 수준으로 개선 노력

국내 다른 가이드라인 사례(2)

- 과기정통부, <인공지능 윤리기준>(2020. 12.)
 - 3대 기본원칙: 인간 존엄성 원칙, 사회 공공선 원칙, 기술의 합목적성 원칙
 - 10대 요건: 인권보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성(활용 상황에 적합한 수준의 투명성과 설명 가능성을 높이려는 노력을 기울여야)
- 방통위, <이용자 중심의 지능정보사회 실현을 위한 원칙>(2019. 11.)
 - 사람 중심, 책임성, 안전성, 참여, 프라이버시와 데이터거버넌스
 - 투명성과 설명가능성: 기업의 정당한 이익을 침해하지 않는 범위에서 이용자가 이해할 수 있도록 관련 정보를 작성, 피해 발생 시 예측, 추천, 결정의 기초로 사용한 주요요인을 설명할 수 있어야
 - 차별금지: 알고리즘 개발과 사용의 모든 단계에서 차별적 요소 최소화 노력

국내 다른 가이드라인 사례(3)

- 방통위, <인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙>(2021. 6.)
 - 핵심 원칙: 투명성, 공정성, 책무성
 - 투명성: 이용 시작 시점에 그 제공사실을 인지하고 서비스의 내용에 영향을 미치는 주된 요인과 효과를 이해할 수 있도록 필요한 정보 제공, 불만 제기시 해당 결과에 도달한 과정을 충실히 설명
 - 실행원칙: 이용자를 위한 정보 공개, 이용자의 선택권 보장, 자율검증 실행, 불만 처리 및 분쟁해결에 필요한 조치, 내부 규칙 제정

인권가이드라인의 의미(다른 기준의 다른점)

- 인권가이드라인은 인공지능 전반에 걸친 포괄적이면서 구체적인 기준 제시. 다른 기준은 특정 영역에 관한 것이고 추상적인 기준으로 제시
- 인권가이드라인은 인공지능의 개발부터 활용에 이르는 모든 과정에 참여하는 사회구성원을 대상으로 하는 포괄적인 기준 제시. 다른 기준은 특정 영역의 특정 단위에 적용
- 인권가이드라인은 강제성은 없으나, 국가인권위의 정책권고를 기반으로 하고 있고, 구제절차에서 해석의 기준이 되므로, 간접적인 방법에 의한 실행력을 갖고 있음. 다른 기준은 권고적 효력
- 즉, 인권가이드라인은 국가인권위가 인공지능 개발/활용의 모든 범위에서 인공지능에 의한 인권침해와 차별을 판단하는 기준이 되며, 관련 제도의 개선을 권고하거나 구제절차를 마련하는 데 적용되는 실행력 있는 기준임

II. 몇 가지 쟁점

인공지능 정의(1)

- 유럽평의회 <알고리즘 시스템의 인권 영향에 대한 회원국 각료위원회 권고>(2020)
 - 알고리즘 시스템이란 수학적 최적화 기술을 사용하여 데이터의 수집, 결합, 정리, 정렬, 분류 및 추론뿐 아니라 선택, 순위지정, 권장 사항 도출 및 의사결정과 같은 업무를 하나 이상 수행하는 응용 프로그램
- EU <인공지능 백서>(2020)
 - 인공지능 시스템이란 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 시스템

인공지능 정의(2)

- 지능정보화기본법 제2조 “지능정보기술”의 하나로 규정
 - “전자적 방법으로 학습·추론·판단 등을 구현하는 기술”
 - 기타: 데이터 수집, 분석, 가공기술, 사물데이터처리기술, 클라우드컴퓨팅, 초연결지능정보통신 기반 기술
- 금융위, <금융분야 AI 가이드라인>
 - “AI 시스템이란 특정 목표가 주어진 상태에서, 데이터를 획득하여 환경을 인식하고, 획득된 데이터를 해석하며, 지식을 추론하거나 정보를 처리하고, 해당 목표를 달성하기 위한 최선의 행동을 결정함으로써 물리적 또는 디지털 차원에서 작동하는 인간이 설계한 소프트웨어 또는 하드웨어 시스템”
- 인공지능 관련 법안(정필모 의원안, 윤영찬 의원안, 윤두현 의원안)
 - “학습, 추론, 지각, 판단, 자연언어의 이해 등 인간이 가진 지적 능력을 전자적 방법으로 구현(하기 위)한 것”

인공지능 정의(3)

- <인권 가이드라인>
 - “12. 본 가이드라인에서 말하는 인공지능은 일차적으로는 학습·추론·판단을 전자적 방법으로 구현하는 알고리즘과 해당 프로세스를 지칭하나, 이차적으로는 빅데이터 등 인공지능을 가능하게 하는 일련의 기술을 포함하고 있습니다.”

