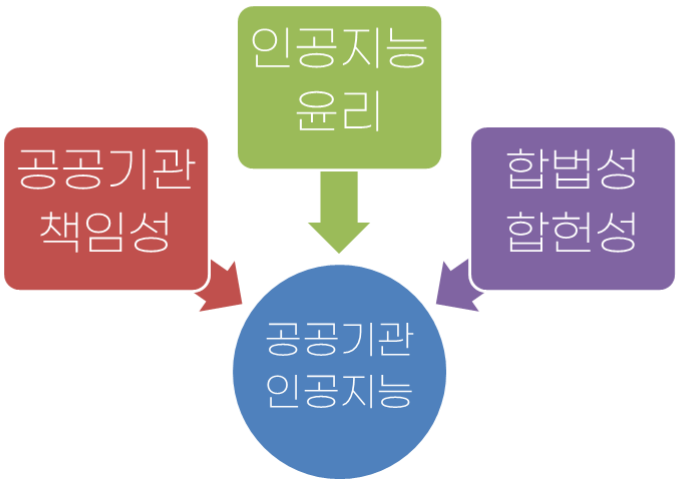
공공기관 인공지능 규범

사단법인 정보인권연구소

2020. 11.

I. 개요



○ 최근 여러 공공기관들이 공공부문의 효율성·합리성을 증대하려는 목적으로 인공지능 기술의 도입을 추진하고 있음. 행정기관이 활용하는 일부 인공지능 알고리즘의 경우 국민에게 법률적이고 행정적인 처분을 내리는 의사결정 절차에 사용될 수도 있음

○ 공공기관 인공지능의 경우 윤리적 책임을 다하여 신뢰가능 인공지능의 발전을 위한 사회적 논의를 이끌고 기술적 혁신에 공공적으로 기여할 필요가 있음

- 유럽연합 집행위원회는 2020년 5월 <인공지능 기반 서비스 및 솔루션 공공조달의 데이터 윤리 백서(이하 '인공지능 공공조달 백서')>에서 인공지능이 정부 및 공공기관과 시민 간의 관계를 변화시키고 형성하고 있다며, 공공기관이 유럽이 지향하는 신뢰가능 인공지능의 혁신을 이끌어야 한다고 선언함

※ 유럽연합 집행위원회는 신뢰가능 인공지능의 3대 요소로 ①합법성(lawful) ②윤리성(ethical) ③기술적·사회적 안전성(robust)을 제시함

○ 특히 자동화된 인공지능 의사결정 알고리즘을 사용하는 자동행정의 경우 기본권 제한에 대한 법률유보, 과잉금지 및 적법절차 등 헌법 원칙을 준수해야 함

인공지능 자동행정 의사결정의 특징[1)]

【기존 행정자동화】
- 교통질서를 위한 신호기, 학교 배정, 공공시설 출입의 전자동화 등
- 의사결정 구조가 정형적, 구조적이고 비교적 단순함
- 자동결정의 기준이 되는 '프로그램'은 비교적 결과예측이 가능함
- 의사결정에는 자연인인 공무원의 질적 개입이 가능함

【인공지능 자동행정】
- 최근 고용 결정, 사회복지 급여 지급, 재범 위험성 평가, 시험 평가 등
- 의사결정 구조가 '알고리즘'을 통해 이루어지며 비정형적, 비구조적이며 매우 복잡함
- 의사결정의 결과에 대한 예측이 어렵고, 알고리즘과 그에 따른 행정행위간 구조적 설명이 곤란함
- 알고리즘의 기술적 우월성으로 공무원의 의사 개입이 사실상 불가능하거나 형식적 개입에 그칠 가능성이 큼

○ 호주 국가인권위원회는 2019년 12월 <인권과 기술> 토론서에서 개인에게 법적 또는 이와 유사하게 중대한 영향을 미치는 의사결정 과정을 인공지능이 실질적으로 지원하는 것을 '인공지능 정보 기반 의사결정(AI-informed decision making)'이라고 정의하면서, 호주 정부에 대하여 인공지능 정보 기반 의사결정의 책무성에 대한 조사를 실시하고 (a)합법성 및 법치주의 원칙 보호 (b)평등 및 비차별 등 인권 증진을 제안함[2)]

○ 그러나 인공지능의 한계와 특성으로 인해 공공기관 업무 수행에서 보장해야 할 책임성 및 합법성·합헌성 보장에 문제가 발생할 우려가 있음

- 스탠퍼드대학교-뉴욕대학교 공동연구는 2020년 2월 미국 정부가 인공지능 시스템을 어떻게 도입하였는지 분석한 후 "성능이나 알고리즘 편향성으로 인해 정부와 시민들 사이의 신뢰가 떨어질 수 있다."고 경고함[3)]

---

1) 김두승 (2019). 인공지능 기반 자동행정과 법치주의. 미국헌법연구, 30(1), 105-138 참고.
2) Australian Human Rights Commission (2019). "Human Rights and Technology : DISCUSSION PAPER"
3) David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey, Mariano-Florentino Cuéllar

ㅇ 최근 세계 여러 나라가 공공 업무에 인공지능을 적용하는 요건과 절차를 마련하고 있으며, 일부 규범은 공공기관이 의무적으로 준수해야 하는 법규적 성격을 띰

ㅇ 이하에서는 각국의 규범을 참고하여, 공공기관으로서 책임을 다하고 윤리적이며 합법적·합헌적인 공공기관 인공지능의 규범을 제안하고자 함

Ⅱ. 공공기관 인공지능의 위험성

ㅇ 영국 국가 전문연구기관인 앨런튜링 연구소는 공공기관 인공지능 시스템이 고려해야 할 해악 우려를 다음과 같이 분류함[4)]

[앨런튜링 연구소] 공공부문 인공지능 시스템의 해악 우려

- ▸편향 및 차별
- ▸개인 자율성, 권리구제, 권리행사 거부
- ▸불투명성, 설명불가능성, 부당한 결과
- ▸프라이버시 침해
- ▸사회적 관계 단절 및 고립
- ▸신뢰할 수 없고, 안전하지 않으며, 품질이 낮은 결과물

ㅇ 영국 공직생활윤리위원회는 2020년 2월 공공영역에서 인공지능 기술을 윤리적이고 안전하게 활용하기 위한 검토 보고서에서, 인공지능이 위협하는 공직생활 원칙을 다음과 같이 진단함[5)]

- (공개성에 대한 도전) 관련 정보를 충분히 제공하지 않을 경우 투명성 저해 위험
- (책임성에 대한 도전) 조직의 책임체계 모호, 공직자 의사결정의 책임소재 불분명, 의사결정 설명 불능 등으로 책임성 불명확화 위험
- (객관성에 대한 도전) 데이터 편향으로 차별의 확산·증폭 위험

(2020). “Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies”. https://law.stanford.edu/education/only-at-sls/law-policy-lab/practicums-2018-2019/administering-by-algorithm-artificial-intelligence-in-the-regulatory-state/acus-report-for-administering-by-algorithm-artificial-intelligence-in-the-regulatory-state/#slsnav-report; 관련 언론보도 https://venturebeat.com/2020/02/19/only-15-of-ai-federal-agencies-use-is-highly-sophisticated-according-to-stanford-and-nyu-report/

4) The Alan Turing Institute (2019). “A guide for the responsible design and implementation of AI systems in the public sector”. p4.

5) The Committee on Standards in Public Life (2020). “Artificial Intelligence and Public Standards”.

○ 특히 공공기관 인공지능 의사결정의 편향성 및 차별적 결과에 대한 논란이 세계 여러 나라에서 일고 있음

- 유럽연합은 2020년 2월 <인공지능 백서>에서 인간의 의사결정에도 편견이 작용하지만 인공지능 의사결정에서 작용하는 편견은 통제 메커니즘 없이 훨씬 더 많은 사람들에게 장기간 영향을 준다고 지적함[6)]

【사례】 형사사법 분야 인공지능 의사결정의 차별 위험성

▸미국 위스콘신주 대법원은 2016년 피고인의 재범 위험성을 평가할 때 참고하는 콤파스(COMPAS) 알고리즘의 평가지수가 법원 결정의 유일한 요소가 되었다면 위법이지만, 보조적인 수단으로 사용되는 경우 적법절차 위반이 아니라고 판결함

▸그러나 언론사 프로퍼블리카에서 2013년부터 2014년까지 콤파스 알고리즘에 의해 법원의 결정이 이루어진 피고인 1200명의 기록을 검증한 결과, 재범률이 높은 것으로 예측되었지만 실제로 2년간 범죄를 저지르지 않은 경우가 흑인의 경우 45%, 백인의 경우는 23.5%이었던 반면, 재범률이 낮은 것으로 예측되었지만 실제로 2년간 범죄를 저지른 경우가 백인이 48%로 흑인 28%보다 훨씬 높았던 것으로 드러남[7)]

【사례】 교육평가 분야 인공지능 의사결정의 차별 위험성[8)]

▸영국 시험감독청(Ofqual)은 2020년 코로나19로 대학수학능력시험에 해당하는 A레벨 시험을 취소하는 대신 인공지능 알고리즘을 통해 학생 성적을 부여함. 이 알고리즘은 각 학생의 A레벨 예비시험과 학교 과제 점수, 교사의 예상치 등을 바탕으로 성적을 산출하고 소속 학교의 역대 학업능력을 고려하여 가중치를 부과함

▸그러나 평가 결과 부유한 지역 학생이 높은 점수를 받은 반면 가난한 지역 학생은 상대적으로 차별을 받은 것으로 나타남. 인공지능이 불평등을 강화한다며 영국 전역에서 시위가 벌어지고 이 사태로 교육부 담당 공무원과 시험감독청장이 사임함. 2020년 8월 영국 교육부 장관과 시험감독청장은 A레벨 알고리즘 성적을 철회한다고 밝히고 교사가 제출한 예상치에 따라 새 성적을 부여한 후 “대학에는 당국과 교사가 산출한 성적 중 더 높은 수치를 제공하겠다”고 밝힘

6) European Commission. 2020. "WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust".
7) 프로퍼블리카 관련 보도 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
8) 가디언 관련 보도 https://www.theguardian.com/education/2020/aug/13/who-won-and-who-lost-when-a-levels-meet-the-algorithm; 한겨레21 관련 보도 http://h21.hani.co.kr/arti/culture/culture_general/49206.html

○ 영국 정부는 2019년 6월 발간한 <공공부문 인공지능 활용 가이드>에서 공공기관이 고려해야 할 위험성과 완화 방안을 다음과 같이 설명함[9)]

[영국 정부] 공공부문 인공지능 프로젝트의 위험과 완화 방법

위험	세부내용
편향 또는 차별의 징후	▸모델의 편향된 결과를 모니터링 하거나 공정하고 설명할 수 있게 만드는 프로세스가 있는지 확인
데이터 사용이 법·제도, 정부 기관의 규정을 준수하지 않음	▸인공지능 데이터 준비에 대한 지침을 참조
기밀 유지 및 데이터 무결성 유지를 보장하는 보안 프로토콜이 존재하지 않음	▸필요한 보안 프로토콜을 정의하기 위해 데이터 카탈로그를 구축
데이터에 접근할 수 없거나 열악한(poor) 데이터 품질	▸내부 및 외부에서 초기 단계에 사용할 데이터셋을 매핑하고, 이후에 데이터를 정확성, 완전성, 고유성, 관련성, 충분성, 적시성, 대표성, 타당성 또는 일관성의 조합에 대한 기준으로 평가
모델 통합 불가능	▸인공지능 모델 구축 초기에 엔지니어를 포함해 개발된 모든 코드가 운용 준비가 되었는지 확인
모델에 대한 책임 프레임워크가 없음	▸인공지능 모델의 서로 다른 영역에서 최종 책임을 지는 책임자를 정의하기 위해 명확한 책임 기록 확립

*요약번역: 한국정보화진흥원(NIA)

○ 특히 공공기관 인공지능이 민간 공급업체 및 영업비밀에 종속될 위험성에 대한 지적이 커지고 있음. 이에 영국 정부 인공지능 조달지침은 입찰 공고 시부터 인공지능 시스템의 ‘블랙박스’ 및 공급업체에 대한 종속(lock-in) 방지를 요구함

