국회토론회

인공지능의 공정성·투명성·책임성 보장을 위한 법제 정비 방안

일시 | 2021년 2월 17일(수) 오후 2시

더불어민주당 정보통신특별위원회(위원장 정필모)

MINBYUN-Lawyers for a Democratic Society 민주사회를 위한 변호사모임 디지털정보위원회

사단법인 정보인권연구소 Institute for Digital Rights

[사]소비자시민모임 CONSUMERS KOREA

진보네트워크센터 Korean Progressive Network 'Jinbonet'

참여연대

순 서

2:00 ~ 2:10	개회	**정필모** 의원 (더불어민주당 정보통신특별위원회 위원장)
	사회	**한상희** 교수 (건국대학교 법학전문대학원, 참여연대 정책자문위원장)
2:10 ~ 3:00	발제	헌법과 인공지능 | **김민우** 박사 (충북대학교 행정학과 BK21사업팀, 박사후연구원)
		공정성, 투명성, 책임성 제고를 위한 인공지능 법제 방향 | **오정미** 변호사 (서울대학교 공익법률센터, 공익펠로우)
3:00 ~ 4:20	토론	인공지능과 기술윤리 | **김병필** 교수 (KAIST 기술경영학부, 민주사회를위한변호사모임 디지털정보위원회 회원)
		인공지능 법제 정비 해외 사례 | **장여경** 이사 (사단법인 정보인권연구소)
		인공지능 시대 이용자 관점을 고려한 법제화가 필요하다 | **윤명** 사무총장 (소비자시민모임)
		인공지능 채용 도구의 공정성과 투명성 | **김민** 정책활동가 (진보네트워크센터)
		과학기술정보통신부 | **김경만** 과장 (인공지능정책과)
		공정거래위원회 | **이동원** 과장 (시장감시총괄과)
		개인정보 보호위원회 | **이한샘** 과장 (데이터안전정책과)
4:20 ~ 4:30	플로어 토론 및 참석자 전체 토론	

인사말

정필모 의원 | 더불어민주당 정보통신특별위원회

반갑습니다. 더불어민주당 정보통신특별위원회 위원장
정필모입니다.

‘인공지능의 공정성, 투명성, 책임성 보장을 위한 법제 정비방안 토론회’에 참여해주신 발제자와 토론자 여러분들을 진심으로 환영합니다.

이번 토론회를 함께 준비하며 수고해주신 민주사회를위한변호사모임 디지털정보위원회, 사단법인 정보인권연구소, 소비자시민모임, 진보네트워크센터, 참여연대 관계자 분들의 노고에 감사드립니다.

인공지능은 많은 분야에서 다양한 제품과 서비스로 상용화되었습니다. 기술 발전으로 획기적 변화는 이미 우리 삶에서 일상적으로 진행되고 있습니다.

또한 코로나19로 인한 사회적 거리두기는 우리 경제 전반의 비대면화와 디지털 전환 등 4차 산업혁명을 가속화하고 있습니다.

정부와 여당은 새로운 기술 변화에 적극적으로 대응하고 있습니다. 한국판 뉴딜은 변화하는 경제사회 구조에 맞춰 사람 투자를 통한 디지털 선도인력 양성을 목표로 디지털 기반 경제혁신 가속화 및 일자리 창출을 추진하고 있습니다.

하지만 새로운 기술은 바람직한 미래만을 보장해주지는 않습니다. 인공지능 알고리즘 오류와 제품 및 서비스 사건·사고는 끊임없이 발생하고 있습니다.

2013년 우리나라 한맥투자증권은 차익거래 자동매매시스템의 알고리즘 오류로 2분 만에 450억원의 손실을 입었고 결국 1년 후 파산했습니다.

2016년 미국의 쇼핑센터에 배치된 보안 로봇이 어린 아이를 공격하는 일도 있었고, 2018년에는 미국 애리조나에서 무단횡단하던 보행자가 시험 운행 중인 우버(Uber)의 자율주행차에 치어 사망하기도 했습니다.

포털 사이트의 인공지능 뉴스 편집에 대한 공정성과 신뢰성에 대한 의문도 지속되고 있습니다. 결국 시장에 모든 것을 맡겨 둔다면 인공지능의 신뢰성은 담보될 수 없습니다.

최근 서비스 중단된 인공지능 챗봇 '이루다'문제가 특히 그렇습니다. 하지만 인공지능이라는 낯선 기술에 대한 사회적 합의의 필요성을 더욱 확실하게 인식하게 되었다는 점은 긍정적입니다.

저는 새로운 산업의 발전과 국민의 권리 보호를 위해 자율과 책임의 원칙을 강조하고 싶습니다.

규제 일변도의 정책으로 인공지능 산업이 세계와의 경쟁에서 뒤처지면 안 된다는 점은 모두가 공감하는 것입니다.

개발자와 기업의 창의력과 경쟁력은 폭넓은 자율성의 보장으로부터 나옵니다. 그러나 자율에는 반드시 그만큼의 책임이 뒤따릅니다.

이 원칙은 투명성과 공정성 확보가 전제되어야 합니다. 특히 공정은 기계적 균형이 아닌 불편부당이라는 개념으로 접근해야 합니다. 기술 발전의 방향은 예측하기 어렵기 때문에, 상황에 맞는 적절한 공정성을 찾아나가야 합니다.

이를 위해서는 현행 법체계에 대한 고려와 사회적 합의를 바탕으로 설명요구권과 이의제기권에 대한 깊이 있는 검토가 필요합니다. 오늘 토론회에서 많은 의견을 부탁드립니다.

저는 더불어민주당 정보통신특별위원장이자 국회 과학기술정보방송통신위원회 소속 위원으로서 여러분의 소중한 의견을 정책으로 적극 반영할 수 있도록 역할과 책무를 다하겠습니다.

이번 토론회에 참석해주신 여러분들의 충분한 의견이 오가길 희망하며, 참석하신 모든 분들의 건강을 기원하겠습니다.
감사합니다. ▢

헌법과 인공지능

김민우*

≪목 차≫

Ⅰ. 서론
Ⅱ. 헌법의 규범과 역할의 변화
Ⅲ. 인공지능사회에서의 기본권 문제
Ⅳ. 인공지능의 헌법 수용 가능성
Ⅴ. 인공지능사회에서의 입법 과제
Ⅵ. 결론

Ⅰ. 서론

우리는 요즘 AI와 인간의 대결에 큰 관심을 갖고 있다. 한 방송에서도 AI 대 인간의 대결을 펼치는 경연 프로그램을 선보이고 있다. 하지만 인간이 빅데이터 기술을 바탕으로 스스로 학습하고 진화하는 AI를 넘어서기는 결코 쉽지 않다. 당연한 결과이며 놀라운 것이 아니다. 초고속 인터넷의 등장은 개인의 삶의 풍경을 바꾸더니 스마트폰 사용의 일반화는 '초연결(超連結)' 사회라는 용어를 만들어 냈다. 현재 우리가 살아가고 있는 초연결사회는 이제 더 이상 특정한 시간에, 특정한 공간에 모여서 의견과 정보를 주고받지 않아도 언제, 어디서나, 원하는 시간에, 원하는 정보를 공유하고 얻을 수 있으며 이 모든 과정이 인터넷상에 기록되어 남아있는 사회인 것이다.[1] 이처럼 지능정보기술에

* 충북대학교 행정학과 BK21사업팀(플랫폼 시대 공공거버넌스 미래인재양성팀) 박사후연구원, 법학박사
1) 황용석 외, 연결사회에서의 소통과 공유, 시간과 물레, 2017, 14면.

기반한 지능정보화의 급속한 진전으로 인해 전 세계적으로 경제시스템과 사회구조를 근본적으로 변화하는 패러다임의 전환기에 있다고 할 수 있다.[2)]

인공지능을 기반으로 한 제4차 산업혁명시대는 장밋빛 미래전망에서부터 냉소적인 종말론에 이르기까지 다양한 견해들이 제시되고 있다. 더군다나 인공지능사회는 계속 진화 중인 기술을 전제로 한다는 점에서 선행적인 법정책을 마련할 수 없을 뿐만 아니라 어떤 경우에는 명확한 문제점이나 정책적 과제를 찾아내는 것조차도 쉽지 않은 문제점을 나타내고 있다.[3)] 하지만 인공지능을 이용한 신기술에 대한 법제도의 불확실성이나 부존(不存)으로 인한 이러한 문제점들은 4차 산업혁명 시대를 살아가야 하는 우리들에게 차선책을 마련하거나 인공지능 사회가 제기하는 법적 이슈들을 검토해야 할 의무를 부과하고 있다.

인공지능 사회는 현재와는 다른 새로운 차원의 이슈를 제기하면서 새로운 규범체계의 정립을 요구하고 있다. 따라서 이러한 준비의 일환으로 인공지능사회에 부합하는 새로운 헌법상의 이념이나 원칙에 대한 수립이 필요하다. 이하의 내용에서는 이러한 문제의식에서 출발하여 인공지능사회에서 새롭게 논의되는 헌법의 규범과 역할을 바탕으로 새로운 헌법으로의 방향에 대해서 논하고자 한다.

Ⅱ. 헌법의 규범과 역할의 변화

1. 헌법상 국가의 책무

헌법 제10조는 “모든 국민은 인간으로서의 존엄과 가치를 가지며 행복을 추구할 권리를 가진다. 국가는 개인이 가지는 불가침의 기본적 인권을 확인하고 이를 보장할 의무를 가진다”라고 규정하여 “있어야 할” 당위로서의 국가의 책무를 정한 헌법의 규범성을 보여주고 있다.[4)] 문제는 이러한 국가의 의무는 권력분립의 원칙에 따라 원칙적으로 국회의 입법을 기다려 행정에 의한 구체화가 이루어진다.[5)] 특히 자본주의를 근간으로 하는 시장질서에 있어서는 ‘소유한 자’와 ‘소유하지 못한 자’간의 대립을 최소화하면서 다층 · 다양한 사회구성원의 조화적 양립 · 발전을 도모할 국가책무의 수행을 위해서 “재산권의 공공복리 적합성”을 규정(제23조 제2항)함과 아울러 “적정한 소득의 분배와 시장지배력의 남용방지를 통한 국가의 규제 · 조정권”을 각각 규정(제119조 제2항)하고 있다.

---

2) 정준현 · 김민호, “지능정보사회와 헌법상 국가의 책무”, 법조 제66권 제3호, 법조협회, 2017, 108면.
3) 조소영, “지능정보사회에서 인격권의 새로운 보호체계 검토”, 공법학연구 제21권 제3호, 한국비교공법학회, 2020. 8, 110면.
4) 정재황, 헌법입문, 박영사. 2021, 6면.
5) 정준현 · 김민호, 앞의 논문, 118면.

헌법 제37조 제2항은 “국가안전보장, 질서유지 또는 공공복리를 위하여 필요한 경우에 한하여 법률로써 국민의 모든 자유와 권리를 제한하되 본질적 내용은 침해하지 못한다”고 규정함으로써, 국민의 기본권 보장을 위해 필요한 법률은 국회의 입법에 의하되, 어떠한 입법이 필요한지 여부는 국회의 판단에 의하도록 하는 의회유보설 내지 본질유보설을 따르고 있다.[6)]

2. 새로운 법적 쟁점

4차 산업혁명의 핵심적인 요소라고 할 수 있는 인공지능기술이 급속도로 발전하고, 그 가능성을 명확히 예측하기 어려운 상황임에도 불구하고, 이러한 기술의 발전으로 야기할 수 있는 법적 쟁점은 매우 다양하며 우리 생활에 직·간접적인 영향을 미칠것으로 보인다. 인공지능의 기본권 주체성과 법인격 주체성, 책임능력, 데이터 활용에 따른 개인정보보호의 문제, 인공지능의 오작동으로 인한 사고 발생 시 책임 귀속의 문제, 인공지능에 의한 의사결정상 투명성과 책임성 확보방안, 인공지능기술의 발전에 따른 노동시장 충격의 완화, 인공지능 알고리즘에 의한 차별 등 인권 침해 방지, 자율무기의 인도법적 규율과 통제, 인공지능 창작물과 저작권 등 보호 문제, 인공지능 기술 격차와 보편적 접근성의 보장 문제 등이 주요 쟁점으로 들 수 있다.[7)]

지능정보기술 등 각종 신기술에 대해 획일규제나 장식적 규제가 되지 않기 위해 개별법을 통해서 대응한다는 것은 한편으로는 위의 쟁점에 대한 해결책이 될 수도 있으나, 이러한 문제를 법에 의해서 잠재적으로 승인하게 된다는 문제를 가진다. 이러한 점에서, 충분한 헌법적 문제의식에도 불구하고 지능정보기술의 발전에서 나타나는 불확실성을 헌법규정의 차원에서, 다시 말해 헌법개정을 통해 사전적으로 해소한다는 견해[8)] 역시 한계가 있을 수 있다.

3. 헌법가치 체계의 변화

과학기술의 발전으로 인해 사회구조가 변화하고, 새로운 생활관계의 출연에 따른 새로운 가치가 등장하게 되면, 이에 대한 규범적 보호의 요청도 새롭게 등장하게 된다.[9)] 예를 들어, 생명공학기술의 발전은 ‘배아의 법적 보호라는 헌법적 가치’의 승인으로 이어지고,[10)] 정보화 사회가 심화되어 감에 따라 ‘개인정보에 대한 자기결정권’이 새로운

---

6) 대법원 2015. 8. 20. 선고 2012두23808 전원합의체 판결.
7) 장민선, “인공지능의 법적 쟁점에 관한 전문가 의견조사”, 법연 Vol. 62, 한국법제연구원, 2019, 36-39면.
8) 정준현·김민호, 앞의 논문, 107면.
9) 같은 견해로는 김선택, “기본권보장의 발전과 기본권학의 과제”, 공법연구 제37집 제2호, 한국공법학회, 2008, 73면; 이장희, “기본권의 개념 및 인정 기준과 법률적 권리의 관계”, 헌법이론과 실무 2015-A-2, 헌법재판연구원, 2015, 7면.

기본권으로 인정되기에 이르렀다.[11] 이후 가상공간에 올려진 수많은 정보로 인해 괴로움을 당하는 사람들이 증가하면서 사회문제화되자 '잊힐 권리'가 새로운 권리가 되었다.[12]

오늘날 전 세계적으로 헌법 상호 간 이념적 지향성과 내용이 일반성과 보편성을 공유하는 경향을 갖게 된 현상의 주요 요인으로 정보화를 들 수 있다.[13] 지능정보사회에서 국가는 국민의 기본권 실현이 과학기술의 발전과 조화될 수 있도록 하는 과제를 갖는다. 이는 헌법과 기타 관련 법률들이 그 핵심적 역할과 기능을 수행하면서도 과학기술의 발전에 따른 새로운 사회변화를 헌법적 차원에서 받아들이는 노력을 지속해야 한다는 것을 의미한다.[14] 국가적 차원에서의 지능정보사회의 새로운 사회질서의 구조와 조건을 밝히는 기초적 판단의 규범적 근거가 헌법이며, 이에 대한 판단기준은 헌법적 가치이다. 다시 말해, 인공지능기술이 보편적 · 일상적으로 활용되기 시작하면서, 당연히 헌법 해석에 대한 논란이 발생할 수밖에 없고, 이때 특정 인공지능기술의 구현과 활용, 그리고 그것이 제시하는 판단 결과가 헌법적 가치와 부합하는 것으로 볼 것인지의 여부는 규범적 판단영역이다.[15] 그리고 이에 대한 규범적 판단 후에 해당 헌법적 가치가 현행 헌법규범을 통해 파악되지 못한다면 헌법의 가치체계에 대한 변화가 필요하다.

Ⅲ. 인공지능사회에서의 기본권 문제

1. 서설

국가에서 개인의 지위를 국가에 대한 권리를 중심으로 규정하는 근본규범인 기본권은 '인간 삶에 있어서 중요한 가치의 헌법적 수용'이라는 맥락에서 이해될 수 있다. 따라서 헌법은 삶의 문화적 가치가 법체계로 수용되고 내재된 국가공동체의 기본질서라고 할 수 있으며, 그 중심이 되는 가치가 '기본권'이라 할 수 있다.[16] 또한 기본권 보장은 국가의 헌법적 과제이며, 민주주의와 법치주의에 기초한 국가작용을 통해 실현된다. 다시 말해, 기본권은 민주적 입법에 의해 그 보장과 실현을 위한 법적 기초가 마련되어야

---

10) 헌법재판소 2010.5.27. 2005헌마346, 판례집 22-1(하), 296면.
11) 헌법재판소 2005.5.26. 99헌마513 등, 판례집 17-1, 682면 등 참조.
12) 허완중, "잊힐 권리에 관한 헌법적 검토", 한국헌법학회 · 한국정보화진흥원 · 전남대 법학연구소 · 충남대 법학연구소 · 부산대 법학연구소 공동학술대회 자료집, 2020, 243-269면 참조.
13) 권영설, "변화하는 헌법과 거버넌스", 연세 공공거버넌스와 법 제3권 제1호, 연세대학교 법학연구원, 2012, 15면.
14) 박기주, "데이터 혁신 시대를 위한 정보기본권 연구", 미디어와 인격권 제4권 제1호, 언론중재위원회, 2018, 62-63면; 같은 견해로는 김배원, "지능정보사회와 헌법-인공지능(AI)의 발전과 헌법적 접근-", 공법학연구 제21권 제3호, 한국비교공법학회, 2020. 8, 79-82면 참조.
15) 심우민, "인공지능 시대의 자유와 민주주의, 그리고 입법", 인공지능과 법, 박영사, 2019, 85-86면.
16) 이장희, 앞의 연구보고서, 70면.

하며, 이후 법의 집행과정 및 사법작용을 통해서도 보장되어야 한다.

인공지능사회에서의 기본권적 문제는 크게 (1) 과학기술의 발전으로 인한 인공지능의 기본권 주체성의 문제와 (2) 변화된 헌법현실에서 발생할 수 있는 기본권의 문제, 즉 인공지능의 발전에 의한 기본권 경합이나 기본권 충돌, 그리고 국가의 기본보호의무, 새로운 생활관계의 출연으로 새로운 기본권 인정의 문제로 나누어 볼 수 있다. 그리고 후자의 경우 먼저, 기본권 경합은 위치정보 혹은 섹스로봇 등과 같은 경우에 개인정보자기결정권과 사생활 비밀의 자유 혹은 성적 자기결정권과 사생활 비밀의 자유 등이 논의될 수 있다.[17] 다음으로, 기본권 충돌의 경우, 이미 경험하고 있는 것과 같이, 코로나19 상황에서 감염병 환자의 이동경로나 접촉자 현황 등의 알권리와 그에 상응하여 침해되는 감염병 환자의 개인정보자기결정권 및 사생활 비밀의 자유 간의 충돌, 빅데이터 사회에서 이에 포함된 개인정보를 이용함으로써 누리는 영업의 자유와 그에 상응하여 침해되는 개인정보 주체의 개인정보자기결정권 간의 충돌이 가장 대표적이다. 기본권 충돌의 경우에는 사인에 의한 기본권 침해의 양상으로 나타날 수도 있기 때문에 국가의 기본권보호의무와 함께 논의되어야 한다. 마지막으로 새로운 기본권의 인정의 문제는 인공지능의 헌법 수용 가능성(Ⅳ)에서 함께 다루기로 한다.

2. 새로운 영역에서의 기본권 충돌

인공지능의 발전에 수반하는 빅데이터의 활용과 개인정보 보호의 문제는 기업에서는 영업의 자유와 개인의 개인정보자기결정권이 상호 충돌하는 영역이다. 이와 관련하여 기존의 개인정보보호법제는 개인정보 개념의 광범위성과 불명확성, 가명정보 활용 범위의 불명확성, 개인정보 처리에 대한 동의요건의 엄격성, 개인정보보호 법체계의 분산·중복성, 개인정보보호 정책 추진체계의 일관성 및 독립성 결여 등이 지적되었다. 이러한 법체제 하에서 데이터 활용이 위축되고 관련 산업발전을 저해한다는 비판이 있어 왔다.[18] 이에 2019년 데이터 3법(개인정보보호법, 신용정보법, 정보통신망법)이 국회를 통과하면서, 그 동안 분산·중복되어 있었던 개인정보보호법제를 일원화하고, 개인정보 개념에 가명처리를 추가하여 개인 식별이 불가능하게 가명처리 할 경우 정보주체의 동의 없이도 개인정보를 사용할 수 있게 되었으며, 통계작성 및 과학적 연구, 공익적 기록보존을 위해 개인정보를 활용할 수 있게 되었다. 그러나 여전히 데이터의 '활용'보다는 '보호'에만 치중한다는 업계의 불만도 제기되고 있으며,[19] 개인정보의 침해 가능성은 여전히 우려되고 있다.[20]

---

17) 김배원, 앞의 논문, 84면.

18) 신용우, "데이터 경제 시대의 개인정보 보호 법제 관련 쟁점 및 개선과제", 이슈와 논점 제1593호, 국회입법조사처, 2019, 1면 이하 참조.

19) "[데이터경제에 길을 묻다] 활용은 없고, 겁겹이 규제만…시행령부터 엉켰다", 아시아경제 2020.7.8.자 기사(https://www.asiae.co.kr/article/2020070209485958233, 2021.2.1. 최종방문) 참조.

20) 인권위, "데이터3법 시행령, 과도한 정보 수집 우려", KBS 2020.6.8.자

한편, 코로나19와 같은 상황에서 감염병 환자의 이동경로나 접촉자 현황 등의 알권리와 그에 상응하여 침해되는 감염병 환자의 개인정보자기결정권 및 사생활 비밀의 자유 간 충돌의 문제도 발생할 수 있다.[21] 감염병 환자의 이동경로나 접촉자 현황 등의 알권리에 대해 「감염병의 예방 및 관리에 관한 법률」 등에서 근거 규정을 마련하고는 있지만, 여전히 정보공개의 적정성 등과 관련하여 개인정보자기결정권이 침해될 가능성이 있다.[22]

이러한 개인정보보호의 문제와 또 다른 측면에서 최근 문제되고 있는 것은 온라인 플랫폼 사업자들의 데이터 활용의 범위가 넓어지면서 나타나고 있는 데이터 집중 현상이다. 우리가 흔히 알고 있는 플랫폼 사업자들은 방대한 양의 데이터를 축적하고 더 진화된 알고리즘 기술을 통해 빅데이터 활용의 주요 주체로서 데이터를 독점하고 있다. 예컨대, 구글이나 넷플릭스 등의 플랫폼 사업자들은 인공지능 서비스, 타겟광고 서비스, 이용자 추천 알고리즘 등과 데이터를 연계시키며 무한한 확장을 꾀하고 있으며, 국내에서도 네이버와 카카오 등을 중심으로 데이터 집중이 심화되고 있다. 온라인 플랫폼 사업자들은 축적되어지는 개인정보에 관한 데이터를 바탕으로 더 혁신적인 이용자 맞춤형 서비스를 제공할 수 있으며, 이를 기반으로 신규서비스 사업을 확장해 나갈 수도 있다.[23] 이처럼 온라인 플랫폼 사업자들을 중심으로 거대한 데이터를 축적하면서 나타나게 되는 데이터 집중 현상은 시장 독점의 문제나 이용자 집중화 현상과 연결되며 이용자들의 권리를 침해할 수 있다. 이에 '열람차단청구권'이나 '사이버 액세스권' 등 액세스권의 개념에 대한 재논의가 필요한 시점이다.[24]

3. 새로운 영역에서의 기본권 침해

(1) 개인의 자유

지능정보화의 진전은 다양한 사회문제를 노출할 가능성도 있다. 정보지능기술을 활용한 대량의 개인 데이터를 수집 · 분석 · 이용함으로써 개인의 사생활이 침해될 수 있다. 또한 지능정보기술이 실생활의 거의 모든 사물과 사람에 직접 연결되면 해킹 등 기존 사이버공간의 보안위험이 현실 세계로까지 확대되어 정보보안의 부담이 더욱 가중될 것이다.[25] 이처럼 AI가 생산해 내는 정보, 지식은 본질적으로 인간이 저장하거나 처리능

뉴스(http://news.kbs.co.kr/news/view.do?ncd=4465085&ref=A, 2021.2.1. 최종방문) 참조.

21) 이에 대하여 권건보, "감염병 위기 대응과 정보인권", 한국비교공법학회 제99회 학술대회 자료집, 2020, 29면 이하 참조.

22) 권건보, 앞의 논문, 47-50면 참조.

23) 유승현, "플랫폼 사업자의 데이터 집중과 액세스권 개념의 재정립", 미디어와 인격권 제6권 제1호, 언론중재위원회, 2020, 54면.

24) 유승현, 앞의 논문, 60면.

25) 정준현 · 김민호, 앞의 논문, 109면.

력의 부족으로 생산해 내지 못하던 정보와 지식이다. 이러한 정보에 기초하여 규범이 정립되거나 집행될 경우 인간은 자신이 인식하지 못하던 이유에 의해서 또는 자신이 의식하지 못하는 메커니즘에 의해서 규율받게 되고, 이는 인간의 자유를 침해할 수밖에 없다.

헌법상 법치주의는 개인의 자유를 보장하는 역할이 크다. AI를 활용하는 결과가 헌법상 자유가 정당하게 제한될 수 있는 범위 내에 있는지 적절한 통제가 필요하다.

(2) 불평등

기술의 발전은 불평등을 심화시킨다. 노동대체, 부의 양극화, 기계종속성 등에 대한 우려와 두려움 또한 크다.[26] 그리고 지능정보기술의 활용에 따라 개인 간에는 심각한 정보의 격차를 야기하고, 이로 인한 고용과 소득의 양극화는 심화될 것이다. 이처럼, 데이터나 알고리즘에 스며든 차별 등과 같이 AI에 내재한 불평등의 문제는 많은 논의가 있었다. 그런데 이에 못지 않게 AI를 통한 법률서비스가 유료로 이루어질 경우, 이를 이용할 경제적 능력 유무에 따른 AI의 외부에 존재하는 불평등의 문제가 있다. 물론 이것은 새로운 문제가 아니다. AI가 아니라 사람인 법률가가 제공하는 서비스도 그 변호사의 전문성과 열정, 그리고 전관예우, 대형법무법인, 인맥과 같은 사회구조적 요소에 의해 차별이 이루어지기 때문이다. 전자가 알고리즘 투명성과 재학습 규제 등을 통해 해결되어야 한다면, 후자는 AI를 통한 공공법률서비스의 확대 등을 통해 해결되어야 할 것이다.

(3) 형평성

AI를 활용한 자동결정 등의 장점은 시간과 자원의 절약 외에 일관성, 중립성 등이 제시된다. 실제로 그렇다면 AI의 활용과 관련하여 개별적 사안의 특수성을 고려하는 기능을 어떻게 확보할 것인지를 고려하여야 한다. 높은 수준의 법치주의는 일관성이 획일성으로 남용되지 않도록, '같은 것은 같게'와 '다른 것은 다르게'의 절묘한 균형점을 실현해 내야 한다. 가령 과거 영미법에서 '보통법(common law)'과 '형평(법)(equity)'을 분리하여 집행하였던 데서 볼 수 있듯이 일관된 법적 기준으로 통치하는 것은 그 일관된 기준을 적용하는 것이 정의의 원칙상 도저히 정당화될 수 없는 사례를 가려내고 이에 대해 적절한 접근을 하는 것으로 완성된다. 이러한 개별적 특수성의 고려가 AI 내부적으로 이루어지기 어렵거나, 신뢰하기 어렵다면, AI를 포함한 외부절차에서 이를 구현하는 방법도 있다. 즉, AI를 통한 1차 결정에 대해 간단한 이의제기로 사람에 의한 추

26) 김민호 · 이규정 · 김현경, "지능정보사회의 규범설정 기본원칙에 대한 고찰", 성균관법학 제28권 제3호, 성균관대학교 법학연구소, 2016. 9, 283면.

가심사가 이루어질 기회를 마련하는 등의 방법이다.

(4) 신뢰성

지능정보기술과 정보·데이터 이용의 신뢰성 확보문제도 이슈로 등장할 수 있다. 이용자에게 적합한 정보수집과 의사결정 과정에서 지능정보기술로 최적화된 결과만 선택받고 나머지 정보는 외면당할 우려가 있다. 인공지능기술에 의한 형식논리적 데이터 분석이 정보를 왜곡하여 소수집단의 인권을 침해하거나 마케팅의 공정성을 저해할 위험성도 제기되고 있다.

(5) 기타

그 밖에도 인간의 지적 능력을 뛰어넘는 지능적 개체들에 대한 인간의 통제 가능성 상실 우려와 윤리성 확보 문제, 자율적 판단기능을 가진 인공지능, 로봇 등의 행위에 대한 책임소재 문제, 인공지능 기술을 활용한 창작물의 저작권 문제 등 지금까지 우리가 겪어보지 못한 다양한 법적 쟁점으로 인하여 개인의 기본권이 침해될 것이다.[27)]

Ⅳ. 인공지능의 헌법 수용 가능성

1. 기본권 주체성

(1) 인공지능의 책임성

인공지능의 자율성·주체성과 관련하여 인공지능을 인간과 동등한 차원에서 논할 수 있는지에 대한 담론과 논쟁은 다양하게 논의되고 있다. 즉 인공지능을 인간과 동등한 차원에서 논할 수 있는지에 대한 담론 등은 철학적·윤리적 분야에서 선도하고 있으며,[28)] 법학에서는 인공지능의 책임성을 중심으로 관련 분야에서 입법정책적 논의가 주

27) 정준현·김민호, 앞의 논문, 110면.

28) 이중원 외 7인, 인공지능의 존재론, 한울 아카데미, 2018, 1면 이하; 이중원 엮음, 이중원 외 8인, 인공지능의 윤리학, 한울 아카데미, 2019, 1면 이하; 고인석, "로봇이 책임과 권한의 주체일 수 있는가?", 철학총론 제67집 제1권, 새한철학회, 2012, 3면; 김진석, "약한 인공지능과 강한 인공지능의 구별의 문제", 철학연구 제117집, 철학연구회, 2017, 111면 이하; 이상형, "윤리적 인공지능은 가능한가?-인공지능의 도덕적, 법적 책임 문제-", 법과 정책연구 제16집 제4호, 한국법정책학회, 2016, 283면; 정대현, "특이점 인문학 특이점 로봇은 인간사회의 성원이다", 철학 제131집, 한국철학회, 2017, 189면; 한희원, "인공지능(AI)의 법인격 주체 가능성의 이론적 기틀에 대한 기초 연구", 중앙법학 제20집

를 이룬다.[29] 지금의 인공지능은 인간에 대한 보조적인 역할에 불과하지만, 앞으로 자율주행자동차나 인공지능의 창작능력이 기술의 발전에 따라 인간의 통제를 벗어나거나 주체적인 역할의 정도가 강해질수록 인공지능의 자율성·주체성을 전제로 한 법적 책임이나 법적 권리가 문제될 수 있다.[30] 이러한 인공지능의 발전은 인공지능과 인간이 공존하는 상황에서의 인공지능의 법적 지위의 문제, 헌법적으로는 기본권 주체성의 문제를 발생시킨다. 현재 인공지능의 기본권 주체성에 대한 본격적인 논의나 판례를 찾아보기는 어렵다.[31] 인공지능을 법적 주체로 인정하고, 법적 책임과 법적 권리를 인정한 경우에는 법률적 차원에서 그칠 것이 아니라 헌법적 차원에서의 논의도 병행되어야 할 것이다. 인공지능의 법적 책임이나 권리를 인정한다고 하더라도, 그와 관련한 분쟁은 결국 현행법 체계상 사법절차를 통하여 해결할 수밖에 없다. 그러나 법적 분쟁의 당사자인 인공지능에 재판청구권이 인정되지 않는다면 사법절차를 통한 법적 분쟁을 해결할 수 없을 것이다. 인공지능에 법적 책임을 부담한다거나 법적 권리를 인정한다는 것은 사법절차에서 적정한 책임과 권리의 확보를 위한 구제수단을 부여할 때 비로소 진정한 법적 의미를 지닐 것이다.[32]

이러한 의미에서 인공지능의 법적 책임과 권리의 인정은 헌법적 차원에서 기본권 주체성과 함께 논의되는 것이 바람직하다. 인간과 인공지능이 공존하는 지능정보사회에서 기본권적 접근은 인공지능의 발전에 의한 '인간 또는 인간 상호간의 기본권 관계'와 '인공지능 자체의 기본권 주체성'으로 나눌 수 있다. 기본권 주체인 자연인과 법인을 비교컨대, 인공지능은 스스로 의사결정 및 판단 능력을 지닌다는 점에서 자연인과 유사하지만 법인과 다르고, 생리적 기능을 갖지 못한다는 점에서 법인과 유사하지만 자연인과 다르다. 인공지능이 판단능력을 가진다고 하더라도 윤리적 판단의 주체가 될 수 있는지 여전히 논쟁의 대상이 되고 있다.[33]

제3호, 중앙대학교 법학연구소, 2018, 375면 참조.

29) 김용주, "인공지능(AI; Artificial Intelligence)의 창작물에 대한 저작물로서의 보호가능성", 법학연구 제27권 제3호, 충남대학교 법학연구소, 2016, 267면; 김윤명, "人工知能(로봇)의 법적 쟁점에 대한 試論的 考察", 정보법학 제20권 제1호, 한국정보법학회, 2016, 141면; 계승균, "법규범에서 인공지능의 주체성 여부", 법조 Vol.724, 법조협회, 2017, 158면; 최재원, "인공지능 창작물에 대한 저작권의 주체", 문화·미디어·엔터테인먼트법 제11권 제1호, 중앙대학교 법학연구원, 2017, 117면; 양천수, "인공지능과 법체계의 변화 -형사사법을 예로 하여-", 법철학연구 제20권 제2호, 한국법철학회, 2017, 45쪽 이하; 손영화, "인공지능(AI) 시대의 법적 과제", 법과 정책연구 제16집 제4호, 한국법정책학회, 2016, 305면.

30) 김배원, 앞의 논문, 83면.

31) 한편, 핸슨 로보틱스(Hanson Robotics)의 휴머노이드 로봇 소피아(Sophia)는 사우디아라비아의 시민권을 취득한 데 이어 휴머노이드도 가족을 가질 자격이 있다고 하였다(사우디, 히잡 안쓴 '여성로봇 소피아'에 시민권 부여, 문화일보, 2017.10.27.자(http://www.munhwa.com/news/view.html?no=2017102701071030307001, 최종방문일 2021.2.3.).

32) 김배원, 앞의 논문, 84면.

33) 김배원, 앞의 논문, 84면.

(2) 인공지능의 재판청구권

인공지능에 대한 정의는 보는 시각에 따라 다양하지만,[34] 헌법적 논의의 대상이 되는 인공지능은 인간의 지시 내지 설계에 의해서만 구동되는 물건의 수준을 넘어서, 자율적 판단이 가능하고, 현실적인 문제 상황에 대처 또는 창의적인 능력을 발휘할 것으로 예상되는 인공지능이다.[35] 따라서 법률적 차원에서 인공지능을 법적 주체로 인정하고 법적 책임이나 권리를 부여한다고 하더라도, 결국 그와 관련된 법적 분쟁에서는 현행법 체계상 궁극적으로는 사법절차를 통해 해결할 수밖에 없다는 점을 고려한다면, 법적 분쟁의 당사자인 인공지능에게 재판청구권이 인정되어야 할 것이다.[36] 즉 인공지능의 법적 책임을 지우거나 법적 권리를 인정한다는 것은 사법절차에서 적정한 책임과 권리의 확보를 위한 구제수단을 부여할 수 있어야 함을 의미하고,[37] 이러한 점에서 인공지능의 법적 책임성과 권리성의 문제는 기본권 주체성과 함께 논의되어야 한다. 이를 위해서는 우선 인공지능에게 일반적인 권리와 의무의 주체가 될 수 있는 법률상의 지위 또는 자격,382) 즉 법인격(法人格)을 부여할 수 있는지를 함께 알아보아야 한다.

2. 법인격

우리의 현행법에서는 인적결합이나 재산에도 법인격을 인정하고 있다. 기업의 법인이나 공공단체도 사람과 마찬가지로 권리와 의무를 부여하고 있으며, 그에 따른 법적 책임을 부과하고 있다. 예컨대, 인공지능이 사람의 능력과 비슷한 수준에 도달한다면 인공지능도 인격권을 가진 존재로서의 가치를 인정할 수 있을 것이다. 다만, 인간처럼 헌법상 부여된 기본권과 대의민주주의의 참여, 민법상 손해배상청구, 형법상 형사처벌의 문제, 노동법상 근로자의 권리, 특허법상 창작물의 권리 등 인간사회에서 발생할 수 있는 여러 가지 권리를 얼만큼, 어디까지 인정하여야 하는지가 관건이다.

일반적으로 인공지능의 법인격 문제는 자율주행자동차나 인간형 로봇 등이 고의 또는 과실로 사람에게 손해를 주는 경우 이들에게 독자적인 법적 책임을 물을 수 있는지가 논의되면서 등장하였다. 인공지능의 법적 책임을 묻기 위해서는 책임능력을 인정할 수 있어야 하고, 이를 위해서는 다시 인공지능의 법인격이 인정되어야 한다. 인공지능

---

34) 대표적으로 인간의 지능적 행동을 수행하도록 공학적 응용을 모색하는 '약한 인공지능'과 인간과 같은 사고체계로 문제를 분석하고 행동할 수 있는 '강한 인공지능'으로 분류하기도 한다(양종모, "인공지능의 위험의 특성과 법적 규제방안", 홍익법학 제17권 제4호, 홍익대학교 법학연구소, 2016, 540-541면.

35) 이에 대해 인공지능기술이 추구하는 최종적인 목표는 인간과 같은, 혹은 인간의 능력을 뛰어넘는 것이지만, 현재의 기술 수준은 이에 미치지 못하므로, 인공지능의 법적 지위 등을 논하기에는 시기상조라는 견해도 있다(장재옥 · 김현희, "인공지능의 법적 지위에 관한 논의", 법학논문집 제43집 제1호, 중앙대학교 법학연구원, 2019, 120면 참조).

36) 김배원, 앞의 논문, 84면.

37) 김배원, 앞의 논문, 84면

의 법적 권리와 의무를 부여하는 법인격을 인정하는 것은 인간과 인공지능이 서로 공존하는 세상에서 조화와 협력을 통해 살아간다는 것을 뜻한다. 이처럼 인공지능의 인격권이 기본권으로서 보장되어야 하는 중요한 헌법적 가치임은 분명한 것이고, 이러한 헌법적 가치는 사회변화에 대한 사실판단과 상호작용 속에서 구현되어야만 한다.

3. 기본소득제의 도입

인공지능사회의 여러 문제점의 대안으로서의 '기본소득' 논의는 한편 조심스럽다. 다시 말해 기본소득에 대해 인플레이션으로 인한 실질 구매력 약화, 노동자의 협상력 약화, 저임금 일자리 양산, 도덕적 정의 문제, 불운과 탕진의 문제, 빈곤해결의 실효성 문제는 별론으로 하더라도, 실제 기본소득이 사회적 기본권 실현의 대안이 될 수 있는가에 대해서는 의문이다. 물론 기본소득의 이념 자체는 국가에 대해 일정한 급부를 제공하도록 요구한다는 점에서 일응 사회적 기본권의 보호영역에 포섭될 수 있고, 경제적 약자에게 일정한 금품이 제공된다는 점에서 공공부조와 유사한 측면이 있다. 그러나 공유재의 관념에 기초하여 공동체 구성원으로서 요구할 수 있는 기본소득의 개념(무조건성, 보편성, 개인성, 현금지급성, 충분성, 정기성) 중 무조건성, 보편성, 충분성을 고려한다면, 막대한 재원이 집중되어야 하는데, 이는 종래 사회적 기본권의 실현을 위해 구축한 사회보장제도에 부정적 영향을 미치게 될 것이다.