인권 기반 접근 v. 위험 기반 접근

- 유엔 최고대표 <디지털 시대 프라이버시권>(2021)
 - (38항)"인공지능에 대한 인권 기반 접근법은 평등과 차별금지, 참여와 책무성, 지속 가능한 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. .. 인공지능 사용과 관련된 인권 침해와 남용 피해를 입은 사람들은 효과적인 사법적 및 비사법적 구제수단을 이용할 수 있어야 한다."는 원칙
- EU <인공지능법(안)>(2021) – 위험 기반 접근
 - 인공지능 시스템을 위험도에 따라 분류하고, 각 분류별로 요구되는 의무를 규정
 - 수용불가 인공지능: 잠재의식조작 등 4개 범주 (원칙적 금지, 법령 등 예외적 허용)
 - 고위험 인공지능: 제3자의 적합성 평가를 받아야 하는 제품 및 생체인증 분야 등 8개 범주 (거너넌스 구축, 기록 유지, 투명성과 정보 제공, 인간의 감독 등 여러 의무 부과)
 - 제한된 위험 인공지능: 사람과 상호작용하는 인공지능 등 3개 범주 (투명성 의무)

‘고위험 인공지능’ 의 규정 여부와 취급(1)

- “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 일반에 공개하고 설명하여야 합니다.”
 - 두 범주(공공기관, 생명/안전/인권에 영향)의 인공지능에 대한 특별한 의무 부과
 - “주요 요소”라는 열린 개념 채용
 - 주요 요소를 “일반”에 공개하도록 하는 기준 제시
 - 다만, ‘고위험’ 개념을 직접 사용하지 아니하고, 국가에 금지 영역, 고위험 영역, 제한위험 영역 등의 위험도 등급에 따른 분류와 그에 따른 규제 제도 마련 권고 (가이드라인 45항)

‘고위험 인공지능’ 의 규정 여부와 취급[2]

- 인공지능 관련 법안 중 윤영찬 의원안
 - “고위험인공지능”이란 국민의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미치는 인공지능으로 다음 각 목의 하나에 해당하는 인공지능을 말한다.”
: 생명 관련, 생체인식 관련, 사회기반시설 관련, 인사평가/직무 배치 관련, 신용평가 등 필수 공공/민간서비스, 수사 등 국가기관의 권한행사, 문서 진위 등
 - 고위험인공지능 심의위에서 고위험 판정,
 - 고위험인공지능 개발사업자에게 설명의무 등 몇 가지 의무 부과
 - 금지 영역과 제한된 위험 영역의 분류 없음

‘고위험 인공지능’ 의 규정 여부와 취급(3)

- 인공지능 관련 법안 중 윤두현 의원안 – 법안소위 통과안
 - “고위험 영역에서 활용되는 인공지능”이란 다음 각 목의 하나로 사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 인공지능을 말한다.”
 - : 에너지, 보건의료, 의료기기, 원자력, 수사에서 생체정보 이용, 권리의무에 영향, 그밖에 안전/건강/기본권에 중대한 영향을 주는 것으로 대통령령으로 규정
 - 고위험 영역 인공지능을 이용하여 제품 또는 서비스를 제공하려는 자에게 고지의 무 부과
 - 고위험 인공지능 개발자에게는 신뢰성, 안전성 확보 의무 부과
 - 고위험 인공지능에 부과되는 의무가 제한적이며, 금지 영역과 제한된 위험 영역의 분류 없음

설명의무의 범위(1)

- “19. 이러한 알 권리의 보장과 인공지능이 미치는 영향력과 중요성을 감안할 때, 인공지능의 판단과정과 그 결과에 대한 적절하고 합리적인 설명이 보장되어야 합니다.”
 - ‘고위험’뿐만 아니라 모든 인공지능에 적용되는 기준
 - “판단과정과 그 결과” 모두에 대한 설명 의무
 - “적절하고 합리적인 설명”의 열린 개념 사용
 - 판단과정과 그 결과에 이른 이유를 설명하기 어려운 경우 등 예외의 범위와 기준 마련 필요
- 신용정보의 이용 및 보호에 관한 법률 제36조의2(자동화 평가 결과에 대한 설명 및 이의 제기 등): 자동화평가 여부, 기준, 결과, 기초정보 등의 제공 의무와 이용자의 재산출 요구권 등