【사례】 공공기관 인공지능과 영업비밀[10)]

▸2017년 5월 미국 휴스턴 지방법원은 민간회사 인공지능 비밀 알고리즘에 기반해서 공립학교 교사의 해고를 결정한 사건에서 “공공기관이 매우 중요한 노동 관련 의사결정을 할 때 민간회사의 비밀 알고리즘에 기반한다면, 이는 최소한의 적법절차를 준수하기 어렵다. 따라서 적법절차와 영업비밀을 모두 지키기 위한 적절한 해결책은 비밀 알고리즘의 공공 도입을 중단하는 것”이라고 설시함

9) Government Digital Service and Office for Artificial Intelligence (2019). “A guide to using artificial intelligence in the public sector”.
10) HOUSTON FED. OF TEACHERS v. HOUSTON INDEPENDENT. https://www.leagle.com/decision/infdco20170530802#

Ⅲ. 인공지능 윤리와 공공기관

○ 유럽연합은 2019년 4월 ‘신뢰가능 인공지능 가이드라인’을 공식 채택하고 모든 시민이 인공지능의 혜택을 누릴 수 있는 인간 중심의 윤리적 목적을 달성하는 동시에 신뢰할 수 있는 인공지능 기술의 발전 기준을 구체적으로 제시함

[유럽연합] 신뢰가능 인공지능 윤리 가이드라인[11)]

3대 요소	핵심 지침
Ⅰ. 신뢰가능 인공지능의 기반	① 인간 존중을 윤리 원칙으로 준수하는 인공지능 시스템 개발·배포·사용 - 자율성, 위해예방, 공정성 등을 고려 ② 어린이·장애인·고용주와 근로자 또는 기업과 소비자 간에 권력이나 정보의 불균형에 대응 - 인공지능 기술이 불이익을 주거나 기술 혜택으로부터 배제 가능성이 있는 취약한 집단 배려 ③ 인공지능 기술이 개인과 사회에 상당한 혜택과 이익을 주지만 특정 위험도 초래할 가능성에 주의 - 위험 강도에 따라 이를 완화하기 위한 적절한 조치 필요
Ⅱ. 신뢰가능 인공지능의 실현	① 인적 관리 및 감독 - 인공지능 시스템은 인간의 기본권을 보장하고 자율성을 저해하지 않는 평등한 사회를 구현해야 함 ② 기술적 견고성 및 안전성 - 인공지능 시스템 알고리즘은 모든 생애주기에서 오류와 오작동 등 처리가 가능한 안전성을 갖추어야 함 ③ 사생활 보호 및 데이터 거버넌스 - 시민은 자신의 데이터(개인정보)를 완전히 삭제할 수 있어야 하며 관련 데이터가 인간에게 해를 입히거나 차별해서는 안 됨 ④ 투명성 - 인공지능 시스템은 설명가능해야 함 ⑤ 다양성, 차별 금지 및 공정성 - 인공지능 시스템은 모든 범위의 인간 능력과 기술 및 요구 사항을 고려하고 접근성을 보장해야 함 ⑥ 사회 복지 및 환경 복지 - 인공지능 시스템은 긍정적인 사회 변화를 주도하고 지속가능한 성장을 이끄는데 활용되어야 함

11) European Commission (2019). “Ethics guidelines for trustworthy AI”.

	⑦ 책임성 - 인공지능 시스템과 그 결과에 대한 책임, 그 책임을 보장하기 위한 구조적 장치를 마련해야 함
III. 신뢰가능 인공지능의 평가	II단계에서 요구 사항을 실제 사례에 적합하게 적용할 수 있는 기틀 마련 - 인공지능 시스템에 대한 요구 사항과 솔루션 평가 기준 확립 - 인공지능 시스템의 생애주기 전반에 걸쳐 성과를 개선하고 이에 대한 이해관계자 참여 등

*요약번역: 한국과학기술기획평가원(KISTEP), 일부수정.

○ 유럽연합은 인공지능의 위험성에 대응하기 위하여 신뢰가능 윤리 원칙에 기반한 인공지능 규제를 추진

- 유럽연합 집행위는 <인공지능 백서(2020. 2)>[12], <인공지능 공공조달 백서(2020. 5)>[13]를 연달아 발표하며 인공지능 규제 프레임워크를 제시함

- 유럽연합의 인공지능 규제 프레임워크는 인공지능 학습 데이터 및 시스템의 설계와 배치가 개인정보보호법, 차별금지법, 제조물 책임 및 소비자 보호 관련법, 조달 관련법 등 관련법을 준수할 것을 요구함

○ 경제협력개발기구(OECD)는 2019년 5월 'OECD 인공지능 권고안'을 공식 채택하고 신뢰가능한 인공지능 구현을 위한 5가지 원칙을 제안함[14]

[OECD] 인공지능 권고안

1. 포용 성장, 지속가능 발전, 복지 증진
2. 인간중심 가치 지향, 공정성 지향
3. 투명성 확보, 설명가능성 확보
4. 보안 및 안전성 확보
5. 책임성 확보

○ 국회 입법조사처는 인공지능의 윤리적 사용을 위해서 정부는 인공지능의 윤리기준 등과 관련된 문제를 조정하고 해결할 수 있는 거버넌스나 사후 감시·감독시스템 등을 도입하고, 기업은 △윤리전문가 채용 △인공지능 윤리강령 제정 △인

12) European Commission (2020). "WHITE PAPER: On Artificial Intelligence - A European approach to excellence and trust".
13) European Commission (2020). "White Paper on Data Ethics in Public Procurement of AI-based Services and Solutions".
14) OECD (2019). "OECD Principles on AI". https://www.oecd.org/going-digital/ai/principles/

공지능 피해보상 방안 등을 마련하며, 소비자 보호를 위해 인공지능으로 인한 피해에 대한 배상책임 제도를 보완할 필요가 있다고 지적함[15)]

○ 영국 인공지능사무국은 2019년 6월 앨런튜링 연구소와 함께 공공부문을 위한 <인공지능의 윤리와 안전을 고려한 시스템 설계 · 구현 가이드>를 발표함. 공공기관은 인공지능 프로젝트 실행 시 책임 있는 데이터 설계와 활용 체계를 지원 · 지지하고 동기를 부여해야 함 (SUM 원칙: Support, Underwrite, Motivate)[16)]

[영국 정부] 공공부문 인공지능 기술 설계 · 활용의 윤리적 가치 체계와 실행 원칙

구분		내용
가치 체계	존중	▸개인의 존엄성 회복: 자유롭고 정보에 입각한 결정을 내릴 수 있는 능력을 보장, 자율성 · 자기표현력 등의 권리를 보호
	연결	▸공개적 · 포괄적 연결: 인공지능 프로젝트 과정의 전 주기에서 다양성과 참여를 활성화, 사회적인 신뢰와 공감, 상호 책임 및 이해의 체계를 강화
	돌봄	▸복지를 위한 돌봄: 인공지능 시스템에 영향을 받는 모든 사람들의 복지와 안전을 증진, 해당 기술의 오용과 남용 위험을 최소화
	보호	▸사회적 가치와 공익 보호: 모든 사람을 동등하게 대우하고 사회적 형평성을 보호, 인공지능 및 디지털 기술을 법에 따라 공정 · 균등하게 보호
실행 원칙	공정성	▸데이터 공정성: 공정한 데이터셋을 사용 ▸설계 공정성: 모델에 합리적인 기능, 프로세스 및 분석 구조를 포함 ▸산출 공정성: 결과물이 차별적 영향을 미치지 않도록 함 ▸시행 공정성: 편파적이지 않은 방법으로 제도를 시행
	책임성	▸프로젝트의 설계 및 구현의 전 과정에 관련된 모든 역할에 책임을 설정 ▸프로젝트 전체 단계에서 검토 및 감독 등의 활동 모니터링 실행
	지속 가능성	▸정확성 · 신뢰성 · 보안성 · 견고성을 포함하여 궁극적으로 안정성을 고려 ▸인공지능 설계자와 사용자는 인공지능 시스템이 개인 · 사회에 미칠

15) 이순기 (2020). "인공지능의 윤리적 사용을 위한 개선과제". 국회입법조사처 이슈와논점 제1759호(2020. 9. 25).

16) Government Digital Service and Office for Artificial Intelligence (2019). “Guidance: Understanding artificial intelligence ethics and safety”

		수 있는 영향 등을 인지해야 함
	투명성	▸인공지능 모델이 처리된 방법과 근거 등을 영향을 받는 이해당사자들에게 공개

*요약번역: 한국인터넷진흥원(KISA). 일부수정.

○ 영국 정부는 윤리적 가치체계를 반영한 <공공부문 인공지능 활용 가이드>에서 공공기관이 인공지능을 활용할 때 6가지 요소를 반드시 고려하도록 함

[영국 정부] 공공기관 인공지능 활용의 고려 요인

구성	세부내용
데이터 품질	▸활용의 성공 여부는 데이터 품질의 우수성이 핵심
공정성	▸인공지능 모델은 관련된 훈련과 테스트가 중요하며, 정확하고 일반화 가능한 데이터셋 활용도 중요 ▸인공지능 시스템이 의도된 목적에 부합해 개발될 수 있도록 전문 지식을 보유한 인력이 개발한 것인가?
책무성	▸인공지능 모델의 각 요소를 담당하는 사람과 인공지능 시스템의 설계자 및 구현자의 최종 책임을 묻는 방법을 고려
개인정보 보호	▸유럽연합 개인정보보호법(GDPR) 및 영국데이터보호법(DPA 2018)과 같은 데이터 법·제도의 준수 여부
설명가능성 및 투명성	▸인공지능 모델이 결론에 도달한 방법을 설명가능한가?
비용	▸인공지능 인프라 구축, 실행 및 유지보수, 관련 인력 훈련 및 교육 등 인공지능 도입 비용과 그에 따른 경제적 효과(혜택, 이익)를 비교

*요약번역: 한국정보화진흥원(NIA)

○ 영국 공직생활윤리위원회는 2020년 보고서에서 공공기관 인공지능 윤리를 구현하기 위한 제도 마련을 다음과 같이 제안함

- 공직생활 7대 원칙(사리사욕 금지, 청렴성, 객관성, 책임성, 공개성, 정직성, 통솔력)이 인공지능 시대에도 충분히 관련성이 있으며 유효함

- 자율 규제를 넘어 모든 공공기관 인공지능에 대하여 현행 법률 준수를 요구하고, 특히 영국 평등인권위원회에 공공부문 인공지능의 평등법 준수지침 개발을 요구

- 정부에 인공지능 공공조달 규칙 마련, 공공기관 인공지능 영향평가의 의무적 실시 및 공개 제도 마련, 일정한 수준에서 공공기관 인공지능에 대한 정보 공개 기준 마련을 요구

- 공공부문 인공지능을 규제하는 전문기구로서 데이터윤리혁신센터 강화를 제안

[영국 공직생활윤리위원회] 인공지능과 공공 윤리

▸정부/국가기관/규제기관, 공공 및 민간 공공서비스 제공자에 대한 권고

1 윤리적 원칙과 지침 마련
- 정부는 공공부문 인공지능 활용에 대한 세 가지 윤리 원칙(FAST SUM 원칙, OECD 인공지능 원칙, 데이터 윤리 프레임워크)의 목적, 적용범위 및 위상을 명확하게 알려야 함

2 인공지능의 법적 근거 명확화
- 모든 공공 부문 조직은 공공서비스에 대한 인공지능 기술 적용이 관련 법률 및 규정을 준수하는지를 발표해야 함

3 데이터 편향 및 차별 금지 지침 마련
- 평등인권위원회는 앨런튜링 연구소 및 데이터윤리혁신센터(CDEI)와 협력하여 공공기관이 평등법을 준수하도록 지침을 개발해야 함