그러나 인공지능에 의한 노동력의 대체로 인한 대안으로 잡 쉐어링과 보편적 복지를 실행하는 기본소득제가 등장했는데, 현실성이 애매해서 논의만 계속되다가 2015년 말에 핀란드에서 기본소득제의 현실성을 알아보기 위한 실험으로 몇 달 동안만 기본소득제를 도입해 보기로 했다. 실험 결과, 핀란드 정부는 기본소득 실험이 삶의 질을 높이는 데는 성공했지만 고용률을 높이는 데는 실패했다는 결과를 발표했다. 기본소득제를 도입은 인공지능이 충분히 노동인력을 대처할 수준까지 오게 된다면 진지하게 사회보장제도에 대해 충분한 논의를 바탕으로 도입을 하는 것을 고려하여야 할 것이다.[38] 이에 대해 기본소득을 사회보장제도의 보완재로 활용하자는 견해도 있으나, 이 경우 기본소득의 충분성을 담보할 수 없고, 특히 기본소득의 재원과 관련하여 매번 함께 거론되는 '관료제의 축소 및 행정비용 감축을 통한 보장급부의 확대'가 사실상 불가능하다는 점에서, 기본소득의 실질적 도입은 지금까지의 사회보장제도를 대체할 경우에나 가능하다고 할 것이다. 그렇다면, 기본소득이 과연 지금의 제도보다 사회적 기본권을 효과적으로 실현하는 제도적 대안이 될 수 있는지에 대한 고민은 앞으로 국민적 합의를 통해서 해결하는 것이 필요하다.

---

38) 김병록, "인공지능의 헌법적 쟁점과 과제", 법학논총 제27집 제2호, 조선대학교 법학연구원, 2020. 8, 100면.

Ⅴ. 인공지능사회에서의 입법 과제[39]

1. 새로운 법규범 필요

인공지능 기술이 발달할수록 인간과 유사하게 사고할 수 있고, 나아가 이들과 인간이 상호 소통하면서 사회질서를 유지하게 될 것이다. 이렇게 된다면 그들을 위한 사회질서를 새롭게 형성하기 위한 법규범이 나타날 것이다. 물론 인공지능이 인간과 동일한 수준의 이성적 판단을 할 수 있는지 기술적으로 명확하지 않더라도, 어쨌든 경우에 따라서는 대화의 상대방을 단순한 기계가 아닌 인간과 유사하게 느끼게 된다는 점은 기존 규범질서의 변화가 불가피하다는 것을 보여준다.[40] 다양한 형태의 인공지능의 등장과 향후 예상되는 상용화로 인해 인공지능의 변화에만 머물러 왔던 것이 책임귀속의 문제까지 방향에 대한 인식변화가 절실해졌다. 이에 따라, 인공지능 등 새로운 지능적 존재의 법적 지위, 권리 및 민·형사 책임의 범위를 어떻게 설정할 것인지 등 현행 법체계를 검토하여 실정법적 접근 뿐만 아니라 기술과 인간의 상호작용에 대한 새로운 법규범의 논의가 필요하다.

인공지능 시스템은 사람과의 경계를 무너뜨리고 다양한 데이터와의 복잡한 결합으로 인해 인간에게 해를 미칠 수 있다는 점에서 법의 역할이 중요하다.[41] 지능정보사회에서 입법자의 재량에 맡기기보다는 헌법의 입법조치 명령을 통해 국가가 보다 적극적으로 보호해야 할 기본권이 무엇인지를 찾아 필요한 입법조치의 방향이나 그 대강을 헌법개정안에 담을 필요가 있다는 견해도 있다.[42] 이는 새로운 과학기술의 출현과 발전에 따른 위험에 대비하고 그로부터 국민의 기본권을 보호하고 안전을 지키는 국가의 역할에 대해 고찰하는 것은 헌법적 대응방안을 구체화하는데 중요한 의미가 있다. 이러한 방안은 안전하면서 풍요로운 인간친화적인 기술, 공동체 전체의 공감대를 형성하면서 지속가능한 발전을 추구하기 위한 규범적 틀을 마련할 수 있다. 인공지능에 대한 법적 논의는 명확한 대상조차 정의할 수 없다는 점과 더불어 오히려 세부 분야별 정의를 각각 내려야 한다는 점에서 매우 가변적이고 복잡하게 전개될 수밖에 없다. 이미 규범적 불안전성을 다분히 지니고 있는 인공지능 기술은 사회에 빠르게 적용되고 있으며 많은 영향을 미치고 있다. 이처럼 인공지능의 핵심이 무엇인지 빨리 파악하고 이를 통한 규범적 논의와 원칙을 마련하는 것이 시급하다.

---

39) 이하의 내용은 김민우, 지능정보사회에서의 인공지능의 현안과 입법 과제, 공법학연구 제21권 제2호, 한국비교공법학회, 2020. 5, 159면 이하의 내용을 정리한 것임.

40) 심우민, "이행기 IT법학의 구조와 쟁점 -가상현실과 인공지능의 영향을 중심으로-", 언론과 법 제15권 제1호, 한국언론법학회, 2016, 197면.

41) Calo, Ryan, "Robotics and the Lessons of Cyberlaw", *California Law Review*, Vol. 103, 2015, p.513.

42) 정준현·김민호, 앞의 논문, 121면.

2. 사회적 논의 기구 필요

인공지능에 대하여 보호할 가치가 있다는 명확한 사회적 합의 내지 인식이 있다면 인공지능에 따른 여러 현안들을 극복하고 인공지능과 인간이 함께 공존하는 사회로 나아갈 수 있을 것이다. 이를 위해, 먼저 인공지능 전문가들의 협업적 네트워크를 구축하여 다양한 사회문제 해결에 적용하는 연구와 개발 활성화를 위해 학제간 융합연구가 전제되어야 한다. 인공지능이 사회에 미치는 영향을 고려해 볼 때, 기술발전은 인류의 기존 가치관과 사회질서에 많은 변화를 초래할 것이며, 이는 법이나 제도뿐만 아니라 윤리적인 혼란을 가져올 수도 있다.[43] 따라서 인공지능 기술이 사회에 미치는 영향력과 그 파급력을 고려할 때, 인공지능 생태계에서 이익을 향유하는 사람들은 그에 걸맞는 윤리적 의식과 책임을 가질 필요가 있다.[44]

일반적으로 인공지능은 개발 단계에서부터 인간에게 우호적인 행동을 하도록 설계되어야 한다. 인공지능의 개발 단계에서부터 원천적으로 인간에 대한 해악적 공격을 차단하는 장치를 적용하는 것이 필요하다. 이제 인공지능은 인간을 보조하는 수단을 넘어 인간처럼 권리능력, 의사능력, 행위능력을 모두 갖춘 하나의 인격체로서의 가능성을 보여주고 있기 때문이다. 소위 'AI 시대'의 도래가 멀지 않았음을 알 수 있다. 따라서 우리는 인공지능에 의한 문제점을 줄이기 위해서는 우리 사회의 윤리 규범에 맞추어 규정되어야 하며, 필요한 경우 능력이 박탈될 수도 있어야 한다. 한편, EU의회는 AI가 인간에게 저항하는 것을 예방하는 규정도 마련했다. 로봇의 움직임을 멈출 수 있는 '킬스위치'를 마련해야 한다는 조항과 함께 '로봇 3원칙'을 마련하였다.[45] EU 뿐만 아니라 유엔에서도 '킬러로봇 개발 제한'등 AI 대응에 관한 공식 의제를 중요 안건으로 다루었다.[46]

이제 우리나라도 기술개발에 대한 지원과 더불어 인공지능을 활용한 기술이 인간에게 미칠 수 있는 다양한 영향력과 그 대응방안에 대한 사회적 논의가 필요하다. 이러한 사회적 기구로서 전문위원회를 설치하여 공감대를 배경으로 한 규제의 방향을 논의해 나가는 것이 타당하다.[47] 인공지능 시스템이 사회와 경제에 미치는 영향, 인간의 행위

---

43) 김유환, "과학기술규제의 특성과 규제거버넌스의 재구성", 행정법연구 제47권, 행정법이론실무학회, 2016, 247면.

44) 윤상필 · 권헌영 · 김동욱, "건전한 인공지능 생태계 형성을 위한 규범적 전략과 법의 역할", 홍익법학 제18권 제2호, 2017, 4면.

45) 1. 로봇은 인간을 해칠 수 없다. 2. ①에 위배되지 않는 한 인간의 명령에 복종한다. 3. ①, ②에 위배되지 않는 한 자기 스스로도 지켜야 한다.

46) EU "킬러 로봇 반대"...자동화된 무기 시스템 금지 촉구, 전자신문 2018. 9. 13자, (https://www.etnews.com/20180913000086, 최종방문 2021. 2. 5)

47) EU에서는 전자적 인격(electronic person) 부여, 로봇 등록제 도입, 로봇기술 헌장, 로봇기술 규제기구 창설, 로봇기술 표준화 등과 같은 로봇 관련 기술 법제화를 위한 논의가 활발하게 이루어지고 있다.

와 인간이 만들어 놓은 각종 데이터가 인공지능 시스템에 미치는 영향을 고려하면 공익적 관점에서 종합적인 거버넌스가 필요하다.

현재 우리나라는 「지능형 로봇 개발 및 보급 촉진법」이 제정되어 있으나 주로 지능형 로봇 기술의 개발 및 산업 촉진에 초점이 맞추어져 있다. 최근 〈지능정보사회기본법안〉과 〈로봇기본법안〉이 발의[48]되기는 하였으나 아직 로봇에게 전자적 인격을 부여한다는 것에 대한 사회적 합의가 이루어지지 않은 상태이다. 즉 인공지능 시대를 맞이하여 우리의 법제는 관련 기술의 발전 및 보급에 치중해 있고, 지능정보기술 활용의 보편화로 인해 발생할 수 있는 사회경제적 영향에 대해서는 아직 구체적인 대응책에 대해 본격적인 논의도 시작하지 못한 단계라고 할 수 있다.[49] 인공지능의 발전은 기술, 산업을 넘어 인간의 삶/생존의 문제와 직결되므로 사회적 논의기구를 통해 이른바 지능정보사회의 본격화에 대비한 중장기 연구 및 국가전략을 협의하여야 한다. 앞에서 라이언 칼로(Ryan Calo) 교수가 제안한 '연방로봇위원회'와 같은 독립적 정부 기구보다는 기술적 이슈외 법적·사회적 관점까지도 포괄적으로 논의할 수 있는 사회적 논의 기구가 지능정보사회를 대비한 한시적 컨트롤 타워의 역할을 수행할 수 있을 것이다.

3. 책임소재의 명확화

향후 인간의 통제권과 인공지능의 자율성에서 비롯될 수 있는 문제점에 대한 책임관계를 명확히 하는 것은 중요한 과제가 될 것이다. 이 경우 인간과 유사한 '인간유사로봇'에 대하여 상대적으로 기존의 규범체계 내에서 통제할 수 있다는 점에서 그 방향을 설정하는 것이 비교적 쉽다고 생각할 수 있다. 하지만 사실상 '인간유사로봇'과 약한 인공지능의 구분이 명확하지 않아서 로봇의 자율성에 따른 사고의 책임과 사고에 따른 입증책임의 문제도 규범적인 차원에서 논의될 수 있다. 즉 약한 인공지능은 민법 제98조에 의거하면 물건에 해당하기 때문에, 인간을 보조하는 수단으로서의 특성이 강하기 때문에 약한 인공지능 자체를 법적인 주체로 보는 것은 문제가 된다. 반면에 강한 인공지능은 상용화 단계에서는 이들의 지각적인 판단과 행위에 대해서 책임을 지울 수 있는 법적 체계 마련이 요구된다. 그렇지 않으면, 행위와 결과는 이에 대한 책임 주체가 상실된 상황이 발생하여 무질서한 상태가 지속될 수 있기 때문이다.[50]

전통적으로 인간의 내면적 동기의 일환이라고 할 수 있는 '고의' 또는 '과실'의 책임법리가 인간이 아닌 로봇에게 적용되는 것이 타당한지 논의하여야 하고, 또한 민사법적 차원에서 '제조물책임'의 인정범위는 어떻게 설정할 것인지 심층적으로 논의되어야 한다.[51] 책임소재의 명확성과 방향에 대한 이론적 논의에 기초하여 새로운 상황에 맞는

48) 로봇기본법안(2017. 7. 19 박영선 의원 대표발의, 의안번호 2008068)
49) 장민선, "인공지능(AI) 시대의 법적 쟁점에 관한 연구", 연구보고 18-10, 한국법제연구원, 2018, 32면.
50) 김나루, "인공지능으로 인한 법적 문제와 그 대안에 관한 연구", 홍익법학 제19권 제2호, 홍익대학교 법학연구소, 2018, 350면.

규범 정립이 필요하다. 또한 로봇 소유자의 권한과 책임에 대한 규제를 명확히 하기 위하여 인공지능 로봇 또는 스마트 로봇은 자동차처럼 등록제로 운영하여, 생성에서 폐기까지 국가차원의 관리와 규제책이 마련되어야 할 것이다.

Ⅵ. 결론

지능정보사회의 도래는 인류에게 다양한 편익을 줄 뿐 아니라 부작용과 도전의 과제들도 동시에 쏟아내고 있다. 지능정보기술의 발전으로 방대한 데이터를 스스로 학습하면서 최적의 판단을 내린다는 기본 전제에서부터 문제상황에 봉착한다. 학습 데이터 그 자체가 중립적이지 못하며 어느 정도의 흠결을 내재하고 있다던지 편향되거나 왜곡될 가능성을 배제할 수 없기 때문이다. 이와 같은 데이터로 학습한 결과는 지능정보기술이 잘못된 판단, 편향과 편견 및 차별이 담긴 판단을 할 수 있을 가능성을 높일 수도 있다. 이러한 편향, 왜곡, 차별의 문제는 민주주의에 대한 위협요인이 될 수 있다. 물론 제한된 정보를 넘어 다양한 정보의 제공원으로서 의견형성의 다양성에 기여하는 긍정적인 측면도 있지만 이러한 플랫폼의 지능정보기술이 이용자에게 편향된 정보를 제공하고, 이로써 확증편향에 빠질 수 있다는 우려 역시 적지 않다. 지능정보기술은 헌법상 기본권에 대한 기회이자 위기이기도 하다. 특히 데이터의 대량이용이라는 점에서 사생활과 인격권, 개인정보 자기결정권의 침해 상황이 더욱 빈번해질 수도 있고 인간의 결정과 판단을 보완하거나 대체한다는 점에서 전혀 새로운 양상으로 나아갈 수도 있다. 이러한 점에서 지능정보기술을 적용한 서비스나 제품으로부터 이용자를 어떻게 보호해야 할 것인지, 지능정보기술을 활용한 서비스가 규제 때문에 불가능해지거나 제한될 경우 이를 어떻게 정당화할 수 있는지, 지능정보사회를 어떻게 촉진할 것인지 등 개별 규제의 문제로부터, 국가의 근간인 헌법질서, 국가조직, 법치주의, 민주주의 등의 규범적 가치에 대한 영향, 평가, 전망에 이르기까지 거시적인 담론이 동시에 다루어져야 한다. 또한 지능정보사회의 발전에 기한 사회현상들에 대한 사실판단을 바탕[52]으로 하는 헌법적 가치 구현의 새로운 체계와 양태들은 그러한 변화를 수용하는 모습으로 나타나길 기대한다. ▢

51) 이원태, “인공지능의 규범이슈와 정책적 시사점”, KISDI Premium Report 15-07, 2015. 12, 19면.
52) 박기주, 앞의 논문, 63면.

공정성, 투명성, 책임성 제고를 위한 인공지능 법제 방향

오정미*

≪목 차≫

Ⅰ. 서론
Ⅱ. 외국의 입법 동향
Ⅲ. 우리나라 입법 현황 검토
Ⅳ. 법제화 방향 제언
Ⅴ. 마치며

Ⅰ. 서론

인공지능 챗봇, 인공지능 비서, 인공지능 카메라 등 요즘 '인공지능(AI)'이라는 단어가 낯설지 않다. 정부가 4차 산업혁명의 주력 사업으로 '데이터'와 '인공지능'을 꼽으며 인공지능 산업 발전을 위한 공격적인 투자를 아끼지 않고 있기 때문이다. 인공지능(AI)은 앞으로 우리의 생활을 더욱 편리하게 변화시킬 것이다. 그러나 최근 이른바 이루다 사태에서 보듯 인공지능 기술을 공정성, 투명성, 책임성 측면에서 살펴보면 반드시 그렇지만은 않다는 것을 쉽게 알 수 있다.

인공지능이 채용 면접을 진행하는 소위 'AI 면접'의 사례를 살펴보자. 정필모 의원실[1]에 따르면 SW마에스트로 연수생 합격자 150명의 면접 점수를 보면, AI가 A와 B+

등급으로 평가했지만, 면접관들은 하위 2등과 1등으로 평가했다. 반대로 면접관들이 1등과 2등으로 평가한 합격자는 AI로부터 B0와 B-로 평가받았다. AI 등급 평균과 면접관들의 평균을 비교하면, 어떤 유의미한 값도 나오지 않았다[2]. 또한 한국방송통신전파진흥원(KCA)의 경우 3년 동안 신입사원 면접에서 AI 면접을 보조수단이 아닌 당락을 결정하는 수단으로 사용해왔다. 정필모 의원실에 따르면 2020년 KCA에 315명 지원자 중 AI가 228명을 떨어트렸지만 KCA는 228명의 불합격자에 대해 AI가 어떤 알고리즘, 즉 어떤 기준을 적용해 불합격시켰는지 전혀 알지 못했다고 밝혔다.

위 AI 면접은 이른바 블랙박스와 같은 절차와 다르지 않았다. 그럼에도 불구하고 인사혁신처는 공공기관의 AI 면접에 대해 장밋빛 기대를 보인 바 있다. 인사혁신처는 사례집을 통해 '채용 프로세스를 개선하여 공평한 기회부여, 우수인재 선발, 비용 절감 등 투명성·공정성·효율성을 높여 전문성과 역량을 갖춘 우수인재를 채용하고 있다'며 인사혁신 사례로 선정하였다[3].

인공지능이 미칠 영향은 개인의 관점에서 뿐만 아니라 사회 전체적인 관점에서 고려되어야한다. 인공지능의 사용은 '지속가능한 발전 목표(Sustainable Development Goals)'를 달성하고, 민주적인 프로세스와 사회적인 권리를 지원하는데 중요한 역할을 수행할 수 있기 때문이다[4]. 이에 따라 순기능을 극대화하기 위하여 진흥 측면에서 혁신지원에 적합한 법제도 구축뿐 아니라, 역기능을 최소화하기 위한 규제 측면의 안전규제 및 위험관리 체계를 균형 있게 마련할 필요가 있다.

온라인으로 진행된 CES 2021에서 "기술에는 양심이 없다"라며 "기술이 세상에 봉사하게끔 만들어가는 일이 앞으로 우리 모두에게 주어지는 책임이 될 것"이라고 언급한 MicroSoft의 브래드 스미스(Brad Smith)[5]의 말처럼, 디지털 신기술의 사용 결과가 어떤 결과를 초래하는지 불확실한 만큼 공공의 안전도 함께 고려되어야 한다.

왜 불합격인지 설명하기 어려운 인공지능의 판단을 당연하게 받아들여야 하는 것인지, 챗봇 이루다 사태에서 보듯 데이터 수집 단계의 개인정보 침해 우려 뿐만 아니라 인간에 대한 혐오, 편견을 조장하는 발언들 역시 초창기 기술이라는 이유로 대책 없이 받아들여야만 하는 것인지 근본적인 질문이 필요한 시점이다.

인공지능의 지속가능한 발전을 위해서는 이용자의 목소리가 반영되는 법제가 필요하며, 그 출발은 공정성, 투명성, 책임성의 시각에서 시작되어야 한다고 본다. 이에 본 발제는 현재 해외의 입법동향 및 우리나라의 인공지능 법제 현황을 살펴보고, 인공지능 법제화 방향에 대한 여러 의견을 제안하는 것으로 마무리 할 예정이다.

---

* 서울대학교 공익법률센터 공익펠로우, 변호사

1) 정필모 의원실 "청년 앞길 막는 AI면접...AI 윤리기준 투명성·공정성 필요" 예산심사(2020. 11. 5.)

2) B+등급의 심층면접 평균이 31.6점인데, D등급은 이보다 높은 32.0점이 나왔다.

3) 인사혁신처, 「인사혁신사례집」" 한국자산관리공사: AI 면접을 통한 차별·편견·제한 없는 인재 채용", 2019.

4) European Commission,, WHITE PAPER - On Artificial Intelligence - A European approach to excellence and trust, 2020.

5) 마이크로소프트(MS)의 CES 2021 기조연설 중 인용

II. 외국의 입법 동향

Association for Computing Machinery(ACM)의 FAT 컨퍼런스는 컴퓨터 과학의 책임성·공정성·투명성 제고를 위한 국제회의다. 이 회의에서는 인공지능이 지켜야 할 원칙으로 공정성(Fairness), 책임성(Accountability), 투명성(Transparency)을 강조하고 있으며 이른바 'FAT 원칙'을 천명하고 있다. 공정성, 책임성, 투명성 원칙은 유럽연합 및 미국의 인공지능 법제에서 큰 틀을 잡고 있는 원칙임을 알 수 있다.

1. 유럽연합(EU)

외국의 경우 일반적인 인공지능 윤리 정립 및 규제와 관련하여 연구와 논의를 진행하고 있으며, 특히 EU가 법제화에 적극적이다. 먼저 유럽연합은 2019년 '신뢰할 수 있는 AI에 대한 가이드라인'에 기반하여 7개의 핵심적인 요구사항을 환영하는 공보(Communication)를 발간하였다. 그 내용은 ① 인간의 주체적 역량과 감독 ② 기술적 견고함 및 안전성[6] ③ 프라이버시 및 데이터 거버넌스 ④ 투명성 ⑤ 다양성, 비차별 및 공정성 ⑥ 사회적·환경적 복지 ⑦ 책임성이다.

이후 EU는 이것을 발전하여 2020년 3월 「인공지능 발전과 신뢰를 위한 백서」를 발표하였는데, 위험 발생 가능성이 높은 분야의 인공지능에 대하여 향후 안전성 요건을 수립하고 사전 적합성 평가를 받도록 하는 방안을 제시하였으며, 법안을 준비하고 있는 것으로 파악된다.[7]

또한 「온라인 플랫폼 시장의 공정성 및 투명성 강화를 위한 2019년 EU 이사회 규칙」[8]을 제정하고 2020년 7월 12일부터 시행하고 있다. 위 규칙은 ① 거래조건을 공정화하기 위한 약관 통제, ② 투명성 강화를 위한 정보공개, ③ 중소판매업체들에 대한

---

6) 안전성을 위한 요건으로서 학습 데이터 관리, 기록 보관, 이용자에게 인공지능 시스템에 관한 정보 제공, 견고성과 정확성 확보, 인간의 감독 개입, 생체인식기술에 대한 보호장치 등을 제시하였음
- 인공지능이 학습하는 데이터에 관하여 EU의 가치나 규정들(안전성 확보, 차별 금지, 사생활 보호 등)을 준수하도록 고려되어야 함
- 알고리즘과 데이터에 관한 기록을 보존하고, 경우에 따라 데이터 그 자체를 보존해야 함
- 인공지능 시스템의 능력과 한계를 고지해야 하고, 인공지능과 상호작용하는 사람에게 상대방이 인공지능이라는 사실을 고지해야 함
- 인공지능 시스템은 견고성(robustness)과 정확성(accuracy)을 갖추어야 함
- 인공지능 시스템은 인간의 감독을 받아야 하며, 감독의 수준은 시스템의 목적과 영향 등을 고려해야 함
- 생체인식기술은 정당하고 비례적이며 적절한 보호장치를 갖춘 경우에만 사용될 수 있음

7) European Commission,, WHITE PAPER - On Artificial Intelligence - A European approach to excellence and trust, 2020.

8) 「Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services」

실효성 있는 피해구제 수단의 확보 등 3개의 주요 과제를 담고 있다[9].

2. 미국

미국의 경우 연방 차원에서는 알고리즘 규제에 관한 법안이 발의된 바 있으며, 주와 지방 차원에서 인공지능 알고리즘 규제 법안이 시행된 사례가 있다. 미국 상원에서 2019년 4월 인공지능 기술과 알고리즘 규제를 위한 「알고리즘 책임 법안(Algorithmic Accountability Act)」이 발의되어 절차가 진행 중이다. 위 법안은 미국 연방거래위원회(FTC)가 고위험(highly sensitive) 자동화 시스템을 평가하는 규칙을 만들고, 기업들이 이 규칙에 따라 알고리즘이 편향적·차별적인지, 프라이버시나 보안 위험이 있는지 여부를 점검하도록 하고 있다.[10]

또한 주(州)와 지방 차원에서 알고리즘의 오남용을 규제하는 내용의 법안들이 통과되었는데, 뉴욕시 의회는 2017년 뉴욕시가 알고리즘 사용에 있어 편향성이 있는지 점검하는 기구를 설립하는 내용의 「알고리즘 책임 법안(algorithmic accountability bill)」을 통과시켰다. 미국 샌프란시스코 시는 2019년 5월 법집행시 수사당국이나 행정당국이 안면인식기술을 사용하지 않도록 하는 조례를 통과시켰으며, 메사추세츠 주의 서머빌(Somerville) 시(市)도 2019년 6월 유사한 내용의 조례를 통과시켰다.[11]

한편, 2020년 4월 미국 연방거래위원회(Federal Trade Commission, 이하 FTC)에서 발표한 'Using Artificial Intelligence and Algorithms[12]'지침은 기업이 AI 및 알고리즘을 사용할 때 ① 투명성 ② 설명 가능성 ③ 공정성 ④ 견고성과 실증적 타당성 ⑤ 책임성을 갖추도록 하였으며, 준수해야 할 내용이 매우 구체적인 점이 눈에 띈다.

> ①'AI 및 알고리즘 활용의 투명성 제고(Be transparent)'는 AI 기반 서비스 제공시 소비자를 기만하지 않고, 민감한 정보 수집 시 투명하게 고지하며, 자동화된 의사 결정으로 발생한 불리한 조치에 대해서는 통지하고 정보 접근 권한과 부정확한 정보를 수정할 권리를 부여하는 것을 말한다.
> ②'AI 및 알고리즘을 활용한 의사결정에 대한 설명(Explain your decision to the consumer)'은 소비자에게 불리한 결정을 내릴 때 그 이유를 구체적으로 설명할 수 있어야 하고, 그 주요한 요인을 공개해야 하며, AI에 의해 거래조건이 변경될 때도 알려야

9) ①'거래조건 공정화' 방안으로는 판매업체의 상품 공급 제한·유보·중단 및 약관 변경시 사전고지, 이용자의 계약해지권 명시 등이 포함됨. ②'투명성 강화' 방안으로는 검색결과 노출순위를 결정하는 알고리즘 주요 매개변수의 공개, 특정 이용자에 대한 차별적 대우나 최고우대고객조항 사용에 대한 설명의무 부과 등이 있음. ③'피해구제의 실효성 확보' 방안으로는 조정절차의 지원, 단체소송제의 도입 등이 있음

10) 국회입법조사처, "주요국의 인공지능 법제화 현황", 2020.

11) 국회입법조사처, "주요국의 인공지능 법제화 현황", 2020.

12) 미국 연방거래위원회는 AI 및 알고리즘의 사용과 관련하여 소비자에 끼칠 부정적인 영향을 방지하는 방법에 대한 새로운 비즈니스 지침을 발표하였다 (2020. 4. 8.)

한다는 것을 말한다.
③‘결과의 공정성에 대한 보장(Ensure that your decisions are fair)’은 결과의 공정성을 보장하면서 특정 집단이나 계층에 대한 차별적 결과가 나오지 않도록 하고, 의사결정에 사용되는 정보에 대한 접근 권한과 수정 기회를 소비자에게 제공하는 것을 말한다.
④‘데이터·모델의 견고성 및 실증적 타당성 보장(Ensure that your data and models are robust and empirically sound)’은 정보의 정확성과 최신성을 유지하면서, AI 모형이 설계 의도에 맞게 작동하고 불법적 차별을 일으키지 않도록 검사하고 재확인해야 한다는 것을 말한다.
⑤‘법령 준수, 윤리, 공정성 및 비차별성에 대한 책임 견지(Hold yourself accountable for compliance, ethics, fairness, and nondiscrimination)’는 AI나 알고리즘 사용 전부터 자가 점검을 하고, 악용 및 무단 이용 가능성과 대비책에 대해 검토하며, 개발한 AI에 대해 책임을 다해야 한다는 것을 말한다.

Ⅲ. 우리나라 입법 현황 검토

1. 국회 계류 중인 법률안[13)]

현재 국회에 인공지능과 관련한 법률안 중 정책추진 및 거버넌스 정립과 관련한 주요 법률안은 아래와 같다. 공통적으로 인공지능 산업의 기반이나 기술 개발을 진흥하는 것을 주요 목표로 하고 있다 보니 산업 육성에 치우쳐 인공지능이 추구해야 할 원칙들을 담보할 내용에 대한 고민이 부족하다. 또한 인공지능과 상호작용하는 국민에 대한 구체적인 내용은 찾아보기 어렵다. 양향자 의원안에서 포괄적 인권보호 의무를 규율하고 있는 것과 이상민 의원안에서 윤리원칙 제정과 관련한 내용만이 그 예외라 하겠다.

거버넌스 구조 역시 산업 진흥에 관한 내용을 주로 규율하고 있어 인공지능산업 육성과 관련된 정부부처와 민간위원만이 참여하는 위원회로 구성되어 있다. 양향자 의원안과 민형배 의원안에서 제한적으로 인권과 관련한 의제도 다루고 있으나 개인정보 보호, 소비자 보호, 인권 보호를 위한 개인정보보호위원회, 국가인권위원회, 공정거래위원회의 역할은 언급되어 있지 않다.

가. 인공지능 연구개발 및 산업 진흥, 윤리적 책임 등에 관한 법률안(제정안) – 이상민 의원 대표발의

▲인공지능 기술개발 및 산업진흥을 위한 기본계획·시행계획 수립 ▲전문인력 양성,

13) 입법 형식에 있어 인공지능에 대한 별도의 법률 제정안과 기존 법률 개정안이 모두 발의되어 있으나, 본 토론에서는 인공지능 제정안을 중심으로 살펴본다.

표준화 지원, 시범사업 추진 ▲인공지능 기술 기반 집적시설 구축 지원 ▲인공지능산업 협회 설립

제3조(국가 및 지방자치단체 등의 책무) ① 국가 및 지방자치단체는 인공지능의 기술개발 및 산업의 진흥을 위하여 필요한 각종 시책을 수립·시행하여야 한다.

② 국가 및 지방자치단체, 인공지능사업자 등은 인공지능 산업에서 **이용자보호를 위한 인공지능 윤리원칙을 제정하고 인간의 기본적 인권과 존엄성이 보호**되도록 하여야 한다.

③ 국가 및 지방자치단체, 인공지능사업자 등은 **인공지능의 개발·제조·생산·유통·활용 등 모든 단계에서 차별과 편향이 발생하지 않도록** 하며 일자리 감소 등 역기능에 대한 대비를 할 수 있도록 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.

② 정책심의위원회는 위원장 및 부위원장 각 1명을 포함한 15명 이내의 위원으로 구성한다.

③ 위원장은 위원 중에서 호선하고, 부위원장은 위원회의 동의를 얻어 위원장이 지명한다.

④ 정책심의위원회의 위원은 인공지능에 관하여 전문성과 경험이 풍부하고 덕망이 있는 사람 중에서 과학기술정보통신부장관이 위촉한다.

⑤ 위원의 임기는 3년으로 하되, 연임할 수 있다.

⑥ 그 밖에 정책위원회의 구성, 운영 등에 필요한 사항은 대통령령으로 정한다.

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.

1. 인공지능 기술개발 및 산업 진흥을 위한 시책의 기본방향
2. 인공지능 산업의 기반조성에 관한 사항
3. 인공지능사업의 창업지원 등 인공지능사업자 육성에 관한 사항
4. 인공지능 전문인력의 양성에 관한 사항
5. 인공지능 산업의 국제협력과 해외진출의 지원에 관한 사항
6. 인공지능의 이용 확산 및 유통 활성화에 관한 사항
7. 인공지능 기술개발 및 산업 진흥을 위한 법·제도개선에 관한 사항
8. 인공지능 기술개발 및 산업 진흥에 필요한 투자계획 및 재원 확보에 관한 사항
9. 인공지능 기술개발 및 보급에 필요한 기반 시설 구축에 관한 사항
10. **인공지능산업에서의 인권보호 및 차별과 편향을 예방하기 위한 윤리강령에 관한 사항**

나. 인공지능산업 육성에 관한 법률안(제정안) – 양향자 의원 대표발의

▲인공지능산업 기반 조성 및 육성 ▲인권보호 의무 ▲국무총리 소속 인공지능산업 육성위원회 설치 ▲정부의 관련 기업에 대한 기술 및 장비 지원, 세금 감면 등 인센티브 ▲인공지능특화단지 지정 및 지원 방안 등

제3조(인공지능산업에서의 인권보호) ① 국가, 지방자치단체 및 인공지능사업을 영위하는 자는 인공지능산업을 육성함에 있어 **인간의 존엄성이 보호**되도록 하여야 한다. ② 누구든지 **인공지능기술의 개발, 생산, 유통, 활용 등 모든 단계에서 차별과 편향이 발생하거나 인권이 침해되지 아니하도록** 하여야 한다. 제7조(인공지능산업육성위원회) ① 정부는 인공지능산업의 육성에 관한 다음 각 호의 사항을 심의하기 위하여 국무총리 소속으로 인공지능산업육성위원회(이하 “위원회”라 한다)를 둔다. 1. 기본계획 및 시행계획의 수립·변경에 관한 사항 2. 인공지능산업 관련 정책의 총괄·조정 3. 인공지능산업 관련 정책의 평가·자문 4. 그 밖에 위원장이 인공지능산업의 육성을 위하여 필요하다고 인정하는 사항 ② 위원회는 위원장 1명을 포함한 15명 이내의 위원으로 구성하되, 위원장은 국무총리가 되며, 위원은 다음 각 호의 사람이 된다. 1. 기획재정부, 교육부, 과학기술정보통신부, 법무부, 산업통상자원부 등 대통령령으로 정하는 관계 중앙행정기관의 장 2. 인공지능산업에 관한 전문지식과 경험이 풍부한 사람으로서 국무총리가 위촉하는 사람 ③ 위원회에 간사위원 1명을 두며, 간사위원은 과학기술정보통신부장관이 된다. ④ 제1항부터 제3항까지에서 규정한 사항 외에 위원회의 구성 및 운영 등에 필요한 사항은 대통령령으로 정한다.

다. 인공지능 기술 기본법안(제정안) – 민형배 의원 대표발의

▲인공지능 기술개발 및 산업진흥에 필요한 기본적인 사항으로서 국가인공지능기술위원회, 지방인공지능기술위원회를 설치 ▲‘국가인공지능기본계획’과 ‘지방인공지능종합계획’을 수립해, 인공지능 기술을 육성하기 위한 국제협력, 민간 참여의 활성화 ▲인공지능 관련 단체의 설립 ▲인공지능 기술에 대한 재정지원 등에 관한 사항

제6조(국가인공지능기술위원회의 설치) 인공지능 기술에 관한 중요 사항을 심의·의결하기 위하여 국무총리 소속으로 국가인공지능기술위원회를 둔다.

제7조(국가인공지능기술위원회의 구성) ① 국가인공지능기술위원회는 위원장 2명을 포함한 30명 이상 50명 이내의 위원으로 구성한다.
② 국가인공지능기술위원회의 위원장은 국무총리와 제3항제3호가목 또는 나목에 해당하는 위원 중에서 대통령이 임명하는 사람으로 한다.
③ 국가인공지능기술위원회의 위원은 다음 각 호에 해당하는 사람으로 한다. 이 경우 공무원이 아닌 위원이 전체위원의 과반수가 되어야 한다.
1. 기획재정부장관·교육부장관·과학기술정보통신부장관·외교부장관·국방부장관·산업통상자원부장관·환경부장관·국토교통부장관·해양수산부장관·중소벤처기업부장관·국무조정실장 및 그 밖에 대통령령으로 정하는 공무원
2. 「공공기관의 운영에 관한 법률」에 따른 공공기관으로서 대통령령으로 정하는 공공기관의 장
3. 다음 각 목의 사람 중에서 대통령이 위촉하는 사람
가. 대학이나 공인된 연구기관에서 인공지능 기술 분야의 부교수 이상 또는 이에 상당하는 직에 10년 이상 재직한 사람
나. 인공지능 기술 관련 단체나 기관에서 10년 이상 종사한 사람
다. 법관, 검사 또는 변호사로 10년 이상 재직한 사람
라. 그 밖에 사회적 신망이 높고 인공지능 기술 분야에 학식과 경험이 풍부한 사람

2. 정부의 '인공지능 법·제도·규제 정비 로드맵'

2020년 12월 24일 발표된 정부의 '인공지능 법·제도·규제 정비 로드맵'(이하 '로드맵')의 특징은 기업의 자율성을 존중하는 시장친화적 법·제도를 마련하겠다는 것이다. 기업 자율의 평가·관리·감독체계를 우선 유도하고, 이후 알고리즘의 편향성 등을 평가·검증할 수 있는 체계를 마련한다는 내용이다. 또한 플랫폼 사업자 공정성 강화를 위해서는 영업비밀을 보장하면서도 알고리즘의 인위적 조작방지와 공정한 운영을 지원할 가이드라인 마련 및 필요시 법률제정을 하겠다는 계획을 발표하였다.

그러나 국민의 기본권이나 기업의 제조물 책임 및 소비자 보호에 관한 법적 규제방안이 누락되어 있다. 또한 기업의 자율성도 중요하지만 온전히 기업 자율에 맡기는 것보다는 유럽연합의 사례처럼 '위험 기반 접근 방식(risk-based approach)'에 따라 위험성에 비례한 규제 개입 및 법적 의무 부담이 필요하다고 본다.

거버넌스 측면에서도 국무조정실이 언급되어 있긴 하지만 '관계부처 합동'이라는 표현이 무색하게 과학기술정통부 단일부처 계획에 가깝다. 개인정보보호위원회(개인정보 관련)와 공정거래위원회(소비자 관련), 그리고 국가인권위원회(차별, 혐오, 편견, 정보인권 관련)등 과의 폭넓은 논의가 요구되나 계획에서 전혀 언급이 없다. 또한 사회적 합의를 통한 상생 포용 기반 법제도, 소비자, 노동자 등 시민사회의 참여도 부족하며, 산업주도와 정부주도적 개발이라는 목표가 뚜렷해 아쉬움이 남는다.

다양한 분야에서 진행되는 인공지능 기술개발과 산업발전을 견인하고 사회변화에 대

비하기 위한 종합적 정책추진 및 거버넌스 정립이 필요하다. 인공지능 기술력 확보를 위한 범정부 차원의 정책 추진이 필요하고, 기술이 산업·사회 전반에 미치는 파급력이 크므로 전체를 조율할 수 있는 컨트롤 타워가 필요하기 때문이다. 신기술의 등장으로 새롭게 추구되는 가치와 기존의 법제도가 보호하고자 하는 가치가 충돌할 수 있어 사회적 논의를 통해 이를 조정하는 역할이 필수적이다. 또한 이러한 거버넌스 체계를 정립하여 국회, 정부·공공기관, 학계, 업계, 시민단체 등이 함께 논의할 수 있는 대화기구가 필요하다.

3. 소결

유럽연합 및 미국은 매우 구체적인 가이드라인을 내놓았다. 단순 제재를 하는 것이 아니라 공정성, 투명성, 책임성이라는 큰 틀 안에 법적 규범 마련을 위한 다양한 방안이 들어있다. 반면, 우리나라는 윤리기준 및 로드맵을 발표하였으나 추상적이고 현실적으로 필요한 내용이 부재하기 때문에 향후 인공지능 법제 마련시에는 유럽연합, 미국의 입법례를 참고할 필요가 있다. 특히 고위험 인공지능이나 공공영역에 있어서는 일정한 강제력이 있는 입법이 필요할 것으로 생각한다. 다음 장에는 공정성, 책임성, 투명성이라는 원칙 아래 구체적으로 법제화 방향을 제언하고자 한다.

IV. 법제화 방향 제언

1. 총론 – 법률 간의 관계 및 정의 규정

가. 지능정보화기본법과의 관계

양향자 의원 안에 대한 과학기술정보방송통신위원회 검토보고서에서도 드러났듯 인공지능기술 관련 제정안들에서 정의되고 있는 인공지능기술은 기존 지능정보화기본법 상 지능정보기술에 대부분 포함되기 때문에 기본계획이나 위원회 구성 등에서 지능정보화기본법과의 중복 규정 여부를 따져보아야 한다.