설명 의무의 범위(2)

- 방통위, <인공지능 기반 미디어 추천 서비스 이용자 보호 기본원칙>(2021. 6.)
 - 투명성: 이용 시작 시점에 그 제공사실을 인지하고 서비스의 내용에 영향을 미치는 주된 요인과 효과를 이해할 수 있도록 필요한 정보 제공, 불만 제기시 해당 결과에 도달한 과정을 충실히 설명
- 해외
 - EU , <P2B 온라인 플랫폼 시장 공정성 및 투명성 강화 규칙>에서 플랫폼이 입점 사업자들에게 검색·배열순위 결정과 관련된 매개변수 및 중요도 등을 약관에 공개하도록
 - EU , <디지털서비스법>에서 모든 온라인 플랫폼은 이용자 맞춤 광고에서 사용한 주요 매개변수를 공개하고 이용자가 변경할 수 있도록 규정
 - 미국 연방거래위, <인공지능과 알고리즘 사용에 대한 지침>을 통해 기업의 알고리즘 의사결정에 대하여 소비자에게 설명하도록 요구
- 위와 같은 기준에 비추어, 인공지능 관련 법안 중 고지 및 설명의무는 매우 불충분

인공지능인 사실 등 사전 고지 의무

- “20. 인공지능과 상호 작용하는 사람에게는 언제나 그 상대방이 인공지능이라는 사실을 알려야 합니다.”
 - 챗봇에 의한 자동상담의 경우의 고지 방법
 - 검색엔진에서 채택된 인공지능 알고리즘의 고지 방법
- “23. 인공지능에 의한 자동화된 의사결정이 예정되어 있는 경우 영향을 받는 당사자들은 사전에 그 사실을 알아야 합니다.”
 - 신용정보의 이용 및 보호에 관한 법률 제36조의2(자동화 평가 결과에 대한 설명 및 이의 제기 등)에 이미 도입
 - 인공지능에 의한 자동화된 의사결정이 예정되어 있는 경우의 범위 구체화 필요
 - 영향을 받는 당사자의 범위 구체화 필요

공공기관 도입 인공지능에 대한 의무 가중

- “21. 공공기관이 인공지능을 통한 의사결정을 할 때 설명할 수 없는 인공지능을 활용해서는 안 되며, 특히 조달의 경우 입찰 단계에서부터 설명가능성이 보장되어야 합니다.”
 - 공공기관의 경우 설명할 수 없는 인공지능의 활용 금지
 - 조달 행정의 경우 가중된 의무
- “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 일반에 공개하고 설명하여야 합니다.”
 - 두 범주(공공기관, 중대 영향)의 인공지능에 대한 특별한 의무
- 국회에 계류중인 인공지능 관련 법안에 공공기관에 대한 특별 취급 기준은 없음

정보주체의 권리 범위

- “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.”
 - 개인정보의 가명화 처리에 대한 정보주체의 정지요구권 인정: 서울중앙지방법원 2023. 1. 19. 선고 2021가합509722 판결
- “27. 정보주체는 인공지능 서비스가 언제, 어디서 자신의 개인정보를 수집하고, 어떻게 개인정보를 처리하여 사용, 보관, 삭제하는지에 대해 알고 참여할 권리가 있습니다.”
 - 인터넷 공개 정보에 대한 무작위 수집과 처리의 경우 권리 확보 방법
 - 영향을 받는 당사자의 확정 문제
 - 구체적인 권리 보호 절차

III. 향후 과제

향후 과제

- 개별 분야 가이드라인의 추가: 안면인식, 노동감시, 살상무기 등
- 설명의무의 범위와 대상 구체화
- 인공지능에 의한 차별 등 권리침해에 대한 특별한 구제수단: 사실조사 방법, 사실인정 기준, 절차적 수단 보완
- 인공지능 분야별 인권영향평가의 적용 기준과 실행수단
- 인권 기반 접근 방법과 위험 기반 접근 방법의 조화
- 인권가이드라인에 부합하지 않는 법률안에 대한 대응

감사합니다.