4 규제 보증기구 설립
- 공공영역 인공지능 사용에 대한 규제 보증기구가 있어야 하며, 위원회는 CDEI가 이 역할을 수행하는 것을 지지함

5 조달 규칙 및 절차 윤리기준 마련
- 정부는 공공부문 인공지능 솔루션을 개발하는 민간기업이 공공기준을 충족하도록 조달 요건을 마련해야 하며, 입찰 및 계약 시 윤리기준에 대한 요건을 명시해야 함

6 국영상업서비스의 디지털 시장에서 윤리기준 마련
- 국영상업서비스의 경우에는 인공지능 상품 및 서비스가 공공표준을 준수하는지 또는 훼손하는지에 대해 평가할 수 있는 다양한 기준을 도입하고, 서비스 제공자는 윤리적 요건에 맞는 인공지능 상품 및 서비스를 찾아야 함

7 의무적 영향평가 및 공개
- 정부는 인공지능이 공공표준에 미치는 잠재적 영향에 대한 평가를 현행 절차에 통합하는 방안을 검토해야 함. 이러한 평가는 의무적으로 실시하고 공개되어야 함

8 투명성 및 공개성
- 정부는 공공기관의 인공지능 시스템 관련 신고 및 정보공개에 관한 명확한 지침을 마련해야 함

▸**공공서비스를 제공하는 공공 및 민간업체에 대한 권고**

9 공공표준에 대한 위험성 평가
- 공공서비스 제공자는 인공지능 시스템이 공공표준에 미칠 잠재적 영향을 평가하고, 시스템 설계가 공공표준에 미칠 위험을 완화하는지 확인해야 함. 인공지능 시스템 설계를 변경할 때마다 표준에 대한 검토가 이루어져야 함

10 다양성 고려
- 공공서비스 제공자는 인구의 다양한 배경, 행동, 관점이 고려되었는지 확인함으로써 편견과 차별이 없는 서비스를 제공하기 위해 노력해야 함

11 책임소재 명확화
- 공공서비스 제공자는 인공지능 시스템에 대한 책임소재를 명확히 해야 함. 인공지능 시스템에 대한 책임을 명확하게 할당하고 문서화해야 하며, 인공지능 시스템 운영자는 책임을 다해야 함

12 모니터링 및 평가
- 공공서비스 제공자는 인공지능 시스템이 원래의 목적에 맞게 운영되고 있는지 항상 모니터링하고 평가해야 함

13 감독 매커니즘 확립
- 공공서비스 제공자는 인공지능 시스템을 적절히 감시할 수 있는 감독 매커니즘을 확립해야 함

14 이의제기 및 배상방법 안내
- 공공서비스 제공자는 시민에게 그들의 권리와 인공지능 기반 결정에 대해 이의제기하는 방법을 알려야 함

15 직원 훈련 및 교육
- 공공서비스 제공자는 인공지능 시스템을 활용하는 직원이 지속적인 훈련 및 교육을 받도록 해야 함

*요약번역: 정보통신정책연구원(KISDI), 일부수정.

○ 유럽평의회 인권위원장은 2019년 5월 인공지능에 대한 인권 규제 준수에 대한 보고서에서 회원국이 인공지능 시스템의 발전과 구현에 있어 특히 중대하게 영향을 받는 이해당사자들의 인권을 보장할 것을 요구함[17)]

[유럽평의회 인권위원장] 인권 규제 준수를 위한 주요 실행 영역

1 인권영향평가
- 다른 영향평가와 유사한 방식으로 인권영향평가를 실시하는 법체제를 수립하고

17) Council of Europe Commissioner for Human Rights (2019). “Unboxing artificial intelligence: 10 steps to protect human rights”.

공공기관은 이를 조달에 반영해야 함

2 공개적인 의견수렴
- 인권영향평가나 조달 등 단계별로 인공지능 시스템의 운영, 기능, 영향 등 세부사항을 공표하고 의견을 수렴해야 함

3 민간기업의 인권 기준 준수
- 모든 인공지능 운용자는 인권 원칙 준수 여부를 확인하고 공표해야 함. 투명한 인권 실사 절차 수용으로 인공지능 시스템의 인권 위험을 확인할 수 있어야 함

4 정보제공과 투명성
- 인공지능 의사결정의 대상 개인들은 이에 대해 고지받고 지체없이 전문가의 조력을 선택할 수 있어야 함. 인공지능 시스템에 인권적 검토나 정밀검사가 허용되어야 함

5 독립적인 감독
- 인공지능의 인권 준수에 대한 독립적이고 효과적인 감독을 위하여 법제도 마련. 독립적인 기구가 준수 여부를 조사하고 영향을 받은 개인 진정을 처리하고 인공지능 시스템의 성능 발전에 따른 정기적인 검토를 수행할 수 있도록 해야 함

6 차별금지 및 평등
- 인공지능 시스템은 차별을 방지하기 위하여 높은 수준의 정밀검사를 받아야 하며, 특히 인공지능으로부터 그 권리에 부당한 영향을 받을 위험성이 큰 사회집단(아동, 노인, 장애인 등)에 대한 차별을 방지해야 함. 이 원칙은 특히 법집행기관의 프로파일링을 방지하기 위해 중요함

7 개인정보 보호 및 프라이버시
- 인공지능 시스템은 개인정보 처리에 대한 법적 근거와 공정하게 비례적이어야 함. 인공지능 시스템이 민감정보를 처리할 경우 높은 수준의 안전조치를 적용해야 함

8 표현의 자유, 집회결사의 자유, 노동권
- 기술적인 독점 형성을 방지하여 인공지능 전문성과 권한이 집중되고 정보의 자유로운 유통에 부정적인 영향이 미치지 않도록 해야 함. 회원국은 인공지능 발전으로 인한 일자리 창출과 실업의 수치와 유형을 추적해서 실업을 완화해야 함

9 권리구제
- 인공지능 시스템은 언제나 인적 통제 하에 속해야 함. 인공지능 인권침해에 대한 책임성과 책무성은 언제나 자연인과 법인이 감당해야 함. 최소 인적 개입을 보장받을 수 있어야 함. 효과적인 권리구제가 시행되어 인공지능 시스템의 결과로 인한 피해를 시정할 수 있어야 함

10 '인공지능 리터러시' 증진
- 인공지능 관련 문제를 자문할 수 있는 정부내 협의기구의 설립을 검토해야 함

○ 국내외에서 지적되어 온 인공지능 의사결정의 여러 위험성과 그에 대응하는 공공기관 인공지능 규범을 다음과 같이 제안할 수 있음. 이 중 일부 규범은 법규화하여 공공기관의 책임성을 담보할 필요가 있음

Ⅳ. 공공기관 책임성과 인공지능

○ 공공기관의 대국민 책임성은 투명성, 공정성, 합법성을 들 수 있으며, 이러한 책임성에 대한 이행 및 감독 방안도 확립할 필요가 있음

- 공공기관은 일부 개인의 소유가 아니라 국민이 주인인 기관이기 때문에 그 사업이 투명하게 이루어져야 함. 공공기관에게 있어 이는 일차적으로 정보비대칭을 해결하기 위한 대국민 정보공개를 의미하지만 의사결정의 투명성 측면에서 보면 그 대상자에 대한 정보제공과 참여보장도 포함할 수 있음

- 공공기관의 절차적 공정성 또한 요구됨. 특히 국가 행정의 경우 사전에는 국민이 예상가능하게, 사후에는 통제가능하도록 적법절차를 준수해야 하며, 행정적인 의사결정에 대해서는 사람이 책임을 져야 함

- 법치국가에서 정부가 수행해야 할 행정을 집행하거나 위탁받은 공공기관으로서는 법적 책임성으로 합법성이 마땅히 요구됨

○ 우리나라의 경우 공공기관 부패영향평가 기준으로, (1)준수부담의 합리성 (2)제재규정의 적정성 (3)특혜발생 가능성 (4)재량규정의 구체성·객관성 (5)위탁·대행의 투명성·책임성 (6)재정누수 가능성 (7)접근의 용이성 (8)공개성 (9)예측가능성 (10)이해충돌가능성 (11)부패방지장치의 체계성 등을 두고 있음

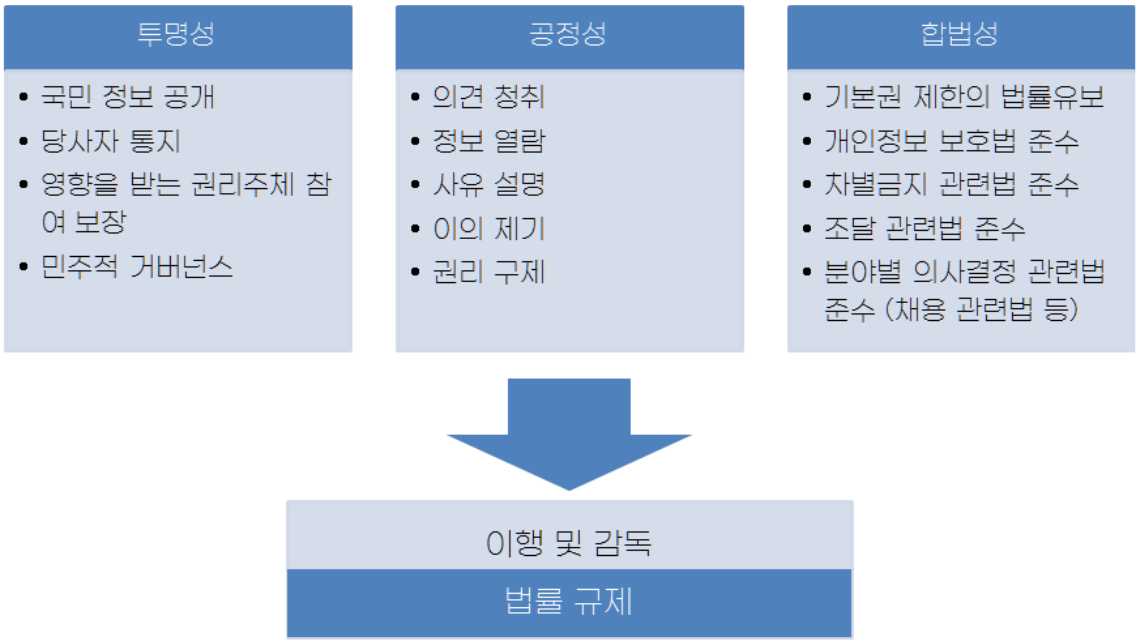
- 여기서 공공기관 인공지능에서 쟁점이 될 기준으로는 재량규정의 구체성· 객관성, 민간업체 위탁·대행의 투명성·책임성, 접근의 용이성, 공개성, 예측가능성 등을 들 수 있음

○ 인공지능의 예측곤란성과 자율성을 명분으로 그 결과에 대한 책임이 모호해지는 사태를 방치한다면 위법한 공권력 행사에 대한 국민 권리구제의 공백이 발생할 가능성이 있음

- 유럽연합은 인공지능의 특성으로 인해 그 의사결정의 관련 법률 준수 여부를 검증하거나 효과적인 이행에 지장을 초래할 수 있고, 당국과 피해자들은 의사결정 과정을 추적·검증하거나 사법적 수단을 비롯한 권리구제 접근에 어려움을 겪을 수 있다고 우려함. 이에 <인공지능 백서>에서 다양한 규제 요건을 모색함

○ 공공기관 인공지능은 인공지능 윤리를 구현해야 할 뿐 아니라, 투명성, 공정성, 합법성 등 공공기관으로서 책임성을 갖추어야 함. 나아가 법규화 등 이행 및 감독 방안도 확립할 필요가 있음

공공기관의 책임성



□ 공공기관 인공지능의 투명성

○ 영국 앨런튜링연구소 및 공직생활윤리위원회는 공공기관 인공지능 시스템의 가장 큰 해악 우려 중 하나로 불투명성, 설명불가능성을 들고 공공기관이 갖추어야 할 투명성 책무 보장을 위해 그 해결을 요구함

○ 유럽연합 집행위원회가 채택한 <신뢰가능 인공지능 가이드라인>은 인공지능의 투명성을 보장하기 위해 설명가능성을 요구하였고, 인공지능 시스템의 생애주기 전반에 걸쳐 이해관계자 참여를 통한 평가를 권장함