제정안 제6조는 과학기술정보통신부장관으로 하여금 인공지능 기술개발 및 산업 진흥을 위하여 중장기적인 기본계획을 수립하고, 기본계획에 따라 연도별 시행계획을 수립·시행하도록 하고 있음. 다만, 제정안의 기본계획에 포함되어 있는 사항 중 다수 내용이 「지능정보화 기본법」 제6조(종합계획) 및 「정보통신 진흥 및 융합」와의 중복 규정이 될 가능성이 있음

사견으로는 지능정보화기본법을 일반법으로 하고 인공지능과 관련한 특별법을 별도로 제정하되, 산업진흥 측면에 천착하기 보다는 이미 제안한 바와 같이 기술 이용에 따른 소비자보호, 인권보장, 개인정보보호 등 영역을 포괄하여 공정성, 투명성, 책임성을 뒷받침하는 내용을 구체적으로 다루는 법안이 되어야 한다고 본다.

나. 인공지능 기술 정의의 확대

이를 위해 인공지능 기술의 정의도 중요한데, EU 인공지능백서에서 인공지능을 “데이터, 알고리즘, 컴퓨팅 연산능력 등을 결합한 기술의 집합”이라고 표현한 것에 주목할 필요가 있다. 현재 발의된 대부분의 제정안과 같이 인공지능을 학습, 추론 등 기능의 구현 방법에 국한하여 인공지능을 좁게 정의하게 되면 앞서 제안한 데이터나 알고리즘과 관련한 규율을 제정안에 포섭하기 어려워진다.

양향자 의원안 제2조(정의) 1. “인공지능”이란 인간의 지능이 가지는 학습, 추론, 지각, 자연언어 이해 등의 기능을 전자적 방법으로 구현하는 소프트웨어나 컴퓨터시스템, 그 밖의 장치를 말한다

따라서 아래 미국의 알고리즘 책임 법안(Algorithmic Accountability Act)의 정의조항과같이 인공지능 기술이 포함되는 “자동화 결정 시스템(automated decision system)”과 같은 보다 폭넓게 정의 규정을 마련하는 것이 타당해 보인다. 이는 지능정보화기본법 상 지능정보기술과도 구분되는 개념으로 이하에서 제안할 여러 법제 내용을 규율하기에 더욱 적합할 것으로 판단된다.

SEC. 2. DEFINITIONS (1) AUTOMATED DECISION SYSTEM.—The term “automated decision system” means a computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making, that impacts consumers.

다. 이용자 정의 추가

현재 발의된 제정안에는 모두 이용자(end user) 개념이 모두 누락되어 있다. 이는 수범자를 주로 인공지능 기술을 개발하는 기업 등 산업계에 한정하여 제정안을 구성하면서 발생한 결과라 할 수 있다. 현재 정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 ‘정보통신망법’)에도 이용자가 정의되어 있고 미국의 알고리즘 책임 법안(Algorithmic Accountability Act)에도 소비자(consumer)가 정의조항에 포함되어 있다.

*정보통신망법
제2조(정의)
4. “이용자”란 정보통신서비스 제공자가 제공하는 정보통신서비스를 이용하는 자를 말한다.

*Algorithmic Accountability Act
SEC.2. DEFINITION
(4) CONSUMER.—The term “consumer” means an individual.

2. 구체적 제도 도입 제안

최근 이슈가 된 이루다 사태의 경우 개인정보 보호법의 준수만으로도 훈련데이터 수집 이용단계의 문제 발생을 예방할 수 있었을 것이다. 그러나 제조물책임법이나 소비자보호법 등 현행 법령과 제도만으로 인공지능 기술 활용에 따른 문제점을 모두 해결할 수 없다는 점에는 이견이 없을 듯 하다. 이하에서는 기존 법령 외에 인공지능 법제에 추가적으로 도입되어야 할 내용을 제안하려고 한다.

가. 책임성 강화를 위한 고위험 인공지능 분류체계 및 영향평가 도입

(1) 고위험 인공지능 분류체계 도입

모든 AI 기술을 동일하게 규율하기 보다는 위험성에 기반한 분류와 해당 분류에 따른 차등적 규제 방법의 도입이 요구된다. 이러한 위험성 분류기준에 따라 고위험 인공지능 영역으로 분류된 영역에 대해서는 다른 영역에 비해 더 강화된 책임을 부여하는 방안을 고려할 필요가 있다. 이하에서는 고위험 인공지능 영역과 일반 영역을 구분하여 수범의 범위를 나누고 고위험 인공지능 영역에 우선적으로 적용되어야 하는 법제 내용을 제안하려고 한다.

고위험 인공지능 분류체계는 유럽연합의 사례를 구체적으로 살펴볼 필요가 있다. 유럽연합은 소비자를 보호하고, 불공정한 상거래 관행에 대처하고, 개인의 데이터와 프라이버시를 보호하기 위한 엄격한 법적 프레임워크를 수립하고 있다. 그리고 원칙적으로 AI에 대한 새로운 규제 프레임워크는 특히 중소기업들에게 필요 이상의 부담을 주지 않도록 지나치게 규범적이지 않으면서 그 목적을 달성하는데 효과적이어야 한다는 목표 아래 ‘위험 기반의 접근방법’을 따르도록 한 점을 주목해야 한다.

AI 중 다음과 같은 두 가지 기준이 충족되는 경우“고위험(high-risk)”으로 간주되며 법률제정의 필요성을 구성한다.

첫째, 중대한 위험이 발생할 것으로 예상되는 분야에서 인공지능 제품 및 서비스가 사용되는 경우로서, 이에 해당하는 분야는 법에 의해 구체적이고 명확하게 나열되어야 한다. (예: 의료, 운송, 에너지, 경찰 및 사법 시스템 등 일부 공공 분야)

둘째, 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 말한다. 이와 함께, 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 간주된다.

유럽연합은 '고위험 인공지능' 시스템으로 분류된 제품 및 서비스에 대하여 법적 규제를 추진 중이며, 법적 의무로는 △훈련 데이터 △기록과 데이터의 보존 △정보 제공 △견고성 및 정확성 △인적 감독 △원격 생체인식 등 특정 애플리케이션 특별 요건 등을 제시하고 있다.

미국의 「알고리즘 책임 법안(Algorithmic Accountability Act)」 역시 연간 5천만 달러 이상의 매출이 있거나 1백만 명 이상의 사람 또는 장치에 대한 정보를 보유한 회사, 데이터 브로커 회사 등에 적용되도록 하고 있다. 해당 법안은 고위험 인공지능 영역에 대해 미국 연방거래위원회(FTC)가 고위험(highly sensitive) 자동화 시스템을 평가하는 규칙을 만들고, 기업들이 이 규칙에 따라 알고리즘이 편향적·차별적인지, 프라이버시나 보안 위험이 있는지 여부를 점검하도록 하고 있다.

따라서 국내 인공지능 법제도에도 위험 기반의 차등적 규제를 도입하고, 고위험 인공지능으로 분류된 제품 및 서비스에 대해서는 이하 규제를 기본적으로 적용하고, 고위험으로 분류되지 않는 경우 EU 백서와 유사하게 자발적 표시와 같은 자율규제 방식을 적용하는 것을 제안한다.

(2) 인공지능 영향평가 도입

(가) 인공지능 인권영향평가

인공지능이 기본권을 침해하는 영향을 완화하기 위하여 국제기구 및 세계 여러나라 정부가 다양한 방식의 인공지능 영향평가를 도입하였거나 도입을 검토 중이며, 인권영향평가는 가장 효과적인 영향평가 방법으로서 권장되고 있다. 최근 유엔 등 국제인권기구들은 인공지능에 대한 인권영향평가를 요구 및 검토하고 있는데 구체적인 내용은 아래와 같다.

유엔 의사표현의 자유 특별보고관은 2018년 인공지능이 인권에 미칠 영향에 대한 보고서에서 인공지능 시스템의 도입에 있어 준수할 절차들을 제안하였고, 인권에 기반한 인공지능 기술을 위하여 인권영향평가 또는 공공기관 알고리즘 영향평가의 실시를 각국 정부에 권고한 바 있다.[14] 한국 정부 역시 이를 수용하고 이행할 필요가 있다.

인공지능 시스템의 활용 전후 및 활용 과정 중에 인권 영향 평가를 할 것(53. Human rights impact assessments)
인권 단체들과 함께 외부에서 감사하고 협의할 것(55. Audits)
개인이 선택할 수 있도록 고지와 동의 절차를 갖출 것(58. Individual autonomy; 59. Notice and consent)
인권침해를 종식시키기 위한 효과적인 구제 절차를 마련할 것(60. Remedy)

62. 인공지능 시스템이나 응용 프로그램을 구하거나 사용할 때, 국가는 공공 부문 기관들이 지속적으로 인권의 원칙을 보장하도록 해야 한다. 그 중에서도 인공지능 시스템의 조달 및 사용 이전에 공공의 협의를 수행하고 인권영향평가 또는 공공기관 알고리즘 영향평가의 착수를 포함한다. 특히 인종 및 종교적 소수자, 정치적 반대 그룹이나 활동가에게 이런 기술이 미칠 수 있는 불평등한 영향에 더 신경을 써야 한다. 인공지능 시스템을 정부에서 사용하는 경우 외부의 독립적인 전문가로부터 정기적인 감사를 받아야만 한다.
63. 국가는 인공지능 시스템의 민간부문에서의 설계, 보급 및 실행에 있어서 인권이 중심에 올 수 있도록 해야만 한다. 이는 인공지능 영역에 대해 현 규제, 특히 개인정보보호 규제를 갱신하고 적용하는 것을 포함하며, 기업에 영향평가와 인공지능 기술에 대한 감사를 실시할 것을 요구하고 효과적인 외부 책임 메커니즘을 보장하도록 설계된 규제 혹은 공동 규제 체제의 추진을 포함한다. (하략)

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 전문가들은 인공지능에 대한 ‘핵심적인 보호조치’로서 인권영향평가 실시를 지지하였다. 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 강조한 바 있다.[15]

(나) 공공부문 인공지능 영향평가

앞서 언급한 포괄적 인공지능 영향평가에 더해 공공부문에만 적용되는 공공부문 인공지능 영향평가를 도입할 것을 제안한다.

캐나다 정부는 2019년 자동화된 의사결정 지침(훈령)을 발표하여 공공기관 인공지능 요건을 법규화하면서 알고리즘 영향평가의 사전 실시 및 공개를 의무화하였다. 시스템 생산 전 영향평가를 완료하고 기능 또는 범위 변경 시 평가를 갱신하도록 하고 있다.

14) “Report of the Special Rapporteur on Promotion and protection of the right to freedom of opinion and expression: Note by the Secretary-General”, UN문서 A/73/348 (2018. 8. 29).
15) https://www.ohchr.org/Documents/Issues/DigitalAge/ExpertSeminarReport-Right-Privacy.pdf

영향평가 결과에 따라 전문가 자문, 고지, 인적 개입, 설명 요건, 검사, 모니터링, 교육 훈련, 비상계획, 시스템 승인 등 부대 의무를 차등화하고 있다.

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.

참고: 캐나다 정부 자동화된 의사결정 지침(훈령)

나. AI 사용 결과의 공정성 보장을 위한 훈련데이터 조치의무 등

아래 의무는 고위험 평가되는 기술 및 시스템에는 필수적으로 적용되어야 할 것이며, 그 외 일반적인 기술 및 시스템까지 수범 대상을 확대해야 할지 여부는 추가적인 논의가 요구된다.

(1) 훈련데이터 등 사전 조치 의무

최근 AI 기술은 사용된 훈련데이터가 개발되는 결과물에 직접적인 영향을 미치는 방향으로 발전하고 있다. 이는 최근 이루다 사태에서 입력 데이터의 편향성이 결과물인 챗봇의 발화(speech)의 편향성에 영향을 주었을 가능성이 매우 크다는 점에서도 잘 드러난다. 따라서 AI 사용 결과의 공정성 보장을 위해서는 우선 인공지능 법제에 훈련데이터 조치에 대한 구체적인 기준이 마련되어야 한다.

(가) 훈련데이터 등의 차별적 요소 통제

AI의 훈련데이터를 수집하는 시점부터 특정 집단이나 계층에 차별이 생기지 않도록 차별적 요소를 통제할 필요가 있다. 이는 훈련데이터 수집과 이용 단계에서 그치는 것이 아니라 결과물 도출 단계까지 적용되어야 하는 바, 실제 해당 기술을 공개하기 전에 엄격한 테스트를 통해 차별적 요소를 철저하게 통제하도록 의무화할 필요가 있다.

미국 FTC ‘AI와 알고리즘 사용에 대한 지침’ 역시 AI를 이용할 때 특정 집단이나 계층에 차별이 발생하지 않도록 알고리즘 사용 전과 후에 항상 엄격하게 테스트하여

알고리즘을 통한 의사결정이 특정 집단에 별다른 영향을 미치지 않도록 관리하여야 한다고 규정하고 있다.

(나) 이용자(end user)의 정보수정 기회 제공

현재는 AI를 개발하는 주체나 운용하는 주체가 AI를 통한 의사결정에 제공되는 정보를 선택하는 재량이 폭넓게 주어져 있는 반면, 판단의 대상이 되는 이용자가 이의를 제기할 수 있는 절차가 마련되어 있지 않다. 이용자가 특정 의사결정 과정에 AI의 적용을 거부할 수 있는 권리에 더하여 자신과 관련한 의사결정에 사용되는 정보에 대해 이의를 제기하고 이와 다른 정보를 직접 제공할 수 있는 절차를 법률로 보장할 필요가 있다.

이 과정에서 제공되는 정보는 개인정보가 대부분일 것이므로 개인정보보호법의 열람, 정정, 삭제권을 준용하는 것도 방법일 것으로 보인다. 결과적으로 이 같은 절차를 통해 이용자는 취업, 신용, 보험 등 자신과 관련한 여러 의사결정이 내려질 때 사용되는 데이터와 관련하여 해당 정보에 접근할 권리를 가지며, 정보가 부정확하다고 여겨질 경우 해당 정보에 대해 이의를 제기할 수 있게 될 것이다.[16)]

(2) 결과의 공정성 유지 의무

훈련데이터 등 조치 의무에 더하여 AI 기술 이용으로 인해 차별적인 결과가 도출되지 않도록 할 필요가 있다. 특히 이용자에게 직접적인 영향을 미치는 차별적 처우(treatment) 영역에서는 결과의 공정성 유지 의무가 더욱 강하게 적용될 필요가 있으며, 차별적 발화(speech)의 경우 표현의 자유의 대원칙 하에 혐오표현 금지법 도입 등 '표현의 자유'의 한계와 함께 고민되어야 한다.

(가) 차별적 처우(treatment)[17)]

EU처럼 차별금지법이 제정되어 있지 않은 상황에서 차별 등을 막기 위해서는 각 영역마다 공정성 개념을 확립해야 하며, 이에 부합하는 기술이 개발되어야 한다. 예를 들어 얼굴인식 음성인식 등 인공지능 기술이 인종, 성별, 출신에 따라 성능차이를 줄이는 기술이 개발되어야 하며 채용, 신용평가, 의료서비스에도 성능의 차별성이나 편향성을 줄여야 한다[18)].

특히 채용면접, 신용평가 등 AI의 판단으로 인하여 이용자에게 구체적인 재산적, 신

16) FTC, 'AI와 알고리즘 사용에 대한 지침(Using Artificial Intelligence and Algorithms)'참고.
17) 박상철, "이루다 사건으로 본 인공지능 거버넌스: AI의 일탈을 어떻게 막을 것인가?"발표에서 '차별적 처우'와 '차별적 발화'에 대한 구분론을 참고함 (2021. 2. 4).
18) 이현규, "인공지능 기술의 신뢰 확보", 과학기술정보통신부, 발표문

분적 이해관계에 영향을 미치는 경우에는 결과의 공정성 의무를 강하게 부여해야 할 것이다. 인공지능 의사결정에서 작용하는 편견으로 차별적 처우를 받을 경우 통제 메커니즘 없이 훨씬 더 많은 사람들에게 장기간 영향을 줄 수 있기 때문이다.

이 같은 의무는 공공부문과 민간부문을 나누어 규율될 필요가 있는데, 특히 공공부문에서 차별적 처우가 발생할 수 있는 영역에는 별도로 규율하는 방안이 필요하다. 공공부문에 관한 규율은 앞서 언급한 공공부문 인공지능 영향평가에 구체적으로 포함되도록 하는 것이 타당하다.

(나) 차별적 발화(speech)

이루다 사태에서 챗봇의 소수자에 대한 편향적인 발화(speech)가 논란의 중심에 있었다. 이루다의 발화가 현행 법령 하에서 자연인에게 금지되어있는 발화인지 여부가 문제되었는바, 현재 포괄적 차별금지법이나 혐오표현을 금지하는 법령이 부재한 상황에서 이 같은 발화를 사전에 금지할 강제적 수단 역시 부재하다.

따라서 AI 결과물이 발화의 형태로 구체화하는 경우 표현의 자유라는 기본권 틀 안에서 포괄적 차별금지법 등 특정 발화를 금지하는 법률의 제정에 대한 사회적 합의를 이끌어 낸 후, 해당 법령에 기초하여 AI의 결과물을 규제하는 순서로 공론화가 진행될 필요가 있다.

다. 책임성 강화를 위한 사람의 감독 의무

고위험 인공지능 시스템의 경우 적절한 사람의 감독을 받아야만 인간의 자율성을 손상되거나 기타 부정적인 결과가 야기되는 것을 방지할 수 있다. 고위험 인공지능과 공공부문의 처분이나 차별적 처우에 적용되어야 한다. EU 인공지능백서에서 사례로 든 바와 같이 무인자동차의 센서 가시성 즉 신뢰성이 낮은 상황에서는 무인 운행을 중단하고 사람이 직접 운행을 해야 하는 경우(아래 표에서 4단계)가 그러하다. EU 백서의 사례를 참고로 위험성에 기반하여 단계별 감독 의무 규정을 도입할 것을 제안한다.

1단계	사람이 검토하고 확인을 해야만 효력이 발생
2단계	사전 감독 의무는 없으나 사후 검토가 필수적
3단계	사람이 실시간으로 개입하여 운영을 정지시킬 수 있음
4단계	설계단계에서 특정 조건 하에서는 시스템 운영 상의 제약을 가함

라. 투명성 보장을 위한 거부권, 통지 및 설명 의무 등

GDPR의 경우 프로파일링 등 법적 효과를 낳는 오로지 자동화된 처리 시 개인정보처리자가 정보주체에게 보장하고 있는 권리로 △구체적인 통지전달, △인적개입을 획득할 수 있는 권리, △자신의 의사를 표현할 권리, △이러한 평가 이후 도달한 결정에 대한 설명을 획득할 권리, △해당 결정에 이의를 제기할 권리 등, △적절한 안전조치를 보장 △아동은 제외해야 함을 명시(전문 71)하고 있다. 이하에서는 이와 유사하게 AI 기술의 적용과 관련하여 이용자에게 보장되어야 할 권리 등을 도출하여 서술하겠다.

(1) AI 이용 여부에 대한 사전 통지 의무

AI 기술을 이용한 자동의사결정의 대상이 되는 경우 이를 이용자에게 알리도록 하는 통지 의무가 도입되어야 한다. EU에서도 인공지능과 상호작용하는 사람에게 그 상대방이 인공지능이라는 점을 분명히 고지하도록 하고 있다[19].

(2) AI 의사결정 대상이 되지 않을 권리 부여

사전 통지 의무에 더해 GDPR 제22조와 유사하게 정보주체가 자동화된 의사결정에 대하여 거부할 수 있는 권리를 도입하여야 한다. 구체적으로는 정보주체가 자신에 관하여 법적 효력을 주거나 이와 유사한 중대한 효과를 미치는 자동화된 처리(프로파일링 포함)에만 근거한 결정에 따르지 않을 권리가 규정되어야 한다.

> EU GDPR 제22조 프로파일링 등 자동화된 개별 의사결정
> 1. 정보주체는 프로파일링 등, 본인에 관한 법적 효력을 초래하거나 이와 유사하게 본인에게 중대한 영향을 미치는 자동화된 처리에만 의존하는 결정의 적용을 받지 않을 권리를 가진다.
> 2. 결정이 다음 각 호에 해당하는 경우에는 제1항이 적용되지 않는다.
> (a) 정보주체와 컨트롤러 간의 계약을 체결 또는 이행하는 데 필요한 경우
> (b) 컨트롤러에 적용되며, 정보주체의 권리와 자유 및 정당한 이익을 보호하기 위한 적절한조치를 규정하는 유럽연합 또는 회원국 법률이 허용하는 경우
> (c) 정보주체의 명백한 동의에 근거하는 경우
> 3. 제2항 (a)호 및 (c)호의 사례의 경우, 컨트롤러는 정보주체의 권리와 자유 및 정당한 이익, 최소한 컨트롤러의 인적 개입을 확보하고 본인의 관점을 피력하며 결정에 대해 이의를제기할 수 있는 권리를 보호하는 데 적절한 조치를 시행해야 한다.

19) 인공지능 시스템의 능력과 한계를 고지해야 하고, 인공지능과 상호작용하는 사람에게 상대방이 인공지능이라는 사실을 고지해야 한다.

4. 제2항의 결정은 제9조(2)의 (a)호와 (g)호가 적용되고, 정보주체의 권리와 자유 및 정당한 이익을 보호하는 적절한 조치가 갖추어진 경우가 아니라면 제9조(1)의 특별 범주의 개인정보를 근거로 해서는 안 된다.

(3) 주요 요인 및 결과에 대해 설명할 의무

AI 기술을 이용한 자동의사결정의 대상이 되는 경우 해당 의사결정의 주요 요인을 사전에 공개할 의무를 부여하고, GDPR과 같이 의사결정의 주요 요인에 대한 설명을 요구할 권리도 함께 규율되어야 한다.

주요 요인 공개	기업이 알고리즘을 사용하여 소비자에게 신용 점수를 매기는 경우 점수에 영향을 미치는 주요 요인을 공개할 것 - 기업이 신용 점수를 근거로 소비자에게 신용 제공을 거부하거나 덜 유리한 조건으로 신용을 제공하는 경우, 신용 점수에 대한 설명과 함께 점수에 부정적인 영향을 끼친 상위 주요 요소를 공개하여야 함

참고: 미국 FTC의 「AI와 알고리즘 사용에 대한 지침」

이와 더불어 의사결정의 결과에 대한 설명할 의무와 이용자가 설명을 요구할 권리가 도입되어야 한다. 미국 FTC ‘AI와 알고리즘 사용에 대한 지침’은 불리한 조치에 한정하여 통지와 설명 의무를 규율하고 있으나[20], 통지와 설명 의무의 대상을 불리한 조치에 한정할 필요는 없어 보인다.

(4) 공공기관의 통지 및 설명 의무

한편, 공공기관이 AI를 의사결정에 사용하는 경우 이에 대한 통지 및 설명의무를 보다 민간부문보다 엄격하게 제정할 필요가 있다. 관련하여 2019년 캐나다 정부는 자동화된 의사결정 지침(훈령)을 발표하여 공공기관 인공지능 요건을 법규화[21]하였는바 구체적인 내용을 참고할만하다.

20) FTC, 「AI와 알고리즘 사용에 대한 지침(Using Artificial Intelligence and Algorithms) 」: 기업이 개인 신용평가사 등 제3자로부터 공급받은 정보를 기반으로 자동결정(automated decision)을 내리는 경우 소비자에게 ‘불리한 조치’에 대해 통지할 것. 소비자는 이를 통하여 자신에 대해 리포팅된 정보를 확인 후 부정확한 정보를 수정할 권리를 행사할 수 있음.

21) 정보인권연구소,“공공기관 인공지능 규범”, 2020. (원문 출처: The Government of Canada. Directive on Automated Decision-Making)

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.

(5) 문서화 및 기록보존 의무

AI 기술 이용의 결과에 대한 사법적 분쟁 등 사후적 구제방법까지 염두에 둔다면 AI 기술에 사용된 데이터에 관한 구체적인 기록을 보존하도록 하거나 경우에 따라서는 데이터 그 자체를 보존하도록 할 필요가 있다.

이는 GDPR 제18조의 '처리제한권(Right to restriction of processing)'과 같은 맥락으로서 정보주체는 자신에 관한 개인정보의 처리를 차단하거나 제한할 권리를 갖는 것이다. 정보주체의 처리제한권은 개인정보의 정확성, 처리의 합법성 등에 대하여 다툼이 있거나 소송 수행 등을 위하여 보존의 필요성이 있는 경우에 이용을 제한하되 삭제를 보류할 수 있도록 요구할 수 있는 권리이다.

이에 따르면 정보의 처리가 불법적으로 이루어지고, 정보주체가 개인정보의 삭제에 반대하면서 대신에 그 개인정보의 이용 제한을 요청한 경우, 더 이상 개인정보가 필요하지 않지만, 정보주체가 법적 청구권의 입증, 행사나 방어를 위하여 그 정보를 요구한 경우 처리를 제한하여야 한다.[22)]

또한 GDPR에서 처리자의 '책임성' 의무 중 하나로 문서화가 포함되어 있는 바, 문서화 및 기록보존 의무는 투명성을 물론 AI 기술 이용 전반에 관한 책임성을 제고하는 데에도 기여할 것으로 보인다.

EU 인공지능 백서 역시 알고리즘과 데이터에 관한 기록을 보존하고, 경우에 따라 데이터 그 자체를 보존해야 한다고 규정하고 있고, 아래에서 보듯 EU 인공지능 공공조달 백서와 캐다나의 지침에서도 문서화 요건을 엄격하게 규율하고 있음을 확인할 수 있다.

EU <인공지능 공공조달 백서> '5단계 실사 절차' 중

2. 공급자 예비 심사 : 설계 절차의 최초 단계서부터 다음과 같은 인공지능 관련 데이터 윤리 요건을 고려하고 정의하고 구현해야 함

- 기술적 안전성은 문서화되어 설명가능성, 공정 커뮤니케이션 및 감사를 보장해야 함

22) 방송통신위원회(2020),「EU GDPR 가이드북」

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

캐나다 <자동화된 의사결정 지침> 알고리즘 영향 수준별 적용 요건 중

수준II이상의 경우 : 시스템의 설계 및 기능에 대한 문서화 의무

마. 책임성 강화를 위한 공공영역 조달 규정, 중소기업에 대한 지원 등

(1) 공공영역 인공지능 조달지침 마련

조달영역의 경우 공공영역의 AI의 사용과 관련하여 민간부문보다 더욱 강화한 조치 의무가 요구된다. 유럽연합도 '인공지능 기반 서비스 및 솔루션 공공조달의 데이터 윤리 백서'에서 신뢰가능한 인공지능은 책임성, 기술적 안전성, 지속가능성에 대한 요구뿐 아니라 데이터 윤리 요소를 포함한 공공조달 체계를 수립하고 이를 현행 법적 의무에 적용함으로써 달성될 수 있다고 지적한 바 있다.[23)]

이러한 측면에서 영국 정부의 인공지능 조달지침을 참고할 만하다. 영국 정부의 인공지능 조달지침은 조달 절차 개시 단계에서 인공지능 영향평가를 수행하고, 조달 절차별로 평가 결과가 반영되는지 반복적으로 평가하도록 하고 있으며 주요 의사결정에서는 위험성 완화 계획을 참고하도록 하고 있다.[24)]

- ▸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

참고 : 영국 정부 인공지능 조달지침 영향평가 항목

(2) 중소기업 및 신생 벤처기업에 대한 지원

마이크로소프트사 같은 글로벌 기업이나 대기업 등은 개인정보 보호 및 인공지능 규

23) 정보인권연구소, "공공기관 인공지능 규범", 2020.
24) Office for Artificial Intelligence (2020). "Guidelines for AI procurement."

범을 준수하기 위한 시스템이 마련되어 있어 자율적 운영이 가능한 반면, 중소기업 및 신생 벤처기업의 경우 이러한 시스템 부재로 인한 자율적 준수가 어려운 경우가 많다. 이들은 인공지능을 이용한 자신들의 프로세스를 수정하거나 인공지능에 대한 높은 수준의 전문성을 갖추기 어렵기 때문이다.

유럽연합은 중소기업의 인공지능 활용을 지원하기 위해 '디지털 혁신 허브(Digital Innovation Hubs)'와 '주문형 인공지능 플랫폼'을 강화하고 있다. 우리나라 역시 기업의 자율성에 모든 것을 맡기는 것보다는 중소기업 등에는 전문성을 고양하고, 인공지능 품질을 개선할 수 있도록 지원하는 창구를 만들어 인공지능에 대한 책임성을 강화할 수 있도록 하는 규정이 필요하다.

(3) 기타: 자율규제 도입의 전제

앞서 고위험 인공지능 시스템이 아닌 경우에는 자율규제 방식의 표시제도 등을 도입하는 것이 타당하다는 의견을 피력한 바 있다. 다만 자율규제의 전제로서 유엔 의사표현의 자유 특별보고관의 2018년 보고서를 참고할 필요가 있다. 그는 아래와 같이 인공지능 윤리는 기업과 공공기관이 법적구속력이 있고 강제력이 있는 인권기반의 규제를 우회하기 위한 포장이 아니라고 지적한 바 있다.[25)]

3. 거버넌스 구조 제안 – 소비자보호, 인권보장, 개인정보보호 측면 보완과 다양한 이해당사자의 참여 보장

가. 반드시 포함되어야 할 정책 분야와 국가기관

현재 인공지능과 관련하여 계류 중인 법률안의 경우 공통적으로 산업진흥을 주된 입법 목적으로 하고 있고, 인권보장이나 개인정보보호 등 인공지능과 상호교류하는 주체로서 국민에 대한 고려가 부족하다. 앞서 살펴본 바와 같이 산업진흥을 위해 정부 부처 역시 과학기술정보통신부가 중심이 되며 국가인권위원회, 개인정보보호위원회, 공정거래

25) "46. (중략)The private sector's focus on and the public sector's push for ethics often imply resistance to human rights-based regulation. While ethics provide a critical framework for working through particular challenges in the field of artificial intelligence, it is not a replacement for human rights, to which every State is bound by law. Companies and governments should ensure that human rights considerations and responsibilities are firmly integrated into all aspects of their artificial intelligence operations even as they are developing ethical codes and guidance." "Report of the Special Rapporteur on Promotion and protection of the right to freedom of opinion and expression: Note by the Secretary-General", UN문서 A/73/348 (2018. 8. 29).

위원회는 아예 그 대상에서 빠져있는 상황이다.

아래 표를 통해 인공지능 정책 관련 거버넌스 구조에서 공정성, 투명성, 책임성을 제고하기 위해 추가적으로 반드시 포함되어야 하는 영역과 국가기관의 사례를 정리하였다.

영역	역할	국가기관
소비자보호	훈련데이터 사전, 사후 규제 AI 사용 관련 소비자보호 결과의 공정성 담보	공정거래위원회
인권보장	차별, 편견, 혐오 방지 인권영향평가	국가인권위원회
개인정보보호	훈련데이터 등에 포함된 개인정보 보호	개인정보보호위원회

(1) 공정거래위원회

공정거래위원회에 해당하는 미국 FTC가 'AI와 알고리즘 사용에 대한 지침'을 제정하고, '알고리즘책임 법안(Algorithmic Accountability Act)'에서 고위험 자동화 시스템을 평가하는 기준을 만드는 등 소비자 보호 측면에서 AI 정책에 깊숙이 개입하고 있음은 이미 앞서 살펴본 바와 같다. 공정거래위원회도 소비자 보호 측면에서 적극적으로 인공지능 문제에 목소리를 낼 필요가 있다.

(2) 국가인권위원회

독립적이고 전문성을 갖춘 국가인권기구는 인권의 원칙에 기반한 인공지능 개발, 사용 등에 있어 중요한 역할을 할 수 있다. 제4차 산업혁명이라는 미명 아래 상업적 분야 등에서 인공지능 기술 등 신기술이 무비판적으로 도입되고 있으며, 개인정보의 상업적 활용이 무분별하게 허용되고 있다. 그러나 정보주체들의 권리를 보장하기 위한 입법적, 행정적 기반은 전무한 상황이다.

국가인권위원회는 위와 같은 공적 보호의 부재로 인해 발생할 수 있는 인권침해 상황을 예방하고 점검하여 권고를 내리는 등 국가인권기구로서의 역할을 해야 한다. 그리고 차별금지법이 없는 우리나라의 경우 차별, 편견, 혐오를 조장하는 인공지능에 대한 관리, 감독이 필요하다. 이루다 사태에서 보듯이 장애인, 성소수자, 유색인종 등 사회적 소수자에 대한 혐오와 편견에 대한 감독을 위해서는 국가인권위원회의 역할이 더욱 중요하다.

(3) 개인정보보호위원회

이루다 사태의 경우 훈련데이터 수집과정에서 개인정보보호법 위반 여부가 가장 큰 쟁점이 되어 개인정보보호위원회가 자연스럽게 AI 정책에 참여하게 되었다. 사후적구제 측면에서 보면 현행 개인정보보호법 상 개인정보보호위원회의 역할로 충분하다고 볼 수도 있으나, AI 기술의 개발과정에서 개인정보의 수집과 이용 제공 등이 차지하는 역할이 큰 만큼 AI 정책을 논의하는 거버넌스 기구에 반드시 개인정보보호위원회에 중요한 역할이 부여되어야 한다.

나. 다양한 이해당사자의 참여와 공론장 필요성

앞서 살펴본 인공지능관련 법률 제정안의 거버넌스 구조는 산업계와 학계 등 관련 전문가들의 참여만 제한적으로 허용되었을 뿐, 소비자단체, 정보인권단체 등 시민 사회의 참여는 거의 고려하고 있지 않다. 거버넌스 구조에 소비자보호, 인권보장, 개인정보보호 의제가 포함되어야 하므로, 해당 영역의 시민사회단체의 참여나 일반 시민의 의사를 반영하는 공론장의 설계가 요구된다. 다양한 이해당사자 참여에 대한 아래 입법조사처의 의견도 참고할 만 하다.

> 외국에서는 인공지능 기술의 파급력에 대한 사회적 논의와 연구를 지속하면서 기술개발·산업진흥과 역기능 대응의 양 측면 모두에서 입법을 시도하고 있으며, 윤리적 기준 마련도 추진하고 있음
> ◦ 주요국들은 인공지능 관련 정책 및 입법에 관한 논의를 일회성으로 끝내지 않고 이해관계자들과 광범위하게 소통하면서 지속적으로 추진하고 있음
> - EU 집행위원회는 인공지능 윤리 가이드라인 마련에 있어 독립적인 전문가 그룹을 구성하였으며, 추후 이해관계자들과 지속적으로 소통할 것을 계획하고 있음
> ◦ 자율주행자동차 운행 기반 마련 등 신기술 수용을 위하여 법제도를 신속하게 정비하면서, 알고리즘 공정성·책임성을 강화하고 인공지능 오남용을 방지하는 입법도 추진하고 있음[26)]

EU 인공지능 백서 역시 거버넌스 기구의 중요성을 강조하고 있고 이해당사자들의 참여를 최대한 보장해야 한다고 언급하고 있다.

> 프레임워크의 구현 및 추가적인 발전에 대해서 이해관계자들(소비자 단체 및 사회적 파트너, 기업, 연구자, 시민사회 단체)와 협의해야 한다.

26) 인공지능 관련 입법 현황 및 전망, 입법조사처, 현안분석 제87호.

(4) 거버넌스 기구 형태 제안

현재 국회에 발의된 제정안의 경우 관련 정책을 심의하는 위원회를 과학기술정보통신부 장관소속(양향자 의원안) 또는 국무총리 소속(민형배 의원안, 이상민 의원안)으로 두도록 하고 있다.

범 부처간 논의가 필요한 성격이므로 국무총리 소속으로 하되, 과학기술정보통신부 외에 이미 제안한 바와 같이 공정거래위원회, 국가인권위원회, 개인정보보호위원회가 당연직 위원으로 포함되는 형태를 제안하고자 한다. 4차산업혁명위원회에 유사한 성격의 거버넌스 기구를 두는 방안도 제안된 상태이나 4차산업혁명위원회의 법률적 근거가 미약하기 때문에 제정안을 통해 별도의 거버넌스 기구를 설립하는 방법이 더 타당해보인다.

호주 국가인권위원회는 2019년 〈인권과 기술〉 에서 호주 정부에 대한 30개 제안 및 9개 질의를 발표하며 인공지능 규제의 법제화를 제안한 바 있는데,[27] 특히 인공지능 안전위원회라는 독립적 법정기구를 설립하는 방안에 주목할 하다.

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

27) Sophie Farthing et al, Human Rights and Technology Discussion Paper, Australian Human Rights Commission, 2019. https://tech.humanrights.gov.au/consultation

Ⅴ. 마치며

인공지능 기술의 발전을 통해 사회의 혁신을 도모하는 것은 장려할만한 일이다. 현재 앞 다투어 발의되고 있는 인공지능 관련 제정안에서는 이러한 '수월성의 생태계' 구축이 중요하다는 점에 방점을 두고 있기도 하다. 그러나 EU 인공지능 백서에서 이와 같은 수월성의 생태계를 달성하기 위해서라도 고유한 '신뢰의 생태계'를 창출할 미래 규제 프레임워크가 중요하다고 강조하고 있다는 점에 다시 주목하지 않을 수 없다.

인사혁신처는 최근 '4차산업혁명 대비 미래형 채용방식 연구' 연구용역을 통해 공정성, 투명성, 책임성에 기반한 신뢰의 생태계가 중요하다는 점을 강조한 것으로 보인다. 이는 최초 별다른 고민 없이 정부혁신 사례로 AI 기술을 이용한 면접을 선정한 것에서 한걸음 더 나아간 것이다. 해당 보고서 요약본에는 인공지능을 채용 과정에 도입하기 위해서는 "자동화된 처리도구(인공지능)의 사용과 활용에 대한 가능성을 확보할 수 있도록 법적 근거가 마련돼야 한다"며 "인공지능 개발·도입·적용과 데이터 생성·저장·활용·공유 등의 채용관리 전반을 반영할 수 있는 '공무원 임용 및 인사관리에 관한 법률' 제정이 필요하다"는 내용이 포함되어 있다.[28)]

이 같은 맥락에서 인공지능 정책과 이용자를 포함한 이해관계 조정을 위한 추진체계와 거버넌스 조직의 법적 근거를 마련하고 공정성, 투명성, 책임성에 기반한 신뢰의 생태계 구축을 위해 제정안에 포함될 규율 내용들을 제안하였다. 이제는 기존의 산업 진흥이나 수월성만을 강조하는 반쪽짜리 법제를 넘어 데이터 수집에서 AI 이용 결과의 공정성까지 AI 기술의 라이프 사이클 전반을 포괄하면서 이용자 보호와 위험성에 기반한 차등적 규제를 담은 입법의 혁신이 요구되는 시점이다. ▢

28) 한겨레, 공무원 채용도 AI로?…"데이터 활용 등 법적 근거 마련부터", 2021. 1. 22. http://www.hani.co.kr/arti/politics/administration/979906.html#csidxabd43d84b7ddb419d5d7e4144b15b17

인공지능과 기술윤리

김병필 교수 |
KAIST 기술경영학부, 민주사회를위한변호사모임 디지털정보위원회 회원

발표자님들의 깊이 있는 고민이 담긴 좋은 발표와 훌륭한 정책 제안에 감사드립니다. 저는 우선 인공지능 공정성의 중요성을 다른 각도에서 강조하고자 합니다. 먼저 우리 삶에 있어 중요한 순간들을 되짚어 보고 싶습니다. 우리가 태어나서 죽을 때까지, 가장 중요한 순간이라면 학교에 가고, 취업을 하고, 승진을 하고, 퇴직하고, 대출을 받고, 병원에 가고, (운이 나쁘면) 경찰이나 법원에 가는 경우 등이 있겠습니다. 이러한 순간들이 우리 삶의 궤적을 결정 짓는 "중요한 결정"을 이룰 것입니다.

이제 이러한 모든 과정에서 인공지능이 사용되기 시작했습니다. 우리 삶은 그래프로 그린다면 오르락 내리락하는 곡선 형태가 될 것입니다. 즉, 삶의 변곡점을 만들어 내는 중요한 결정들에 있어 인공지능이 사용되는 것입니다. 개별 인공지능이 특정 집단의, 특정 출신의 개인에게 약간씩 불리하게 이루어진다면, 그 결과 그 개인의 삶에 누적된 효과는 매우 커질 수 있습니다. 그래서 인공지능의 활용 사례에 있어 공정성을 보장하는 것이 매우 중요해 집니다.

인공지능의 활용 범위가 굉장히 넓은 만큼, 저는 인공지능의 활용 분야를 크게 3가지 유형으로 나누어 보고 싶습니다. 첫째로, 원래 인간은 잘하지 못하고, 컴퓨터는 잘하는 영역이 있습니다. 수학 문제를 푸는 것이 대표적입니다. 인공지능이 자연과학 연구에의 활용되는 경우도 마찬가지입니다. 얼마 전 구글의 Alpha Fold 2가 딥러닝을 이용해 뛰어난 정확도로 단백질의 구조를 예측하는 모델을 개발했다는 소식이 전해졌습니다. 원칙적으로 이러한 분야는 더욱 활성화되고 장려될 필요가 있겠습니다.

다음으로 인간은 잘하는데 컴퓨터는 잘하지 못하던 영역이 있습니다. 대표적인 것이 컴퓨터 비전, 언어 이해와 같은 분야입니다. 인간은 누구라도 매우 쉽게 할 수 있는 일인데, 컴퓨터로 처리하기 매우 어려웠습니다. 최근 10년간의 딥러닝의 성장은 이 영역에서의 발전이라고 해도 과언이 아닙니다. 이러한 인공지능의 1차적 목표는 인간만큼 잘하게 되는 것입니다. 인간보다 더 잘하게 되는 것은 다음 목표입니다. 이러한 분야에서는 일단 정확도를 높이는 것이 중요합니다. 예컨대, 자율주행차라면 주변 사물을 잘 인식하고 안전하게 가는 것이 무엇보다 중요하겠습니다. 이러한 분야의 인공지능에서는 정확도를 높임으로써 그 안전성에 대한 사회적 우려가 해소될 수 있는 경우도 적지 않을 것입니다.

마지막으로 남은 부분이 가장 유의해야 할 영역입니다. 즉, 인간도 잘못하고, 컴퓨터도 잘못하는 영역입니다. 오늘 주요하게 논의된 채용 심사를 생각해 보면, 인간도 사람을 뽑는 일을 잘못합니다. 저도 회사 다닐 때 직원 선발 과정에 참여했고, 대학에 와서는 학생 선발에 참여하는데 매번 어려운 일이라고 느낍니다. 인공지능에게 맡기면 인간보다 잘할까하는 의문이 들 수밖에 없습니다. 그 이유는 '잘한다'는 것이 무엇인지를 정의하는 것부터가 어려운 문제이기 때문입니다. 예컨대 채용에 있어 공정성이 무엇을 의미하는지는 쉽사리 합의에 이르기 어려운, 철학적이고 윤리적인 복잡한 문제입니다.

최근의 시도는 이처럼 인간도 잘하지 못하던 문제를 해결하기 위해, 인공지능을 이용해서 더 나은 결정을 내려보자는 접근으로 정리할 수 있다고 생각합니다. 굳이 인공지능이 아니더라도 그 이전부터 통계적 방법, 과학적 방법, 증거에 기반한 의사결정에 의해 인간 의사결정을 개선하려는 시도가 이어져 왔습니다. 최근 10년 간의 인공지능의 발전은 이러한 기대를 한 단계 더 높였다고 생각합니다. 앞서 이야기한 우리 삶의 중요한 변곡점들, 우리 삶의 궤적을 만들어 내는 결정들에 인공지능을 활용하는 시도는, 인간도 정답을 모르지만, 인공지능을 이용하면 아무튼 더 잘할 수 있겠지라는 사고가 전제되어 있습니다. 하지만, 인공지능의 공정성 문제는 결국 기존 인간 판단의 어려움과 그 한계와 마찬가지의 문제를 제기합니다. 인공지능을 활용했다고 해서 이러한 문제가 그저 사라지는 것은 아닙니다.

몇 년 전만 하더라도 인공지능 개발자들은 "세상이 공정해야 인공지능이 공정하지"라고 이야기했습니다. 연구자들은 인공지능에게 뭐라고 하지 말고, 공정한 사회를 만들어다 주면, 우리가 공정한 인공지능을 만들어 주겠다는 태도였습니다. 일리가 있는 말이지만, 이제는 더이상 그대로 수용할 수는 없는 주장이 되어 버렸습니다. 이제 "인공지능이 공정해져야 세상이 공정하게 된다"는 마인드로 변화되어야 하고, 실제로도 그렇게 변화하고 있습니다.

한 가지 긍정적인 점은 인공지능 덕분에 우리가 인간 의사결정의 한계와 문제에 대해 다시 되짚어 보고, 이를 인공지능을 이용해서 평가하고 극복하려는 시도가 진행 중이라는 것입니다. 오정미 발표자님께서 말씀해 주신 ACM FAccT 컨퍼런스의 경우 'AI 감사'가 중요한 연구 주제입니다. 즉, 'AI 감사'를 통해 데이터를 통해 기존 인간 결정자의 판단이 소수자에게 얼마나 차별적이었는지, 혹은 자의적이거나 부정확한 판단이 내려지지는 않았는지 검토해 볼 수 있게 되는 것입니다. 이러한 측면에서 인공지능을 활용해서 기존의 인간 결정의 한계를 더 잘 이해하고, 더 공정한 사회를 만들어 낼 방법을 찾아내는 것이 중요한 과제라고 생각합니다.

이러한 점에서 인공지능 개발 기업들이 흔히 제기하는 인공지능 '규제'가 이제 막 싹트기 시작한 인공지능 산업을 위축시킬 것이라는 주장은 재고될 필요가 있다고 생각합니다. 인공지능 산업 발전에 있어 가장 중요한 점은 인공지능에 대한 사회적 신뢰를 형성하는 것입니다. 적절한 규제는 사회적 신뢰를 구축하는 기초가 될 수 있습니다. 비록 현재의 인공지능 기술 수준에 비추어서는 인공지능 공정성이나 설명 가능성을 달성하기가 쉽지 않다는 어려움이 있는 것은 사실입니다 하지만, 환경 분야 사례에서 보는 것처럼 규제가 오히려 관련 기술 발전을 추동할 수도 있습니다. 앞으로 인공지능에 있어서도 적절한 공정성, 투명성, 설명가능성에 관한 적절한 수준의 규제가 마련됨으로써 이러한 신뢰가능한 인공지능 기술 분야에 대한 관심과 투자가 증가하고 그 결과 인공지능에 대한 사회적 신뢰가 높아지는 선순환의 고리가 만들어지기를 기대합니다. ▢

인공지능 법제 정비 해외 사례

장여경 이사 | 사단법인 정보인권연구소

○ 발제자들의 취지에 대체로 공감함

- 인공지능과 같은 새로운 과학기술의 발전으로 인한 위험에 대비하고 국민의 기본권과 안전을 지키기 위해 국가적 역할과 규범이 수립되어야 하며, 이를 위해서는 헌법적 고찰과 사회적 논의가 수반되어야 함

- 국민으로부터'신뢰받는 생태계'를 구축하기 위하여 인공지능 정책과 이용자를 함께 포함하는 거버넌스 체계를 마련하고 공정성, 투명성, 책임성에 기반한 법률적 규율을 도입할 필요가 있음

- 국민의 인권과 안전을 보호하는 인공지능 기술 혁신과 생태계 구축을 위하여 인공지능 규율 관련 해외 법제도 소개를 통해 보완 토론을 하고자 함

○ 정부는 산업중심적 인공지능 정책을 성찰하고 최근 유엔 등 국제기구가 각국 정부에 인공지능과 관련하여 인권규범 준수를 권고하는 바를 이행할 필요가 있음

> 사회권의 실현에 있어 신기술의 역할에 대한 유엔 사무총장 보고서(2020. 3. 4. 유엔문서번호 A/HRC/43/29. 62문.)
>
> (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 한다.
>
> (b) 국가가 민간 부문 활동에 관한 조치를 포함하여 입법 조치를 취해야 할 의무를 재확인하고 준수함으로써, 신기술은 경제·사회·문화적 권리를 포함한 모든 사람들의 인권에 대한 완전한 향유에 기여하고 인권에 미치는 부작용이 방지되어야 한다.

(c) 국가 간 및 국가 내적으로 정보 격차 및 기술 격차를 해소하기 위한 노력을 가속화하고, 신기술의 접근성, 가용성, 경제성, 적응성 및 품질을 개선하기 위한 포괄적인 접근 방식을 촉진해야 한다.

(d) 기술 변화 등에 의해 야기되는 변화와 불안정성으로부터 탄력성을 구축할 수 있는 사회적 보호의 권리에 투자하고, 모든 고용 형태의 노동권을 보호해야 한다.

(e) 공공부문에서 신기술, 특히 인공지능의 이용에 관한 정보를 대중에게 전파하기 위한 노력을 대폭 증진해야 한다.

(f) 신기술의 개발 및 도입에 관한 의사결정에 모든 관련 이해당사자의 참여를 보장하고, 특히 공공부문에서 인공지능이 지원하는 의사결정에 대하여 적절한 설명가능성이 보장될 필요가 있다.

(g) 인권의 향유에 중대한 영향을 미칠 수 있는 신기술 시스템, 특히 인공지능 시스템의 전체 생애주기 동안 체계적으로 인권 실사를 실시해야 한다.

(h) 신기술이 사용되는 상황에서 완전한 책임을 보장하는 적절한 법률 체계와 구조를 창출해야 하며, 이는 국내 법제도의 공백을 검토 및 평가하고, 필요한 경우 감독 체제를 수립하고, 신기술로 인한 피해에 대해 접근 가능한 구제 수단을 마련하는 것이 포함된다.

(i) 신기술의 개발 및 사용, 특히 경제·사회·문화적 권리의 향유에 필수적인 제품 및 서비스에 대한 접근에 있어서 차별과 편견을 해소해야 한다.

(j) 정례인권검토(UPR)와 인권조약기구 하에서 이루어지는 보고 및 검토에 있어 신기술이 경제·사회·문화적 권리에 미치는 영향에 특히 주의를 기울여야 한다.

○ 공공과 민간의 모든 인공지능 제품과 서비스는 헌법을 비롯하여 국민의 기본권과 관련한 현행 법률들 – 소비자 안전과 보호에 관련한 법률들, 개인정보 보호법, 평등과 차별금지에 관한 법률들을 준수할 의무가 있으며, 포괄적 차별금지법 등 보호가 미진한 부야에 대해서는 관련 법제도를 신속히 마련해야 함

○ 나아가 국민의 인권과 안전을 보호하기 위해 인공지능 제품과 서비스에 특별한 규율을 도입하였거나 도입을 검토하고 있는 해외 법제도를 참고하여 관련 법제도를 정비할 필요가 있음

○ 캐나다 정부

– 2019년 캐나다 정부는 〈자동화된 의사결정 지침〉(훈령)을 발표하여 공공기관 인공지능 요건을 법규화함[1]. 이 훈령은 공공기관이 의사결정에 사용하는 인공지능 알고리즘에 대하여 영향평가를 실시하고 위험성 수준별로 4단계로 나누어 의무를 차등 적용함

1) The Government of Canada. Directive on Automated Decision-Making

【캐나다 정부】 〈자동화된 의사결정 지침〉 알고리즘 영향평가에 따른 수준[2)]

수준	세부사항
I	의사결정이 다음 사항에 거의 영향을 미치지 않을 것으로 보이는 경우 – 개인 또는 공동체의 권리, – 개인 또는 공동체의 건강 또는 복리, – 개인, 단체 또는 공동체의 경제적 이익 – 생태계의 지속가능성 통상 수준 I의 결정은 복구되고 일시적인 영향을 미칠 수 있음
II	의사결정이 다음 사항에 중간 정도의 영향을 미칠 것으로 보이는 경우 – 개인 또는 공동체의 권리, – 개인 또는 공동체의 건강 또는 복리, – 개인, 단체 또는 공동체의 경제적 이익 – 생태계의 지속가능성 통상 수준 II의 결정은 복구시킬 수 있고 단기적인 영향을 미칠 수 있음
III	의사결정이 다음 사항에 큰 영향을 미칠 것으로 보이는 경우 – 개인 또는 공동체의 권리, – 개인 또는 공동체의 건강 또는 복리, – 개인, 단체 또는 공동체의 경제적 이익 – 생태계의 지속가능성 통상 수준 III의 결정은 복구되기 어려울 수 있고 지속적인 영향을 미칠 수 있음
IV	의사결정이 다음 사항에 매우 큰 영향을 미칠 것으로 보이는 경우 – 개인 또는 공동체의 권리, – 개인 또는 공동체의 건강 또는 복리, – 개인, 단체 또는 공동체의 경제적 이익 – 생태계의 지속가능성 통상 수준 IV의 결정은 복구가 불가능하고 영구적인 영향을 미칠 수 있음

– 캐나다 훈령은 알고리즘 영향별 수준에 따라 전문가 검토, 공지, 인적 개입, 설명, 검사, 모니터링, 교육훈련 비상 계획, 시스템 구동 승인 의무 등 훈령의 요건을 차등 적용함

2) (Appendix B) Impact Assessment Levels

【캐나다 정부】 〈자동화된 의사결정 지침〉 알고리즘 영향 수준별 적용 요건[3)]

요건	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
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 - 훈련데이터에 대한 설명, 또는 이 데이터를 공개적으로 사용할 수 있는 경우 익명화된 훈련데이터에 대한 링크	

3) (Appendix C) Impact Level Requirements

주요 의사결정에 대한 인적 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 요건	해당 법정 요건에 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 요건에 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, 훈련 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

○ 뉴질랜드 정부

- 뉴질랜드 정부는 2020년 7월 〈아오테아로아 뉴질랜드 알고리즘 헌장〉을 발표하고 공공기관들이 서약하도록 함[4].

【뉴질랜드 정부】 아오테아로아 뉴질랜드 알고리즘 헌장

투명성

의사결정이 알고리즘에 의해 어떻게 영향을 받았는지 명확하게 설명함으로써 투명성을 유지한다. 이는 다음을 포함한다.

▸알고리즘을 평이한 영어로 문서화

▸데이터와 처리과정에 관한 정보를 접근가능하게 만들기(법적인 제한이 있지 않는 이상)

▸데이터가 어떻게 수집되고, 저장되고, 보호받는지에 관한 정보 공개

파트너쉽

다음과 같은 조약의 약정을 통해 명확한 공익을 제공한다.

▸와이탕이 조약의 원칙에 부합하는 알고리즘의 사용 및 마오리족 세계관(Te Ao Māori)의 관점을 포함하는 알고리즘 개발

사람

다음과 같은 방법으로 사람에 초점을 맞춘다.

▸알고리즘에 관심이 있는 사람, 집단, 커뮤니티를 찾고 적극적인 참여 조직, 알고리즘 사용에 영향을 받는 사람들의 의견 수렴

데이터

다음과 같은 방법으로 데이터가 그 목적에 부합하는지 확인한다.

▸그 한계를 이해하기

▸편향을 식별하고 관리하기

개인정보 보호, 윤리 그리고 인권

다음과 같은 방법으로 개인정보, 윤리 그리고 인권에 대한 보호를 보장한다.

▸의도치 않은 결과를 평가하고 이에 대한 조치를 취하기 위한 정기적인 전문가 검토

인간의 감독

다음과 같은 방법으로 인간의 감독을 유지한다.

▸알고리즘에 대한 공개적인 조사(public inquiry)를 위한 담당자 지명

▸알고리즘에 영향을 받은 결정에 대해 불만을 접수하거나 이의를 제기하기 위한 수단 제공

▸알고리즘에 영향을 받은 결정에서 인간의 역할에 대한 명확한 설명

4) Algorithm charter for Aotearoa New Zealand.

- 뉴질랜드 헌장은 각 기관이 도입 및 사용하는 인공지능 알고리즘의 위험성 발생가능성과 영향을 평가하도록 하고, 위험성 수준별로 헌장의 의무적 적용 여부를 결정하도록 함

【뉴질랜드 정부】 알고리즘 영향평가와 위험성 매트릭스

위험성 매트릭스

발생가능성			
통상 있음 표준적인 작동 중에 자주 발생할 수 있음			
때때로 있음 표준적인 작동 중에 간혹 발생할 수 있음			
거의 없음 표준적인 작동 중에 발생할 가능성이 낮지만 발생할 수는 있음			
영향 정도	낮음 의사결정의 영향이 독자적이며, 심각하지 않음	중간 의사결정의 영향이 중간 규모의 사람들에게 미치며, 어느 정도 심각성이 있음	높음 의사결정의 영향이 광범위하며, 매우 심각함

위험성 등급

낮음 알고리즘 헌장이 적용될 수 있음	중간 알고리즘 헌장이 적용됨	높음 알고리즘 헌장이 반드시 적용돼야함

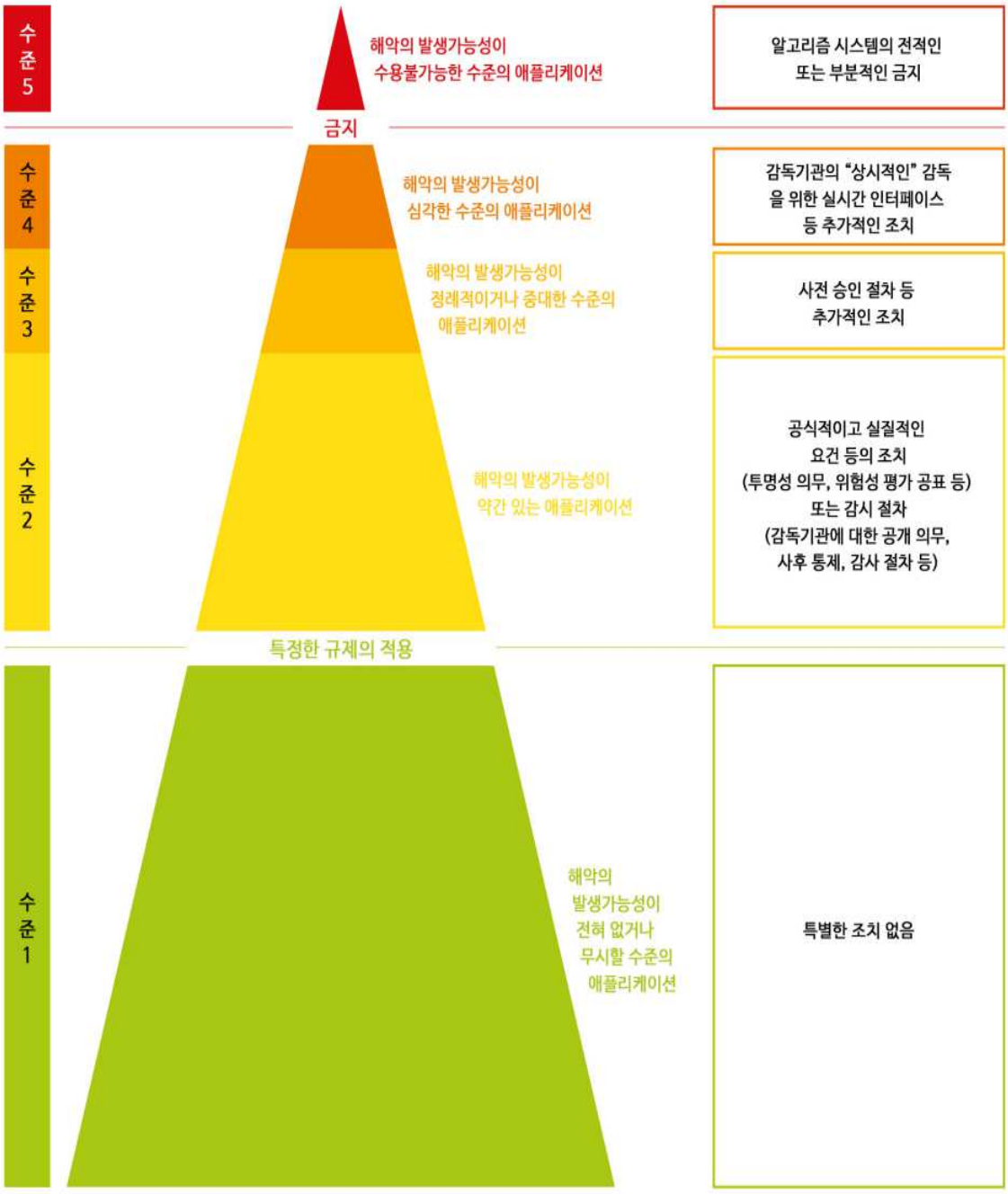
○ 독일 정부

- 독일 연방정부 데이터윤리위원회는 2019년 10월 5단계 '위험도 피라미드' 등 위험성에 기반한 알고리즘 시스템의 규제 모델을 제안함[5)]

5) Opinion of the Data Ethics Commission. https://assets.contentstack.io/v3/assets/blt3de4d56151f717f2/blt300ce23c9789e0f3/5e5cfe13fa08326331360f93/191023_DEK_Kurzfassung_en_bf.pdf

【독일 정부】

알고리즘 시스템 사용에 대한 위험도 피라미드 및 위험성 기반 규제 시스템



○ 유럽연합

- 2020년 유럽연합 집행위원회는 〈인공지능 백서(2020. 2)〉[6], 〈인공지능 공공조달 백서(2020. 5)〉[7]를 연달아 발표하며 인공지능 규제 프레임워크를 제시함

- 유럽연합은 〈인공지능 백서〉에서 위험기반 인공지능 규제 프레임워크를 제안함. 즉, '고위험 인공지능' 시스템으로 분류된 제품 및 서비스에 대하여 법적 의무를 부과하고 이에 대한 준수를 검증하기 위한 사전 적합성 검사의 실시를 의무화함

【유럽연합】 고위험 인공지능의 범주

구분	범주
일반	①일반적으로 수행되는 활동의 특성이 상당한 위험이 발생할 것으로 예상되는 분야에서 사용되고(의료, 운송, 에너지 및 공공 부문 등) ②해당 인공지능 애플리케이션이 해당 분야에서 상당한 위험이 발생할 가능성이 높은 방법으로 사용되는 경우(개인이나 기업의 권리에 법적인 영향 또는 유사하게 상당한 영향을 미치는 경우. 또는 부상·사망 또는 상당한 물질적·비물질적 손상을 초래하거나, 개인이나 법인이 합리적으로 피할 수 없는 영향을 미치는 경우)
특별	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 사용

- 고위험 인공지능에 대해서 기존 법률 규제에 더한 요구사항이 법적 의무로 적용됨. 이러한 법적 의무로는 △훈련 데이터 △기록과 데이터의 보존 △정보 제공 △견고성 및 정확성 △인적 감독 △원격 생체인식 등 특정 애플리케이션 특별 요건 등이 제시됨

【유럽연합】 고위험 인공지능 알고리즘에 대한 법적 의무

항목	요구사항
훈련 데이터	위험 시나리오를 고려하고 충분히 광범위한 데이터셋에 기반해 훈련하는 등 안전을 합리적으로 보장할 것. 시스템의 사용이 금지된 차별을 수반하는 결과로 이어지지 않도록 합리적인 조치를 취할 것. 제품과 서비스를 사용하는 동안 사생활과 개인정보를 적절히 보호할 것
기록과 데이터의 보존	인공지능 시스템의 훈련 및 테스트에 사용된 데이터셋의 특성 및 선택 절차에 대하여 정확히 기록할 것, 정당한 경우 데이터셋 그 자체를 보존할 것, 시스템의 구축/테스트/검증에 사용된 프로그래밍 및 훈련의 방법론에 대하여 문서화할 것

6) European Commission (2020). "WHITE PAPER: On Artificial Intelligence - A European approach to excellence and trust".

7) European Commission (2020). "White Paper on Data Ethics in Public Procurement of AI-based Services and Solutions".

정보 제공	시스템의 역량과 한계, 특히 시스템이 의도한 목적, 의도한 대로 기능할 것으로 기대할 수 있는 조건 및 특정 목적을 달성하는 데 예상되는 정확도 수준에 대한 명확한 정보를 제공할 것, 시스템과 상호작용하는 시민들에게 사람이 아니라 인공지능 시스템이라는 사실을 알릴 것
견고성 및 정확성	시스템의 모든 생애주기 단계에서 견고하고 정확하거나 최소 자기 정확성의 수준을 올바르게 반영하도록 보장할 것, 결과를 재현할 수 있도록 보장할 것, 시스템의 모든 생애주기 단계에서 오류 또는 불일치를 적절히 처리할 수 있도록 보장할 것, 공개적인 공격 및 데이터 또는 알고리즘 자체를 조작하려는 교묘한 시도 모두에 대해 탄력성을 가지고 완화 조치를 취하도록 보장할 것
인적 감독	시스템의 결과물이 사전에 사람에 의해 검토되고 검증되지 않은 경우 효력이 없는 경우(사회보장급여 신청에 대한 거부 등), 시스템의 결과물에 즉시 효력이 발생하지만 사후 인적 개입이 보장되는 경우(신용카드 신청에 대한 거부 등), 설계단계에서 시스템의 작동을 제약하는 경우(무인자동차의 센서 가시성이 저하되는 경우 작동을 중지시키거나 선행 차량과 일정 거리를 유지하는 등)
원격 생체인식 등 특정 애플리케이션에 대한 특별 요건	GDPR은 자연인을 고유하게 식별하려는 목적으로 생체인식 정보를 처리하는 것을 원칙적으로 금지하고 상당한 공익상의 이유로 법률에 따라 처리하는 경우 등에서만 예외적으로 인정함

- 고위험 인공지능에 대한 의무적 요구사항이 준수되는지 검증하고 보장하기 위해 객관적이고 사전적인 적합성 평가(Prior Conformity Assessment)를 의무적으로 실시하도록 함. 사전 적합성 평가는 인공지능 시스템에 대한 테스트, 검사 또는 인증 절차로 이루어지며 개발 단계에서 사용되는 알고리즘과 데이터셋에 대한 점검이 포함됨. 사전 적합성 검사 부적합 판정 시 인공지능 시스템을 재교육하도록 하고, 중소기업에 대해 온라인 검사 도구 등을 지원하도록 함. 시스템으로부터 부정적 영향을 받은 당사자에 대한 효과적인 사법적 보상을 보장하도록 함

○ 특히 국민과 영향을 받는 당사자들에게 투명하고 적법한 절차를 보장하는 법률 규율을 마련하는 것이 중요함

【사례】 휴스턴 공립학교 고용 평가 알고리즘의 투명성과 적법 절차[8)]

▸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

8) HOUSTON FED. OF TEACHERS v. HOUSTON INDEPENDENT.

【사례】 네덜란드 사회복지 부정수급 탐지 알고리즘의 투명성[9)]

▸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

- 공공 부문 인공지능의 경우 알고리즘을 이용한 의사결정이 이루어지기 전에 국민 일반에게 정보와 설명을 공개하고 특히 공청회 등의 방식으로 영향을 받는 주체들의 의견을 수렴하는 절차가 보장되어야 함. 더불어 공공과 민간을 아울러 정보주체에게 법적이거나 상당한 영향을 미치는 자동화된 의사결정의 경우 영향을 받는 주체에게 그 사실과 주요 로직, 장래의 영향에 대하여 사전에 통지하고 거부권을 보장할 필요가 있음

【사례】 암스테르담과 헬싱키 시, 알고리즘 등록부 공개[10)]

▸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

- 한편 인공지능으로 자동화된 의사결정으로 공공 처분이 이루어지는 경우 청문권, 문서열람권, 결정의 이유제시요구권 등 헌법상 적법 절차를 보장해야 함. 더불어 공공과 민간을 아울러 정보주체에게 법적이거나 상당한 영향을 미치는 자동화된 의사결정의 경우 정보주체의 동의나 법률, 계약에 의하여 적법하게 처리하는 경우라 하더라도 인적 개입 요구권, 의견 진술권, 이의제기권 및 권리구제를 보장해야 함. 이러한 처리에서 아동은 제외하는 것이 바람직함 ▢

---

https://www.leagle.com/decision/infdco20170530802#

9) 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/

10) 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/

인공지능 시대 이용자 관점을 고려한 법제화가 필요하다

윤명 사무총장 | 소비자시민모임

인공지능이라는 단어가 우리의 일상에서 점차 보편적으로 사용되고 있지만, 국민 개개인이 생각하는 인공지능의 수준이나 정의는 아마 다를 것이라 생각한다. 우리는 그동안 데이터이용과 관련하여 여러 이슈들을 고민해 왔다. 그러나 인공지능(AI)는 그간의 데이터 활용과는 좀 다르게 어떤 면에서는 기대감을 주지만, 또 한편으로는 두려움을 갖게 된다.

인간과 AI의 대결을 논하고 아직은 초기단계이지만 인공지능을 기반으로 한 서비스를 실제 이용하면서 오히려 기술이 가져다주는 혜택과 이익 보다는 알고리즘이라고 명명되는 AI의 운용체계의 실체를 알지 못하는 일반적인 이용자로서는 '모른다는 것', '내가 결과 도출 과정이나 데이터의 내용을 확인하거나 선택할 수 없다는 것'에 대한 두려움(불안함)이 있다.

국제소비자기구(Consumers International)에서는 2019년 4월에 인공지능 관련 세계 각국의 소비자단체(소비자)를 대상으로 설문조사를 진행하였다. 그 결과를 보면, 긍정적인 답변으로는 AI 서비스를 통해 개인 맞춤형 추천 등 새로운 것을 발견하게 도와주는 등 정보를 얻고, 의사를 결정하는 데 유용하고, 다른 사람에게 의지할 필요가 없고(낯선 사람에게 길을 묻거나, 매장의 직원의 도움을 받거나 등), 오프라인 매장에 가지 않아도 되어 생활에 유용하다는 의견이 많았고, 특히 인도나 아랍 등에서는 여성들에게 독립성과 권한을 부여한다는 의견도 있었다. 부정적인 답변으로는 AI 서비스를 다른 사람들이 이용하니까 최신 경향을 따라가기 위해서 이용한다. 개인의 데이터정보 사용되는 방식에 대한 우려와 이에 대한 투명성(공개)가 부족해 불안하다. 내 사적인 영역까지 생활 전반에 대해 AI가 알고 있는 것에 대해 소름끼친다. 소비자가 AI 기술을 통제할

수 없는 경우가 많다.(약관에 동의하지 않으면 서비스를 사용하지 못하기 때문에 불편해도 동의해야 한다)

국제소비자기구(CI)에서는 AI 서비스의 소비자 위험을 최소화하고 잘 활용하기 위한 6가지 방법을 제시하였다. 소비자단체들의 논의 내용을 공유하고, 이를 인공지능 개발과 규제를 위한 입법과정에서 함께 고민해 보기를 바라면서 소개한다.

1. 책임성

문제가 생겼을 때 책임질 대상이 있어야 한다. 공공정책 기준에서 책임의 틀이 확립되고 감시되는지가 중요하다.

책임감, 신뢰성은 핵심적인 사안으로 투명성의 목적을 명확히 하는 것이 필요하다.

기업이 아는 것과 소비자가 아는 것 사이의 정보 불균형 때문에 투명성을 강화하는 것은 중요하다.

문제가 생겼을 때 책임의 대상을 명확히 해야 한다. 책임을 지는 역할을 누가 하고 그렇게 하도록 누가 효과적으로 지원할지를 확립한다.(정부, 시민사회, 이익집단 등)

투명성의 목적을 명확히 정의한다.

알고리즘 영향 평가

2. 권한과 통제권

소비자들은 선택권이 없이 디지털 서비스를 이용하고 따라가는 추세이다. AI는 또 다른 측면에서 힘의 집중을 가져왔고, 또 다른 의미의 독점적 구조가 형성된다. 따라서 경쟁과 새로운 참여자가 필요하며, 과거 시장을 기준으로 정립된 독점의 정의도 디지털 사회에 맞게 다시 정립되어야 한다.

정책 결정을 위해 디지털 경제 사회에서 소비자들이 의미하는 선택권, 권한, 통제권이 무엇인지를 정의하고 이해하기 위해 다양한 이해관계자 참여하는 프로세스를 확립한다.

소비자들이 자신의 데이터를 갖고 다른 서비스로 이동시킬 수 있는 데이터 이동권

기업들 간에 정보 공유에 대한 새로운 규정

AI 환경에서 잠재적 장애와 경쟁의 특성 모색

3. 규제적 접근

AI관련 문제는 지금의 규제로는 적용이 안 되고, 다른 분야의 개념을 적용하여 해결을 하려고 하지만, 소비자보호를 위해서는 충분하지 않고, 제대로 적용될 수 없다. 따라서 새로운 규제를 마련할 필요성을 느끼지만, 새로운 규제를 시험해 보면서 더 효과적인 관리 규제 방안을 모색해야 한다. 기업들도 자체적으로 규제하기 위해 노력하고 있지만 자체 규제에는 한계가 있으므로 시스템을 만들고 규제 툴을 만드는 것이 정책 입안자

의 책임이다.

AI와 관련된 대부분의 문제가 윤리적 결정과 관련되어 있는데 사회적으로 세계적으로 공유되는 윤리적 원칙과 규정이 개발되어야 한다.

4. AI 정의

거의 모든 사람들이 AI라는 용어가 불분명하고 혼란스러워 AI가 주는 기회와 위험성을 어떻게 관리할지를 건설적으로 논의하기가 어렵다. 게다가 지역별로도 다른 AI에 개념은 AI가 사회에 미치는 영향이나 AI 규제 틀에 대한 인식에도 영향을 끼친다.

소비자 관점에서 AI에 대한 이해가 명확하고 공유되어야 하고 AI의 위험과 영향에 대해서도 공유되어야 한다.

5. 새로운 시스템(구조) 만들기

기술 전문성과 국가적 투자와 관련해서 정부가 대 기업들에게 의존하고 있고 이 기업들은 AI의 좋은 면을 강조하며 개입이 필요 없다고 주장하고 있다. 정책 입안자들이 AI 이해력을 높여야 한다. AI 관리를 도와줄 수 있는 신뢰할 수 있는 다양한 이해관계자 전문가 기구가 설립되어야 한다. 기업이 정책 결정 규정의 중요성, 즉 어떤 사람을 고용할지 측면에서도 다양한 사람들을 채용할 필요가 있다.

· 소비자 보상을 위해 AI 옴부즈만을 설립한다.

· AI에 대한 공유된 국제적 원칙 개발

· AI에 대한 믿을 만한 전문성을 갖추고 규제도 하고 다른 나라의 동일 기관들과 협력할 수 있는 국가적 중앙 기구를 만든다.

6. AI 이해력(literacy) 키우기

모든 사람이 AI가 어떻게 계획되고, 사용되는지에 대한 이해력을 키워야 한다. 소비자, 정책 입안자. 기술자, 디자이너, 기업은 AI 관련된 능력을 키우는 방법을 배워야 한다.

· AI 이해력 프로그램 : AI에 관련된 모두를 위한 개발법과 사용법

· AI관련 능력을 키우기 위해 대학과 리더십 프로그램과 협력한다.

· 전문가 기구가 제공하는 소비자에 대한 AI의 위험과 영향 관련 명확하고 중립적인 정보

· 변호사, 기술자, 소비자 단체 등 다양한 이해관계자가 참여한 프로세스를 이용해 AI 제품과 서비스에 ‘디자인 단계부터 인권’을 적용하는 방안을 개발

인공지능 기술이 우리의 생활에 긍정적으로 잘 활용되기 위해서는 개발 초기 단계에서부터 많은 논의를 통한 정책 및 입법 과제를 발굴해야 한다. 특히 오정미 변호사님의

발제에서 언급하고 있듯이 우리나라 규제적 접근에 있어서는 기술적 발전을 통한 인공지능의 활용 등의 산업 진흥적 측면에서 규제방안이나 법제도가 논의되고 있다. 그러나 인공지능은 아직 무한한 개발의 단계를 앞도고 있고, 인공지능이 미칠 영향은 개인뿐만 아니라 사회 전체적인 관점에서 고려되고 논의되어야 할 것이다. 또한 개발 초기 단계에서부터 역기능을 최소화하기 위한 안전규제 및 위험관리 소비자보호 체계를 균형 있게 마련하려는 노력이 매우 중요하다.

규제나 법은 위험의 요소를 최소화하고 이용자를 보호하는 측면이 우선되어야 함에도 불구하고 현재 우리의 인공지능과 관련한 규제나 법제도 마련에 있어서는 이러한 부분이 거의 나타나고 있지 않다. 문제가 발생한 다음에 법을 만드는 것이 아니라 개발 단계에서부터 올바른 활용을 위한 전제, 사회적 규범, 규제를 논의하고 마련해야 한다.

인공지능에 있어서 소비자 입장에서는 정보의 비대칭성이 다른 어떤 산업분야보다도 크게 나타날 것이다. 정보의 비대칭성은 수많은 소비자 문제를 야기하였고, 문제를 해결하거나 피해보상을 받는데 있어서도 소비자에게 많은 어려움이 있다. 따라서 AI의 공정성, 투명성, 책임성은 매우 중요한 가치이다.

특히 발제에서 언급하고 있는 내용은 그동안 국제소비자기구에서 논의하고 있는 방향과도 일맥 상통한다. 인공지능사회는 발전하고 있는 단계로 어떠한 결과를 만들어 낼지, 또 우리 사회에 어떤 영향을 미칠지에 대해서 단언하기는 어렵다. 그러나 현재 우리가 갖고 있는 규제나 법체계로는 기술이 발달하면 할수록 우리 사회질서를 지키는 데 새로운 법규범 체계가 필요하다고 생각된다. 이러한 법규범을 세우는 데 있어서 다양한 가치가 상충할 수도 있을지 모르겠다. 또 인공지능은 데이터의 결합이며, 데이터의 관리 분석이 집약된 형태로 이해할 수 있는데, 과연 이러한 과정에서 모두의 이익을 공정하게 대변할 수 있을지에 대해서도 고민해봐야 한다. 어쩌면 한쪽의 다수의 이익을 대변하거나 다수의 가치가 우선시 되어 소수의 기본권은 무시되지는 않을지도 소비자로서는 매우 관심이 많고 고민이 되는 부분이다.

국가나 정부, 입법기관에서는 인공지능 시대, 디지털 시대에 따른 사회적 논의기구를 통한 끝임 없는 가치판단 윤리기준을 다시 정립해야 한다.

앞으로 발전될 인공지능 사회가 어떠한 모습일지는 지금 우리가 논의하고 있는 논의의 중심이 이용자의 관점에서 더 많이 고민되고 고려되어야 할 것이다.

또한 투명성, 공정성 책임성에 대한 산업에서의 책임에 대해 좀 더 우리는 명확하게 정의하고 그 방향을 설정해야 할 필요가 있다. 인공지능시대 산업과 이용자 사이의 정보, 지식의 비대칭을 극복하기 위한 이용자 보호 측면이 더욱 중요시 되길 기대한다. ▢

인공지능 채용 도구의 공정성과 투명성

김민 정책활동가 | 진보네트워크센터

1. 개요

인공지능 기술이 향상되고 보급되며 서비스의 비용을 절감하고 품질을 개선하는 등 긍정적인 모습과 동시에 차별, 개인정보 자기결정권 침해 등 새로운 문제와 위험 또한 지속적으로 발견되고 있음

채용 과정 또한 인공지능 기술을 기반으로 한 도구들이 널리 도입되고 있는 추세임. 2019년 한국경제연구소 조사 자료에 따르면 인공지능을 활용한 신규 채용을 진행하고 있는 기업이 11.4%, 활용 계획이 있는 기업이 10.7%였으며 공공기관의 경우에도 최소 20여곳 이상의 기관에서 AI면접 등의 인공지능 시스템을 도입하고 있는 것으로 알려짐.

인공지능 기술을 기반으로 한 채용 도구는 주로 서류검사, 온라인 게임, 영상 분석으로 나뉘며 역량, 직무수행에 필요한 인성, 인지능력 보유 여부, 성장 가능성, 잠재력, 주의력 등을 평가한다고 주장함.

채용 담당자인 고용주들은 많은 양의 서류, 지원을 효과적이고 빠르게 처리하기 위해 이러한 도구를 활용하며 기술을 통해 객관적이고 공정한 채용 절차를 마련했다고 생각하지만, 인공지능 기술의 코드와 알고리즘에는 드러나지 않는 차별의 위험성이 항시 존재함.