- 이어 유럽연합 <인공지능 공공조달 백서>는 공공조달 인공지능 시스템으로 하여금 추적가능하고 설명가능하고 이해관계자를 수용하도록 함

※ 미국국립표준기술연구소가 2020년 8월 설명가능 인공지능 시스템을 위한 원칙을 제안하는 등[18] 설명가능 인공지능 발전을 위한 각국의 노력이 계속되고 있음

18) National Institute of Standards and Technology (2020). "Four Principles of Explainable Artificial Intelligence". https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf

○ 영국 정부는 <공공부문 인공지능 활용 가이드>에서 설명가능성 및 투명성을 반드시 보장하도록 함

- 나아가 영국 공직생활윤리위원회는 정부가 공공기관의 인공지능 시스템 관련 신고 및 정보공개에 관한 명확한 지침을 마련할 것을 제안함

○ 캐나다 자동화된 의사결정에 대한 지침(훈령)에서는 투명성을 보장하기 위해 △의사결정 전 공지 △의사결정 후 설명 △구성 요소에 대한 접근 권한 △소스 코드 공개 등을 명시함

○ OECD 인공지능 권고안 역시 투명성과 설명가능성을 명시하였고, 이에 기반하여 한국정보화진흥원은 공공기관 인공지능의 투명성을 위한 실행원칙으로 이해관계자들에게 인공지능 모델이 처리된 방법과 근거 등을 공개하도록 함

○ 유럽평의회 인권위원장의 2019년 보고서는 인공지능 의사결정의 대상이 된 개인들이 정보제공과 투명성 원칙에 따라 이를 고지받고 지체없이 전문가의 조력을 선택할 수 있어야 한다고 지적함

○ 호주 국가인권위원회는 인공지능 정보 기반 의사결정이 이루어진 경우 영향을 받은 사람에게 정보를 제공하고, 그 의사결정의 설명가능성을 보장하는 법안 마련을 정부에 제안함. 이 설명은 의사결정 사유를 포함해서 개인 또는 관련 기술 전문가가 의사결정의 기반을 이해하고 이의 제기가 가능한 근거를 이해할 수 있어야 함. 개인의 인권을 침해할 수 있는 의사결정에 대해 합리적인 설명을 제공하지 않는 경우 인공지능 정보 기반 의사결정 시스템을 도입해서는 안 됨

- 나아가 정부가 인공지능 정보 기반 의사결정 시스템의 도입을 계획할 경우 가장 영향을 받을 가능성이 높은 사람들에 초점을 맞춘 공청회를 개최하고 법률에 명시되고 적절한 인권 보호가 이루어진 경우에만 시스템을 도입할 것을 제안함

【사례】 공공기관 인공지능 의사결정과 투명성[19)]

▸네덜란드 사회복지 위험발견시스템(SyRI)은 중앙정부 및 지자체가 본래 분리보관되어 있던 데이터들을 광범위하게 결합하여 이를 비공개 인공지능 “위험 모델”에 기반해 분석 후 부정수급 소지가 있는 사람들을 발견하려는 시스템이었음

▸네덜란드 헤이그 지방법원은 2020년 2월 SyRI 관련 법률의 프라이버시 침해 보호조치가 충분치 않고 그 작동 원리에 대한 “투명성이 중대하게 결여되어 있다”며 사용 중단을 명령함. 법원은 이 시스템이 추구하는 사회복지 부정수급자 발견이라는 목표가 사생활권 침해와 비례적이지 않아 위법하다고 판시함

▸유엔 빈곤과 인권에 관한 특별보고관은 이 시스템이 적법절차 보장 없이 저소득층, 이민자 및 소수민족에게 불리하게 사용되었다면서 "사회복지 분야 신기술은 효율성 뿐 아니라 공정성과 정의에도 부합해야 한다. 이를 위해 투명성이 필요하고 취약계층을 부당하게 배척하지 않도록 보장해야 한다"고 지적함

【사례】 암스테르담과 헬싱키 시, 알고리즘 등록부 공개[20)]

▸네덜란드 암스테르담과 핀란드 헬싱키 시는 인공지능의 투명성을 최대한 보장하고 시민의 신뢰를 확보하기 위해 시에서 사용하는 인공지능 알고리즘에 대해 등록하고 공개하는 '알고리즘 등록부'를 2020년 9월 공개함

▸알고리즘 등록부는 각 인공지능 시스템의 △훈련 데이터셋에 대한 정보 △데이터 처리에 대한 정보 △차별 방지에 대한 정보 △인간 감독에 대한 정보 △위험성에 대한 정보 등을 읽기 쉬운 평문으로 공개함

▸또한 알고리즘 등록부는 시에서 사용하는 알고리즘의 도입을 책임지는 공직자의 이름, 부서 및 연락처를 공개하고 시민들이 의견을 제출할 수 있도록 함

□ 공공기관 인공지능의 공정성

○ 인공지능을 이용한 행정의 대상이 된 개인들에 대하여 의견 청취, 정보 열람, 사유 설명, 이의 제기, 권리 구제 등의 적법절차가 보장되어야 함

○ 유럽연합은 2009년 제정 기본권헌장에서 '좋은 행정에 관한 권리(The right to good administration)'를 규정하고(제41조), 청문권, 문서열람권, 결정의 이유제시 요구권 등 공정한 행정절차에 관한 권리를 보장함. 유럽연합은 특히 행정기관이 도입하는 인공지능에 대하여 공정한 절차 보장의 기본권 측면에서 검토하고 있음

○ 유엔 의사표현의 자유 증진 및 보호를 위한 특별보고관(David Kaye)은 2018년 10월 인공지능이 인권에 미칠 영향에 대한 보고서[21)]에서 인공지능 시스템으로부터 반인권적인 영향을 받은 개인들에 대한 구제 수단이 확보되어야 한다고 강조함

19) 가디언 관련 보도 https://www.theguardian.com/technology/2020/feb/05/welfare-surveillance-system-violates-human-rights-dutch-court-rules; 유엔 빈곤과 인권에 관한 특별보고관 보도자료 https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25152 &LangID=E ; 공익소송단 https://pilpnjcm.nl/en/landslide-victory-in-syri-case-dutch-court-bans-risk-profiling/

20) 암스테르담 알고리즘 등록부 https://algoritmeregister.amsterdam.nl/en/ai-register/; 헬싱키 알고리즘 등록부 https://ai.hel.fi/en/ai-register/; 관련 언론보도 https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/

21) David Kay (2018). "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression". A/73/348.

□ 공공기관 인공지능의 합법성

【개인정보보호법 준수】

○ 영국 개인정보 보호 감독기구(ICO)는 2019년 앨런튜링 연구소와 함께 “인공지능 의사결정에 대하여 설명하기” 지침 초안을 발표함[22].

- 유럽연합 개인정보보호법(GDPR) 및 영국 개인정보보호법(DPA 2018)에 따르면 정보주체는 개인정보 자동처리에 대하여 설명을 들을 권리를 보유하며, 행정 및 공공기관 자동처리 의사결정에서도 설명할 권리를 의무적으로 보장하도록 함. 정보주체인 국민에게 이러한 권리가 보장된다면 적법절차에 대한 권리 및 이의제기권과 더불어 공공기관 인공지능의 책임성과 투명성을 보장할 수 있을 것으로 기대됨

- GDPR(제24조)에 따르면 개인정보 처리자인 공공기관들은 개인정보를 보호하기 위하여 책임성 요건을 충족하는 기술 및 관리적 조치를 취하여야 함. 이러한 조치들에는 개인정보보호 정책의 이행, 개인정보 보호 중심 설계 및 설정(Data protection by design and by default), 처리 업무의 문서화, 개인정보 보호 영향평가 등이 있음

○ 영국 <공공부문 인공지능 활용 가이드> 또한 인공지능 활용 의사결정이 개인에게 법률적으로 영향을 줄 수 있는 경우라면 반드시 유럽연합 GDPR 및 영국 개인정보보호법 규정에 따라 다음과 같은 보호 조치를 필수적으로 갖추도록 함

- 자동화된 의사결정 절차에 대하여 구체적이고 쉽게 접근할 수 있는 정보 제공

- 인공지능 의사결정에 대해 인간이 검사하고 결정을 변경하는 등 개입할 수 있는 명확한 방안 마련

【사례】 인공지능 수사와 개인정보보호법 준수

개인정보보호위원회는 2019년 5월 서울특별시 민생사법경찰단의 인공지능 수사관이 피내사자 또는 피의자를 특정하지 않고 온라인에 공개된 전국 불특정 다수의 게시물을 광범위하게 감시·분석하는 것은 개인정보보호법 제15조에 위반된다고 보았음
(개인정보 보호위원회 2019. 5. 13. 결정 제2019-09-130호)

22) ICO and the Alan Turing Institute (2019), “Explaining decisions made with AI: Draft guidance for consultation”.

【차별금지 관련법 준수】

○ 유럽연합 신뢰가능 인공지능 규제 프레임워크는 차별금지 관련법 준수를 요구함

○ 영국 공직생활윤리위원회는 평등인권위원회・앨런튜링 연구소・데이터윤리혁신센터가 협업하여 공공기관의 평등법 준수 지침을 개발하도록 제안함

【조달 관련법 준수】

○ 유럽연합은 <인공지능 공공조달 백서>에서 데이터 윤리, 민주주의 및 기본권에 부합하는 공공조달을 구현하고자 함

[유럽연합] 공공조달에 있어 위험기반·체계적 접근법

▸신뢰가능 인공지능은 책임성, 기술적 안전성, 지속가능성에 대한 요구 뿐 아니라 데이터 윤리 요소를 포함한 공공조달 체계를 수립하고 이를 현행 법적 의무에 적용함으로써 달성될 수 있음
▸이를 위한 5단계 실사 절차를 권장함

1 사전적인 위험 영향평가 : 사람과 집단, 권리와 자유, 민주적 조직과 절차, 사회와 환경에 부작용을 미치는지 살핌

2 공급자 예비 심사 : 설계 절차의 최초 단계서부터 다음과 같은 인공지능 관련 데이터 윤리 요건을 고려하고 정의하고 구현해야 함
- 인공지능이 이용자와 직접적으로(챗봇, 가상비서 등) 또는 간접적으로(자동화된 의사결정) 상호작용한다면 이는 필히 인간이 아니라는 점을 밝혀야 함
- 인공지능 시스템이 추적가능하고, 설명가능하고 이해관계자를 수용해야 함
- 인공지능 시스템이 편향을 방지하고 보편적 설계를 따라야 하며 검토 절차를 포함해야 함
- 기술적 안전성은 문서화되어 설명가능성, 공정 커뮤니케이션 및 감사를 보장해야 함

3 계약 : 입찰 선정의 품질 기준은 정보보안, 데이터 윤리, 환경 측면, 프라이버시, 보편적 설계 등에 적용되는 표준 및 관리시스템에 관한 기술사양을 반영해야 함

4 계약 이행 조건 : 발주 공공기관은 계약이행조건에 지속가능성, 기본권 존중, 데이터 윤리에 대한 조항을 포함하고 제재 조항 및 문서화 요건을 명시해야 함

5 계약 집행 : 공급자는 데이터 윤리, 법적 준수, 책임, 기술적 안전성 및 지속가능성의 다섯 가지 표제에 따라 공공 계약에 명시된 요건을 충족해야 함

○ 영국 인공지능청은 2020년 6월 정부의 인공지능 조달지침을 발표하고 공공조달을 통하는 인공지능에서 10대 원칙을 따르도록 함[23)]

23) Office for Artificial Intelligence (2020). “Guidelines for AI procurement”.