동시에 인공지능 기술의 불투명하고 배제적인 특성으로 인해 구직자들의 부담이 증가하고 있음. 2021년 구인구직 매칭 플랫폼 사람인의 조사[1]에 의하면 구직자 64.4 퍼센트

가 인공지능 채용에 부담을 느끼고 있으며 그 주된 이유로 '무엇을 준비해야 할지 몰라서(58.6%), '관련 정보 자체가 부족해서(53.4%)', '평가 기준이 모호해서(36%)', 'AI전형을 위한 준비 시간, 비용이 늘어서(22%)'를 꼽았음. 더불어 인공지능 채용에 대비하기 위해 월 평균 7만 5000원을 지출하고 있는 것으로 집계되었음.

현재의 노동 인구는 성별, 성적 지향과 성 정체성, 장애, 나이, 인종과 민족, 학력, 지역 등 다양한 요소에 기반한 오랜 기간의 차별이 반영되어 있음. 성별에 따른 통계를 보면 그 격차가 지속적으로 감소해 왔음에도 불구하고 2019년 여성 고용률은 51.6%, 남성 고용률은 70.7%였으며 남성 임금 대비 여성 임금의 비율 또한 69.4%에 불과했음. 이보다 격차가 더 큰 과거 데이터에 의존하는 인공지능 기술은 불평등과 차별이 반영된 데이터를 학습하는 것이며 결과적으로 이를 확대하고 재생산할 위험성이 있음.

국내의 채용 공정화를 위해 마련된 법안들을 살펴보면 공통적으로 채용 과정에서 성별, 연령, 신체조건, 용모, 출신지역 등으로 차별하지 말 것을 규정하고 있으나[2] 인공지능 기술이 개입된 채용 절차가 이러한 규정을 준수하고 있는지 확인하고 있지 않음. 이를 검증하는 일은 가장 큰 피해를 받는 구직자에게 뿐만 아니라 이에 대한 책임을 갖고 있는 정부 기관에게도 어려운 일임. 공정 채용의 중요성에 비추어 볼 때, 채용 과정에서의 인공지능 기술 사용은 다른 영역에서의 인공지능 기술 사용에 비해 보다 엄격한 기준 및 조치가 필요할 것임.

2. 채용 과정의 인공지능

인공지능 기술을 활용한 도구는 편향을 가진 인간 대신, 규칙과 데이터에 기반하고 있으므로 객관적이고 공정할 것이라는 인상을 주지만 알고리즘과 모델이 어떻게 설계될 것인지 결정하는 것은 인간이며, 일반적으로 실제 사회의 데이터를 사용하여 훈련하므로 사회의 편견이 반영되어있을 가능성이 있음.

검증되지 않은 모델 기반의 채용 도구는 채용 과정에서 부정행위를 저지르는 소수의 인간 채용 담당자보다 훨씬 큰 영향력을 갖고 차별적 결과를 내놓을 수 있으며 이에 대한 문제 제기, 구제 절차 또한 명확하지 않음.

채용 과정은 누구나에게 자신과 가족을 부양하는 중요한 경제적 기회를 제공할 수 있는 절차이므로 공정함을 유지하며 차별을 식별하고 피해에 대한 구제절차를 마련해두는 것이 적합함.

현재 활용되는 인공지능 채용 도구는 크게 다음과 같이 나눠볼 수 있음[3].

---

1) "구직자 10명 중 6명, AI 채용 부담스러워!", 사람인 https://www.saramin.co.kr/zf_user/help/live/view?idx=108088&list_idx=20&listType=news&category=10&keyword=&menu=1&page=2

2) 채용의 공정성과 인공지능, 오정미

3) 현재 국내에서 활용되고 있는 대표적인 특정 기업의 인공지능 도구를 기반으로 작성됨. AI 서류평가의 경우 무하유 카피킬러 HR, AI 면접 및 역량검사의 경우 마이다스 아이티(마이다스 인), 제네시스 랩

(1) AI 서류 평가

지원자의 이력서, 자기소개서 등의 채용서류를 인공지능 기술을 통해 분석하는 도구임. 문장 및 맞춤법 오류 분석, 표절 검사, 주요 내용 요약과 같은 기본적 검수 기능 제공과 동시에 'AI평가' 및 '직무 적합도 평가' 기능을 제공하고 있음.[4)]

AI평가의 경우 고득점자 및 저득점자 자기소개서의 패턴을 기반으로 지원자 자기소개서의 점수를 매김. 예를 들어 지원자 자기소개서 속 단어와 문장을 능력, 경험, 신념, 가치관, 포부, 지원동기 등의 분류된 특성을 찾아내 이를 고성과자 및 저성과자의 특성과 유사 여부를 판단하는 것임[5)].

50만 건 이상의 자기소개서를 학습한 AI모델을 활용, AI 서류 평가를 도입하려는 회사의 합격자 자기소개서를 기반으로 우수 인재의 패턴을 학습 후 지원자 자기소개서를 평가[6)]

(2) AI 면접

온라인 영상 면접을 통해 지원자의 표정, 감정, 안구 움직임 등의 얼굴 정보와 목소리 톤, 크기, 속도, 음색 등의 음성정보를 추출하여 매력도, 의사표현, 감정 전달력, 호감도 등 외형적인 특성을 분석함.

이러한 도구는 실제 직군별 고성과자 및 저성과자의 AI면접 응시데이터와 성과데이터, 실제 지원자의 AI면접 응시데이터와 실제 성과데이터, 지원자 인터뷰 영상의 얼굴 정보 및 음성 정보에 대한 100여명의 우수 면접관이 판단한 평가 결과 데이터를 기반으로 하며 이를 통해 직군 적합도, 응시자의 특성을 요약해 보여주고 등급을 매겨 최종 결과로 도출해내는 것으로 알려짐.[7)]

얼굴 및 음성 분석을 통해 인지능력, 심리적 기질, 사회성 등을 평가한다고 주장해온 미국의 AI 채용 기술 기업 HireVue의 경우 검증되지 않은 얼굴 분석 알고리즘의 비과학성, 부정확성 등의 문제가 불거지자 얼굴 분석 기능을 중단한다고 밝힌 바 있음.[8)]

(3) AI 게임테스트 (역량검사)

인적성 검사의 게임화된 형태로 지원자의 게임 수행 과정에서 정당과 오답 뿐만 아니라 응답 반응 시간, 의사결정, 학습 속도 등을 평가하고 분석하는 것으로 알려져 있음.

---

4) "AI가 자기소개서를 읽는다. 어떻게? 이렇게! " , 전자신문 https://m.etnews.com/20191002000324?obj=Tzo4OiJzdGRDbGFzcyI6Mjp7czo3OiJyZWZlcmVyIjtOO3M6NzoiZm9yd2FyZCI7czoxMzoid2ViIHRvIG1vYmlsZSI7fQ%3D%3D

5) (주) 무하유, 카피킬러HR에 적용된 AI기술 https://www.hr.copykiller.com/technology

6) https://www.hr.copykiller.com/board/faq

7) 마이다스 인 https://www.midashri.com/intro/process/ai-interview/anal

8) HireVue, Facing FTC Complaint From EPIC, Halts Use of Facial Recognition, EPIC https://epic.org/2021/01/hirevue-facing-ftc-complaint-f.html

이 또한 실제 직군별 고성과자 및 저성과자의 응시 데이터, 사용 기업 현 재직자의 응시 데이터 및 성과 데이터를 기반으로 비교하여 평가 결과를 도출해냄.

3. 인공지능 기반 채용 절차의 문제와 활용 현황

인공지능 기술에 필요한 학습 데이터는 적용 대상이 되는 모집단을 충분히 반영하는 대표성이 확보되어야 함. 허나 기존에 존재하여 이용가능한 데이터는 모집단을 충분히 골고루 대표하지 않을 가능성이 있으며 이에 따라 특정 집단이 과잉대표되거나 과소대표되는 문제가 발생하고 차별을 재생산하는 기본적인 원인이 될 수 있음.[9] 또한 수집된 데이터에 이미 차별이 반영되어 있는 경우도 고려해야 할 것임.

인공지능 채용 도구는 일반적으로 성과가 좋은 것으로 판단된 노동자의 데이터, 흔히 말하는 '고성과자' 재직자의 데이터를 기반으로 인공지능을 개발함. 이렇게 고성과자 데이터의 특성과 패턴을 학습하는 경우, 그와 유사한 특성을 보여준 지원자에게 유리한 결과로 돌아가는데 해당 학습 데이터가 차별적으로 구성되어 있다면 이를 기반으로 한 평가 모델 또한 차별적인 결과를 내놓을 수 있음. 여성에 대해 차별적 결과를 내놓아 폐기된 아마존의 채용 프로그램의 사례는 차별적인 과거 데이터를 활용하여 섣불리 설계된 인공지능 채용 도구가 차별적 결과를 재생산한다는 것을 보여줌.

AI 서류평가, AI 면접, AI 게임테스트 모두 기존 고성과자의 패턴을 추출하고 학습하여 지원자의 결과와 비교하고 점수를 매기며 평가하는 것으로 확인되는데 위와 같은 지점이 충분히 고려되었는지 의문임.

아울러 인공지능 채용 도구의 통계적 정확성이 높은 것이나 평가 결과 인간의 평가보다 상관계수가 높은 것이 공정함을 담보하지 않음. 인공지능 기술은 기본적으로 특정한 패턴을 찾아내는 것에 효과적이고 효율적이지만 채용 과정에서 중요한 것은 해당 특성이 실제 직무 역량과 어떤 관련이 있는지, 실제 의도와는 다른 대리 데이터가 판단의 기준이 되는 것은 아닌지, 학습 데이터가 차별적으로 구성되어 있는 것은 아닌지 살펴보며 공정성을 설계하는 것임

더 큰 문제로, 인공지능 채용 도구의 사용자인 고용주는 채용 과정에서 해당 도구가 정확히 무엇을 테스트하는지, 구체적으로 지원자의 어떤 특성들이 측정되는지, 측정되는 특성들이 직무 수행을 위해 필수적인지 고려하거나 평가하지 않은 것으로 보임. 또한 검증되지 않았음에도 보조 및 참고용 도구로 사용하는 것이 아닌, 채용 여부의 당락을 결정하는 직접적인 수단으로 사용하는 경우 또한 존재함.

다음은 정보공개 청구, 언론보도 및 채용 공고 등을 통해 확인한 공공기관 채용 과정의 일부 사례임

9) 빅데이터 알고리즘의 성차별 가능성에 관한 실증적 분석과 개선방안, 한국여성정책연구원 고려대학교 노동문제연구소, 2019.12

SW마에스트로 연수생 선발 과정 AI면접 및 역량검사 도입 사례[10)]

-AI면접의 결과를 면접 참고자료로 활용하였으나 AI면접의 결과와 실제 채용 결과가 유의미한 상관관계를 보이지 않았음

< AI면접 및 심층면접 평가 비교 >

AI 결과	심층면접 결과 (40점 만점)	
B+	26	하위 1등
A	26.6	하위 2등
B-	27.6	하위 3등

AI 결과	심층면접 결과 (40점 만점)	
B	36.4	상위 1등
B-	36.2	상위 2등
C	36	상위 3등

< AI면접 등급별 심층면접 평균값 현황 >

AI 등급	A	B+	B	B-	C	D
심층면접 평균값	32.2	31.6	32.0	31.7	31.4	32.0

한국공항공사, 인천공항공사 AI면접 및 역량검사 도입 사례[11)]

-공사의 경우 AI면접의 결과를 면접 참고자료로 활용하였으나, AI면접의 결과에 대한 사전교육 등이 면접위원에게 진행되지 않았음.

-인천공항의 경우 AI면접과 실제 채용 결과가 일치하지 않는 등 유의미한 결과를 보이지 않았음

<인천국제공항공사 AI면접 결과와 실제 최종합격자 비교>

AI면접 평가등급	S	A	B+	B	B-	C	D
등급 내 최종합격자 비율	0%	51%	43%	35%	27%	15%	35%

-AI면접을 도입하며 측정방법과 알고리즘에 대한 기술적 검토 또는 외부 자문은 물론이고 차별과 편향성에 대한 사전 논의 또한 없었음

-AI면접 미이행 시 전형 탈락

한국방송통신전파진흥원 AI면접 및 역량검사 도입 사례[12)]

-기술적 신뢰도 등의 문제가 있음에도 불구하고, 한국방송통신전파진흥원은 AI면접을 참고자료 등 보조수단이 아닌 채용 여부의 당락을 결정하는 직접수단으로 사용

-2020년의 경우, 315명의 지원자 중 228명이 AI면접의 단독 평가에 의해 불합격함

-진흥원은 AI면접이 어떤 알고리즘을 통해 이들을 불합격시켰는지 파악하고 있지 못함

중소벤처기업진흥공단 AI면접 및 역량검사 도입 사례

-2020년 상반기 체험형 청년인턴 모집 도중 일반 면접 전형을 AI면접으로 변경

-AI면접이 최종 전형이었으며 각 지역을 합해 3배수의 1차 서류 평가를 통과한 101명 중 64명이 AI면접 전형으로 불합격함

한국자산관리공사, 한국전력기술 등의 사례

-한국자산관리공사의 경우 2018, 2019, 2020 체험형 청년인턴 모집 과정에서 전부 AI면접을 최종 전형으로 진행.

-한국전력기술의 경우 2020 상반기 하반기 체험형 청년인턴 모집과정에서 AI면접을 최종 전형으로 진행

중소벤처기업진흥공단 AI서류 평가 도입 사례

-2020년 7월 비대면 디지털 일자리 채용 과정에서 AI 서류평가 전형을 도입. 해당 채용 과정은 서류 전형만으로 이뤄졌으며 그외의 일체 전형은 생략되었음[13)]

-AI평가 (50%), 직무적합도 평가 (50%)로 구성된 서류 전형에서 단독적인 AI 서류 평가 결과로 선발. 이에 대한 해당 전형에서 채용 담당자의 개입 및 관리 유무에 대한 정보 없음

-각 지역을 합해 4000명이 넘는 지원자 중 200명 선발, 3800명이 넘는 지원자가 AI 서류 평가 단독 평가로 불합격함.

한국수자원공사 AI 서류 평가 도입 사례

-2019년부터 AI 서류 평가 시스템을 도입하여 분석된 서류 평가 결과를 면접 전형에 활용하였음[14)]

-채용 공고 상 2차 전형 : 직무역량면접평가 (PT 면접(40%), 자기소개서 기반 경험역량면접(60%))로 진행되었으나 해당 자기소개서가 AI 평가를 기반으로 진행된다는 언급 또는 공지 없음[15)]

한국전력거래소 AI 서류 평가 도입 사례

-2020년 6월 일반직 공개채용 과정에 1차 전형으로 AI 서류 평가를 도입하였고 30배수를 선

10) 청년 앞길 막는 AI면접… AI윤리기준 투명성 공정성 필요, 과기정통부 예산심사 질의서- 정필모 의원실)
11) 인천공항・한국공항공사, 인공지능의 차별 학습 및 편향성 대비 없이 무책임 AI면접 도입, 심상정 의원실
12) 청년 앞길 막는 AI면접… AI윤리기준 투명성 공정성 필요, 과기정통부 예산심사 질의서, 정필모 의원실

발하였음

-채용 공고 상 서류 전형 : 자기소개서 NCS 및 직무역량 평가로 진행되었으나 해당 자기소개서가 AI평가를 기반으로 진행된다는 공지는 없음[16)]

한국산업단지공단 AI 서류평가 및 AI 역량검사 도입 사례

-2021년 상반기 체험형 청년 인턴 채용 과정에서 AI 서류 평가와 AI 역량 검사 절차 도입

-1차 AI 서류 전형 (3배수, 인공지능 시스템 심사 100점), 2차 AI 면접 전형 (1배수, AI역량평가 80점)로 이뤄지며 다른 전형 없이 AI역량평가 일반행정 직군 고득점자 순 최종 합격자를 선발함.

4. 해외 인공지능 채용 기술 관련 법안

해외의 경우, 채용 및 노동 영역의 인공지능 기술을 규제하기 위한 구체적인 법안이 통과되거나 발의되고 있음.

유럽연합 집행위원회가 발행한 EU 인공지능 백서는 사전규제가 필요한 고위험 분야로 일반적으로 위험이 발생할 가능성이 가장 높다고 판단되는 영역 (의료, 운송, 에너지 및 공공부문)과 더불어 채용 과정과 노동자의 권리에 영향을 미치는 상황에서의 AI 어플리케이션 사용은 항상 '고위험'으로 간주될 수 있으며 특정한 요구사항이 항상 적용될 것이라 밝힘.

이러한 '고위험 AI 어플리케이션'에 부과되는 법적 의무 요구사항으로는 '학습용 데이터 관리 및 기록 보관, 사용자에게 인공지능 시스템에 관한 정보(능력과 한계 등) 제공, 견고성 및 정확성 확보, 시스템의 목적과 영향을 고려한 인간의 감독 등이 제시됨.[17)18)]

미국의 경우 일부 주와 시에서 AI면접 또는 채용 목적으로 활용되는 인공지능 기술을 규제하는 법안이 통과되거나 발의되었음

---

13) 중진공, AI 서류 평가 통해 비대면 디지털 청년 인력 채용
http://www.muhayu.com/%eb%89%b4%ec%8a%a4%ed%94%84%eb%a6%ac%ec%a1%b4%e3%85%a3%ec%a4%91%ec%a7%84%ea%b3%b5-ai-%ec%84%9c%eb%a5%98-%ed%8f%89%ea%b0%80-%ed%86%b5%ed%95%b4-%eb%b9%84%eb%8c%80%eb%a9%b4-%eb%94%94%ec%a7%80%ed%84%b8/

14) 한국수자원공사, AI 기반 서류평가 결과 면접에 활용
http://www.muhayu.com/%eb%89%b4%ec%8a%a4%ed%94%84%eb%a6%ac%ec%a1%b4%e3%85%a3%ed%95%9c%ea%b5%ad%ec%88%98%ec%9e%90%ec%9b%90%ea%b3%b5%ec%82%ac-ai-%ea%b8%b0%eb%b0%98-%ec%84%9c%eb%a5%98%ed%8f%89%ea%b0%80-%ea%b2%b0%ea%b3%bc/

15) 2020년 상반기 일반직 신입 인턴[체험+채용형] 사원 일반 공채, 한국수자원공사

16) 2020년도 전력거래소 공개채용(일반직_경력직), 한국전력거래소

17) 유럽연합 인공지능 백서 - 5장 신뢰 생태계 구축 : AI규제 프레임워크

18) AI면접의 공정성·투명성 제고와 청년 취업준비생들의 부담 감소 방안 검토, 정필모 의원실-국회입법조사처

일리노이 인공지능 영상면접 법 Artificial Intelligence Video Interview Act

미국 일리노이 주는 고용자가 채용 과정에서의 인공지능 영상 면접 사용을 규제하는 법안을 제정하였음. 해당 법에 따라 영상 면접을 녹화하고 인공지능 기술을 이용해 지원자의 직무적합성을 분석하려는 고용자에게는 다음과 같은 의무가 부여됨

- 통지 제공 : 면접 전, 인공지능 기술이 면접 평가에 사용될 수 있음을 지원자에게 통지
- 정보 제공 : 인공지능이 어떻게 작동하는지, 그리고 지원자를 평가하는 데 사용되는 일반적인 유형의 특징을 설명하는 경위서를 지원자에게 제공
- 제3자 공유 금지 : 직무적합성을 판단하기 위해 필수적인 전문지식이나 기술이 있는 자 외 면접 영상의 공유를 제한

또한 다음과 같은 지원자의 권리가 있음

- 동의권 : 인공지능 기술로 평가받는 것에 대해 지원자로부터 사전 동의를 받아야 함
- 삭제권 : 지원자가 자신의 면접 영상 삭제를 요청한 경우 고용자는 30일 이내에 이를 파기해야 하며 영상을 공유 받은 제3자 또한 이를 삭제해야 함. 이는 전자적으로 생성된 모든 사본에 해당함

뉴욕시 자동화된 채용 결정 도구 판매에 관한 법안 Sale of automated employment decision tools (Intro. No. 1894-2020)

미국 뉴욕시는 알고리즘 책무성 법안이라 불리우는 지방법 49조를 통과시키며 인공지능 대책위원회를 구성했음. 해당 위원회는 시 기관의 자동화된 의사결정 시스템이 연령, 인종, 종교, 성별, 성적 지향 등의 여부에 따라 시민들을 차별하는지 조사함. 이후 채용과정에서의 자동화된 의사결정 도구 판매를 규제하는 법안이 2020년 2월 발의됨.

해당 법안은 다양한 알고리즘 방법론을 사용하여 채용 후보를 결정하거나 그와 비슷한 결정을 내리는 시스템을 정의하며, 그러한 도구의 판매를 위해 개발자에게 다음과 같은 요건을 요구함

- 나이, 인종, 신념, 피부색, 국적, 성별, 장애, 성적 지향, 시민권 상태 등 뉴욕 시 차별금지 조항과 같은 지역 고용법을 준수하고 있는지 점검하기 위한 편향성 감사의 대상이 됨.
- 해당 프로그램이 지난 해의 실시된 편향성 감사의 대상이었음을 확인할 수 있어야 함
- 추가 비용 없이 매년 편향성 감사를 진행하며, 그 감사 결과를 구매자에게 제공

또한 고용자는 지원자에게 다음과 같은 사항을 통지해야 함

- 법에서 감사를 요구로 하는 자동화된 채용 결정 도구가 지원자 평가에 사용되었음을 통지
- 해당 도구가 평가에 활용하는 지원자의 특성과 자격을 공개

*별첨. 공공기관의 AI면접 활용 현황과 정보공개 청구 문안 및 답변 요약

공공기관의 AI면접 활용 현황

- 2020년 10월 기준, AI면접을 도입했거나 계획중인 공공기관은 20여곳에 달함
- 각 공공기관마다 AI면접을 활용하는 방식이 다름. 점수에 반영하지 않고 대면 면접 시 참고용 자료로 활용하는 기관과 적극적으로 채용 점수에 반영하는 기관이 있으며, 일반 직원 채용에선 사용하지 않지만 시범적으로 인턴 선발에는 적극적으로 사용하는 기관 등이 있음.
- 일부 공공기관의 경우 AI면접의 결과를 대면 면접 시 면접관에게 제공되는 참고자료로 활용. 그러나 AI면접을 응시하지 않으면 탈락 처리되는 등 지원자는 필수적으로 AI면접을 응시해야 함.
- 서민금융진흥원, 한국방송통신전파진흥원, 한전KDN의 경우 AI면접의 결과를 합격자 채용을 위한 점수에 반영하고 지원자를 탈락시키는 용도로 활용하는 등 채용 결과에 직접적으로 큰 영향을 미치는 방식으로 AI면접을 활용하고 있음.
- 중소벤처기업진흥공단, 한국남동발전, 한국동서발전, 한국자산관리공사, 한국전력기술의 경우 체험형 청년인턴의 채용과정에 있어 AI면접을 대면 면접을 대체하여 활용하고 있으며 AI면접의 결과로 최종 합격 여부를 결정함

정보공개 청구 문안 및 답변 요약

정보공개 청구 문항은 인사혁신처의 [공정채용 가이드북], 고용노동부의 [개인정보 보호 가이드라인 – 인사·노무 편], 소프트웨어 공공조달 관련 법규를 참고하여 작성되었으며 이하는 정보공개 청구 질문과 답변에 대한 요약임.

정보공개 청구 기관 : 국민건강보험공단 일산병원, 서민금융진흥원, 인천국제공항공사, 중소벤처기업진흥공단, 한국국제협력단, 한국남동발전, 한국동서발전, 한국방송통신전파진흥원, 한국보훈복지의료공단, 한국자산관리공사, 한국전력거래소, 한국전력기술, 한전KDN

1) AI면접 관련 도입 근거와 계획, 절차, 사용된 목적 및 결과와 관련한 사항

- 채용절차에 AI면접을 도입한 근거 규정에 대하여, 한국방송통신전파진흥원, 한국자산관리공사, 한국국제협력단을 제외한 9곳의 공공기관은 AI면접 관련한 사항이 채용세칙 또는 채용업무세칙 등 규정에 존재하지 않으며 공개되지 않은 내부 계획 등을 근거로 들음.
- AI면접 사용 계획 및 구체적인 목적과 방법에 대하여는 대부분의 기관이 정보를 비공개하여 구체적인 사항을 확인할 수 없었으며, 부분공개한 기관의 경우에도 일반적

인 채용공고문 또는 AI면접 업체에서 제공한 단순 설명자료 등 간략한 정보에 그쳤음.

- AI면접 관련 공공기관이 채용담당자 또는 면접위원에게 제공하도록 되어 있는 사전교육의 경우 대부분의 기관이 비공개하였으며, 부분공개한 기관의 경우에도 AI면접과 관련한 면접위원 교육을 진행하지 않았음.
- AI면접 후 진행된 사후 점검과 평가와 관련하여 6곳의 기관은 비공개하였으며, 4곳의 기관은 AI면접의 내용과 관련한 사후 점검을 시행하지 않았다고 답변하였음.
  중소벤처기업진흥공단의 경우 감사인 입회를 통해 AI면접의 정상작동 여부, 조작가능성 여부, 면접 진행과정의 적절성 여부 등을 평가하였으나 이에 자세한 내용은 확인할 수 없었음. 한국방송통신전파진흥원의 경우 시스템 임차 계약에 관한 단가, 금액 등 형식적 계약사항에 관한 사후 점검만을 진행하고 AI프로그램의 사용적절성 등에 관하여는 점검을 진행하지 않은 것으로 보임.

2) AI면접 관련 용역 또는 위탁 계약 및 프로그램에 대한 정보에 관한 사항

- AI면접 계약 관련 입찰공고문, 용역 및 위탁 계약서 등의 경우 서민금융진흥원, 인천국제공항공사, 중소벤처기업진흥공단, 한국국제협력단, 한국동서발전, 한국자산관리공사, 한국전력기술, 한전KDN은 전부 비공개하였으며, 그 외 기관의 경우 감리, 감독, 검사의 현황 및 대가의 지급 현황 등의 정보를 제외한 용역 제안요청서, 입찰공고문 등을 부분공개하였음.
- AI면접 프로그램이 실제 사용될 당시 프로그램의 기능별 오류 및 오차율에 관하여는 12곳의 기관 전부 비공개 또는 이와 관련한 평가 자체를 진행하지 않아 정보 부존재로 답변하였음.
- AI면접 프로그램의 기술자료 (소스코드, 매뉴얼, 설계서, 기능명세서, 유지보수자료, 플로우차트 등)의 경우 12곳의 기관 전부 비공개하였음.
- AI면접 프로그램의 학습 데이터와 채점 결과의 통계 (성별, 학력, 지역, 장애유무에 따른 구분)와 관련하여 12곳의 기관 전부 비공개 또는 부존재 처분하였음.

3) AI면접 관련 개인정보의 수집과 처리에 관한 사항

- AI면접 과정에서 처리되고 수집된 개인정보의 처리를 위탁한 경우 수탁업체 명과 제공된 개인정보 항목 및 목적과 관련하여 인천국제공항공사, 한국국제협력단, 한국동서발전, 한국자산관리공사, 한전KDN의 경우 비공개 또는 부존재 처분하였음.
- AI면접 프로그램이 지원자로부터 구체적으로 수집하는 사항 (표정, 감정, 안구 움직임 등 얼굴 관련 개인정보와 목소리 톤, 크기, 음색, 속도 등 음성 정보 및 IP주소, 쿠키 등) 및 지원자로부터 받은 동의서 양식과 관련하여 8곳의 기관은 비공개하였고, 국민건강보험공단 일산병원, 서민금융진흥원, 한국동서발전, 한국방송통신전파진흥원, 한국전력거래소는 개인정보 제공 동의서 등을 공개하였으나 그 항목 등이 AI

면접 프로그램에 대하여 구체적이지 않았으며 민감정보에 대한 명시와 별도 동의 또한 제대로 이뤄지고 있지 않았음.

- 재직자의 AI면접 평가 실시 또는 AI면접 프로그램 사용을 위한 재직자 데이터 제공 유무와 관련하여 서민금융진흥원을 제외한 11곳의 기관에서 정보 부존재(재직자에게는 적용하지 않음) 또는 비공개(유무 확인할 수 없음)하였음. 서민금융진흥원의 경우 재직자 일부도 평가를 실시하였다고 답하였으나 재직자의 면접 결과, 또는 재직자의 성별, 학력별, 지역별 등의 요소에 대하여는 비공개하였음 ▢

과학기술정보통신부 토론문

김경만 과장 | 인공지능정책과

Ministry of Science and ICT

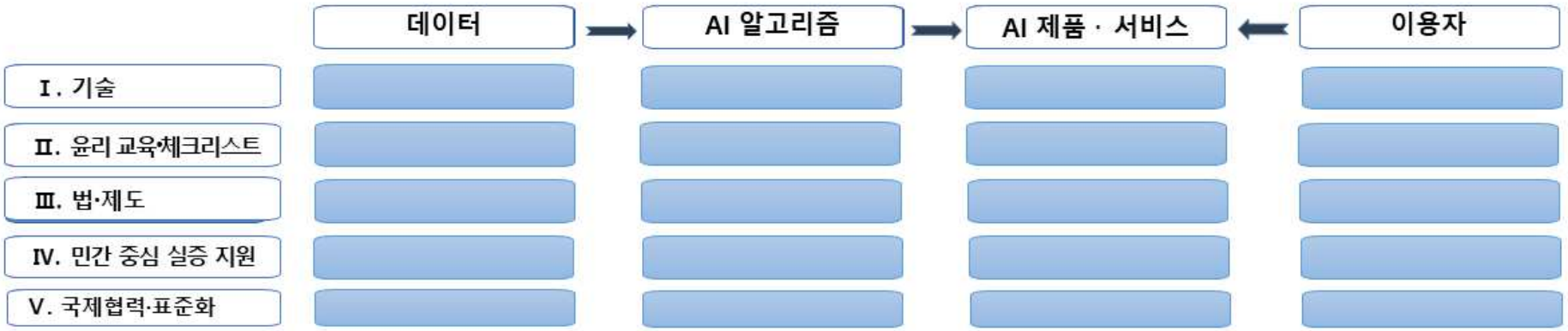
김경만
과기정통부 인공지능기반정책과장

사람 중심 인공지능 구현을 위한 도전과 과제- 정부의 관점

▪ 인공지능 기술이 일상생활의 일부가 되면서 윤리 문제 등 신뢰성이 점차 중요
▪ 인공지능 신뢰성 확보를 위한 기술개발, 윤리교육, 법제도 등 정책 필요

- (**체크리스트**) 인터넷기업협회·지능정보산업협회 등 **개발자와 기업 등을 대표할 수 있는 민간단체** 중심으로 주체별로 준수해야할 **체크리스트 마련 지원**
- (**윤리 교육**) 데이터·알고리즘 편향 등 인공지능 개발~활용 단계에서 발생할 수 있는 윤리 이슈에 대해 **일반 시민·대학원생 등 단계별로 교육 방안** 마련
- (**기술 개발**) 인공지능이 이용하는 **데이터 편향성을 완화하거나 윤리기준을 위배하는 알고리즘을 검증**할 수 있는 **기술개발 투자**
- (**법제도·인증**) **기업 자율의 알고리즘 검증**을 우선권고하고, **민간 중심의 자율인증제** 검토

<정책 영역(예시)>



공정거래위원회 토론문

이동원 과장 | 시장감시총괄과

[발제1] 헌법과 인공지능 (김민우 박사) 관련

□ 지적하신 바와 같이, 주요 온라인 플랫폼 사업자들에게 데이터가 집중되면서 거대 플랫폼의 지배력 강화 등 시장에 쏠림현상이 우려되는 상황임

ㅇ 특히 핵심 시장을 선점한 플랫폼 사업자가 연관 시장으로 지배력을 확대하기 위해 자사 상품·서비스를 우대하는 방식으로 검색알고리즘을 조정하는 경우, 시장의 경쟁이 훼손될 수 있음

□ 최근 공정위의 네이버 쇼핑·동영상 건 조치 이후, 검색 알고리즘의 공정성·투명성 확보가 필요하다는 사회적 공감대가 더욱 확산되고 있음

ㅇ 공정위는 네이버가 검색 알고리즘을 인위적으로 조정하여 자사 상품·서비스는 상단에, 경쟁사 상품·서비스는 하단에 배치한 행위를 조사해 시정하였음
* (쇼핑) 약 265억 원, (동영상) 2억 원 과징금 부과(`20.10월 보도자료)

ㅇ 이로 인해 검색결과 노출순위가 객관적이라고 믿는 소비자를 기만하고 오픈마켓, 동영상 플랫폼 시장의 경쟁을 왜곡하였다고 판단하였음

□ 주요 온라인 플랫폼 상의 검색결과 노출순위는 소비자의 구매여부, 입점업체의 매출을 좌우하는 핵심적인 요소임

ㅇ 이에 따라 EU 및 일본 등 각국에서도 온라인 플랫폼 상의 노출순위 결정기준에 대한 정보공개를 강화하는 법안을 제정하여 이미 시행하고 있음

□ 공정위 또한 총 12차례의 이해관계자 간담회, 관계부처 협의를 거쳐 정부입법절차를 통해 지난 1월 온라인 플랫폼 공정화법을 국회에 제출하였음

ㅇ 해당 법안은 상품노출순서 결정기준 등 입점업체의 이해관계와 직결되는 주요 거래조건을 계약서 기재사항에 포함하도록 하고 있음

ㅇ 단, 법안은 플랫폼 사업자의 영업비밀에 해당하는 검색알고리즘 자체를 공개하도록 하는 것은 아님

ㅇ 입점업체의 예측가능성을 확보하면서도 플랫폼 사업자의 영업비밀 유출이나 혁신 저해 우려는 최소화하는 균형있는 규율방안을 설계하는데 중점을 두고 법안을 마련하였음

□ 법안이 통과되면 플랫폼 사업자의 인위적인 검색알고리즘 조정을 예방하고 온라인 플랫폼 거래의 투명성·공정성을 강화하는데 기여할 것으로 기대됨

[발제2] 인공지능 법제정비 제안 (오정미 변호사) 관련

□ 오정미 변호사님께서는 과기부 중심의 산업진흥에 초점을 둔 현재 인공지능 법제에 공정성, 투명성, 책임성을 제고하기 위하여 공정위, 인권위, 개보위 3개 기관을 포함하는 거버넌스 구축을 제안해 주셨음

ㅇ 발제취지에 깊이 공감하며, 특히 미FTC 사례 설명과 함께 이에 대응하는 한국 공정위의 적극적 역할이 필수적이라고 말씀하신 부분은 해당부처 과장으로써 더욱 무겁게 받아들이고 있음

□ 이와 관련하여 현재 공정위는 디지털 공정거래 질서 확립을 최우선 과제로 하여 온라인 플랫폼 공정화법 제정, 전자상거래법 전면개정, 심사지침 제정 등 다양한 법제정비를 추진하고 있다는 점을 먼저 설명드림

ㅇ 해외 입법례로 언급해주신 EU 온라인 플랫폼 이사회 규칙에 상응하는 법안이 현재 공정위에서 국회에 제출한 온라인 플랫폼 공정화 법임

구분	온라인 플랫폼 공정화 법(정부안)	EU 온라인 플랫폼 이사회 규칙
적용대상	-판매금액 1,000억원 또는 매출액 100억원 이상 중개플랫폼	-모든 중개플랫폼 (규모요건 없음)
절차적 규제	-상품노출순서 결정기준 등 필수기재사항을 포함한 계약서 작성·교부 의무 -계약내용 변경 등 사전통지의무	-상품노출순서 결정기준, 차별취급 여부 등을 약관에 기재할 의무 부여 -계약내용 변경 등 사전통지의무
기타	-분쟁조정협의회 설치, 공정거래협약, 표준계약서 도입 등	-내부분쟁해결 시스템 마련의무, 단체소송 근거 마련 등

ㅇ 또한 전자상거래법 전면개정을 추진하여 검색결과 및 검색순위의 투명성·공정성을 확보함으로써 소비자 선택권을 보호하고자 함

ㅇ 아울러 온라인 플랫폼 분야의 경쟁제한행위에 대한 심사지침을 제정하여 공정위 법집행 기준에 대한 업계의 예측가능성을 제고하고 법위반 행위를 예방하려는 노력도 병행하고 있음

ㅇ 주요 법안들이 국회 논의를 거쳐 입법성과로 이어질 수 있도록 오늘 참석해주신 전문가 분들의 지속적인 관심도 부탁드림

□ 한편, 발제자의 거버넌스 제안과 같이, AI분야의 종합적인 법제 정비를 위해서는 관계부처 간 협업이 중요하다고 봄

ㅇ 이러한 측면에서 지난 12월 관계부처 합동으로 발표된 「인공지능 법제도 규제정비 로드맵」이 의미있는 방향을 제시하고 있다고 생각함

ㅇ 일례로 AI 산업 진흥을 주관하는 과기부와 디지털 시장의 불공정거래 규율을 주관하는 공정위가 공동으로 (가칭)알고리즘 공개 및 설명가이드라인 제정을 추진할 계획임

- 플랫폼 알고리즘 운영의 투명성·공정성을 확보하면서도 기업의 영업비밀 공개 우려를 최소화할 수 있는 방안을 마련하기 위해 과기부와 적극적으로 협업하여 성공적인 부처협력 사례를 만들어가겠음 □

개인정보 보호위원회 토론문 :
AI 서비스 관련 개인정보보호 대응방안

이한샘 과장 | 데이터안전정책과

1. 추진 배경

○ 지능정보사회에서 **인공지능(AI)** 기술이 적용된 다양한 서비스의 개발·확산으로 **사생활 침해 등 개인정보 침해위험** 우려가 제기되어,

○ 이에 개인정보위는 AI 서비스의 개발·운영 시 개인정보의 안전한 처리·보호를 위한 **수칙 개발 및 정보주체 권리 강화**를 위한 입법을 추진하고 있습니다.

2. 「AI 환경의 개인정보보호 수칙」 마련 중

○ 개인정보위는 **「AI 환경의 개인정보보호 수칙」**을 마련 중에 있으며, 검토 중인 **주요 원칙 및 실천수칙**(예시)은 다음과 같습니다.

<주요 원칙 및 실천수칙(예시)>

- **(적법성)** 이용자가 개인정보 수집·이용 목적 등을 명확히 인지하도록 사전동의
- **(안전성)** 개인정보의 비식별처리 활용 및 암호화, 유·노출 방지 등 안전조치
- **(투명성)** 개인정보의 활용 범위 및 보유기간, AI 서비스 작동흐름 등 공개

○ AI 개발자 및 서비스 운영자가 「개인정보 보호법」 등에 따라 지켜야 할 **실천수칙**, AI 서비스 이용자 **안내사항**, **참고사례** 등을 수록할 예정이며,

○ 향후 전문가 및 이해관계자 등 **의견수렴**을 거쳐 발표할 예정입니다.

3. 자동화된 의사결정 대응권 도입 추진

○ AI 관련 **‘자동화 의사결정에 대한 배제 등의 권리 도입*’**에 대한 「개인정보 보호법」 개정을 추진하고 있습니다.(입법예고 1.6.~2.16.)

* 자동화 의사결정 등에 대한 거부, 이의제기, 설명요구권 등 도입(개정안 제37조의2)

자료

EUROPEAN
COMMISSION
(유럽연합집행위원회)

Brussels, 2020년 2월 19일
COM(2020) 65 final

인공지능 백서
- 수월성과 신뢰를 위한 유럽의 접근방법

번역제공 : 한국지능정보사회진흥원(NIA)

인공지능(Artificial Intelligence: AI)이 급속하게 발전하고 있다. AI는 의료 서비스를 개선하고(예: 정확한 진단과 질병의 예방을 향상), 영농 효율성을 높이고, 기후 변화의 경감과 적응에 기여하고, 예측 유지보수를 통해 생산 시스템의 효율성을 향상시키고, 유럽인들의 보안을 제고하고, 우리가 현재는 상상만 할 수 있는 다른 많은 방법으로 우리의 삶을 변화시킬 것이다. 이와 동시에 AI에는 몇 가지 잠재적인 위험이 수반된다. 여기에는 불투명한 의사결정, 성별 기반의 또는 기타 유형의 차별, 우리의 사생활 침해, 범죄 목적에 사용되는 것 등이 포함된다.

격심한 글로벌 경쟁 환경에 대비해서, 2018년에 발표된 유럽의 AI 전략[1]을 기반으로 하는 건실한 유럽의 접근방법이 필요하다. AI가 제공하는 기회와 과제를 해결하기 위해 EU는 AI의 개발과 배치를 촉진할 수 있도록 유럽의 가치를 바탕으로 하나와 같이 행동을 취하고, 자신들만의 방법을 정의해야 한다.

EU는 과학적인 혁신을 가능하게 하고, EU의 기술 리더십을 보존하고, 새로운 과학기술을 모든 유럽인들이 마음대로 이용할 수 있도록 하고, 이를 통해서 유럽인들의 권리를 존중하면서 이들의 삶을 개선하는데 전념하고 있다.

본 위원회 의장인 Ursula von der Leyen은 자신의 정치 '가이드라인[2]'에서 AI의 인간적/윤리적 파급효과에 대해 하나의 통합적인 유럽의 접근방법, 그리고 혁신을 위한 빅데이터의 활용에 대한 자신의 생각을 발표하였다.

이에 따라 본 위원회는 AI의 도입 촉진, 그리고 이 신기술의 특정한 활용에 관련된 위험의 해결이라는 두 개의 목적과 함께, 규제 및 투자 지향적인 접근방법을 지지한다. 본 백서의 목표는 이러한 목적을 달성하는 방법에 대한 정책적 대안을 제시하는 것이다. 여기에서 군사 목적의 AI 사용 및 개발은 다루지 않는다. 본 위원회는 회원국, 기타 유럽의 기관들, 산업/사회적 파트너/민간단체/연구자/일반 대중/모든 관심을 가진 당사자들을 포함한 모든 이해관계자에게 아래에서 설명할 대안에 반응을 보이고, 이 분야에 대한 위원회의 미래 의사결정에 기여하기를 요청하는 바이다.

1. 서론

디지털 기술이 사람들의 삶의 모든 면에서 핵심적인 부분이 되므로, 사람들은 이것을 신뢰할 수 있어야 한다. 또한 신뢰는 수용의 전제 조건이다. 유럽은 가치와 법규에 대해 강력한 애착을 가지고 있고, 항공 분야에서부터 에너지, 자동차, 의료 장비에 이르기까지

1) AI for Europe, COM/2018/237 final
2) https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

안전하고, 신뢰성 있고, 정교한 제품/서비스를 구축할 수 있는 입증된 역량을 가지고 있다는 점을 감안하면, 이것은 유럽에게는 큰 기회이다.

유럽의 현재 및 미래의 지속가능한 경제 성장과 사회 복지는 데이터에 의해 창출된 가치를 바탕으로 하고 있다. AI는 데이터 경제의 가장 중요한 애플리케이션 중의 하나이다. 오늘날 대부분의 데이터는 소비자와 관련이 있고, 중앙의 클라우드 기반 인프라에 저장되고 처리된다. 이에 비해서 훨씬 더 풍부한 미래의 데이터의 대부분은 산업, 기업, 공공 부문에서 비롯될 것이고, 다양한 종류의 시스템, 그중에서도 특히 네트워크의 종단부(edge)에서 작동되는 컴퓨팅 기기들에 저장될 것이다. 이것은 유럽에 새로운 기회를 열어준다. 왜냐하면 유럽은 디지털화된 산업과 B2B 애플리케이션에 강점을 가지고 있지만, 소비자 플랫폼에는 비교적 약하기 때문이다.

간단히 말하면, AI는 데이터, 알고리즘, 컴퓨팅 연산능력 등을 결합한 기술의 집합이다. 컴퓨팅의 발전과 데이터의 가용성 증가는 현재의 AI 급증의 핵심 동인이다. 유럽은 유럽 데이터 전략[3]에 명시되어 있는 바와 같이 **데이터 경제 및 그 응용에서 혁신의 글로벌 리더가 되겠다**는 기본적인 가치를 바탕으로, 기술적/산업적인 강점을 고품질의 디지털 인프라 및 규제 프레임워크와 결합할 수 있다. 이를 기반으로 유럽은 전 유럽 사회 및 경제에 다음과 같은 기술의 혜택을 가져다주는 AI 생태계를 개발할 수 있다.

- **시민들**은 의료 서비스의 향상, 가전 기기들의 고장 감소, 교통 시스템의 안전성 및 청정성 증가, 공공 서비스의 향상 등의 새로운 효과를 실현할 수 있다
- **기업들**은 유럽이 강점을 가진 영역(기계, 운송, 사이버 보안, 영농, 녹색 및 순환(green & circular) 경제, 의료 서비스, 패션과 관광과 같은 고부가가치 산업)에서 새로운 세대의 제품/서비스를 개발할 수 있다.
- 서비스(교통, 교육, 에너지, 폐기물 관리)의 제공 비용 감소, 제품의 지속가능성[4] 향상, 시민들의 권리와 자유를 존중할 수 있는 적절한 안전 조치와 함께 법집행 기관들이 시민들의 안전을 보장하기 위한 적절한 도구 구비[5] 등을 통해서 **공익 서비스**를 제공할 수 있다.

AI가 우리 사회에 미칠 수 있는 주요한 파급효과, 그리고 신뢰를 구축해야 할 필요성을 감안하면, 유럽의 AI가 인간의 존엄성과 프라이버시 보호 등과 같은 우리의 가치와 기본적

3) COM(2020) 66 final.
4) AI와 디지털화는 유럽의 그린딜(Green Deal) 목표의 핵심 동인이다. 그러나 ICT 산업의 현행 환경 발자국은 전세계 총 배출량의 2% 이상으로 예측되고 있다. 본 백서에 동봉된 유럽 디지털 전략에서는 디지털에 대한 그린 변환 지표를 제안하고 있다.
5) AI 도구는 EU 시민들을 범죄와 테러행위로부터 더 잘 보호할 수 있는 기회를 제공할 수 있다. 이러한 도구들은 예를 들면 온라인 테러범들의 선전을 식별하고, 위험한 제품의 판매에서 의심스러운 거래를 발견하고, 숨겨진 위험한 물건이나 불법 물질/제품을 식별하고, 비상 상황의 시민들에게 지원을 제공하고 최초의 대응자들을 가이드하는 것을 도울 수 있다.

인 권리에 기반을 두도록 하는 것이 매우 중요하다.

그뿐만 아니라, AI 시스템의 영향은 개인의 관점에서뿐만 아니라 사회 전체적인 관점에서 고려되어야한다. AI 시스템의 사용은 '지속가능한 발전 목표(Sustainable Development Goals)'를 달성하고, 민주적인 프로세스와 사회적인 권리를 지원하는데 중요한 역할을 수행할 수 있다. '유럽의 그린딜(European Green Deal)'[6]을 통한 최근의 제안으로 유럽은 기후와 환경 관련 과제의 해결을 선도하고 있다. AI와 같은 디지털 기술은 '그린딜'의 목표를 달성하기 위한 핵심적인 동인이다. AI의 중요성이 증가하고 있다는 것을 감안하면, AI 시스템이 환경에 미치는 영향은 시스템의 수명주기와 공급망 전반에 걸쳐서(예: 알고리즘의 훈련과 데이터 저장을 위한 자원의 사용에 상관없이), 적절하게 고려되어야 한다.

충분한 규모를 달성하고, 단일 시장의 분열을 방지하기 위해서는 AI에 대한 하나의 공통적인 유럽의 접근방법이 필요하다. 국가별로 이니셔티브들을 도입하는 것은 법적인 확실성을 위태롭게 하고, 시민들의 신뢰를 약화시키고, 역동적인 유럽 산업의 등장을 저해할 수 있는 위험이 있다.
본 백서에서는 유럽에서 EU 시민들의 가치와 권리를 완전하게 존중하면서, AI가 신뢰성 있고 안전하게 개발될 수 있도록 하는 정책 대안을 제시한다. 본 백서의 주요한 구성요소는 다음과 같다.

- 유럽, 국가 및 지역 수준의 노력들을 연계하는 조치들을 제시하는 정책 프레임워크. 이 프레임워크의 목표는 민관 제휴를 통해서 연구와 혁신에서 출발하여 전체 가치 가치에 걸쳐서 **'수월성의 생태계'**를 달성하기 위한 자원들을 동원하고, 중소기업을 포함하여 AI 기반의 솔루션의 도입을 촉진하기 위한 적절한 인센티브를 창출하는 것이다.
- 고유한 **'신뢰의 생태계'**를 창출할 AI에 대한 유럽의 미래 규제 프레임워크의 핵심 요소. 이를 위해서는 기본적인 권리와 소비자들의 권리, 그중에서도 특히 EU에서 운영되고 있는 고위험 AI 시스템[7]에 대한 권리를 보호하기 위한 규칙들을 포함해서 EU의 규칙들을 준수하도록 해야 한다. 신뢰의 생태계를 구축하는 것은 그 자체로 하나의 정책 목표이고, 시민들에게 AI 애플리케이션의 수용에 대한 자신감을 심어주고, 기업과 공공 기관들에게 AI를 이용한 혁신에 대한 법적인 확실성을 제공해야 한다. 본 위원회는 '인간 중심적인 AI에서의 신뢰 구축에 관한 공보(Communication on Building Trust in Human-Centric AI)'[8]를 기반으로 하고 있는 인간 중심적인 접근방법을 강력히 지지하는 바이다. 'AI 고위전문가그룹(High-Level Expert Group on AI)'이 작성한 '윤리 가이드라인(Ethics

6) COM(2019) 640 final.
7) AI가 범죄 목적으로 오용되는 것을 방지하고 여기에 대응하기 위한 추가적인 조치들이 필요할 수 있지만, 이것은 본 백서의 범위에 포함되지 않는다.
8) COM(2019) 168.

Guidelines)'의 시범 사업 단계에서 입수한 의견들을 감안할 것이다.

본 백서에 동봉된 유럽의 데이터 전략의 목표는 전세계에서 유럽이 가장 매력적이고, 안전하고, 역동적인 데이터 애자일 경제(data-agile economy)가 되는 것이 가능하도록(데이터를 통해서 유럽이 의사결정을 향상시키고, 모든 시민들의 삶을 향상시킬 수 있도록) 하는 것이다. 이 전략은 목표 달성에 필요한 민간과 공공의 투자 동원을 포함하여 몇 가지 정책적인 조치들을 제시하고 있다. 마지막으로 AI, 사물인터넷(IoT), 기타 디지털 기술들이 안전과 책임에 관한 법률에 미치는 파급효과는 본 백서에 동봉된 '위원회 보고서(Commission Report)'에 분석되어 있다.

2. 산업 및 전문 시장의 강점 활용

유럽은 AI 기술의 사용자로서뿐만 아니라 생산자와 창작자로서 AI의 잠재력으로부터 효과를 실현하기에 좋은 위치에 있다. 유럽은 우수한 연구센터, 혁신적인 신생 벤처기업, 로봇 공학에서 세계 최고의 위치, 자동차에서부터 의료 서비스, 에너지, 금융 서비스, 농업 등에 이르는 경쟁력 있는 제조/서비스 산업을 보유하고 있다. 유럽은 AI의 작동에 필수적인 강력한 컴퓨팅 인프라(예: 고성능 컴퓨터)를 구축하고 있다. 또한 유럽은 막대한 양의 공공 및 산업 데이터를 보유하고 있는데, 그 잠재력을 현재 다 사용하지 못하고 있다. 유럽은 AI를 더욱 발전시키는데 필수적인 저전력을 소비하는 안전하고 보안이 확보된 디지털 시스템 분야에서 널리 인정받고 있는 산업적인 강점을 가지고 있다.

EU가 차세대 기술과 인프라, 그리고 데이터 해독능력과 같은 디지털 역량에 투자할 수 있는 역량을 갖추면, 데이터 경제를 위한 핵심적인 동인 기술과 인프라에 대한 유럽의 기술적 주권이 강화될 것이다. 인프라는 신뢰할 수 있는 AI(예: 유럽의 가치와 규칙을 기반으로 하는 AI)를 가능하게 하는 유럽의 데이터 풀의 창출을 지원해야 한다.

유럽은 자신들의 강점을 활용하여, 특정한 하드웨어 제조 산업으로부터 소프트웨어 산업과 서비스 산업에 이르기까지 생태계에서 그리고 가치체인 전반에 걸쳐서 그 지위를 확대해야 한다. 이것은 어느 정도 이미 일어나고 있다. 유럽은 모든 산업 및 전문 서비스 로봇(예: 정밀 농업, 보안, 의료, 물류)의 1/4 이상을 생산하고 있고, 기업과 조직을 위한 소프트웨어 애플리케이션(ERP, 설계 및 엔지니어링 소프트웨어 등과 같은 B2B 애플리케이션), 그리고 전자정부와 "지능형 기업(intelligent enterprise)"을 지원하는 애플리케이션을 개발하고 사용하는데 중요한 역할을 수행하고 있다.

유럽은 AI를 제조에 효율적으로 사용하는 것을 선도하고 있다. 일류 제조업체의 1/2 이상이 제조 운영에 적어도 한 건 이상의 AI를 구현하고 있다.[9)]

9) 그다음으로 일본(30%), 미국(28%)의 순.(자료원: CapGemini(2019))

유럽이 연구 분야에서 강력한 위치를 차지하고 있는 이유 중의 하나는 회원국들의 조치를 한데 모으고, 중복을 방지하고, 공공 및 민간 투자를 활용하는데 그 중요성이 입증된 EU의 재정 지원 사업 때문이다. 지난 3년 동안 AI 연구 및 혁신을 위한 EU의 재정 지원은 15억 유로로 증가하였고, 이것은 전기에 비해 70%나 증가한 수치이다.

그러나 유럽의 연구 혁신 투자는 여전히 세계 다른 지역의 공공 및 민간 투자액의 일부에 불과하다. 2016년에 AI에 대한 유럽의 투자액은 32억 유로였는데, 북미의 투자액은 121억 유로, 아시아는 65억 유로에 달했다.[10] 이에 대응하여 유럽은 투자 수준을 크게 증가시킬 필요가 있다. 회원국들과 함께 개발한 'AI에 대한 통합 계획(Coordinated plan on AI)'[11]은 유럽에서 AI에 대한 긴밀한 협력 관계를 구축하고, AI 가치 체인에서 투자를 극대화할 수 있는 시너지 효과를 창출하기 위한 좋은 출발점인 것이 입증되고 있다.

3. 다가올 기회의 포착: 다음의 데이터 물결(Wave)

소비자 애플리케이션과 온라인 플랫폼 분야에서 현재 유럽의 위치는 취약하고, 이로 인해 데이터의 접근 측면에서 경쟁적인 열세에 있지만, 가치의 큰 전환과 산업 전반에 걸친 데이터의 재사용이 진행되고 있다. 세계에서 생산되고 있는 데이터의 양은 급속하게 증가하여, 2018년의 33 제타바이트에서 2025년에는 175 제타바이트로 증가할 것으로 예측되고 있다.[12] 새로운 데이터 물결은 유럽이 데이터 애자일 경제에서 자리를 잡고, 이 분야에서 세계의 리더가 될 수 있는 기회를 가져다준다. 그뿐만 아니라, 데이터가 저장되고 처리되는 방법은 향후 5년 동안에 크게 변경될 것이다. 오늘날 클라우드에서 이루어지고 있는 데이터 처리 및 분석의 80%는 데이터센터와 중앙집중식 컴퓨팅 시설에서 이루어지고 있고, 나머지 20%는 자동차, 가전 제품, 제조 로봇 등과 같은 연결된 스마트한 객체, 그리고 사용자에게 근접한 컴퓨팅 설비("에지 컴퓨팅")에서 이루어지고 있다. 2025년까지 이러한 비중은 크게 변경될 것이다.[13]

유럽은 차세대 AI 전용 프로세서의 핵심인 저전력 전자장치 분야의 글로벌 리더이다. 이 시장은 현재 비EU 국가들이 지배하고 있다. 에지 및 차세대 고성능 컴퓨팅을 위한 저전력 컴퓨팅 시스템 개발에 초점을 맞추고 있는 '유럽 프로세서 이니셔티브(European Processor Initiative)'와 같은 이니셔티브, 그리고 2012년에 시작 예정인 '핵심 디지털 기술 공동 프로젝트(Key Digital Technology Joint Undertaking)' 등의 도움으로 이러한 현상은 변화하고 있다. 또한 유럽은 산업 프로세스(인더스트리 4.0)와 운송 모드를 자동

10) AI와 자동화 시대에 유럽의 10대 과제(10 imperatives for Europe in the age of AI and automation). McKinsey(2017).
11) COM(2018) 795.
12) IDC (2019).
13) Gartner (2017).

화하는데 매우 적합한 뉴로모픽 솔루션[14]을 선도하고 있다. 이것들은 에너지 효율성을 수십 배 향상시킬 수 있다.

최근 양자 컴퓨팅의 발전은 처리 용량을 기하급수적으로 증가시킬 것이다.[15] 양자 컴퓨팅에 대한 학술적인 강점, 그리고 양자 컴퓨팅을 위한 양자 시뮬레이터와 프로그래밍 환경에서 유럽 산업의 강력한 위치 덕분에 유럽은 이 기술의 선두에 설 수 있다. 양자 테스트 및 실험 시설의 가용성을 높이기 위한 목적의 유럽의 이니셔티브들은 이러한 새로운 양자 솔루션을 몇 가지 산업 및 학문 분야에 적용하는 것을 도울 것이다.

이와 동시에, 유럽은 과학 분야의 수월성을 바탕으로 AI 알고리즘 분야에서 계속해서 발전을 선도해 나갈 것이다. 기계학습과 딥러닝(제한적인 상호운영성, 모델을 훈련시키고 상관관계를 통해 학습하기 위해 필요한 많은 양의 데이터 등의 특성을 가진), 그리고 상징적(symbolic) 접근방법(규칙들이 사람의 개입을 통해 생성) 등과 같이 현재 별도로 작업이 진행되고 있는 분야들을 서로 연결할 교량을 구축할 필요가 있다. 상징적 추론과 딥신경망의 결합은 우리가 AI의 결과를 설명하는 것을 향상시키는 것을 도울 수 있다.

4. 수월성의 생태계

EU의 경제 및 공공 행정 전반에 걸쳐서 AI를 개발하고 수용하는 것을 지원할 수 있는 수월성의 생태계를 구축하기 위해서는 다음과 같이 여러 가지 수준에서 조치를 취할 필요가 있다.

A. 회원국들과의 공조

2018년 4월에 채택된 AI에 대한 전략[16]의 이행 차원에서 2018년 12월 본 위원회는 회원국들과 함께 작성한, 유럽에서 AI의 개발과 사용을 촉진하기 위한 '통합 계획'[17]을 발표하였다.

이 계획에서는 연구, 투자, 시장의 수용, 스킬 및 재능, 데이터, 국제적인 협조 등과 같은 핵심 영역에서 회원국들과 위원회 간에 보다 긴밀하고 효율적인 협력을 위한 약 70개의 공동 조치를 제안하고 있다. 이 계획은 정기적인 모니터링과 검토 하에 2027년까지 시행될 예정이다.

---

14) 뉴로모픽(neuromorphic) 솔루션은 신경 체계에 존재하는 신경 생물학적 아키텍처를 흉내내는 모든 초대규모 직접회로 시스템(very large-scale system of integrated circuits)을 의미한다.
15) 양자(quantum)컴퓨터는 오늘날의 최고 성능 컴퓨터에 비해 1초 이하에 몇 배나 많은 데이터 세트를 처리할 수 있는 용량을 가질 수 있다.
16) Artificial Intelligence for Europe, COM(2018) 237.
17) Coordinated Plan on Artificial Intelligence, COM(2018) 795.

목표는 연구, 혁신, 배치에 대한 투자의 영향을 극대화하고, 국가별 AI 전략들을 평가하고, 회원국들과 AI '통합 계획'을 확장하는 것이다.

- 조치 1: 본 위원회는 백서에 대한 공개 협의의 결과를 감안하여, 회원국들에게 2020년 말에 채택될 '통합 계획'의 수정을 제안할 것이다.

AI에 대한 EU 차원의 재정 지원은 한 회원국이 달성할 수 있는 것 이상의 조치가 필요한 영역에 대한 투자를 유치하고 결집해야 한다. 목적은 향후 10년간에 걸쳐서 EU에서 AI에 대해 매년 200억 유로[18] 이상의 투자를 유치하는 것이다. 민간과 공공의 투자를 촉진하기 위해서 EU는 '디지털유럽사업(Digital Europe Programme)', 'Horizon Europe', '유럽구조및투자기금(European Structural and Investment Funds)' 등으로부터의 자원을 활용하여, 저개발 지역과 지방의 니즈를 충족시킬 것이다.
또한 '통합 계획'은 AI의 핵심 원칙으로 사회적/환경적 복지를 포함시킬 수 있다. AI 시스템은 기후 변화와 환경 악화를 포함하여 가장 시급한 우려사항들의 해결을 도울 것을 약속하고 있다. 또한 이것이 환경친화적인 방법으로 이루어지는 것이 중요하다. AI는 자원의 사용과 에너지의 소비를 냉정하게 검토하고, 환경에 긍정적인 대안을 선택하도록 훈련받아야 하고, 또한 그렇게 할 수 있다. 본 위원회는 회원국들과 함께 이러한 일을 할 수 있는 AI 솔루션을 장려하고 촉진하기 위한 대안들을 고려할 것이다.

B. 연구 및 혁신 커뮤니티의 노력을 집중화

유럽은 현재와 같이 역량 개발을 위한 센터들이 분열되어서 어느 하나도 세계적인 선도 기관들과 경쟁하는데 필요한 규모를 갖추지 못한 상황을 계속 유지할만한 여유가 없다. 유럽에 있는 여러 개의 AI 연구센터들 간에 더 많은 시너지와 네트워크를 창출하고, 수월성을 향상시키고, 최고의 연구자들을 유치하고, 최고의 기술을 개발할 수 있도록 이들의 노력을 서로 연계시키는 것이 시급하다. 유럽은 이러한 노력들을 조정하고, AI 분야에서 전 세계적으로 가장 우수한 참조 사례 되고, 투자와 최고의 인재들을 유치할 수 있는, 등대의 역할을 수행할 연구/혁신/전문성 센터를 필요로 한다.

이러한 센터와 네트워크들은 산업, 의료, 교통, 금융, 농식품 가치 체인, 에너지/환경, 임업, 지구 관측 및 우주 등과 같이, 유럽이 글로벌 챔피언이 될 수 있는 잠재력을 가진 부문에 집중해야 한다. 이러한 모든 부문에서 글로벌 리더십을 위한 경주가 진행되고 있고, 유럽은 상당한 잠재력, 지식, 전문성 등을 제공하고 있다.[19] 이와 동등하게 중요한 것은 참신한 AI 애플리케이션의 개발 및 배치를 지원하는 테스트 및 실험 사

---

18) COM(2018) 237.
19) 미래의 '유럽국방기금 및 영구구조화협력(European Defence Fund and Permanent Structured Cooperation: PESCO)' 또한 AI에 대한 연구 개발을 위한 기회를 제공할 것이다. 이러한 프로젝트들은 AI에 초점을 맞추고 있는 보다 광범위한 EU의 민간사업들과 동기화되어야 한다.

이트를 창출하는 것이다.

- 조치 2: 본 위원회는 유럽, 국가 및 민간 투자를 결합할 수 있는 수월성 및 테스트 센터의 설립을 촉진할 것이다. 본 위원회는 '디지털유럽사업'을 통해서 전세계에서 참고할 수 있는 테스트 센터를 유럽에 설립하는 것을 지원하기 위한 상당한 양의 야심찬 투자를 제안하였고, 필요한 경우, 2021년~2027년까지 '다년도재무프레임워크(Multiannual Financial Framework)'의 일환으로 'Horizon Europe'의 연구 및 혁신 조치로 보완하였다.

C. 스킬

AI에 대한 유럽의 접근방법은 역량 부족을 메우기 위한 스킬에 초점을 맞출 필요가 있다.[20] 본 위원회는 조만간 '스킬 과제(Skills Agenda)'의 강화안을 발표할 예정인데, 이것의 목표는 유럽에서 모든 사람들이 EU 경제의 그린화와 디지털 트랜스포메이션으로부터 혜택을 볼 수 있도록 하는 것이다. 또한 이니셔티브에는 관련 규칙들을 효과적이고 효율적으로 시행하기 위해 자신들의 AI 스킬을 향상시키려는 부문별 규제기관에 대한 지원이 포함될 수 있다. 갱신된 '디지털교육실행계획(Digital Education Action Plan)'은 교육 및 훈련 시스템을 개선하고, 이러한 시스템들이 디지털 시대에 적합하도록 만드는 것을 목표로 하여, 학습과 예측 분석과 같은 데이터 및 AI 기반의 기술들을 보다 잘 사용하도록 도울 것이다. 또한 이 계획은 AI에 의해 점차 더 많은 영향을 받게 될 정보에 입각한 의사결정에 대해 시민들이 준비를 갖출 수 있도록 모든 수준의 교육에서 AI에 대한 인식을 제고할 것이다.

AI로 일하는데 필요한 스킬을 개발하고 AI가 주도하는 트랜스포메이션에 적합하도록 업스킬(upskill) 시키는 것은 회원국들과 개발할 AI에 대한 수정 '통합 계획'의 최우선 순위일 것이다. 여기에는 윤리 가이드라인의 평가 목록을 AI 개발자들을 위한 "교과과정"으로 변환하는 것이 포함될 수 있는데, 이러한 교육과정은 훈련 기관들을 위한 자원으로 이용 가능하게 될 것이다. 이 분야에서 교육을 받고 채용된 여성의 수를 늘리는데 특별한 노력을 기울여야 한다.

그뿐만 아니라 등대의 역할을 수행할 유럽의 AI 연구혁신 센터는 이 센터가 제공할 수 있는 가능성 때문에 전 세계로부터 재능 있는 인력들을 유치할 것이다. 또한 이 센터는 유럽에 뿌리를 두고 유럽 전반에 걸쳐서 성장하고 있는 우수한 스킬을 개발하고 확산할 것이다.

- 조치 3: '디지털유럽사업(Digital Europe Programme)'의 스킬 고도화 과제를 통해 최고의 교수와 과학자들을 유치하고, 세계 최고의 AI 분야 석사 과정을 제공하는 선도적인 대학 및 고등교육 기관들의 네트워크를 수립하고 지원한다.

---

20) https://ec.europa.eu/jrc/en/publication/academic-offer-and-demand-advanced -profiles-eu

노동자들과 고용인들은 업스킬뿐만 아니라 직장에서의 AI 시스템의 설계와 사용에 의해 직접적인 영향을 받고 있다. 사회적인 파트너들의 참여는 직장에서 AI에 대한 인간 중심의 접근방법을 확보하는데 핵심적인 요인일 것이다.

D. 중소기업에 초점

중소기업들이 AI를 접근하고 사용할 수 있도록 해 주는 것 또한 중요할 것이다. 이를 위해서 '디지털혁신허브(Digital Innovation Hub)'[21]와 AI 주문형(AI-on-demand) 플랫폼[22]은 더욱 강화되어야 하고, 중소기업들 간의 협동을 촉진해야 한다. '디지털유럽 사업'은 이것을 달성하는데 중요할 것이다. 모든 '디지털혁신허브'들이 중소기업들이 AI를 이해하고 도입하는 것을 지원해야 하지만, 회원국별로 적어도 하나 이상의 혁신 허브가 AI에 대한 높은 수준의 전문성을 보유하도록 하는 것이 중요할 것이다.

중소기업들과 신생 벤처기업들이 AI를 이용하여 자신들의 프로세스를 수정하거나 혁신하기 위해서는 자금이 필요할 것이다. 본 위원회는 AI와 블록체인에 대해 예정되어있는 1억 유로의 시범 투자 기금을 기반으로, 'InvestEU'[23] 하에서 AI에 대한 지원 자금을 더 늘릴 계획이다. 'InvestEU' 자금의 사용 자격이 있는 영역에 AI가 명시적으로 언급되어 있다.

- 조치 4 : 본 위원회는 회원국들과 공조해 회원국별로 적어도 하나 이상의 디지털 혁신 허브가 AI에 대한 높은 수준의 전문성을 보유하도록 할 것이다. '디지털 혁신 허브'는 '디지털유럽사업'을 통해 지원받을 수 있다.
- 본 위원회와 '유럽투자기금(European Investment Fund)'은 2020년 1/4분기에 1억 불 상당의 시범 체계에 착수하여, AI 분야의 혁신적인 개발에 자금을 제공할 것이다. MFF와의 최종 합의가 남아 있지만, 본 위원회는 'InvestEU'를 통해서 2021년부터 금액을 크게 증가시키려고 한다.

E. 민간 부분과의 제휴

민간 부분이 연구 및 혁신 과제를 설정하는데 충분히 참여하고, 필요한 수준의 공동 투자를 제공하도록 하는 것 또한 필수적이다. 이를 위해서는 광범위한 민관 파트너십을 수립하고, 기업의 고위 경영진의 의지를 확보하는 것이 필요하다.

- 조치 5 : 본 위원회는 'Horizon Europe'의 관점에서 AI, 데이터 및 로봇 공학 분야에

---

21) ec.europe.eu/digital-single-market/en/news/digital-innovation-hubs-helping-companies-across-economy-make-most- digital-opportunities.
22) www.Ai4eu.eu.
23) Europe.eu/investeu.

새로운 민관 파트너십을 수립하여, 노력들을 결합하고, AI에 대한 연구와 혁신을 조정하고, 'Horizon Europe'에서 또 다른 민관 파트너십으로 협동하고, 위에서 언급한 테스트 시설과 '디지털 혁신 허브'들과 공조할 것이다.

F. 공공 부문의 AI 도입 촉진

공공기관, 병원, 공익(utility) 및 운송 서비스, 금융 감독 기관, 공공의 이익에 관련된 기타 영역들이 AI에 의존하는 제품과 서비스를 자신들의 활동에 신속하게 배치하기 시작하는 것이 매우 중요하다. 대규모 배치가 가능할 정도로 기술이 성숙한 의료 서비스와 운송 분야에 특별히 초점을 맞출 것이다.

- 조치 6 : 본 위원회는 개발, 실험 및 도입을 촉진하기 위한 실행계획을 작성하기 위해 개방적이고 투명한 부문별 대화에 착수할 것이다(의료 서비스, 지방 행정 및 공공 서비스 운영 기관들에 우선순위 부여). 부문별 대화를 활용하여 'AI도입사업(Adopt AI programme)'을 준비할 예정이다. 이 사업은 AI 시스템의 공공 조달을 지원하고, 공공 조달 프로세스 그 자체를 변혁하는 것을 도울 것이다.

G. 데이터와 컴퓨팅 인프라에 대한 접근 보장

본 백서에서 제시하는 조치들은 유럽의 데이터 전략에서 제시한 계획을 보완한다. 데이터에 대한 접근과 관리를 향상시키는 것은 필수적인 일이다. 데이터가 없다면, AI와 기타 디지털 애플리케이션의 개발은 가능하지 않다. 앞으로 생성될 막대한 양의 새로운 데이터는 유럽이 데이터와 AI 트랜스포메이션의 선두에 설 수 있는 기회가 된다. 책임성 있는 데이터 관리 프랙티스와 데이터의 FAIR 원칙[24] 준수를 촉진하는 것은 신뢰를 구축하는데 기여하고, 데이터가 재사용되도록 할 것이다. 이와 동등하게 중요한 것은 핵심 컴퓨팅 기술과 인프라에 대한 투자이다.

본 위원회는 에지 컴퓨팅 및 AI, 데이터, 클라우드 인프라 등을 포함하여, 고성능 및 양자 컴퓨팅 분야에 '디지털유럽사업'을 통해서 40억 유로 이상을 지원하기로 제안하였다. 유럽의 데이터 전략은 이러한 우선 지원 부문을 추가로 개발하고 있다.

H. 국제적 측면

유럽은 공유 가치를 중심으로 제휴를 구축하고, AI의 윤리적인 사용을 촉진하는 데 있어서 글로벌 리더십을 발휘하기에 좋은 위치에 있다. AI에 관한 EU의 작업은 이미 국제적인 논의에 영향을 미치고 있다. '고위 전문가 그룹(High-Level Expert Group)'이 윤리 가이드라인을 개발할 때, 몇 개의 비EU 조직들과 여러 명의 정부 참관인들을 참여시켰다. 이와 동시에 EU는 OECD의 AI에 대한 윤리적 원칙[25]을 개발하는 데 긴밀하게 참여

---

24) 2018년의 FAIR 데이터에 관한 위원회 전문가 그룹의 최종 보고서와 실행 계획에서 언급하고 있는 바와 같이, 발견 가능하고(Findable), 접근 가능하고(Accessible), 상호운영성이 있고(Interoperable), 재사용 가능함(Reusable). https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

하였다. G20은 그 이후 2019년 6월에 '무역과 디지털 경제에 대한 장관 선언문(Ministerial Statement on Trade and Digital Economy)'에서 이 원칙들을 지지하였다.

이와 동시에 EU는 AI에 대한 중요한 작업들이 기타 다자간 포럼에서 진행되고 있다는 사실을 인지하고 있다. 여기에는 Council of Europe, United Nations Educational Scientific and Cultural Organization (UNESCO), Organisation for Economic Co-operation and Development's (OECD), World Trade Organisation, International Telecommunications Union (ITU) 등이 포함된다. UN에서 EU는 AI에 대한 권고사항을 포함하여 '디지털 협력에 관한 고위 패널(High-Level Panel on Digital Cooperation)' 보고서의 후속 작업에 참여하고 있다.

EU는 EU의 규칙과 가치(예: 규제의 상향 수렴을 지원, 데이터를 포함한 핵심 자원의 접근, 공평한 경쟁의 장 창출)를 기반으로 하고 있은 접근방법을 바탕으로 AI에 대해 비슷한 생각을 가지고 있는 국가들, 그리고 세계적인 국가들과 계속해서 협력해 나갈 것이다. 본 위원회는 데이터 흐름을 제한하는 제삼 국가들의 정책을 긴밀하게 모니터링하고, 양자간 무역 협상에서 그리고 세계무역기구(WTO)의 관점에서 조치를 통해 과도한 제한을 해결할 것이다. 본 위원회는 AI 관련 사항에 대한 국제 협력은 인간의 존엄성, 다원주의, 포용, 비차별, 프라이버시와 개인 데이터의 보호 등을 포함하여 기본적인 권리의 존중을 촉진하는 접근방법을 기반으로 해야 한다고 확신하고 있고[26], 이러한 가치를 전 세계에 전하기 위해 노력할 것이다.[27] 또한 AI의 책임성 있는 개발과 사용은 '지속 가능한 발전 목표'를 달성하고, 2040과제를 전전시키는 원동력이 될 수 있다.

5. 신뢰의 생태계: AI에 대한 규제 프레임워크

다른 모든 신기술과 마찬가지로 AI의 사용은 기회와 위험 두 가지 모두를 가져다준다. 시민들은 알고리즘 기반 의사결정의 정보 비대칭성을 직면할 때 자신들의 권리와 안전을 보호할 수 있는 힘을 잃을까봐 두려워하고 있고, 기업들은 법적인 불확실성에 대해 우려하고 있다. AI가 시민들의 안전을 보호하고, 자신들의 기본권을 향유할 수 있게 도울 수 있지만, 시민들은 AI가 의도하지 않은 효과를 내거나 악의적인 목적으로 사용될 수 있다는 사실을 걱정하고 있다. 이러한 우려는 해결되어야 한다. 그뿐만 아니라, 투자와 스킬 부족에 더하여 신뢰 부족은 AI의 광범위한 도입을 저해하는 주요한 요인이다.

이것이 바로 본 위원회가 2018년 4월 25일에 AI 전략[28]을 제시한 이유이다. 이 전략은

---

25) https://www.oecd.org/going-digital/ai/principles/

26) 본 위원회는 EU의 AI 가이드라인을 홍보하고, 공통의 원칙과 운영 결론을 도입하도록 하기 위해, Partnership Instrument를 통해서 생각이 같은 파트너들과의 협력을 촉진할 프로젝트에 250만 유로의 자금을 지원할 것이다.

27) President Von der Leyen, A Union that strives for more - My agenda for Europe, page 17.

EU 전반에 걸친 연구, 혁신, AI 역량에 대한 투자 증가와 함께 사회경제적인 측면을 다루고 있다. 위원회는 회원국들과 전략들을 서로 연계하기로 하고, 통합 계획[29]에 합의하였다. 또한 본 위원회는 '고위전문가그룹'을 수립하여, 2019년 4월에 '신뢰할 수 있는 AI에 대한 가이드라인'을 발간하였다.[30]

본 위원회는 '고위전문가그룹'의 '가이드라인'에서 식별된 다음과 같은 7개의 핵심적인 요구사항을 환영하는 공보(Communication)[31]를 발간하였다.

- 인간의 주체적 역량과 감독(Human agency and oversight)
- 기술적 견고함(robustness) 및 안전성
- 프라이버시 및 데이터 거버넌스
- 투명성
- 다양성, 비차별 및 공정성
- 사회적/환경적 복지
- 책임성

그뿐만 아니라, 이 '가이드라인'은 기업들의 실질적으로 사용할 수 있는 평가 목록을 담고 있다. 2019년 하반기 동안, 350여개 이상의 조직들이 이 평가 목록을 테스트해보고 피드백을 보내왔다. '고위전문가그룹'은 이러한 피드백을 반영하여 가이드라인을 수정 중에 있고, 2020년 6월까지 이 작업을 마무리할 예정이다. 피드백 프로세스의 최종 결과는 몇 가지 요구사항들은 이미 기존의 법규 체제에 반영되어 있지만, 투명성/추적성/사람의 감독 등에 관련된 요구사항들은 많은 경제 부분에서 현행 법규로는 커버되지 않고 있다는 것이다.

이러한 '고위전문가그룹'의 구속력 없는 가이드라인 위에, 그리고 의장의 정치적인 가이드라인과 일관성을 가지고 있는, 명확한 유럽의 규제 프레임워크는 소비자들과 기업들의 AI에 대한 신뢰를 구축할 것이고, 이에 따라 AI의 도입을 가속화할 것이다. 이러한 규제 프레임워크는 이 분야에서 유럽의 혁신 역량과 경쟁력을 제고하기 위한 기타 조치들과 일관성을 가지고 있어야 한다. 그 뿐만 아니라, 이것은 사회적/환경적/경제적으로 최적의 결과, 그리고 EU의 법규, 원칙, 가치 등의 준수가 이루어지도록 해야 한다. 이것은 특히 시민들의 권리가 가장 직접적으로 영향을 받는 영역(예: 법률의 집행과 재판에 AI를 적용하는 사례)에서 관련성이 높다.

AI를 개발하고 배치하는 사람들은 이미 기본권(예: 데이터 보호, 프라이버시, 비차별)에 관

28) COM(2018) 237.
29) COM(2018) 795.
30) https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
31) COM(2019) 168.

한 유럽의 법규, 그리고 소비자 보호, 제품 안전 및 책임에 관한 규칙의 적용을 받고 있다. 소비자들은 제품이나 서비스가 AI에 의존하고 있느냐의 여부에 상관없이 동일한 수준의 안전과 자신들의 권리에 대한 존중을 기대한다. 그러나 AI의 일부 특성(예: 불투명함)은 이러한 법규의 적용과 집행을 더 어렵게 만들 수 있다. 이러한 이유에서 현행 법규가 AI의 위험을 다루고 있고 효과적으로 집행될 수 있는지, 법규의 수정이 필요한지, 새로운 법규가 필요한지 등을 조사할 필요가 있다.

AI가 얼마나 빨리 진화하고 있는지를 감안하여, 규제 프레임워크는 추가적인 발전을 수용할 수 있는 여유를 남겨놓아야 한다. 모든 변경은 실현 가능한 해결방안이 존재하는 명확하게 식별된 문제에 국한되어야 한다.

회원국들은 현재 공통적인 유럽의 프레임워크가 부재하다는 것을 지적하고 있다. '독일 데이터윤리위원회(German Data Ethics Commission)'는 5단계의 위험 기반의 규제 시스템을 요구하고 있다. 5단계는 대부분의 무해한 AI 시스템에 대해서는 아무런 규제를 가하지 않은 것에서부터 대부분의 위험한 AI 시스템은 완전히 금지하는 것으로 구성되어 있다. 덴마크는 이제 막 '데이터윤리인증(Data Ethics Seal)'의 프로토타입에 착수하였다. 말타(Malta)는 AI에 대한 자율 인증 시스템을 도입하였다. EU가 EU 전반에 걸친 접근방법을 제공하지 못한다면, 내부 시장이 분열될 위험이 존재하고, 이것은 신뢰, 법적인 확실성, 시장의 수용이라는 목적을 손상시킬 것이다.
신뢰성 있는 AI를 위한 건실한 유럽의 규제 프레임워크는 모든 유럽 시민들을 보호하고, AI의 추가적인 발전과 수용이 이루어질 수 있는 마찰 없는 내부 시장의 창출을 돕고, AI 분야에서 유럽의 산업 기반을 강화할 것이다.