[영국 정부] 인공지능 조달지침

1 인공지능 도입 계획에 조달을 포함할 것

2 다학제적 팀을 구성하여 의사결정을 수행할 것. 낙찰 공급업체에 대해서도, 적합한 기술력을 갖춘 팀을 구성하고 인공지능 시스템의 편향성을 완화하기 위해 다양성 수용을 요구할 것

3 조달절차 개시 전에 데이터 평가를 실시할 것. △조달 절차의 개시 단계부터 데이터 거버넌스 메커니즘이 가동될 수 있도록 확보할 것 △프로젝트에 관련 데이터를 사용할 수 있는지 여부를 평가할 것 △시장에 출시하기 전에 데이터 내부의 결함 및 편향 가능성을 해결할 것. 데이터 문제를 직접 해결할 수 없는 경우 이를 해결하기 위한 계획을 수립할 것 △조달 계획 및 후속 프로젝트를 위해 공급업체와 데이터를 공유할 것인지 여부 및 방법을 정의할 것

4 인공지능 도입의 혜택과 위험성을 평가할 것. △제안서를 평가할 때 공익이 의사결정 절차의 주요 동인이라는 점을 조달 문서에 설명할 것. 〈사회적 가치 지침〉에 따라 인공지능 시스템이 인간과 사회 경제에 미치는 영향 및 편익을 고려할 것. 조달되고 있는 사업이 공익적 목표와 관련이 있어야 하며, 차별금지, 동등한 대우 및 비례성의 원칙을 준수해야 함 △당면한 문제와 관련하여 인공지능을 고려한 배경을 조달 문서에 명확히 설명하고 대안적 솔루션에 대하여 열린 태도를 취할 것 △조달 절차 개시 단계에서 인공지능 영향평가를 수행하고, 중간 조사 결과가 조달에 반영되는지 확인할 것. 주요 의사결정 단계에서 평가 결과를 재차 살펴볼 것

5 시장 형성 초기단계부터 효과적으로 개입할 것. 다양한 인공지능 공급자들과 다양한 방식으로 관계를 맺고, 인공지능 생태계에 개방적인 경쟁 환경을 구축할 것

6 올바른 시장 경로를 구축하고 특정 솔루션보다 해결하고자 하는 과제를 제시할 것

7 거버넌스 및 정보인증을 위한 계획을 수립할 것. 인공지능 시스템에 대한 철저한 검사를 위한 관리감독 메커니즘을 구축하고, 기존 법과 표준을 준수할 것. 인공지능 의사결정의 투명성을 최대화하여 사용자에게 인공지능 시스템이 잘 기능한다는 확신을 부여할 것

8 블랙박스 알고리즘 및 공급업체 종속을 방지할 것. 알고리즘의 설명/해석 가능성을 중요 기준으로 설정하고, 특정 공급업체에 고착되지 않도록 여러 다른 공급자들의 인공지능 시스템 참여를 유도할 것

9 평가 단계에서 인공지능 도입의 기술적/윤리적 한계를 해소 필요성에 집중할 것. 데이터 편향 문제는 없는지, 기존 서비스/기술과 통합 과정에 충분한 검토가 이루어졌는지, 적절한 기술적 표준을 준수하고 있는지 등

10 인공지능 시스템의 생애주기 관리를 고려할 것. △인공지능 조달 과정에서 일회성이 아니라 생애주기에 걸친 검사 필요성을 고려할 것 △지식 이전 및 교육훈련을 요구사항에 포함할 것 △인공지능 시스템을 이해해야 하는 비전문가 대상 교육훈련 및 설명을 요구사항에 포함할 것 △적절하고 지속적인 고객지원 및 호스팅 협의를 보장할 것

*요약번역: 한국과학기술기획평가원(KISTEP), 일부수정.

□ 공공기관 인공지능의 이행 및 준수 책임

○ 유엔 인권최고대표실은 2018년 <인공지능 기술과 표현의 자유> 자료에서 국제인권법에 기반한 인공지능 규제 체제를 다음과 같이 제안함[24)]

[유엔 인권최고대표실] 인공지능 인권법 규제 체제

▸인권 원칙: 인공지능은 모든 다른 기술과 마찬가지로 국제인권법에 따른 국가의 의무와 민간기업의 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

○ 2019년 캐나다 정부는 자동화된 의사결정 지침(훈령)을 발표하여 공공기관 인공지능 요건을 법규화함[25)]

[캐나다 정부] 자동화된 의사결정에 대한 지침 (요건)

▸자동 결정 시스템 사용 프로그램을 주무하는 부처의 실장이 지명하는 사람 또는 차관보는 다음을 책임진다.

24) Artificial Intelligence Technologies and Freedom of Expression(2018)
25) The Government of Canada. Directive on Automated Decision-Making

1. 알고리즘 영향평가

1.1. 자동화된 의사결정 시스템을 생산하기 전에 알고리즘 영향평가를 완료한다.
1.2. 알고리즘 영향평가에 의해 결정이 내려진 경우 부록 C에 규정된 관련 요건을 적용한다.
1.3. 자동화된 의사결정 시스템의 기능 또는 범위가 변경될 시 알고리즘 영향평가를 갱신한다.
1.4. 알고리즘 영향평가의 최종 결과를 〈정부 개방 지침〉에 부합하도록 캐나다 정부 웹사이트 및 캐나다 재정위원회가 지정한 기타 서비스를 통해 일반 접근이 가능한 형식으로 공개한다.

2. 투명성

의사결정 전 공지

2.1. 해당 의사결정이 부록 C에 규정된 바대로 자동화된 의사결정 시스템에 의해 전체 또는 부분적으로 수행된다는 내용을 관련 웹사이트에 공지한다.
2.2. 〈Canada.ca 컨텐츠 스타일 가이드〉에 부합하는 뚜렷하고 쉬운 용어를 이용하여 공지한다.

의사결정 후 설명

2.3. 부록 C에 규정된 대로 결정이 내려진 방법과 이유에 대해 영향을 받는 개인들에게 이해가능하게 설명한다.

구성 요소에 대한 접근 권한

2.4. 소프트웨어 구성요소에 대하여 〈정보 기술 관리 지침〉의 C.2.3.8장에 명시된 요건에 따라 적절한 라이선스를 정한다.
2.5. 독점 라이선스를 사용하는 경우 다음을 보장한다.
2.5.1. 자동화된 의사결정 시스템에 사용되는 독점 소프트웨어 구성요소의 모든 공개 버전을 해당 부서에 전달하고 보호한다.
2.5.2. 캐나다 정부는 특별 감사, 조사, 검사, 심사, 집행 조치 또는 사법 절차에 필요한 경우 자동화된 의사결정 시스템에 대하여 접근하고 시험할 권리를 보유한다. 이는 독점 소프트웨어의 모든 공개 버전에도 적용되며, 이때 인가되지 않은 공개에 적용되는 안전조치를 준수한다.
2.5.3. 이러한 접근권의 일부로서, 캐나다 정부는 필요한 경우 외부인에게 이러한 구성요소를 검토하고 감사할 수 있는 권한을 부여할 수 있다.

소스 코드 공개

2.6. 다음과 같은 경우를 제외하고, 〈정보 기술 관리 지침〉 C.2.3.8장에 명시된 요건에 따라 캐나다 정부가 소유한 사용자 정의 소스 코드를 공개한다.
2.6.1. 소스 코드가 1급 비밀, 2급 비밀, C급 대외비로 분류된 데이터를 처리하는 경우
2.6.2. 정보공개법에 따라 공개가 면제되거나 제외되는 경우
2.6.3. 캐나다 정보관리 최고책임자에 의해 면제되는 경우
2.7. 공개된 소스 코드에 대하여 적절한 접근 제한을 결정한다.

3. 품질보증

테스트 및 모니터링 결과

3.1. 생산에 착수하기 전, 자동화된 의사결정 시스템이 사용하는 데이터와 정보에 대하여 의도하지 않은 데이터 편향 및 결과에 부당하게 영향을 미칠 수 있는 기타 요소를 검사할 수 있는 절차를 개발한다.

3.2. 자동화된 의사결정 시스템을 의도하지 않은 결과로부터 보호하고 본 지침뿐만 아니라 기관 및 프로그램 관련 법률의 준수를 확인하기 위하여 그 결과를 정기적으로 모니터링하는 절차를 개발한다.

데이터 품질

3.3. 자동화된 의사결정 시스템을 위해 수집되고 사용되는 데이터가 정보관리 정책 및 개인정보보호법에 따라 관련성이 있고, 정확하며, 최신인지 검증한다.

전문가 검토

3.4. 부록 C에 규정된 바대로 자동화된 의사결정 시스템을 검토하기 위해 적절한 자격을 갖춘 전문가의 자문을 받는다.

직원 교육

3.5. 부록 C에 규정된 바대로 자동화된 의사결정 시스템의 설계, 기능 및 구현에 대한 적절한 직원 교육훈련을 실시하여 그 운영을 검토, 설명 및 감독할 수 있도록 한다.

비상 기획

3.6. 부록 C에 따라 비상 시스템 또는 절차를 수립한다.

보안

3.7. 〈정부 보안 정책〉에 따라 시스템의 개발 주기 동안 위험 평가를 실시하고 적절한 안전조치 적용을 확립한다.

합법성

3.8. 자동화된 의사결정 시스템의 사용이 해당 법률 요건을 준수하도록 보장하기 위해 기관 법무부서와 협의한다.

인적 개입 보장

3.9. 적절한 경우 부록 C에 따라 자동화된 의사결정 시스템이 인적 개입을 허용하도록 보장한다.

3.10. 부록 C에 따라 자동화된 의사결정 시스템을 생산하기 전에 적절한 수준의 승인을 획득한다.

4. 상환 청구

4.1. 고객이 행정 결정에 이의를 제기할 때 사용할 수 있는 상환 청구 적용 옵션을 제공한다.

5. 보고

5.1. 프로그램 목표 달성에 있어 자동화된 의사결정 시스템의 효과와 효율성에 관한 정보를 캐나다 재무부가 지정한 웹사이트 또는 서비스에 게시한다.

ㅇ 호주 국가인권위원회는 2019년 <인권과 기술> 토론서에서 호주 정부에 대한 30개 제안 및 9개 질의를 발표하며 인공지능 규제의 법제화를 제안함

- 호주 인권위는 정부에 인공지능 정보 기반 의사결정이 이루어진 경우 영향을 받은 사람에게 정보를 제공하고 그 설명가능성을 보장하는 법안 마련을 제안함

- 정부는 인공지능 정보 기반 의사결정 시스템의 도입을 계획할 시 (a)인공지능 사용에 대한 비용 편익 분석을 수행하며 특히 인권 보호 및 책무 보장과 관련하여 살펴 보고 (b)가장 영향을 받을 가능성이 높은 사람들에 초점을 맞춘 공청회를 개최하고 (c)법률에 명시되고 적절한 인권 보호가 이루어진 경우에만 시스템을 도입해야 함. 호주 정부 조달 규칙은 정부가 조달하는 인공지능 정보 기반 의사결정 시스템에 적절한 인권 보호를 포함하도록 요구해야 함

- 인공지능 정보 기반 의사결정과 관련하여 호주에 적용되는 모든 표준은 인권 준수에 대한 지침을 포함해야 하고, 인권 중심 설계(human rights by design) 및 자율적·법적 인증제도를 검토하며, 인권영향평가의 개발 및 법규화를 제안함