A. 문제 정의

제품과 프로세스를 보다 안전하게 만드는 것을 포함하여 AI는 많은 도움을 줄 수도 있지만, 또한 해를 끼칠 수도 있다. AI가 끼칠 수 있는 피해는 유형적(생명의 손실을 포함하여 개인의 안전과 건강, 재산의 손상) 또는 무형적(프라이버시의 침해, 표현의 자유의 제약, 인간의 존엄성, 고용 상의 차별)일 수 있고, 다양한 여러 가지 종류의 위험과 관련될 수 있다. 규제 프레임워크는 잠재적인 피해, 그중에서도 특히 중대한 피해에 대한 다양한 위험을 최소화하는 방법에 집중해야 한다.

AI의 사용에 관련된 주요한 위험은 기본권(개인 데이터 및 프라이버시 보호, 비차별 포함), 안전[32], 그리고 책임 관련 이슈를 보호하기 위한 규칙의 적용과 관련이 있다.
개인 데이터 및 프라이버시 보호, 비차별 등을 포함하여 기본권에 대한 위험
AI의 사용은 EU의 기반이 되는 가치에 영향을 미칠 수 있고, 표현의 자유, 집회의 자유, 인간의 존엄성, 성별/인종/종교/신념/장애/연령/성적 지향 등을 기반으로 한 비차별

32) 사이버보안, 핵심 인프라에서 AI 애플리케이션에 관련된 이슈, AI의 악의적인 사용 등이 포함된다.

을 포함하여 기본적인 권라[33], 개인 데이터 및 사생활 보호[34], 효과적인 사법적 구제 및 공정한 재판, 그리고 소비자 보호의 위반이 발생할 수 있다. 이러한 위험은 AI 시스템의 전반적인 설계 오류나 편향된 데이터를 수정하지 않고 사용함으로부터 야기될 수 있다 (예: 남성의 데이터만으로(혹은 주로 남성의 데이터로) 시스템을 훈련시킴으로써, 여성에 관하여 최적이 아닌 결과가 도출되는 경우).

AI는 이전에는 사람들만이 수행했던 많은 기능을 수행할 수 있다. 그 결과, 시민들과 법적 실체들은 AI 시스템에 의해서 혹은 AI 시스템의 지원으로 내려진 조치와 의사결정의 적용을 점차 더 많이 받게 될 것이고, 이러한 조치와 결정은 때때로 이해하기 어렵고 필요한 경우 효과적으로 이의를 제기하는 것이 어려울 수 있다. 그뿐만 아니라, AI는 사람들의 일상적인 습관을 추적하고 분석할 수 있는 가능성을 증가시킨다. 예를 들면, EU의 데이터 보호 및 기타 규칙을 위반하면서 AI가 정부 당국이나 기타 기관들에 의해서 대중 감시에 사용되거나, 고용주들에 의해서 고용자들이 행동하는 방법을 관찰하는 데 사용될 수 있는 잠재적인 위험이 존재한다. 많은 양의 데이터를 분석하고, 이들 간의 연결 관계를 식별함으로써, AI는 개인에 대한 데이터를 추적하고 익명성을 없애는 데 사용되어, 개인 데이터를 포함하고 있지 않은 데이터 세트에서조차 개인 데이터 보호에 관한 새로운 위험이 창출될 수도 있다. 또한 AI는 온라인 중개업체에 의해 사용되어, 사용자들을 위한 정보의 우선순위를 결정하고, 콘텐츠 조정(content moderation)을 수행할 수도 있다. 처리된 데이터, 애플리케이션이 설계된 방법, 사람의 개입 범위 등은 표현의 자유, 개인 데이터의 보호, 프라이버시, 정치적 자유에 영향을 미칠 수 있다.

> 특정한 AI 알고리즘은 범죄자의 재범 예측에 이용될 때, 성별과 인종에 대한 편향성을 나타낼 수 있다. 예를 들면, 여성과 남성, 혹은 내국인과 외국인의 재범 예측 확률이 달리 나올 수 있다.
>
> (자료원: *Tolan S., Miron M., Gomez E. and Castillo C. "Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia", Best Paper Award, International Conference on AI and Law, 201)*
>
> 특정한 안면 분석 AI 프로그램은 성별과 인종에 대한 편향성을 나타내고 있다. 즉, 밝은 피부의 남성의 경우에는 성별 결정 오류가 낮지만, 검은 피부의 여성의 경우에는 성별 결정 오류가 높게 나온다.
>
> (자료원: *Joy Buolamwini, Timnit Gebru; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018.*)

33) Council of Europe의 연구에 의하면, AI의 사용으로부터 많은 수의 기본권이 영향을 받을 수 있다. https://rm.coe.int/algorithms-and-human-rights-en-rev/ 16807956b5.

34) 일반데이터보호규정(General Data Protection Regulation: GDPR)과 e프라이버시 지침(ePrivacy Directive) (협상 중인 새로운 e프라이버시 규정)은 이러한 위험을 다루고 있지만, AI 시스템들이 추가적인 위험을 제기하는지의 여부를 조사할 필요가 있을 수 있다. 본 위원회는 GDPR의 적용을 지속적으로 모니터링하고 평가할 것이다.

편향성과 차별은 모든 사회적/경제적 활동에 내포되어 있는 원천적인 위험이다. 인간의 의사결정이 실수와 편향성에서 자유로울 수 없다. 그러나 동일한 편향성이 AI에 나타날 때에는 그 영향이 더욱 커서, 인간의 행동을 관장하는 사회적인 통제 메커니즘 없이 많은 사람에게 영향을 미치고 차별하게 된다.[35] 이것은 AI 시스템이 작동하는 동안에 '학습할' 때에도 발생할 수 있다.

설계 단계에서 결과를 방지하거나 예측할 수 없었던 경우, 위험은 시스템의 설계 결함에서 비롯되는 것이 아니라, 시스템이 대규모의 데이터 세트에서 식별한 상관관계나 패턴의 실제적인 영향으로부터 기인한다.

불투명성('블랙 박스 효과'), 복잡성, 예측 불가능성, 부분적 자율 행동 등을 포함하여, 많은 AI 기술의 특성은 기본권을 보호하기 위한 목적으로 제정된 기존 EU 법률의 규정을 준수하고 있는지를 검증하는 것을 어렵게 만들 수 있고, 이를 효과적으로 집행하는 것을 저해할 수도 있다. 법집행 당국과 피해자들은 AI의 관여로 이루어진 의사결정이 어떻게 내려졌는지, 따라서 관련 규칙들이 존중되었는지의 여부를 검증할 수 있는 수단이 부족할 수 있다. 개인과 법적 실체들은 이러한 의사결정이 자신들에게 부정적인 영향을 미치는 경우, 정의에 대한 효과적인 접근이 어려워지는 상황에 봉착할 수도 있다.

안전과 책임 체계의 효과적인 작동에 관한 위험
AI 기술은 이것이 제품과 서비스에 내장되는 경우, 사용자들에게 안전에 관한 새로운 위험을 제기할 수 있다. 예를 들면, 자율주행차가 물체 인식 기술의 오류로 도로상의 물체를 잘못 식별하여, 부상이나 물질적인 손상이 관련된 사고를 유발할 수 있다. 기본권에 대한 위험과 같이, 이러한 위험은 AI 기술의 설계상의 오류에 의해 야기되고, 데이터의 가용성과 품질에 관련된 문제나 기계학습으로부터 비롯되는 기타 문제와 관계가 있을 수 있다. 이러한 위험의 일부는 AI에 의존하는 제품과 서비스에만 국한되는 것은 아니지만, AI의 사용은 위험을 증가시키거나 악화시킬 수 있다.

관련된 사람들에게 대한 위험뿐만 아니라 이러한 위험에 대처하기 위한 명확한 안전 규정의 미흡은 AI가 관련된 제품을 EU에서 마케팅하는 기업들에게 법적인 불확실성을 창출한다. 시장 감시 및 집행 당국은 자신들이 개입할 수 있는지의 여부에 대해 확신을 가지지

35) 본 위원회의 양성평등자문위원회(Advisory Committee on Equal Opportunities for Women and Men)는 현재 AI가 양성 평등에 미치는 영향을 분석한 "AI에 대한 의견(Opinion on Artificial Intelligence)"을 준비 중에 있고, 이것은 2020년 초에 본 위원회에 의해서 채택될 예정이다. 또한 EU 양성평등전략 2020-2024(EU Gender Equality Strategy 2020-2024)도 AI와 양성 평등 간의 연결 관계에 대해 언급하고 있다. European Network of Equality Bodies (Equinet)도 다음과 같은 보고서를 2020년 초에 발간할 예정이다. 저자: Robin Allen & Dee Masters), 제목: "Regulating AI: the new role for Equality Bodies - Meeting the new challenges to equality and non-discrimination from increased digitalisation and the use of AI",

못할 수도 있다. 왜냐하면 자신들이 조치를 취할 수 있는 권한을 부여받지 못했거나, 시스템을 검사하는데 필요한 적절한 기술 역량을 보유하고 있지 못하고 있기 때문이다.[36] 따라서 법적인 불확실성은 전반적인 안전 수준을 낮추고, 유럽 기업들의 경쟁력을 손상시킬 수 있다.

안전에 관한 위험이 현실화되면, 명확한 요구사항의 미흡과 위에서 언급한 AI 기술의 특성 때문에 AI 시스템이 관여하여 이루어진 잠재적으로 문제의 소지가 있는 의사결정을 역추적하는 것이 어렵다. 또한 이것은 피해를 입은 사람들이 현재의 EU 및 국가별 책임 법규하에서 보상을 받는 것을 어렵게 만들 수 있다.[37]

> '제품책임지침(Product Liability Directive)' 하에서, 제조업체는 결함이 있는 제품에 의해 야기된 피해에 대해 책임이 있다. 그러나, 자율주행차와 같은 AI 기반의 시스템의 경우, 제품의 결함, 발생한 피해, 둘 간의 인과관계 등을 입증하는 것이 어려울 수 있다. 그뿐만 아니라, 특정한 유형의 결함(예: 결함이 제품의 사이버 보안상의 취약점에서 비롯된 경우)에 대해서는 제품책임지침이 어떻게, 어느 정도나 적용되는지에 대한 불확실성이 존재한다.

따라서 AI 시스템에 의해 내려진 잠재적으로 문제의 소지가 있는 의사결정을 역추적하는 것이 어려운 것은 안전과 책임에 관련된 이슈에도 동일하게 적용된다. 피해를 입은 사람들이 소송을 하는데 필요한 증거를 효과적으로 접근하지 못하고, 전통적인 기술에 의해 손상이 발생한 경우에 비해서 효과적인 배상의 가능성이 낮아질 수 있다. AI의 사용이 확산됨에 따라 이러한 위험은 증가할 것이다.

B. AI 관련 기존 EU 규제 프레임워크의 조정

산업 고유의 규칙을 포함하여 국가별 법규로 보완되고 있는 광범위한 기존의 EU의 제품 안전과 책임에 관한 법규[38]는 최근에 등장하고 있는 몇 가지 AI 애플리케이션과 관련이 있고, 적용될 수 있는 잠재력을 가지고 있다.

기본권과 소비자 권리의 보호의 경우, EU의 규제 프레임워크에는 '인종평등지침(Race

---

36) 한 예로 아동들을 위한 스마트 시계를 들 수 있다. 이 제품은 시계를 차고 있는 아동에게 직접적인 해를 끼치지 않지만, 최소 수준의 보안이 미흡하면 아동을 접근할 수 있는 도구로 쉽게 사용될 수 있다. 시장감시 당국은 이처럼 위험이 제품에 직접 연결되지 않는 경우 개입하기 어렵다고 판단할 수 있다.

37) AI, IoT, 기타 디지털 기술 등이 안전 및 책임 법규에 미치는 파급효과는 본 백서에 동봉된 '위원회보고서(Commission Report)'에 분석되어 있다.

38) 제품 안전에 관한 EU의 법규 프레임워크는 안전망으로서 일반제품안전지침(General Product Safety Directive - Directive 2001/95/EC), 그리고 높은 수준의 건강 및 안전을 제공하기 위한 기계, 비행기, 차에서부터 장남감, 의료기기에 이르기까지 몇 개의 산업 고유의 규칙으로 구성되어 있다. 제품 책임에 관한 법률은 제품이나 서비스에 의해 야기된 피해에 대한 여러 가지 민사 책임 시스템으로 보완되고 있다.

Equality Directive)'[39], '고용 및 직업 평등지침(Directive on equal treatment in employment and occupation)'[40], '고용 및 제품/서비스 접근에 관련된 양성 평등 지침(Directives on equal treatment between men and women in relation to employment and access to goods and services)'[41], 몇 가지 소비자 보호 규칙들[42], 그리고 '일반데이터보호규정(General Data Protection Regulation: GDPR)', '데이터보호법집행지침(Data Protection Law Enforcement Directive')[43]과 같은 개인 정보 보호에 관한 산업별 법규 등과 같은 개인 정보보호 및 프라이버시 관련 규칙들이 포함된다. 여기에 추가하여, 2025년부터 '유럽접근성법(European Accessibility Act)'에서 명시하고 있는 제품과 서비스에 대한 접근성 요구사항에 대한 규칙들이 적용될 것이다.[44] 또한 금융 서비스, 이주, 온라인 중개업체의 책임 등의 분야를 포함하여 기타 EU 법규를 시행할 때 기본권을 존중해야 한다.

EU 법규는 원칙적으로 AI의 관여 여부와 상관없이 완전하게 적용 가능한 상태로 남아 있지만, 이것이 AI 시스템이 창출하는 위험을 처리할 수 있도록 적절하게 집행될 수 있는지의 여부 또는 특정한 법적 수단을 수정하는 것이 필요한지의 여부를 평가하는 것은 중요한 일이다.

예를 들면, 경제 주체들은 AI가 소비자 보호를 위한 기존의 규칙들을 준수하도록 하는데 여전히 책임이 있고, 기존 규칙을 위반하면서 소비자 행동을 알고리즘으로 부당하게 이용하는 것은 허용되지 않고, 위반 행위는 처벌받을 것이다.

본 위원회의 의견은, 법규 프레임워크는 위험과 상황을 처리할 수 있도록 다음과 같이 개선될 수 있다는 것이다.

- 기존 EU 및 국가별 법규의 효과적인 적용 및 집행: AI의 핵심 특징 때문에 EU 및 국가별 법규들을 적절하게 적용하고 집행하기 위해 해결해야할 과제가 창출된다. 투명성 부족(AI의 불투명성)은 기본권을 보호하고, 책임을 묻고, 보상을 청구할 조건을 충족시키는 법률 조항을 포함하여, 법률의 위반을 식별하고 입증하는 것을 어렵게 만든다. 이에 따라 효과적인 적용과 집행이 이루어지도록 하기 위해서는 특정 영역, 예를 들면, 본 백서에 동봉된 보고서에서 자세하게 설명하고 있는 바와 같이 책임에 관한 기존의 법규를 수정하거나 명확화하는 것이 필요할 수

39) Directive 2000/43/EC.
40) Directive 2000/78/EC.
41) Directive 2004/113/EC; Directive 2006/54/EC.
42) 예: 불공정상거래프랙티스지침(Unfair Commercial Practices Directive - Directive 2005/29/EC), 소비자권리지침(Consumer Rights Directive - Directive 2011/83/EC).
43) 2016년 4월 27일의 유럽 의회와 본 위원회의 Directive (EU) 2016/680으로서, 범죄 행위의 예방/조사/적발/처벌이나 형사 처벌의 집행 목적으로 권한을 가진 당국에 의한 개인 데이터의 처리에 관해서 자연인을 보호하는 것, 그리고 이러한 데이터의 자유로운 이동에 대한 내용을 담고 있다.
44) 제품과 서비스에 대한 접근성 요구사항에 대한 Directive (EU) 2019/882 .

있다.

- 기존 EU 법규의 범위의 한계: EU의 제품 안전에 관한 법규의 기본적인 초점은 시장에 제품을 출시하는 데 있다. EU의 제품 안전에 관한 법규에서 소프트웨어가 최종 제품의 일부인 경우에는 관련된 제품 안전에 관한 규칙들을 준수해야 하지만, 독립적인 소프트웨어가 EU의 제품 안전에 관한 법규의 적용을 받는지의 여부는 아직 정답이 정해지지 않은 문제이다 (일부 산업에는 명시적인 규칙이 존재함).[45] 현재 시행 중인 일반적인 EU의 안전에 관한 법규는 제품에는 적용되지만, 서비스에는 적용되지 않는다. 따라서 원칙적으로는 AI 기술을 기반으로 하고있는 서비스(예: 의료 서비스, 금융 서비스, 운송 서비스)에는 원칙적으로 적용되지 않는다.
- AI 시스템의 기능 변화: AI를 포함하여 소프트웨어를 제품에 통합시키는 것은 제품과 서비스의 수명주기 동안에 기능을 변경시킬 수 있다. 이것은 빈번한 소프트웨어 업데이트가 필요한 시스템이나 기계학습에 의존하는 시스템의 경우 특히 그러하다. 이러한 특성 때문에 시스템이 출시하였을 때는 없던 위험이 새로이 나타날 수 있다. 출시 시점에 존재하는 안전에 관한 위험에 주로 초점을 맞추고 있는 기존의 법규에서는 이러한 위험이 적절하게 다루어지지 않고 있다.
- 공급망에서 서로 다른 경제 주체들 간의 책임 할당에 관한 불확실성: 일반적으로 봤을 때, 제품의 안전에 관한 EU의 법규는 모든 구성요소들(예: AI 시스템)을 포함하여 시장에 출시된 제품의 생산자에게 책임을 할당한다. 그러나 제품이 출시된 이후에 생산자가 아닌 제삼자에 의해서 AI가 추가되면, 규칙은 불분명해질 수 있다. 그 뿐만 아니라 EU의 제품 책임에 관한 법규는 생산자의 책임만을 규정하고 있고, 공급망의 기타 주체의 책임은 국가별 책임 규칙으로 관장하도록 하고 있다.
- 안전의 개념 변화 : 제품과 서비스에 AI를 사용함으로써, EU의 법규가 현재 명시적으로 다루고 있지 않은 위험이 발생할 수 있다. 이러한 위험은 사이버 위협, 개인의 보안 위험(가전제품과 같이 새로운 AI 애플리케이션에 연관된), 접속 불능으로 야기되는 위험 등에 연관될 수 있다. 이러한 위험은 제품을 시장에 출시하는 시점에 존재하거나, 제품을 사용하고 있는 중에 소프트웨어 업데이트나 자가 학습의 결과로 발생할 수 있다. AI 위협 상태 평가에 대한 'EU사이버보호원(Cybersecurity Agency: ENISA)'의 경험을 활용하는 것을 포함하여, EU는 AI 애플리케이션에 관련된 잠재적인 위험에 대한 증거 베이스(evidence base)를 제고하기 위해 동원할 수 있는 모든 도구들을 활용해야 한다.

앞에서 언급한 바와 같이, 여러 회원국들은 AI에 의해 창출되는 해결과제를 다루기 위한 국가 수준의 법규 도입을 위한 대안들을 이미 탐색하고 있다. 이것은 단일 시장이

45) 예를 들어서, 제조업체에 의해서 의료 목적으로 사용되는 소프트웨어는 의료기기규정 (Medical Device Regulation - Regulation (EU) 2017/745) 하에서는 의료 기기로 간주된다.

분열될 수 있는 위험을 제기한다. 국가별로 서로 다른 규칙들은 단일 시장에 AI 시스템을 판매하고 운영하기를 원하는 기업들에게 장애요인이 될 가능성이 높다. EU 차원의 하나의 공통적인 접근방법을 제공하는 것은 유럽 기업들이 단일 시장을 원활하게 접근함으로써 혜택을 볼 수 있도록 하고, 이들이 세계 시장에서 경쟁력을 가지는 것을 지원할 것이다.

AI, IoT, 로봇공학이 안전과 책임에 미치는 파급효과에 관한 보고서

본 백서에 동봉된 이 보고서는 관련 법률 프레임워크를 분석한다. 이 보고서는 AI 시스템과 기타 디지털 기술들에 의해 야기되는 특정한 위험에 관련하여 이 프레임워크를 적용하는데 있어서의 불확실성을 식별하고 있다.

본 보고서의 결론은, 제품 안전에 관한 현행 법규는 제품으로부터 야기되는 모든 종류의 위험으로부터 보호한다는 확장된 개념의 안전을 이미 지원하고 있다는 것이다. 그러나 법적인 확실성을 높일 수 있도록, 새로이 출현하는 디지털 기술에 의해 야기되는 새로운 위험을 명시적으로 다루는 다음과 같은 조항들이 도입될 수 있다.

- 특정한 AI 시스템의 수명주기 동안에 발생하는 자율적인 행동에는 안전에 영향을 미치는 중요한 제품의 변경이 수반될 수 있고, 이것은 새로운 위험 평가를 필요로 할 수 있다. 그뿐만 아니라 AI 제품과 시스템의 설계에서부터 수명주기 전반에 걸쳐서 안전장치로서 사람의 감독이 필요할 수 있다.
- 필요한 경우(예: 인간형(humanoid) 로봇과의 협력), 사용자들의 정신적인 안전에 대한 위험에 관련하여 생산자에 대한 명시적인 의무 또한 고려될 수 있다.
- EU의 제품 안전에 관한 법규는 설계 단계에서 잘못된 데이터의 안전에 관련된 위험, 그리고 AI 제품과 시스템을 사용하는 동안 내내 데이터의 품질이 유지되도록 하는 메커니즘에 관한 특정한 요구사항에 대비할 수 있다.
- 알고리즘 기반 시스템의 불투명성은 투명성 요구사항을 통해 처리될 수 있다.
- 시장에 출시되어 있거나, 출시 이후에 제품에 다운로드된 독립적인 소프트웨어의 경우, 안전에 영향을 미치는 경우, 기존 규칙들은 수정되거나 명확화되어야 할 필요가 있을 수 있다.
- 새로운 기술의 공급망의 복잡성이 점차 커지고 있다는 점을 감안하면, 공급망의 경제 주체들과 사용자들 간의 협력을 의무화하는 조항들은 법적인 확실성을 제공할 수 있다.

AI, IoT, 로봇 공학 등과 같이 새로이 등장하는 기술의 특성은 책임 프레임워크의 특정한 측면에 의문을 제기하고, 이것의 효과성을 낮출 수 있다. 이러한 특성 중의 일부는 피해를 특정한 한 사람에게로 역추적하는 것을 어렵게 만드는데, 이러한 역추적은 대부분의 국가별 규칙에 따라 과오 기반의 청구(fault-based claim)에 필요할 수 있다. 이것은 피해자의 비용을 크게 증가시키고, 생산자 이외에 사람들에게 대한 책임을 청구하거나 입증하는 것이 어려울 수 있다는 것을 의미한다.

- 지속적으로 발전할 수 있도록 기술 혁신은 허용되어야 하지만, AI 시스템의 관여로 말미암아 피해를 입은 사람들은 다른 기술에 의해 피해를 받은 사람들

과 동일한 수준의 보호를 받을 수 있어야 한다.

- ‘제품책임지침(Product Liability Directive)’의 수정, 그리고 국가별 책임 규칙들의 일치화를 포함하여 이 목적의 달성을 위한 모든 대안들을 주의 깊게 고려해야 한다. 예를 들면, 본 위원회는 AI 애플리케이션에 의해 야기된 피해에 대해 국가별 책임 규칙들이 요구하고 있는 입증 책임을 수정함으로써 복잡성의 결과를 경감할 필요가 있는지의 여부, 경감한다면 어느 정도나 경감해야 할지에 대한 견해를 들으려고 노력하고 있다.

위와 같은 논의를 바탕으로 본 위원회가 내린 결론은 EU의 법률 프레임워크가 현재 및 예상되는 기술적/상업적 발전상황에 적합하도록 만들기 위해서는 기존 법규의 수정뿐만 아니라 특히 AI에 대한 새로운 법규가 필요할 수 있다는 것이다.

C. 미래 EU 규제 프레임워크의 범위

AI에 대한 미래 규제 프레임워크의 핵심 이슈는 AI의 적용 범위를 결정하는 것이다. 가정(assumption)은 규제 프레임워크가 AI에 의존하고 있는 제품과 서비스에 적용될 것이라는 것이다. 따라서 본 백서 그리고 미래의 정책 수립 이니셔티브의 목적을 위해 AI를 명확하게 정의해야 한다.

‘유럽의 AI에 대한 공보(Communication on AI for Europe)’에서 본 위원회는 AI의 첫 번째 정의를 제시하였다.[46] 이 정의는 ‘고위전문가그룹’에 의해 추가적으로 정제되었다.[47]

새로운 법적 수단에서 AI의 정의는 필요한 법적 명확성을 제공할 수 있을 정도로 정확하면서도, 기술의 진보를 수용할 수 있을 정도로 충분히 유연해야 할 필요가 있을 것이다.

예를 들면, 자율주행의 경우, 알고리즘은 실시간으로 자동차로부터의 데이터(속도, 엔진 소비, 충격흡수장치 등), 그리고 자동차의 전체 환경을 살피는 센서로부터의 데이터(도로, 신호, 다른 차량, 보행자 등)를 사용하여, 정한 목적지로 가기 위해 어

46) COM(2018) 237 final, p. 1: “AI는 특정한 목표를 달성하기 위해 환경을 분석하고 조치를 취함으로써 (일정 부분은 자율로) 지능적인 행동을 나타내는 시스템을 말한다. AI 기반의 시스템들은 전적으로 소프트웨어 기반으로 가상 세계에서 작동되거나 (예: 음성 보조원, 이미지 분석 소프트웨어, 검색엔진, 음성 및 안면 인식 시스템) 혹은 AI가 하드웨어에 내장될 수 있다(예: 고급(advanced) 로봇, 자율주행차, 드론 또는 IoT 애플리케이션).”

47) ‘고위전문가그룹’, AI의 정의, p.8: “AI 시스템은 인간에 의해 설계된 소프트웨어 (하드웨어일 수도 있음) 시스템으로서, 주어진 복잡한 목표 하에서, 데이터 획득을 통해서 환경을 인식하고, 수집된 구조화/비구조화 데이터를 해석하고, 이러한 데이터로부터 도출된 지식을 바탕으로 추론하거나 정보를 처리하고, 주어진 목표를 달성하기 위한 최선의 조치를 결정함으로써 물리적 또는 디지털 차원에서 작동한다. AI 시스템은 상징적(symbolic) 규칙을 이용하거나, 수치 모델을 학습할 수 있고, 자신들의 이전 조치에 의해 환경이 어떻게 영향을 받는지를 분석함으로써 자신들의 행동을 수정할 수 있다.”

느 방향으로 어느 정도의 속도로 가야하는지를 도출한다. 관찰한 데이터를 기반으로 알고리즘은 다른 운전자들의 행동을 포함하여 도로의 상황과 외부 조건에 맞게 수정하여, 가장 안전하고 편안한 주행을 도출한다.

본 백서, 그리고 정책 이니셔티브 대한 미래의 논의를 위해서 AI를 구성하는 주요한 요소("데이터"와 "알고리즘")를 명확화하는 것이 중요하다고 판단된다. AI는 하드웨어에 통합될 수 있다. AI의 하위 집합인 기계학습 기법의 경우, 알고리즘은 주어진 목표를 달성하는데 필요한 조치를 결정할 수 있도록 일련의 데이터를 기반으로 특정한 패턴을 추론하도록 훈련된다. 알고리즘은 사용되는 동안 지속적으로 학습할 수 있다. AI 기반의 제품들은 미리 정해진 지시를 따를 필요 없이 자신들의 환경을 인식함으로써 자율적으로 행동할 수 있지만, 이들의 행동은 대부분 개발자들 의해서 정의되고 제약을 받는다. 인간이 목표를 결정하고 프로그래밍하고, AI 시스템은 이를 최적화해야 한다.

EU는 소비자를 보호하고, 불공정한 상거래 프랙티스에 대처하고, 개인의 데이터와 프라이버시를 보호하기 위한 엄격한 법적 프레임워크를 수립하고 있다. 그뿐만 아니라, 판례법은 특정 산업(예: 의료 서비스, 운송)에 대한 구체적인 규칙을 담고 있다. 디지털 트랜스포메이션과 AI의 사용을 반영하기 위해서 일정한 수정이 필요할 수도 있지만, 이러한 EU 법률의 기존 조항들은 AI에 계속해서 적용될 것이다 (B항 참조). 그 결과, 기존의 수평적 또는 산업별 법규(예: 의료 기기[48], 운송 시스템)에 의해서 이미 다루어지고 있는 측면들은 이러한 법규에 의해서 계속해서 관장될 것이다.

원칙적으로 AI에 대한 새로운 규제 프레임워크는 특히 중소기업들에게 필요 이상의 부담을 주지 않도록 지나치게 규범적이지 않으면서, 그 목적을 달성하는데 효과적이어야 한다. 이러한 균형을 달성하기 위해 본 위원회의 견해는 새로운 규제 프레임워크는 위험 기반의 접근방법을 따라야 한다는 것이다.

위험 기반의 접근방법은 규제를 통한 개입이 적절한 수준이 되도록 도와주는데 중요하다. 그러나 이것은 서로 다른 AI 애플리케이션들을 구분할 수 있는, 특히 이들이 "고위험"인지의 여부를 결정할 수 있는 명확한 기준을 필요로 한다.[49] 무엇이 고위험 AI 애플리케이션인지에 대한 결정은 명확하고, 쉽게 이해 가능하고, 모든 관련 당사자들에게 적용 가능해야 한다. 그렇지만, AI 애플리케이션이 고위험군으로 분류되지 않더라도, 이것은 여전히 전적으로 이미 존재하는 EU 규칙의 적용을 받는다.

---

48) 예를 들면, 의사들에게 전문적인 의료 정보를 제공하는 AI 시스템, 환자에게 직접 의료 정보를 제공하는 AI 시스템, 환자에게 직접 의료 작업을 스스로 수행하는 AI 시스템 등에 관련하여, 서로 다른 안전에 대한 고려사항과 법적인 파급효과가 존재한다. 본 위원회는 의료 서비스에 고유한 이러한 안전과 책임에 관련된 해결과제를 조사하고 있다.

49) EU 법규는 영역에 따라 (예를 들면 제품 안전 분야), "위험"을 여기에서 설명한 것과는 다르게 분류할 수 있다.

본 위원회의 의견은 특히 안전, 소비자 권리, 기본권의 관점에서, AI 애플리케이션이 사용되는 부문과 AI의 예상되는 활용 두 가지 모두에 큰 위험이 수반되는지의 여부를 고려하여, 해당 AI 애플리케이션은 위태로움의 관점에서 고위험군으로 간주되어야 한다는 것이다.

보다 구체적으로 보면, 다음과 같은 두 가지의 누적 기준을 충족하는 경우 AI 애플리케이션은 고위험군으로 간주되어야 한다.

- 첫째, 흔히 수행하는 활동의 특성을 고려해볼 때, AI 애플리케이션이 상당한 위험이 발생할 것으로 예상되는 부문에 사용되고 있다. 이 첫 번째 기준은 규제를 통한 개입이 위험이 발생할 가능성이 가장 높을 것으로 생각되는 영역을 목표로 하도록 해준다. 규제에 포함되는 영역들은 새로운 규제 프레임워크에 구체적이고 빠짐없이 나열되어야 한다. 예를 들면, 의료 서비스, 운송, 에너지, 공공 부문의 일부[50]를 들 수 있다. 이 목록은 주기적으로 검토되고, 실무에서의 발전 상황에 따라 필요한 경우 수정되어야 한다.
- 둘째, AI 애플리케이션이 문제가 되는 영역에서 상당한 위험이 발생할 가능성이 높은 방법으로 사용되고 있다. 이 두 번째 기준은 선정된 부문에서 사용되는 모든 AI에 반드시 상당한 위험이 수반되는 것은 아니다 라는 사실을 반영하는 것이다. 예를 들면, 의료 서비스 분야는 일반적으로 관련 산업이지만, 병원의 약속 일정관리 시스템의 결함은 법규의 개입을 정당화할 만큼 심각한 위험을 제기하지는 않을 것이다. 주어진 사용의 위험 수준을 평가하는 것은 피해 당사자들에게 미치는 파급효과를 기반으로 할 수 있다. 예를 들면, 개인이나 기업의 권리에 법적인 혹은 이와 유사한 상당한 결과를 생성하는 AI 애플리케이션의 사용하는 경우, 부상, 사망 또는 상당한 유형/무형의 피해의 위험을 제기하는 경우, 개인이나 법적 실체가 합리적으로 피할 수 없는 효과를 생성하는 경우 등을 들 수 있다.

이러한 두 가지의 누적 기준을 적용하는 것은 규제 프레임워크의 범위가 목표를 대상으로 하고, 법적인 확실성을 제공하도록 한다. AI에 대한 새로운 규제 프레임워크에 담긴 의무적인 요구사항들(아래의 D항 참조)은 원칙적으로 이러한 두 가지 누적 기준에 따라 고위험군으로 식별된 애플리케이션에만 적용된다.

위에서 언급한 내용에도 불구하고, 예외적인 상황이 있을 수도 있다. 걸려 있는 위험 때문에, AI 애플리케이션을 특정한 목적에 사용하는 것이 고위험군으로 간주되는 경우, 즉 관련 산업에 상관없이, 아래의 요구사항이 여전히 적용되는 경우를 말한다.[51] 한

---

50) 공공 부분에는 망명, 이주, 국경 통제 및 재판, 사회적 보안, 고용 서비스 등과 같은 영역이 포함될 수 있다.

51) 다른 EU 법규도 적용될 수 있다는 것을 강조하는 것이 중요하다. 예를 들면, 소비자 제품에 포함되었을 경우에는 AI 애플리케이션의 안전에 '일반제품안전지침(General Product Safety Directive)'이 적용될 수

예로서 다음과 같은 상황을 생각해 볼 수 있다.

- 개인에 대한 중요성, 그리고 고용 평등에 관한 EU의 판례의 관점에서 보면, 고용 프로세스, 그리고 노동자들의 권리에 영향을 미치는 상황에 AI 애플리케이션을 사용하는 것은 항상 '고위험'으로 간주된다. 따라서 아래의 요구사항들이 항상 적용될 것이다. 소비자의 권리에 영향을 미치는 추가적인 애플리케이션들도 고려될 수 있다.
- 원격 생체인식 식별[52], 그리고 기타 침입 감시 기술의 목적으로 AI 애플리케이션을 사용하는 것은 항상 "고위험"으로 간주되고, 이에 따라 아래의 요구사항들이 항상 적용될 것이다.

D. 요구사항의 유형

AI에 대한 미래의 규제 프레임워크를 설계할 때, 관련 주체에 가할 의무적인 법적 요구사항의 종류를 결정하는 것이 필요할 수 있다. 이러한 요구사항들은 기준을 통해서 보다 구체적으로 명시될 수 있다. C항에서 언급한 바와 같이, 그리고 이미 존재하고 있는 법규에 추가하여, 이러한 요구사항들은 고위험 AI 애플리케이션에만 적용될 것이고, 이에 따라 법규를 통한 개입이 초점을 가지고 해당 상황에 상응하도록 할 것이다.
'고위전문가그룹'의 가이드라인과 본 백서의 앞부분에서 제시한 내용들을 감안하면, 고위험 AI 애플리케이션에 대한 요구사항들은 다음과 같은 핵심적인 특성으로 구성될 수 있다. 다음에서는 여기에 대해 보다 자세하게 설명하도록 한다.

- 훈련 데이터
- 데이터 및 기록 관리
- 제공될 정보
- 견고성(robustness) 및 정확성
- 사람의 감독(human oversight)
- 원격 생체인식 식별의 목적으로 사용되는 AI 애플리케이션과 같은 특정한 AI 애플리케이션에 대한 요구사항

법적인 확실성을 보장하기 위해서, 이를 준수해야 하는 모든 주체들에게 명확한 벤치마크를 제공할 수 있도록 이러한 요구사항들은 추가적으로 구체화될 것이다.

a) 훈련 데이터

EU의 가치와 규칙, 그중에서도 특히 시민들이 EU의 법률로부터 비롯된 권리를 고취하

있다.

52) 원격 생체인식 식별(Remote biometric identification)은 생체인식 인증(biometric authentication)과 구별되어야 한다. 후자는 사람의 고유한 생물학적인 특성에 의존하여, 그 사람이 진짜 그 사람인지를 검증하는 보안 프로세스이다. 원격 생체인식 식별은 생채인식 식별자(지문, 안면 이미지, 홍채, 혈관 패턴 등)의 도움으로, 원격으로 공공 장소에서 지속적으로 데이터베이스에 저장되어있는 데이터에 대비해서 식별자들을 확인함으로써 여러 사람의 신원을 입증하는 것이다.

고, 강화하고, 수호하는 것이 어느 때보다 더 중요하게 되었다. 이러한 노력들은 의심의 여지없이 EU에 출시되어 사용되고 있는 고위험 AI 애플리케이션에도 적용된다.

앞서 설명한 바와 같이 데이터가 없으면 AI도 있을 수 없다. 많은 AI 시스템의 기능, 그리고 이러한 시스템이 야기하는 조치와 의사결정은 시스템이 훈련을 받은 데이터 세트에 크게 의존한다. 따라서 AI 시스템을 훈련시키는데 사용된 데이터가 EU의 가치와 규칙, 그중에서도 특히 기본권 보호를 위한 안전 및 기존 법규를 존중하도록 필요한 조치를 취해야 한다. AI 시스템을 훈련시키는데 사용되는 데이터에 관련하여 다음과 같은 요구사항들을 생각해 볼 수 있다.

- AI 시스템이 가능하게 해 준 제품/서비스의 사용이 안전하다, 즉, 적용되는 EU의 안전 규칙(현행 규칙 및 이를 보완하는 규칙)에서 규정하고 있는 기준을 충족시키고 있다는데 대한 합리적인 보증을 제공하기 위한 목적의 요구사항. 예를 들면, AI 시스템이 충분히 광범위하고, 위험한 상황을 피하는데 필요한 모든 관련 시나리오를 포괄하는 데이터 세트로 훈련받도록 하는 요구사항을 들 수 있다.
- AI 애플리케이션의 사용이 금지되고 있는 차별을 수반하는 결과를 야기하지 않도록 하기 위한 목적의 합리적인 조치를 취하라는 요구사항. 이러한 요구사항은 충분히 대표성을 갖춘 데이터 세트를 사용하는 의무, 특히 성별, 인종, 기타 금지된 차별의 근거 등의 모든 관련 요소들이 해당 데이터 세트에 적절하게 반영되도록 하는 의무를 수반할 수 있다.
- AI로 가능해진 제품/서비스를 사용하는 동안 프라이버시와 개인 정보를 적절하게 보호하도록 하는 목적의 요구사항. 이 범주에 속하는 이슈에 대해서는 '일반데이터보호규정(GDPR)'과 '법집행지침(Law Enforcement Directive)'이 이러한 문제를 규제하고 있다.

b) 기록 관리 및 데이터

많은 AI 시스템의 복잡성과 불투명성, 그리고 적용되는 규칙의 준수를 효과적으로 검증하고 집행하는데 관련하여 존재하는 어려움 등과 같은 요소들을 감안하여, 요구사항은 알고리즘의 프로그래밍, 고위험 AI 시스템을 훈련시키는데 사용되는 데이터, 어떤 경우에는 데이터 자체의 관리 등에 관련된 기록 관리를 요구한다. 이러한 요구사항들은 기본적으로 AI 시스템에 의해 이루어진 잠재적으로 문제의 소지가 있는 조치나 의사결정을 역추적하고 검증할 수 있도록 해 준다. 이것은 감독과 집행을 촉진할 뿐만 아니라, 관련 경제 주체들이 초기 단계부터 이러한 규칙의 준수 필요성을 고려하게 만드는 동기를 제공할 수 있다.

이를 위해서 규제 프레임워크는 다음과 같은 것을 유지하도록 규정할 수 있다.