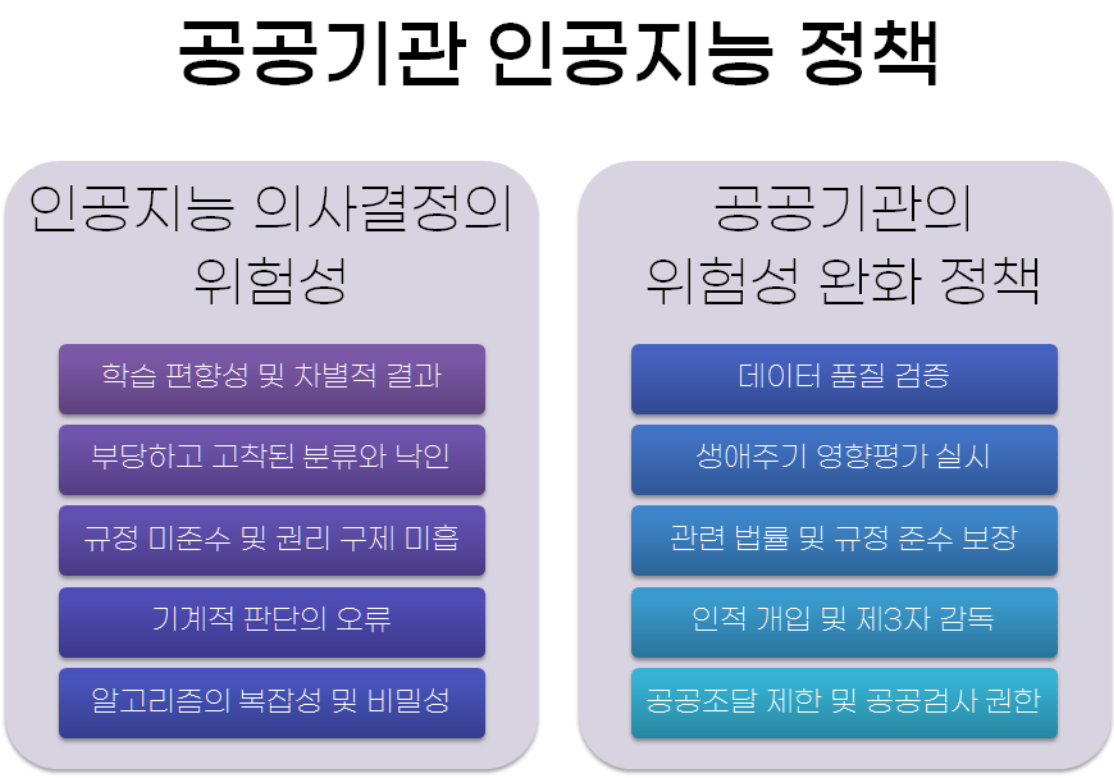
- 전문적이고 독립적인 법정기구로 인공지능 안전위원회(AI Safety Commissioner)를 설립하여 개인 및 지역사회의 피해를 방지하고 인권을 보호하고 증진하는데 주력할 것을 제안함

- 특히 호주 인권위는 논란이 커지고 있는 얼굴인식기술의 사용에 대하여, 적절한 법제도가 마련될 때까지 개인에게 법적 또는 이와 유사하게 중대한 영향을 미치는 의사결정에서 그 사용을 법적으로 유예할 것(moratorium)과 이 법제도는 강력한 인권 보호를 포함하고 호주 국가인권위원회 및 호주 개인정보보호위원회 등 전문 기관과 협의하여 추진할 것을 호주 정부에 제안함

V. 공공기관 인공지능의 위험성 완화 정책

ㅇ 유럽연합은 <인공지능 백서>에서 공공 부문을 포함한 고위험 인공지능에 요구될 규제로 △훈련용 데이터 규제 △추적·검사를 위한 데이터 및 기록 보존 △정보 제공 △모든 생애주기 견고성 및 정확성 △인적 감독 △원격 생체인식 등의 특별 요건을 제시하였음

ㅇ 세계 여러 나라에서 공공기관 인공지능 의사결정의 위험성을 완화하기 위하여 추진 중인 다음의 정책들을 주목해볼 필요가 있음



□ 데이터 품질 검증

○ 인공지능이 학습하는 데이터의 편향성으로 차별의 확산·증폭 위험 문제를 해결하기 위하여 데이터 검증이 필요함

○ 유럽연합 공공조달 접근법은 인공지능 시스템이 편향을 방지하기 위하여 검토 절차를 보장하도록 함

○ 영국 정부는 <인공지능의 윤리와 안전을 고려한 시스템 설계·구현 가이드>에서 공공기관은 공정한 데이터셋을 사용해야 할 의무가 있다고 명시하고 <공공부문 인공지능 활용 가이드>에서 모델의 편향된 결과를 모니터링 하거나 공정하고 설명가능하도록 보장하는 절차를 확보하도록 함

○ 영국 인공지능 조달지침의 경우, 시장 출시 전 데이터 내부의 결함 및 편향 가능성을 해결하도록 하고, 평가 단계에서 데이터 편향의 해소를 검토하도록 함

○ 캐나다 자동화된 의사결정에 대한 지침(훈령)은 공공기관으로 하여금 의도하지 않은 데이터 편향 및 결과에 부당하게 영향을 미칠 수 있는 요소를 검사할 수 있는 절차를 개발하도록 함

【사례】 채용 인공지능의 학습 데이터와 여성 차별[26)]

▸아마존은 2014년 인공지능을 이용한 채용시스템을 활용하였지만, 여성을 차별하는 알고리즘이 발견되어 2015년도에 해당 시스템을 폐기함

▸시스템은 “여성‘ 또는 '여성 체스 클럽 장” 등의 단어를 포함한 이력서에 불이익을 주었고, 여자 대학을 졸업한 여성 2인의 점수를 가감하고, 남성 엔지니어의 이력서에서 흔히 사용되는 동사인 “executed” 및 “captured”등의 단어를 사용한 후보자를 선호한 것으로 나타남. 성에 따른 편향 외에도, 일부 자격 없는 후보자가 모든 업무 방식에 대해 추천되기도 함

▸인공지능 채용 시스템은 지난 10년 동안 회사에 제출된 이력서 패턴을 관찰하여 구직자를 조사하도록 훈련되었는데, 제출된 대부분의 구직 서류가 남성으로부터 제출된 것과, 기술 산업 전반에 미치는 남성 지배력이 채용 의사결정에 반영된 결과로 보임

□ 생애주기 영향평가 실시

ㅇ 영국 공직생활윤리위원회는 2020년 보고서에서 공공기관 인공지능 의사결정에서 의무적으로 영향평가를 실시하고 공개하도록 권고함

- 영국 공직생활윤리위원회는 공공기관 인공지능에 의무적 영향평가가 필요한 이유를 다음과 같이 설명함. ①인공지능 시스템 관리 미흡은 공공기준 훼손 우려가 있음. 영향평가는 공공기관에 자기관 인공지능의 위험 수준을 인식시키고 그에 따른 위험관리를 실시하도록 함 ②인공지능에 대한 경험이 부족한 공공기관에게 영향평가가 데이터 편향 등 친숙하지 않은 위험성 문제를 다룰 수 있도록 함 ③영향평가는 책임성 측면에서 중요함. 영향평가가 자기관 인공지능의 위험성을 인식하고 완화 조치를 취하도록 하여 공공기관의 적절한 책임성을 구현할 수 있음 ④인공지능 기술은 국민 다수에 영향을 미치는 바 영향평가에서 그 이해관계 및 권리 보장 여부를 확실히 할 필요가 있음. 이때 영향평가는 상당한 주의 의무에 부합하는 요소가 될 수 있음

- 공공기관 인공지능 영향평가의 핵심 요소로는 ①영향평가를 공공기관 인공지능 배치 전과 후에 의무적으로 실시해야 함 ②영향평가는 당사자 공공기관으로부터 분리된 제3자가 실시해야 함 ③영향평가 결과는 공개되어야 한다는 점을 들음

ㅇ 영국 인공지능 조달지침은 조달 절차 개시 단계에서 인공지능 영향평가를 수행하

26) 요약번역: 국가생명윤리정책원. http://www.nibp.kr/xe/news2/123168

고, 조달 절차별로 평가 결과가 반영되는지 반복적으로 평가하도록 함. ‘계속진행/중단’ 등 주요 의사결정에서는 위험성 완화 계획을 참고해야 함

- 인공지능 조달지침은 영향평가가 개인정보 보호 영향평가 및 평등 영향평가를 참고하도록 하고 특히 다음 사항에 미치는 영향을 평가하도록 함

[영국 정부] 인공지능 조달지침 영향평가 항목

▸인공지능 시스템에 대한 사용자 요구사항과 그 공익
▸인공지능 시스템의 인적 및 사회 경제적 영향 - 이는 인공지능이 사회적 가치 편익을 제공할 수 있도록 보장함
▸기존의 기술적, 절차적 환경에 미친 결과
▸데이터 품질 및 부정확하거나 편향될 가능성
▸의도하지 않은 결과가 나올 가능성
▸지속적인 지원 및 유지보수 요구사항을 비롯해 전체 생애주기에 대한 비용적 고려사항

ㅇ 유럽연합 <인공지능 공공조달 백서>는 사전 위험 영향평가를 권장하고, 사람과 집단, 권리와 자유, 민주적 조직과 절차, 사회와 환경에 부작용을 미치는지 살피도록 함

ㅇ 캐나다 자동화된 의사결정 지침(훈령)은 알고리즘 영향평가의 사건 실시 및 공개를 의무화함. 시스템 생산 전 영향평가를 완료하고 기능 또는 범위 변경시 평가를 갱신하도록 함

- 영향평가 결과에 따라 전문가 자문, 고지, 인적 개입, 설명 요건, 검사, 모니터링, 교육훈련, 비상계획, 시스템 승인 등 부대 의무를 차등화함

ㅇ 영국 앨런튜링 연구소는 공공부문에 이해당사자 영향평가(Stakeholder Impact Assessment)를 제안함. 이 평가는 공공기관들이 인공지능으로 영향을 받는 이해당사자를 확인하고 그 공정성 및 바람직한 결과물을 분석하며 인공지능 시스템이 개인과 사회에 미칠 수 있는 영향에 대해 검사하도록 함

ㅇ 유럽연합 개인정보 보호법(GDPR)의 경우 다음과 같은 고위험 개인정보 처리에 대하여 개인정보 보호 영향평가를 의무화함[27)]

27) 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

[유럽연합 GDPR] 개인정보 보호 영향평가 의무화 대상

▸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

○ 최근 유엔 등 인권기구들은 인공지능에 대한 인권영향평가를 요구 및 검토함

- 유엔 의사표현의 자유 특별보고관은 2018년 보고서에서 인권에 기반한 인공지능 감독을 위하여 인권영향평가 또는 공공기관 알고리즘 영향평가의 실시를 각국 정부에 권고함

- 유엔인권최고대표실이 2020년 5월 28일 “인공지능, 프로파일링, 자동화된 의사결정, 머신러닝 기술이 적절한 보호조치가 없을 경우 프라이버시권의 향유에 미치는 영향”에 대하여 개최한 온라인 전문가 세미나에서, 다수의 인권 전문가들이 인공지능에 대한 ‘핵심적인 보호조치’로서 인권영향평가 실시를 지지함.
  특히 테러리즘에 대응 시 인권과 기본적 자유의 증진 및 보호에 관한 특별보고관(Ni Aolain)은 각국 정부에 대하여 인권영향평가의 엄중한 실시를 권고하였으며, 데이터 집중 시스템은 법적 목적 달성을 위한 필요성과 비례성이 입증되었을 때에만 도입될 수 있다고 강조함[28)]

- 호주 국가인권위원회는 2019년 <인권과 기술> 토론서에서 호주 정부에 인권영향평가의 개발 및 법규화를 제안함

28) https://www.ohchr.org/Documents/Issues/DigitalAge/ExpertSeminarReport-Right-Privacy.pdf

○ 유럽평의회 인권위원장은 2019년 보고서에서 인권영향평가의 실시를 요구

- 유럽평의회 회원국은 인권영향평가 수행을 위한 법체제를 수립할 것. 인권영향평가는 GDPR 개인정보 보호 영향평가 등 다른 영향평가와 유사한 방식으로 실시되어야 함. 인권영향평가는 인공지능 시스템을 검사하여 인권에 미치는 영향 및 위험성을 발견하고 조치하고 규명해야 함. 공공기관은 인권영향평가의 공표나 수행이 가능하지 않는 공급자로부터 인공지능 시스템을 조달해서는 안 됨

※ 유럽평의회는 인공지능 인권영향평가 권고 추진(자동화된 개인정보 처리 및 인공지능의 인권 문제 전문위원회, 2019년 11월 초안 발표)[29]

□ 관련 법률 및 규정 준수 보장

○ 영국 공직생활윤리위원회는 2020년 보고서에서 모든 공공기관 인공지능이 현행 법률을 준수하고 이를 공표하도록 권고함

○ 영국 정부 <공공부문 인공지능 활용 가이드> <인공지능 조달지침> 및 캐나다 자동화된 의사결정 지침(훈령)은 개인정보 보호법을 비롯한 현행법률 준수를 권고함

○ 호주 국가인권위원회는 정부 인공지능 정보 기반 의사결정 시스템이 법률에 명시되고 적절한 인권 보호가 이루어진 경우에만 도입할 것을 제안함

□ 인적 개입 및 제3자 감독

○ 알고리즘에 대한 통제와 관리를 소수의 전문가(특히 민간 기술전문가)에게만 맡겨두어서는 안 됨. 알고리즘이 가지는 복잡성, 불명확성 그리고 위험성을 고려할 때 이에 관한 강력하고 실효적인 전문기관의 설치가 필요하며, 이러한 전문기관은 사후 사법적인 조치보다 사전적 규제에 참여할 수 있어야 함[30]

○ 유엔 의사표현의 자유 특별보고관은 2018년 보고서에서 정부가 인공지능 시스템을 사용하는 경우 외부의 독립적인 전문가로부터 정기적인 감사를 받을 것을 권

---

29) 권고 초안 참조. Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence MSI-AUT. 2019. "Addressing the impacts of Algorithms on Human Rights: Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems". https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf

30) 향후 인공지능을 활용한 행정이 확대될 경우 이러한 인공지능행정 알고리즘이 법령의 내용을 충실히 반영하고 있는지, 데이터 내용이나 알고리즘 프로그램이 합헌성을 유지하고 있는지에 대한 전문화된 검증 기관이 필요하다는 주장이 있음. 김두승(2019: 129) 참고

고함

ㅇ 유럽평의회 인권위원장은 2019년 보고서에서 회원국들에 인공지능의 인권 준수에 대한 독립적이고 효과적인 감독을 위한 법제도 마련을 권고함. 독립적인 기구가 인권 준수 여부를 조사하고 영향을 받은 개인 진정을 처리하고 인공지능 시스템의 성능 발전에 따른 정기적인 검토를 수행할 수 있도록 해야 함

ㅇ 영국 공직생활윤리위원회는 2020년 보고서에서 공공부문 인공지능을 전문적으로 감독·규제하는 전문기구의 필요성을 확인하고 데이터윤리혁신센터 강화를 권고함

ㅇ 호주 국가인권위원회는 2019년 토론서에서 호주 정부의 인공지능에 대하여 새로운 기관 또는 기존 기관이 다음 사항에 대하여 검토할 것을 제안함

[호주 인권위] 정부 인공지능에 대한 감독 사항

▸호주 정부의 의사결정 시 인공지능의 사용 여부 확인
▸인공지능 사용에 대한 비용편익 분석을 수행하며, 이때 인권 보호 및 책무성 보장을 특별히 살펴볼 것
▸인권영향평가를 비롯하여, 정부가 인공지능을 사용하는 의사결정 시스템을 채택하기로 결정하기까지 절차의 개괄
▸의사결정으로 영향을 받는 사람에게 인공지능 사용에 대해 설명하는지 여부, 가장 영향을 받을 가능성이 높은 사람들에 초점을 맞춘 공청회를 실시하는지 여부 등 설명 방법에 대한 확인
▸의사결정 인공지능 사용에 대한 감시 및 평가 체계 검토

- 더불어 호주 인권위는 전문적이고 독립적인 법정기구로 인공지능 안전위원회(AI Safety Commissioner)를 설립하여 국가 전반적으로 인공지능으로 인한 개인 및 지역사회의 피해를 방지하고 인권을 보호하고 증진하는데 주력하도록 제안함

ㅇ 유럽연합 기본권청은 법집행기관의 얼굴인식기술 활용 등 공공부문의 인공지능 기술에 대하여 유럽 기본권 헌장 좋은 행정의 권리를 보장하기 위하여 독립적인 감독 등 책임성 체계 구축을 요구함

[유럽연합 기본권청] 좋은 행정의 권리와 인공지능[31)]

▸인공지능 기반 법집행 목적의 얼굴인식기술을 도입할 때 좋은 행정의 권리의 하나인 효과적인 구제를 보장하기 위해서는 독립적인 책임성 체계(independent accountability mechanisms) 마련이 중요하다. ▸기관 내외부 감독 기능을 모두 사용하고 얼굴인식기술의 사용 이전, 사용중, 사용후 등 과정 단계별로 감독 기능을 활성화하는 것이 기본권을 적절하고 효과적으로 보호하는 데 필요하다.

□ 공공조달 제한 및 공공검사 권한

○ 캐나다 자동화된 의사결정 지침(훈령)은 캐나다 정부가 특별 감사, 조사, 검사, 심사, 집행 조치 또는 사법 절차에 필요한 경우 자동화된 의사결정 시스템에 대하여 접근하고 시험할 권리를 보유한다는 점을 명시하고, 이를 민간 독점 소프트웨어의 공개 버전에도 적용하도록 함

Ⅵ. 공공기관 인공지능에 대한 권고

□ 공공기관 인공지능 윤리 측면의 검토

○ 한국정보화진흥원(NIA)은 2019년 12월 OECD 권고안을 적용한 <공공기관 신뢰가능 인공지능 구현 실행가이드>(이하 ‘NIA 공공기관 가이드’)를 발표

[한국정보화진흥원] 공공기관 신뢰가능 인공지능의 구현 실행가이드

원칙	실행가이드
포용성장, 지속가능 발전, 복지증진	공공성의 확인 - 인공지능 시스템의 기관 미션 연계성과 사회경제적 영향평가
	사회적 차별요소 배제 - 데이터, 모델로부터 성, 인종 등 차이로 인한 근원적 차별 배제
인간중심 공정성	인간중심 가치와 공정성 촉진 - 인권영향평가, 인권실사, 윤리 행동 강령, 품질인증 조치

31) European Union Agency for Fundamental Rights (2019). “Facial recognition technology: fundamental rights considerations in the context of law enforcement”. p31 발췌.

	인간중심 가치 내재화 - 적절한 안전장치 (Kill Switch, Human in the loop 등)
투명성 설명가능성	인공지능 시스템에 대한 투명한 정보공개 인공지능에 관한 일반 정보, 개발/훈련/ 운영/활용의 방식에 관한 정보
	인공지능 시스템 결과에 대한 설명 요인, 데이터, 알고리즘 등 의사결정 요인과 전후 맥락 설명
보안 및 안전성	인공지능 시스템의 추적 가능성 보장 - 데이터 세트, 알고리즘, 프로세스 및 의사결정 관련 추적 가능성
	체계적인 위험관리 접근 - 가능한 위험 및 확률, 관리방안
책임성	인공지능 시스템 원칙의 실현 - 라이프사이클에서 발생한 의사결정과 행동 문서화

*출처: 한국정보화진흥원[32)]

○ NIA 공공기관 가이드의 기초인 ‘OECD 인공지능 권고안’(이하 ‘OECD 윤리’)은 유럽연합의 ‘신뢰가능 인공지능 가이드라인’(이하 ‘유럽연합 윤리’)과 함께 인공지능 윤리에 관한 대표적인 국제 규범으로 꼽힘. 그러나 OECD 윤리와 유럽연합 윤리 간에는 크게 보아 △준수 의무 △영향을 받는 사람들의 권리보장 측면에서 차가 있음

- 영국 세계디지털파트너와 미 스탠퍼드대학교 세계디지털정책인큐베이터는 2020년 4월 각국의 인공지능 국가 전략을 분석한 공동보고서에서 한국을 인권 준수 체계 대신 윤리적 / 사람중심 접근을 취한 국가로 분류함[33)]

【준수 의무】

○ 유럽연합 윤리에서 언급한 인공지능의 기술적 견고성 및 안전성, 사생활 보호 및 데이터 거버넌스, 차별 금지 원칙은 유럽연합 및 각국에서 이미 일정 수준의 법률적 준수 의무가 부여되어 있기 때문에 유럽연합 윤리의 일부는 그 이행 의무가 법률적으로 부여되어 있음

○ 유럽연합 집행위원회가 2020년 2월 발간한 후속 <인공지능 백서>에서는 유럽연합 윤리에 대한 350개 이상 기관들의 피드백 수령 결과, 가이드라인의 요구사항 다

32) 한국정보화진흥원. 2019. “공공기관 신뢰가능 AI 구현 실용가이드: OECD 권고안의 적용”. NIA 「DNA플러스 2019」(2019. 12.)

33) 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

수는 이미 기존 법률 또는 규제 체제에 반영되어 있다고 요약함. 다만 투명성, 추적성 및 인적 감독에 관한 요구사항은 현행 법률에서 구체적으로 다루고 있지 않다는 의견이었음

- 백서는 유럽연합 시민들이 기본권(개인정보 보호, 프라이버시, 차별금지), 소비자 보호, 제품 안전 및 책임 규율에 관한 유럽 법률의 보호수준을 인공지능에 대해서도 기대하고 있기 때문에 향후 인공지능 규제프레임워크에 있어서 인공지능 특성에 따른 변경은 필수불가결한 경우에 그쳐야 한다고 지적함

- 인종 평등 지침, 고용 및 직업에서 동등한 처우에 관한 지침, 재화 및 서비스에 대한 접근과 고용에서 남녀 동등 처우에 관한 지침, 일련의 소비자 보호 규정들은 물론, GDPR로 대표되는 개인정보 보호와 프라이버시에 대한 규정들 및 법집행기관 개인정보 보호 지침 등 유럽연합의 법률들은 인공지능 여부와 관계없이 원칙적으로 충분히 적용 가능한 상태임. 다만 백서는 기존 법률의 효과적인 적용 및 집행, 적용 범위의 불확실성, 행위자 간의 책임 배분 등의 문제를 해결하기 위한 법률 체계 개선을 모색함

- 백서는 특히 고위험 인공지능에 대하여 기존 법률적 규제에 더한 규제 요건을 모색함. 이때 고위험 인공지능이란 △의료, 운송, 에너지 및 공공 부문 등 일반적으로 수행되는 활동의 특성이 상당한 위험이 발생할 것으로 예상되는 분야에서 △해당 인공지능 애플리케이션이 법적인 영향 또는 유사하게 상당한 영향을 미치거나, 부상·사망 또는 상당한 물질적·비물질적 손상을 초래하거나, 개인이나 법인이 합리적으로 피할 수 없는 영향을 미치는 경우를 말함. 그밖에 추가적으로 △채용 과정과 근로자의 권리에 영향을 미치는 인공지능 애플리케이션의 사용 △소비자 권리에 영향을 미치는 인공지능 애플리케이션의 사용 △원격 생체인식 및 기타 침입 감시 기술 목적의 인공지능 애플리케이션의 사용이 고위험으로 간주됨

- 고위험 인공지능에 요구될 규제 요건으로 △훈련용 데이터 규제 △데이터 및 기록 보존 △정보제공 △견고성 및 정확성 △인적 감독 △원격 생체인식 등 특정 애플리케이션 특별 요건 등이 제시됨

ㅇ 또한 유럽연합은 <인공지능 공공조달 백서>에서 신뢰가능 인공지능은 책임성, 기술적 안전성, 지속가능성에 대한 요구 뿐 아니라 데이터 윤리 요소를 포함한 공공조달 체계를 수립하고 이를 현행 법적 의무에 적용함으로써 달성될 수 있다고 지적함

ㅇ OECD 윤리 역시 “AI행위자는 AI시스템의 생애주기 전반에 걸쳐 법치주의, 인권,

민주적 가치를 존중해야 함" 을 원칙적으로 명시함(원칙 2. 인간중심 가치와 공정성 원칙).

- 그러나 NIA 공공기관 가이드는 인간중심 가치와 공정성 원칙의 실행 가이드로 인권영향평가, 인권실사, 윤리 행동 강령, 품질인증 조치 등을 제시하고 국내 관련 법률 준수를 요구하지 않음. 이 실행 정책들은 각각의 긍정적 가치에도 불구하고 자율적인 조치를 전제하고 있다는 점에서 한계가 있음.

ㅇ 많은 사람에게 상당한 영향을 미치거나 공공기관이 운용하는 인공지능에 대하여 영향평가나 감사 실시를 법규적으로 규정하였거나 고려 중인 세계 여러 나라 사례를 참고하여 인간중심 가치와 공정성 실현을 위한 이행 의무를 제도적으로 수립할 필요가 있음

【영향을 받는 사람들의 권리보장】

ㅇ 최근 인공지능 윤리 규범은 일반 시민에 대한 정보공개 원칙과 구분하여 인공지능 의사결정으로 영향을 받는 사람들(affected individuals) 및 정보주체의 권리를 정의하고 보장하려는 추세로 발전해 옴

- 유럽연합 윤리는 신뢰가능 인공지능 윤리 실현을 위한 인공지능 생애주기 평가에 이해관계자 참여를 제안함. 후속 <인공지능 백서>에서는 영향을 받는 이해당사자에 대한 정보제공 의무를 명시함

- 영국 정부 또한 공공부문을 위한 <인공지능의 윤리와 안전을 고려한 시스템 설계·구현 가이드>에서 인공지능 모델이 처리된 방법과 근거 등을 영향을 받는 이해당사자들(affected stakeholders)에게 공개하는 투명성 원칙을 명시함

- 유럽연합 GDPR의 경우 프로파일링 등 자동화된 처리시 컨트롤러(개인정보 처리자)가 정보주체(data subject)에게 구체적인 통지전달, 인적개입을 획득할 수 있는 권리, 의사를 표현할 권리, 이러한 평가 이후 도달한 결정에 대한 설명을 획득할 권리, 해당 결정에 이의를 제기할 권리 등, 적절한 안전조치를 보장하도록 하고, 아동은 제외해야 함을 명시함(전문 71).

- 유럽평의회 인권위원장은 인공지능 의사결정으로 중대하게 영향을 받는 이해당사자들(affected individuals)의 의견수렴, 정보제공, 권리구제를 요구함.

- 호주 국가인권위원회는 인공지능 의사결정으로 영향을 받는 사람들에 대한 공청회 등 의견수렴 및 정보 제공과 설명가능성 의무를 법적으로 요구함. 또한 개인

의 인권을 침해할 수 있는 의사결정에 대해 합리적인 설명을 제공하지 않는 경우 인공지능 정보 기반 의사결정 시스템을 도입해서는 안 된다고 강조함

ㅇ NIA 공공기관 가이드는 공공기관 인공지능의 투명성 원칙을 위한 실행 가이드로 이해관계자들에게 인공지능 모델이 처리된 방법과 근거 등을 공개하고 설명하도록 하였음. 또 인권실사에 있어 마지막 단계에 이해관계자와의 협의를 포함함. 또한 거버넌스 프레임워크에 있어 공공기관이 의사결정 프로세스에서 소비자 또는 고객이 AI의 작동방식, 특정 결정이 내려진 방법 및 결정의 이유, 결정의 영향 및 결과에 대해 설명을 요청할 수 있는 제도적 장치를 마련하는 것이 필요하다고 지적함

- 그러나 이 가이드가 상정하고 있는 '이해관계자'는 AI시스템의 사용자, 소비자 등 AI 시스템과 연관되어 있거나 영향을 받는 모든 사람 또는 기관을 포괄하는 폭넓은 개념이고 이 이해관계자가 '영향을 받는 사람들'(affected individuals)을 반드시 의미하지는 않음. 따라서 그 투명성과 설명가능성의 의무 또한 형식적인 고지에 그칠 우려가 있으며 법적 또는 상당한 영향을 받는 개인들에 대한 구체적인 설명의 의무를 두었는지 모호함

ㅇ 많은 사람에게 상당한 영향을 미치거나 공공기관이 운용하는 인공지능의 경우 그 의사결정으로 영향을 받는 사람들의 권리를 명확하게 정의하고 법적으로 보장할 필요가 있음

- 특히 법적 또는 상당한 영향을 미치는 행정기관 의사결정의 경우 당사자에게 헌법이 보장하는 적법절차를 보장해야 함

□ 공공기관 인공지능 조달 측면의 검토

ㅇ 한국정보화진흥원은 공공기관 인공지능 윤리에 관한 <공공기관 신뢰가능 인공지능 구현 실행가이드>와 별도로 <공공부문 AI 시스템 도입에 따른 조달 분야 이슈 분석> 보고서(이하 'NIA 조달 보고서')를 발표함

- 그러나 NIA 조달 보고서는 NIA 공공기관 가이드에서 제안한 인공지능 윤리를 소극적으로 반영함. 편향성 방지 등 인공지능 윤리를 공급기업 선정 평가체계와 오작동 유지보수 이슈에 포함하는 수준에 그침. 설명가능성, 추적가능성 등 구체적인 조달 요구사항은 명시하지 않음

[한국정보화진흥원] AI 조달 단계별 주요 이슈

단계별 주요 이슈	세부 내용
1 AI 시스템 발주 전(前) 단계의 고려 사항	제안요청서를 작성하기 전, 명확한 목표와 해결하고자 하는 문제점을 명확히 인지하고 현재 보유 데이터의 상태 확인 필수
2 AI 시스템 제안사 선정을 위한 평가체계 개선	스타트업, 중소기업이 활성화될 수 있는 방안을 모색하고, AI의 특수성을 고려한 기술적 측면 및 윤리적 측면의 평가 필요
3 AI 시스템 구축·운영·유지보수 단계의 주요이슈	AI 시스템 운영을 관리할 전문 인력이 필요하고, AI 구축기업과 유지보수 기업 간 발생할 수 있는 이슈 해결 필요

*출처: 한국정보화진흥원[34]

○ 유럽연합은 <인공지능 공공조달 백서>에서 인공지능 윤리 의무 이행을 위하여 5단계에 걸친 실사를 권장함

- 5단계 실사는 다음을 포함함. 1 사전: 위험 영향평가 2 공급자 심사: 시스템의 추적가능성・설명가능성・이해관계자 수용 여부, 시스템의 편향 방지・보편적 설계・검토 절차 포함 여부, 문서화 여부 등 데이터 윤리 요건 심사 3 입찰 선정: 데이터 윤리 표준 및 관련 기술사양을 반영한 품질 기준 4 계약 이행 조건: 데이터 윤리에 대한 조항, 제재 조항 및 문서화 요건 명시 5 계약 집행: 데이터 윤리, 법적 준수, 책임, 기술적 안전성 및 지속가능성 요건에 대한 충족.

○ 영국 정부는 인공지능 조달지침에서 데이터 윤리 실현을 위한 원칙을 포함

- 이 원칙은 다음을 포함함. △공급업체에 편향성 완화 요구 △조달절차 개시 전 데이터 편향성 해결 원칙 △제안서 평가시 공익적 목표 및 차별금지・동등한 대우・비례성 원칙 준수에 대한 평가, 조달절차 개시 전 영향평가 수행 및 결과 반영과 반복적 검토 △시스템에 대한 검사 및 기존 법과 표준 준수에 대한 확인, 의사결정의 투명성 최대화 △블랙박스 알고리즘 및 공급업체 종속 방지, 알고리즘의 설명/해석 가능성 기준 수립 △평가시 데이터 편향 등 윤리적 한계 해소 검토 △일회성이 아니라 생애주기 관리 및 검사.

○ 캐나다 정부는 자동화된 의사결정 지침(훈령)에서 인공지능 윤리 준수를 법규적으로 보장함

- 이 훈령은 다음을 포함함. ①알고리즘 영향평가: 시스템 생산이나 변경 전 알고리

34) 한국정보화진흥원. 2019. "공공부문 AI 시스템 도입에 따른 조달 분야 이슈 분석". NIA, 「IT & Future Strategy 보고서」(2019. 12. 31.)

즘 영향평가 실시 완료 및 공개 ②투명성: 의사결정 전 공지, 의사결정 후 영향을 받는 개인들에 대한 설명, 소프트웨어 구성요소에 대한 접근 및 검사 권한 행사, 소스코드의 원칙적 공개 ③품질보증: 생산 전 데이터 편향 및 결과에 부당하게 영향을 미칠 수 있는 요소 검사 및 법률 준수 여부의 정기적 모니터링을 위한 절차 개발, 데이터 품질 검증, 전문가 검토, 직원 교육, 비상 기획, 보안, 합법성, 인적 개입 보장 ④상환 청구: 행정 결정에 대한 이의제기 보장 ⑤시스템 효과성 공개 보고.

□ 공공기관 인공지능 모범 규범 권고

○ NIA의 <공공기관 신뢰가능 인공지능 구현 실행가이드>를 공공기관 인공지능 규범의 기초로 삼음. 다만 세계 여러 나라의 보다 모범적인 공공기관 인공지능 규범에서 시사점을 취하여 다음과 같이 보완할 것을 권고함

- NIA는 공공기관 가이드 및 인공지능 관련 조달 기준 마련에서 공공기관의 현행 법률 및 당해 지침의 준수 의무, 영향을 받는 사람들의 권리를 정의하고 보장해야 함

원칙	실행가이드
포용성장, 지속가능 발전, 복지증진	공공성의 확인 - 인공지능 시스템의 기관 미션 연계성과 사회경제적 영향평가
	합법성의 확인 - 대한민국 헌법, 개인정보 보호법, 공공기관의 정보공개에 관한 법률, 조달사업에 관한 법률 등 현행 법률 및 관련 지침의 준수
	차별 행위 방지 의무 - 성별, 종교, 장애, 나이, 사회적 신분, 출신 지역, 출신 국가, 출신 민족, 용모 등 신체 조건, 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) 등 국가인권위원회법에 명시된 차별 행위를 방지하기 위해 충분히 대표적인 데이터와 모델의 사용
인간중심 공정성	인간중심 가치와 공정성 촉진을 위한 생애주기 관리 - 인권영향평가, 인권실사, 윤리 행동 강령, 품질인증 조치의 단계별 실시 및 개선
	인간중심 가치 내재화 - 적절한 안전장치 (Kill Switch, Human in the loop 등) 및 인적개입 의무

	권리구제 - 인공지능 의사결정으로 영향을 받은 이해당사자가 이의 또는 진정을 제기하거나 시정을 요구할 수 있는 독립적인 기관 또는 절차의 수립
투명성 설명가능성	사전 의견수렴 - 영향을 받는 이해당사자가 참여하는 공청회·설명회 개최로 의견 수렴
	인공지능 시스템에 대한 투명한 정보공개 - 일반적으로 인공지능에 관한 일반 정보, 개발/훈련/ 운영/활용의 방식에 관한 정보
	인공지능 시스템 결과에 대한 설명 의무 - 영향을 받는 이해당사자에게 요인, 데이터, 알고리즘 등 의사결정 요인과 전후 맥락 설명 의무
보안 및 안전성	인공지능 시스템의 추적 가능성 보장 - 데이터 세트, 알고리즘, 프로세스 및 의사결정 관련 추적 가능성
	체계적인 위험관리 접근 - 가능한 위험 및 확률, 관리방안
책임성	인공지능 시스템 원칙의 실현 - 생애주기에서 발생한 의사결정과 행동 문서화
	기록 및 데이터의 보존 - 검증과 실사를 위하여 데이터 세트, 프로그래밍 및 훈련 방법론, 시스템 구축, 테스트 및 검증 절차와 기법에 대한 문서화 또는 그 자체

○ 정부, 지방자치단체, 국회는 공공기관 인공지능 규범을 법규화하여 국민 앞에 그 이행과 준수의 책임성을 보장해야 함

○ 정부와 지방자치단체는 공공기관 인공지능 규범을 입찰 요구사항 및 계약 이행 조건 등 공공 조달 규범에 반영하여 공공기관으로 하여금 국민이 신뢰할 수 있는 인공지능의 혁신을 이끌도록 해야 함

○ 국가인권위원회는 공공기관 인공지능의 차별 방지 및 인권영향평가를 위한 기준과 감독 체계를 마련해야 함

○ 정부는 전문적이고 독립적인 법정기구로 '인공지능 감독관'을 설립하여 특히 공공기관 인공지능의 위험영향평가(risk impact assessment)를 주무하고 국민의 피해 방지 및 인권 보호를 보장해야 함. 특히 이 감독관은 산업 육성 부처로부터 독립적으로 설치되고 충분한 권한 행사를 통하여 이를 견제할 수 있어야 함

〈끝〉