- 데이터 세트의 주요한 특징, 데이터 세트를 선정한 방법 등에 대한 설명을 포함하여,

AI 시스템을 훈련시키고 테스트하는데 사용된 데이터 세트에 관한 정확한 기록

- 그럴만한 이유가 있는 경우, 데이터 세트 자체
- 해당하는 경우 안전과 금지되어 있는 차별을 야기할 수 있는 편향성 회피에 관한 문서를 포함하여, 프로그래밍[53] 및 훈련 방법, AI 시스템을 구축/테스트/확인하는데 사용된 프로세스 및 기법 등에 대한 문서

관련 법규의 효과적인 집행이 이루어질 수 있도록 기록, 문서, 그리고 해당하는 경우 데이터 세트는 한정된 합리적인 기간 동안 보관되어야 한다. 요청받았을 때, 특히 권한을 가진 당국이 테스트하거나 조사를 위해 요청할 때, 이를 이용 가능하도록 하기 위한 조치를 취해야 한다. 필요한 경우, 기업 비밀과 같은 기밀 정보가 보호되도록 하기 위한 준비를 갖추어야 한다.

c) 정보 제공

위의 B항에서 설명한 기록 관리 요구사항 이상의 투명성이 요구된다. 추구하는 목적 (특히 책임성 있는 AI의 사용 촉진, 신뢰 구축, 필요한 경우 보상의 용이한 처리)을 달성하기 위해서는 고위험 AI 시스템의 사용에 대한 적절한 정보를 능동적으로 제공하는 것이 중요하다.

이에 따라, 다음과 같은 요구사항들이 고려될 수 있다.

- AI 시스템의 역량과 한계에 대한 명확한 정보, 그 중에서도 특히 시스템이 의도하는 목적, 시스템이 계획대로 작동하는 것으로 예상할 수 있는 조건, 명시된 목적을 달성하는데 있어서 예상되는 정확도 등에 대한 정보가 제공되도록 한다. 이러한 정보는 시스템을 배치하는 사람들에게는 특히 중요하지만, 권한을 가진 당국과 피해 당사자들과도 관련이 있을 수 있다.
- 이와는 별도로, 시민들이 사람이 아니라 AI 시스템과 상호작용할 때 이러한 사실을 명확하게 통보받아야 한다. EU의 데이터 보호 법규가 이미 이러한 종류의 규칙을 담고 있지만[54], 위에서 언급한 목적을 달성하기 위해서는 추가적인 요구사항이 필요할 수 있다. 만일 그렇더라도 불필요한 부담을 피해야 한다. 따라서 예를 들면, 자신들이 AI 시스템과 상호작용하고 있다는 것이 시민들에게 즉각적으로 분명한 경우에는 그러한 정보를 제공할 필요가 없다. 제공하는 정보가 객관적이고, 간결하고, 쉽게 이해할 수 있는 것이 더욱 중요하다. 정보를 제공하는 방법은 상황에 맞게 맞춤화되어야 한다.

---

53) 예를 들면, 모델이 최적화하려는 것, 초기에 특정 파라미터에 부여한 비중 등을 포함하여 알고리즘에 대한 문서를 들 수 있다.

54) 특히, GDPR 13(2)(f)에 따라 관리자(controller)들은 개인 정보를 입수할 때 데이터 주체에게 자동화된 의사결정의 존재에 대해 공정하고 투명한 처리가 이루어지도록 하는데 필요한 추가적인 정보를 제공해야 한다.

d) 견고성(Robustness) 및 정확성

AI 시스템, 그중에서도 특히 고위험 AI 애플리케이션이 신뢰를 받기 위해서는 기술적으로 견고하고 정확해야 한다. 이것은 그러한 시스템은 책임성 있는 방법으로, 그리고 발생할 수 있는 위험을 사전에 충분하고 적절하게 고려하면서 개발되어야 한다는 것을 의미한다. AI 시스템이 계획대로 신뢰성 있게 행동하도록 하기 위해서는 시스템의 개발과 작동이 그러해야 한다. 야기될 수 있는 피해의 위험을 최소화하기 위한 모든 합리적인 조치를 취해야 한다.

따라서 다음과 같은 요소들을 고려해야 한다.

- AI 시스템이 견고하고 정확한 상태를 유지하거나, 적어도 수명주기의 모든 단계에서 자신들의 정확성 수준을 정확하게 반영하도록 하는 요구사항
- 결과를 재생산 가능하도록 하는 요구사항
- AI 시스템이 수명주기의 모든 단계에서 오류나 모순을 적절하게 처리할 수 있도록 하는 요구사항
- AI 시스템이 명백한 공격, 그리고 데이터나 알고리즘 자체를 조작하려는 보다 미묘한 시도, 두 가지 모두에 대해 복원력을 가지고, 그러한 경우 경감 조치를 취하도록 하는 요구사항

e) 사람의 감독

사람의 감독은 AI 시스템이 인간의 자율성을 손상하거나, 기타 부정적인 결과를 야기하지 않도록 하는 것을 도와준다. 신뢰할만하고, 윤리적이고, 인간 중심적인 AI라는 목적은 고위험 AI 애플리케이션의 경우, 인간의 적절한 참여로만 달성될 수 있다.

본 백서에서 특정 법률 체계에 대해 고려된 AI 애플리케이션은 모두 고위험으로 간주되지만, 사람의 감독의 적절한 유형과 정도는 사례별로 다를 수 있다. 이것은 시스템의 계획된 사용 방법, 그리고 시스템의 사용이 영향을 받는 시민들과 법적 주체에 미치는 결과에 의해 결정된다. 또한 AI 시스템이 개인 데이터를 처리할 때 GDPR에 의해 수립된 법적인 권리를 침해해서는 안된다. 예를 들면, 사람의 감독은 다음과 같이 나타날 수 있다 (이것이 모든 것을 나열하고 있는 것은 아니다).

- 사람에 의해서 사전에 검토되고 확인되지 않았다면 AI 시스템의 출력은 효력을 발휘할 수 없다(예: 사회보장 지원금의 신청에 대한 기각은 사람에 의해서만 될 것이다).
- AI 시스템의 출력은 즉각적으로 효력을 나타낼 수 있지만, 나중에 사람의 개입이 이루어져야 한다(예: 신용카드 신청에 대한 거절은 AI 시스템에 의해 처리될 수 있지만, 나중에 사람의 검토가 가능해야 한다).

- 운영 중인 AI 시스템의 모니터링, 실시간으로 개입하여 정지시킬 수 있는 능력(예: 무인 자동차에서 사람이 자동차의 운행 상태가 안전하지 않다고 판단했을 때 정지 버튼 또는 절차가 이용 가능한 경우).
- 설계 단계에서 AI 시스템에 운영상의 제약을 가함으로써 (예: 무인 자동차는 센서의 신뢰성이 낮아질 수 있는 가시성이 낮은 상황에서 운행을 중단하고, 주어진 조건에서 앞의 차량과 일정한 거리를 유지해야 한다).

f) 원격 생체인식 식별에 대한 요구사항

예를 들면, 공공장소에 안면 인식 기술을 배치하여 원격 식별[55]을 위해 생체인식 데이터[56]를 수집하고 사용하는 것은 기본권에 대한 위험[57]을 수반한다. 원격 생체인식 식별 AI 시스템의 사용이 기본권에 미치는 파급효과는 사용의 목적, 상황, 범위에 따라 크게 달라질 수 있다.

EU의 데이터 보호 규칙들은 특별한 경우를 제외하고는 생체인식 데이터를 자연인의 식별 목적으로 처리하는 것을 원칙적으로 금지하고 있다.[58] GDPR 하에서 그러한 처리는 제한적인 근거에서만 수행될 수 있는데, 주요한 이유는 중대한 공공의 이익이다. 그러한 경우, 처리는 EU나 국가별 법률을 기반으로 수행되고, 비례성의 요구사항의 적용을 받고, 데이터 보호에 대한 권리의 본질과 적절한 보호장치를 존중해야 한다. '법집행지침' 하에서 이러한 처리를 위해서는 엄격한 필요성, 즉, 원칙적으로 EU나 국가별 법률에 의한 승인, 그리고 적절한 안전 장치가 존재해야 한다. 자연인을 고유하게 식별하기 위한 목적으로 생체인식 데이터를 처리하는 것은 EU 법률에 명시되어 있는 금지사항의 예외와 관련되어 있지만, 이것은 EU의 '기본권헌장(Charter of Fundamental Rights)'의 적용을 받을 것이다.

---

55) 안면 인식에서 식별은 개인의 안면 이미지의 템플릿을 데이터베이스에 저장되어 있는 많은 다른 템플릿과 비교하여, 해당 개인의 이미지가 거기에 저장되어 있는지를 알아보는 것을 의미한다. 한편, 인증(또는 검증)은 흔히 1-대-1 매칭이라고도 불린다. 이것은 동일한 사람에게 속하는 것으로 추정하는 두 개의 생체인식 템플릿의 비교를 가능하게 해준다. 두 개의 생체인식 템플릿은 비교되어, 두 개의 이미지에 나타난 사람이 동일인인지를 결정한다. 이러한 절차는 예를 들면 공항의 국경 통과에 사용되는 자동국경통제(Automated Border Control: ABC) 게이트에 사용되고 있다.

56) 생체인식 데이터는 "안면 이미지나 지문 데이터와 같이, 자연인의 물리적, 생리적 또는 행동적 특성에 관련된 특정한 기술 처리로부터 야기되는 개인 데이터로서, 해당 자연인의 고유한 인증이나 식별을 가능하게 하거나 확인한다."라고 정의된다(법집행지침(Law Enforcement Directive), Art. 3 (13); GDPR, Art. 4 (14); Regulation (EU) 2018/1725, Art. 3 (18).

57) 예를 들면 인간의 존엄성을 들 수 있다. 안면 인식 기술을 사용할 때 사생활의 존중 및 개인 데이터의 보호는 기본권에 대한 우려의 핵심이다. 아동, 노인, 장애인 등과 같은 특별한 집단의 비차별과 권리에 미치는 잠재적인 영향도 존재한다. 표현/결사/집회의 자유가 기술의 사용으로 손상되어서는 안된다. 참조: 안면인식 기술: 법집행 관점에서 기본권에 관한 고려사항, https://fra.europa.eu/en/publication/2019/facial- recognition.

58) 참조: Article 9 GDPR, Article 10 Law Enforcement Directive. Article 10 Regulation (EU) 2018/1725 (EU의 기관과 조직에 적용됨)

현행 EU의 데이터 보호 규칙과 '기본권헌장'에 따라 AI는 그러한 사용이 정당한 사유가 있고, 비례적이고, 적절한 안전장치의 적용을 받는 경우 원격 생체인식 식별 목적으로만 사용될 수 있다.

공공장소에서 그러한 목적으로 AI를 사용하는 것에 관한 사회적인 우려에 대응하고, 내부 시장의 분열을 피하기위해 본 위원회는 그러한 사용을 정당화할 수 있는 상황, 그리고 공통적인 안전장치에 대해 유럽의 광범위한 토의에 착수할 예정이다.

E. 대상자 (Addressees)

위에서 언급한 고위험 AI 애플리케이션에 관한 법적 요구사항의 대상자들에 관련하여, 고려해야 할 다음과 같은 두 가지 이슈가 있다.

첫째, 관련된 경제 주체들 간에 어떻게 의무를 분배하는지에 대한 의문이다. AI 시스템의 수명주기에 많은 경제 주체들이 관련되어 있다. 여기에는 개발자, 배치자(AI를 갖춘 제품/서비스를 사용하는 사람), 기타 (생산자, 유통업체/수입업체, 서비스 제공자, 전문/민간 사용자) 등이 포함된다.

본 위원회의 견해는 미래 규제 프레임워크에서 각 의무는 잠재적인 위험을 다루는데 가장 좋은 위치에 있는 주체에게 돌아가야 한다는 것이다. 예를 들면, AI의 개발자들이 개발 단계에서 발생하는 위험을 다루는데 가장 좋은 위치에 있을 수 있지만, 이들이 사용 단계에서의 위험을 통제할 수 있는 능력은 제한적일 수 있다. 그러한 경우, 배치자가 관련 의무를 맡아야 한다. 이것은 피해를 입은 최종 사용자나 기타 당사자의 책임, 그리고 정의의 효과적인 접근 목적에서 누가 야기된 피해에 책임을 져야하는지의 여부에 대한 의문을 침해하지 않는 것이다. EU의 제품 책임에 관한 법률하에서 결함을 가진 제품에 대한 책임은 생산자에게 속하는데, 이것은 다른 당사자들로부터 복구를 허용하는 국가별 법률을 침해하지 않는다.

둘째, 법규를 통한 개입의 지리적인 범위에 대한 의문이 존재한다. 본 위원회의 견해는 요구사항들은 EU에서 AI로 가능해진 제품/서비스를 제공하는 모든 관련 경제 주체들(EU에 수립되어 있느냐의 여부에 상관없이)에게 적용되는 것이 무엇보다 중요하다는 것이다. 그렇지 않으면, 앞에서 언급한 법규를 통한 개입의 목적을 충분히 달성할 수 없다.

F. 준수 및 집행

AI가 신뢰할만하고, 안전하고, 유럽의 가치와 규칙을 준수하도록 하기 위해서는 적용되는 법적 요구사항들이 실제로 준수되고, 권한을 가진 국가별 및 EU 당국, 그리고 피해 당사자들에 의해 효과적으로 집행되어야 한다. 권한을 가진 당국은 개별 사례를 조사할

뿐만 아니라 사회에 대한 영향을 평가할 수 있는 위치에 있어야 한다.

특정한 AI 애플리케이션이 시민들과 우리 사회에 높은 수준의 위험을 제기할 수 있다는 점을 고려하여, 이 단계에서 본 위원회는 위에서 언급한 고위험 애플리케이션에 적용되는 의무적인 요구사항(D항 참조)의 준수를 검증하고 준수되도록 하기 위해서 객관적이고 사전적인 준수 평가가 필요하다고 생각한다. 사전적인 준수 평가에는 테스트, 검사 또는 인증 절차가 포함될 수 있다.[59] 여기에는 개발 단계에서 사용된 알고리즘과 데이터 세트의 확인이 포함될 수 있다.
고위험 AI 애플리케이션에 대한 준수 평가는 EU의 내부 시장에 출시되고 있는 많은 수의 제품에 대해 이미 존재하고 있는 준수 평가 메커니즘의 일부이어야 한다. 그러한 의존할 수 있는 기존 메커니즘이 없는 경우, 베스트 프랙티스, 이해관계자 및 유럽 표준화 조직의 의견을 바탕으로 이와 유사한 메커니즘을 수립할 필요가 있다. 그러한 새로운 메커니즘은 국제 의무를 준수하는데 비례적이고, 비차별적이어야 하고, 투명하고 객관적인 기준을 사용해야 한다.

사전 준수 평가에 의존하는 시스템을 설계하고 구현할 때, 다음과 같은 사항들을 특별히 고려해야 한다.

- 위에서 제시한 모든 요구사항들이 사전 준수 평가를 통해 검증하는데 적합하지 않을 수도 있다. 예를 들면, 제공되어야 하는 정보에 대한 요구사항은 일반적으로 그러한 평가를 통한 검증에 적합하지 않다.
- 특정한 AI 시스템은 진화하고 경험으로부터 학습할 가능성이 있고, 해당 AI 시스템의 수명기간 동안 반복적인 평가가 필요할 수 있다는 것에 특별히 주의를 기울여야 한다.
- 훈련과 관련 프로그래밍에 사용된 데이터, AI 시스템을 구축/테스트/확인하는데 사용된 훈련 방법론, 프로세스 및 기업을 검증할 필요성
- 준수 평가 결과, AI 시스템이 예를 들면, 이것을 훈련하는데 사용된 데이터에 관련된 요구사항을 충족시키지 못하는 것으로 나타나면, 식별된 취약점은 예를 들면, 모든 적용되는 요구사항을 충족시킬 수 있는 방법으로 시스템을 재훈련시킴으로써 해결되어야 할 것이다.

준수 평가는 설립 장소에 상관없이 요구사항의 적용을 받는 모든 경제 주체들에게 의무적일 것이다,[60] 중소기업들의 부담을 제한하기 위해서는 '디지털혁신허브'를 통하는 것을 포함해서 특정한 지원 구조를 고려할 수 있다. 그 뿐만 아니라 표준과 전담 온라인 도구들은 준수를 촉진시킬 수 있다.

---

59) 참조: AI의 특성을 고려하여 Decision 768/2008/EC 또는 Regulation (EU) 2019/881 (Cybersecurity Act), 참조: Blue Guide on the Implementation of EU product rules, 2014.
60) 준수 평가를 수행하기로 지정된 기구를 포함하여 관련 거버넌스 기구에 대해서는 아래의 H항 참조.

사전 준수 평가는 권한을 가진 국가별 당국에 의한 모니터링 준수와 사후 집행을 침해해서는 안된다. 문제가 되는 애플리케이션이 고위험이라는 것이 권한을 가진 국가별 당국이 전자에 특별히 주의를 기울이는 이유일 수 있지만, 이것은 고위험 AI 애플리케이션뿐만 아니라 법적 요구사항의 적용을 받는 기타 AI 애플리케이션에 대해서도 유효하다. 사후 통제는 관련 AI 애플리케이션의 적절한 문서(위의 E항 참조), 그리고 해당하는 경우, 권한을 가진 당국과 같은 제삼자가 그러한 애플리케이션을 테스트할 가능성에 의해서 가능해져야 한다. 기본권에 위험이 발생하는 경우 이것은 특히 중요할 수 있다. 그러한 준수 모니터링은 지속적인 시장 감사 체계의 일부이어야 한다. 거버넌스에 관련된 측면은 아래의 H항에서 보다 자세하게 설명한다.

그뿐만 아니라, 고위험 AI 애플리케이션과 기타 AI 애플리케이션, 두 가지 모두에 대해, AI 시스템에 의해 부정적인 영향을 받은 당사자들에 대한 효과적인 사법적 보상이 이루어져야 한다. 책임에 관련된 이슈들은 본 백서에 동봉된 안전과 책임 프레임워크에 관한 보고서에서 보다 자세하게 설명한다.

G. 고위험 AI 애플리케이션에 대한 자발적인 표시(labelling)

'고위험'으로 분류되지 않고(위의 C항 참조), 따라서 위에서 설명한 의무적인 요구사항의 적용을 받지 않는(위의 D, E, F항 참조) AI 애플리케이션의 경우, 대안은 적용되는 법규에 추가하여 자발적인 표시(labelling) 체계를 수립하는 것이다.

이 체계하에서 의무적인 요구사항에 적용을 받지 않는 경제 주체는 자발적으로 그러한 요구사항의 적용을 받기로 하거나, 그렇지 않으면 자발적 체계의 목적을 위해 특별히 수립된 유사한 요구사항의 적용을 받기로 결정할 수 있다. 관련된 경제 주체는 그리고 나서 자신들의 AI 애플리케이션에 품질 표시를 부여할 것이다.

자발적 표시는 관련 경제 주체들이 AI로 가능해진 자신들의 제품/서비스가 신뢰할만하다는 것을 나타낼 수 있도록 해 줄 것이다. 이것은 사용자들이 문제의 제품/서비스가 특정한 목적과 표준화된 EU 전반의 벤치마크를 준수하여, 일반적으로 적용되는 법적 의무를 초과하고 있다는 것을 쉽게 인지하도록 해 줄 것이다. 이것은 AI 시스템에 대한 사용자들의 신뢰를 제고하고, AI의 전반적인 수용을 촉진하는 것을 도울 것이다.
이 안은 고위험으로 간주되지 않는 AI 시스템의 개발자 및/혹은 배치자를 위한 자발적인 표시 프레임워크를 제시하는 새로운 법적 수단의 창출을 수반할 것이다. 표시 체계에 참여하는 것은 자발적이지만, 개발자나 배치자가 일단 표시를 사용하기로 선택하면, 요구사항은 구속력을 가질 것이다. 모든 요구사항들이 준수되도록 하기 위해서는 사전(ex ante) 및 사후(ex post) 집행의 결합이 필요할 것이다.

H. 거버넌스

책임의 분편화를 피하고, 회원국들의 역량을 높이고, 유럽이 AI로 가능해진 제품/서비스를 테스트하는데 필요한 역량을 갖추도록 하기 위해서는 권한을 가진 국가별 당국들의 협력을 위한 프레임워크 형식의 AI에 대한 유럽의 거버넌스 기구가 필요하다. 이러한 관점에서 권한을 가진 국가별 당국들이 AI가 사용되는 곳에서 자신들의 임무를 완수할 수 있도록 이들을 지원하는 것이 유익할 것이다.

유럽의 거버넌스 기구는 다양한 과업을 수행할 수 있다. 여기에는 정보와 베스트 프랙티스를 정기적으로 교환하기 위한 포럼, 새로이 등장하는 추세의 식별, 표준화 활동과 인증에 대한 자문 등이 포함된다. 또한 이 기구는 지침, 의견, 전문성 등의 발표를 통해서 법적 프레임워크의 구현을 촉진하는데 핵심적인 역할을 수행해야 한다. 그러한 취지에서 이 기구는 국가별 당국의 네트워크, 국가 및 EU 수준의 사회적 네트워크와 규제 당국에 의존해야 한다. 그 뿐만 아니라 전문가들로 구성된 위원회가 본 위원회를 지원할 수 있다.

이 거버넌스 기구는 이해관계자들의 참여를 최대한 보장해야 한다. 프레임워크의 구현 및 추가적인 발전에 대해서 이해관계자들(소비자 단체 및 사회적 파트너, 기업, 연구자, 시민사회 단체)과 협의해야 한다.

금융, 제약, 항공, 의료기기, 소비자 보호, 데이터 보호 등과 같은 분야에 기구들이 이미 존재하고 있다는 사실을 감안하면, 제안된 거버넌스 기구는 기존의 기능과 중복되어서는 안된다. 다양한 부문의 기타 EU 및 국가별 당국들과 긴밀한 연결 관계를 수립하여, 기존의 전문성을 보완하고, 기존 당국들이 AI 시스템과 AI로 가능해진 제품/서비스에 관련된 경제 주체들의 활동을 모니터링하고 감독하는 것을 도와야 한다.
마지막으로 이 안을 추진한다면, 준수 평가의 수행은 회원국들이 지정하는 인증기관에 위임할 수 있다. 테스트 센터들은 위에서 설명한 요구사항에 따라 AI 시스템에 대한 독립적인 감사와 평가가 가능하도록 해야 한다. 독립적인 평가는 신뢰를 높이고 객관성이 확보되도록 할 것이다. 또한 이것은 권한을 가진 관련 당국의 작업을 촉진할 수 있다.

EU는 우수한 테스트 및 평가 센터들을 향유하고 있고, 그 역량을 AI 분야에도 개발해야 한다. 내부 시장에 진입하기를 원하는 제삼국에 수립한 경제 주체들은 EU에 수립된 지정된 기관들을 이용하거나, 제삼국과의 상호인정협정의 적용을 받아 그러한 평가를 수행하기로 지정된 제삼국의 기관들을 이용할 수 있다.

AI에 관련된 거버넌스 기구, 그리고 여기에서 문제가 되는 준수 평가는 기존의 EU 법률하에서 특정한 산업이나 특정한 이슈(금융, 제약, 항공, 의료 기기, 소비자 보호, 데이터 보호 등)에 관련된 권한을 가진 당국의 힘과 책임에 영향을 미치지 않을 것이다.

6. 결론

AI는 인간 중심적이고, 윤리적이고, 지속 가능하고, 기본적인 권리와 가치를 존중한다면, 시민, 기업, 사회 전반에 많은 혜택을 제공하는 전략적 기술이다. AI는 유럽 산업의 경쟁력을 강화하고, 시민들의 복지를 향상시킬 수 있는 중대한 효율성 및 생산성의 향상을 제공한다. 또한 AI는 기후 변화 및 환경 악화와의 싸움, 지속가능성 및 인구통계적 변화에 연관된 난제, 민주주의의 보호, 범죄와의 싸움 등을 포함하여, 가장 긴급한 사회적인 난제에 대한 해결방안을 발견하는데 기여할 수 있다.

유럽이 AI가 제공하는 기회를 완전하게 포착하기 위해서는 필요한 산업적/기술적 역량을 개발하고 강화해야 한다. 동봉된 유럽의 데이터 전략에서 명시하고 있는 바와 같이, 이를 위해서는 EU가 데이터의 글로벌 중심(hub)이 될 수 있도록 하기 위한 조치가 필요하다.

AI에 대한 유럽의 접근방법의 목표는 AI 분야에서 유럽의 혁신 역량을 육성하고, 이와 동시에 EU 경제 전반에 걸쳐서 윤리적이고 신뢰성 있는 AI를 개발하고 수용하는 것을 지원하는 것이다. AI는 사람들을 위해 일해야 하고, 사회의 선을 위한 힘이 되어야 한다.

본 백서와 여기에 동봉된 안전과 책임 프레임워크에 관한 보고서를 가지고 본 위원회는 AI에 대한 유럽의 접근방법에 대한 구체적인 제안에 대해 회원국들의 시민 사회, 산업 및 학계 등과 광범위한 협의에 착수할 예정이다. 여기에는 연구 및 혁신에 대한 투자를 확대하고, 스킬의 개발을 향상시키고, 중소기업들의 AI 도입을 지원하기 위한 정책 수단, 그리고 미래 규제 프레임워크의 핵심 요소들에 대한 제안이 포함된다. 이러한 협의를 통해 모든 관련 당사자들과의 종합적인 대화가 가능할 것이고, 이를 통해 본 위원회가 취해야 할 다음 단계에 대한 정보를 제공할 것이다. ▢

자료

EUROPEAN
COMMISSION
(유럽연합집행위원회)

Brussels, 2020년 2월 19일
COM(2020) 64 final

유럽위원회가 유럽의회에 제출한 보고서
위원회 및 유럽경제사회소위원회
(REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL AND THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE)

인공지능(AI), 사물인터넷(IoT), 로보틱스(robotics)의 안전과 책임의 함의에 관한 보고서

(Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics)

번역제공 : 한국지능정보사회진흥원(NIA)

1. 서론

인공지능(Artificial Intelligence: AI)[1], 사물인터넷(Internet of Things: IoT)[2], 로보틱스(robotics)는 우리 사회에 새로운 기회와 혜택을 제공할 것이다. 본 위원회는 이러한 기술의 중요성/잠재력, 그리고 이러한 영역에 상당한 투자가 필요하다는 사실을 인식하고 있다.[3] 본 위원회는 AI, IoT 및 로보틱스 분야에서 유럽을 세계의 리더로 만들 의지를 가지고 있다. 이러한 목표를 달성하기 위해서는 기술적인 도전에 대응할 수 있는 명확하고 예측가능한 법적 프레임워크가 필요하다.

1.1. 안전 및 책임에 관한 기존 프레임워크

안전 및 책임에 관한 법적 프레임워크의 전반적인 목적은 새로이 출현하는 디지털 기술들을 집약하고 있는 제품/서비스를 포함하여, 모든 제품/서비스가 안전하고/신뢰성 있고/일관성 있게 운영되고, 발생되는 피해가 효율적으로 구제되도록 하는 것이다. 새로운 디지털 기술을 집약하고 있는 제품/시스템에 대한 높은 수준의 안전과 발생한 피해를 해결하는 견고한 메커니즘(즉, 책임 프레임워크)은 소비자들을 보다 잘 보호하는데 기여한다. 또한 이것은 이러한 기술에 대한 신뢰를 창출하고, 이러한 신뢰는 산업과 사용자들이 이러한 기술을 수용하는데 있어서의 전제 조건이다. 또한 이러한 기술의 수용은 우리 산업의 경쟁력을 활용하고, 유럽연합의 목적 달성에 기여할 것이다.[4] 소비자 보호와 기업에 대한 법적 확실성을 보장하기 위한 목적을 가진 명확한 안전 및 책임에 관한 프레임워크는 AI, IoT, 로보틱스와 같은 새로운 기술이 출현할 때 특히 중요하다.

유럽연합은 견고하고 신뢰성 있는 안전 및 제품 책임에 관한 법규 프레임워크, 그리고 국가별로 일치화되지 않은 책임 법규들로 보완되고 있는 건실한 안전 기준들을 가지고 있다. 이것들은 단일 시장에서 우리 시민들의 복지를 보장하고, 혁신과 기술의 도입을 장려하고 있다. 그러나 AI, IoT, 로보틱스는 많은 제품/서비스의 특성을 변혁시키고 있다.

2018년 4월 25일에 채택된 유럽의 AI에 대한 공보(Communication on Artificial Intelligence for Europe)[5]에서는 본 위원회가 새로이 출현하는 디지털 기술들이 안전과

---

1) AI 고위전문가그룹의 AI에 대한 정의는 다음을 참조. https://ec.europa.eu/ futurium/en/ai-alliance-consultation/guidelines
2) Recommendation ITU-T Y.2060에서 제시하는 IoT의 정의는 다음을 참조. https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=y.2060
3) SWD(2016) 110, COM(2017) 9, COM(2018) 237 and COM(2018) 795.
4) http://ec.europa.eu/growth/industry/policy_en
5) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN. 여기에 동봉된

책임에 관한 기존 프레임워크에 미치는 파급효과를 평가하는 보고서를 제출할 것이라고 발표하였다. 이 보고서의 목적은 AI, IoT, 로보틱스에 대한 책임 및 안전에 관한 프레임워크의 광범위한 함의와 잠재적인 갭을 식별하고 조사하는 것이다. AI에 대한 백서(White Paper on Artificial Intelligence)에 동봉된 본 보고서에서 제시하는 방향은 논의를 위해 제공되었고, 이해관계자들과 광범위하게 협의한 결과의 일부이다. 안전에 관한 절의 내용은 기계지침(Machinery Directive)[6] 의 평가 결과[7]와 관련 전문가 그룹들과의 작업 결과[8]를 기반으로 하고 있다. 책임에 관한 절의 내용은 제품책임지침(Product Liability Directive)[9]의 평가 결과[10], 관련 전문가 그룹들의 의견[11], 이해관계자들과의 접촉 결과를 바탕으로 하고 있다. 본 보고서의 목적은 안전과 책임에 대한 기존 규칙들을 모두 분석하는 것이 아니라, 지금까지 식별된 핵심 이슈에 초점을 맞추는 것이다.

1.2 AI, IoT, 로보틱스의 특성

AI, IoT, 로보틱스는 많은 특성들을 공유한다. 이것들은 **연결성(connectivity)**, **자율성(autonomy)**, **데이터 의존성(data dependency)**을 결합하여, 인간의 통제나 감독이 거의 없거나 전혀 없이 과업을 수행할 수 있다. 또한 AI를 갖춘 시스템들은 경험으로부터 학습함으로써 자신들의 성능을 향상시킬 수 있다. 이들의 **복잡성**은 **공급망**에 관여하는 경제 주체의 복수성, 그리고 구성요소/부품/소프트웨어/시스템/서비스의 복수성에서 나타나는데, 이들이 결합하여 새로운 기술 생태계를 형성한다. 여기에 추가되는 것은 이들이 출시된 이후에 이루어지는 업데이트와 업그레이드에 대한 **개방성(openness)**이다. 관련된 데이터의 엄청난 양, 알고리즘에 대한 의존성, AI 의사결정의 **불투명성(opacity)**

Staff Working Document (2018) 137 (https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52018SC0137)에서는 새로이 출현하는 디지털 기술의 상황에서 발생하는 책임 문제에 대한 최초의 매핑을 제시하였다.

6) Directive 2006/42/EC

7) SWD(2018) 161 final.

8) 일반제품안전에 관한 Directive 2001/95/EC, Machinery Directive 2006/42/EC, Radio Equipment 2014/53/EU에 의해서 수립된 소비자 안전 네트워크(Consumer Safety Network). 지침전문가그룹은 회원국, 산업, 소비자단체 등과 같은 기타 이해관계자들로 구성되어 있다.

9) Directive 85/374/EEC

10) COM(2018) 246 final

11) 본 위원회에 제품책임지침과 국가별 민사책임 규칙의 적용성에 대한 전문성, 그리고 신기술에 관련된 법률의 수정을 위한 지침이 되는 원칙을 개발하는데 지원을 제공하기 위해 책임및신기술 전문가 그룹(Expert Group on Liability and New Technologies)이 수립되었다. 이 그룹은 '제품책임 분과'와 '신기술 분과' 등 두 개의 분과로 구성되어 있다. 참조: http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.group Detail&groupID = 3592&NewSearch=1&NewSearch=1.
AI 및 기타 새로이 출현하는 기술에 대한 책임에 관한 '신기술 분과'의 보고서는 다음을 참조: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=63199.

등은 AI로 가능해 진 제품의 행태를 예측하고, 피해의 잠재적인 원인을 파악하는 것을 더욱 어렵게 만들고 있다. 마지막으로 연결성과 개방성은 AI와 IoT 제품을 **사이버 위협**(cyber- threats)에 노출시킬 수도 있다.

1.3 AI, IoT, 로보틱스에 의해 창출되는 기회

새로이 출현하는 기술에 대한 사용자들의 신뢰와 사회적인 수용 증가, 제품/프로세스/비즈니스 모델의 향상, 유럽 제조업체들이 보다 효율화 되도록 도와주는 것 등은 AI, IoT, 로보틱스에 의해 창출되는 기회의 일부에 불과하다.

AI는 생산성 및 효율성의 향상을 넘어서 인간이 아직까지 도달해보지 못한 지능의 개발을 약속함으로써, 새로운 발견을 위한 문을 열고, 전세계적으로 가장 큰 해결과제들(예: 만성질환 치료에서부터 질병 발생 예측이나 교통사고로 인한 사망률 감소, 기후변화에 대한 대응이나 사이버 보안 위협 예측에 이르기까지)을 해결하는 것을 도와주고 있다.

이러한 기술들은 제품의 안전을 향상시키고, 특정 위험에 덜 노출되게 함으로써 많은 혜택을 제공할 수 있다. 예를 들면, 연결되어 있고 자동화된 차량은 도로의 안전을 향상시킬 수 있다. 왜냐하면, 현재 대부분의 교통 사고는 인간의 오류로 인해서 발생하고 있기 때문이다.[12] 그 뿐만 아니라, IoT 시스템은 서로 다른 소스로부터 막대한 양의 데이터를 접수하고 처리하도록 제작되고 있다. 이러한 증가된 수준의 정보는 제품이 스스로 적응하고, 결과적으로 더 안전하게 될 수 있도록 사용된다. 새로운 기술들은 예를 들면, 안전 문제를 피할 수 있도록 제품이 사용자들에게 경고를 보내는 것과 같이 제품 리콜의 효과성을 향상시키는데 기여할 수 있다.[13] 연결되어 있는 제품을 사용하는 동안에 안전에 관한 이슈가 발생하면, 생산자는 사용자와 직접 의사소통하여, 한편으로는 사용자에게 위험에 대해 경고를 보내고, 또 다른 한편으로는 예를 들면, 안전 업데이트를 제공함으로써 문제를 직접 해결할 수도 있다. 예를 들면, 한 스마트폰 생산자는 2017년에 자신들의 한 기기를 리콜하면서, 사용자들이 위험한 기기를 사용하지 못하도록 리콜한 전화기의 배터리 용량을 제로로 만드는 소프트웨어 업데이트를 수행하였다.[14]

---

12) 교통사고의 90% 정도는 인간의 오류에 의해 발생하는 것으로 추산되고 있다. 본 위원회의 다음 보고서 참조 - 인명 구조: EU에서의 차량 안전성 향상(Saving Lives: Boosting Car Safety in the EU: COM(2016) 0787 final).

13) 예를 들면, 전방에 사고가 난 경우 자동차 운전자는 감속하라는 경고를 받을 수 있다.

14) OECD (2018), "Measuring and maximising the impact of product recalls globally: OECD workshop report", OECD Science, Technology and Industry Policy Papers, No. 56, OECD Publishing, Paris, https://doi.org/10.1787/ab757416-en.

그 뿐만 아니라, 새로운 기술은 제품의 추적성을 향상시키는데 기여할 수 있다. 예를 들면, IoT의 연결 기능은 기업과 시장 감독기관들이 위험한 제품을 추적하고, 공급망 전반에 걸쳐서 위험을 식별할 수 있게 해 준다.[15)]

AI, IoT, 로보틱스는 우리 사회 및 경제에 기회를 제공할 뿐만 아니라, 법적으로 보호받고 있는 이익(유형 및 무형)을 해칠 수 있는 위험을 창출하기도 한다. 이러한 해가 발생할 위험은 적용 분야가 넓어지면서 더욱 증가할 것이다. 이러한 관점에서 안전과 책임에 관한 현재의 법적 프레임워크가 여전히 사용자들을 보호하는데 적합한지의 여부와 적합한 정도를 분석하는 것은 필수적인 일이다.

2. 안전

본 위원회의 "인간 중심적 AI에서의 신뢰 구축(Building Trust in Human-Centric Artificial Intelligence)" 공보에서는 *관련된 모든 당사자들의* ***물리적/정신적 안전을 감안하여, AI 시스템이 모든 단계에서 검증 가능하게 안전하도록, AI 시스템은 안전과 보안 내재화(security-by-design) 메커니즘을 통합해야 한다***고 언급하고 있다.[16)]

본 절의 유럽연합 제품안전에 관한 법규의 평가에서는 유럽연합의 현재 법규 프레임워크가 새로이 출현하는 기술, 그 중에서도 특히 AI 시스템이 안전과 보안 내재화를 통합하도록 관련 요소들을 담고 있는지의 여부를 분석한다.

본 보고서는 주로 일반제품안전지침(General Product Safety Directive)[17)], "새로운 접근방법(New Approach)"[18)] 및/혹은 "새로운 법규 프레임워크(New Legislative Framework)" (이하 "유럽연합 제품안전 법규 또는 프레임워크")[19)]의 수평 규칙을 따르는 일치화된 제품 법규를 검토한다. 수평 규칙들은 제품 안전에 관한 산업별 규칙들 간의 일관성을 보장한다.

---

15) OECD (2018), "Enhancing product recall effectiveness globally: OECD background report", OECD Science, Technology and Industry Policy Papers, No. 58, OECD Publishing, Paris, https://doi.org/10.1787/ef71935c-en.

16) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Building Trust in Human-Centric Artificial Intelligence, Brussels, 8.4.2019 COM(2019) 168 final

17) 유럽의회와 2001년 12월 3일자 위원회의 일반제품안전에 관한 Directive 2001/95/EC(OJ L 11, 15.1.2002, p. 4-17).

18) OJ C 136, 4.6.1985, p. 1.

19) Regulation (EC) No. 2008/765, Decision (EC) No. 2008/768

유럽연합 제품안전 법규의 목적은 유럽연합 시장에 출시된 제품들이 건강/안전/환경에 관한 요구사항을 충족시키고, 그러한 제품들이 유럽연합 전체에 자유롭게 유통될 수 있도록 하는 것이다. 산업별 법규[20]는 일반제품안전지침[21]으로 보완되고 있는데, 이 지침은 모든 소비자 제품은 유럽연합의 산업별 법규에 의해서 규제되지 않더라도 안전해야 하는 것을 의무화하고 있다. 안전 규칙들은 시장감시규정(Market Surveillance Regulation)[22] 및 일반제품안전지침[23] 하에서 국가별 당국에 부여된 시장 감시 및 공권력으로 보완되고 있다. 운송 분야의 경우, 자동차[24]/항공기/선박의 서비스 개시에 대한 유럽연합 및 국가별 추가적인 규칙들이 존재하고, 운영자의 과업과 당국의 감시 활동을 포함하여 운영 과정에서 안전을 거버넌스하는 명확한 규칙들이 존재한다.

또한 유럽의 표준화는 유럽연합 제품안전 법규의 핵심 요소이다. 디지털화와 새로이 출현하는 디지털 기술의 글로벌한 특성을 감안하면, 표준화에 대한 국제 협력은 유럽 산업의 경쟁력과 특히 관련성이 높다.

유럽연합 제품안전 프레임워크의 많은 부분은 AI, IoT, 로보틱스와 같은 디지털 기술의 출현 이전에 작성되었다. 따라서 이러한 프레임워크들이 이처럼 새로이 출현하는 기술의 새로운 해결과제와 위험을 명확하게 다루는 조항을 항상 담고있는 것은 아니다. 그러나 기존의 제품안전에 관한 프레임워크는 기술 중립적이므로, 이러한 기술들을 포함하고 있는 제품에 적용되지 않을 것이라는 것을 의미하는 것은 아니다. 그 뿐만 아니라, 이러한 프레임워크의 일부인 이후의 법규 (예: 의료 기기 또는 자동차 산업)는 이미 디지털 기술의 등장에 관한 일부 측면들(예: 자동화된 의사결정, 별도 제품으로서의 소프트웨어, 연결성)을 명시적으로 고려하고 있다.

---

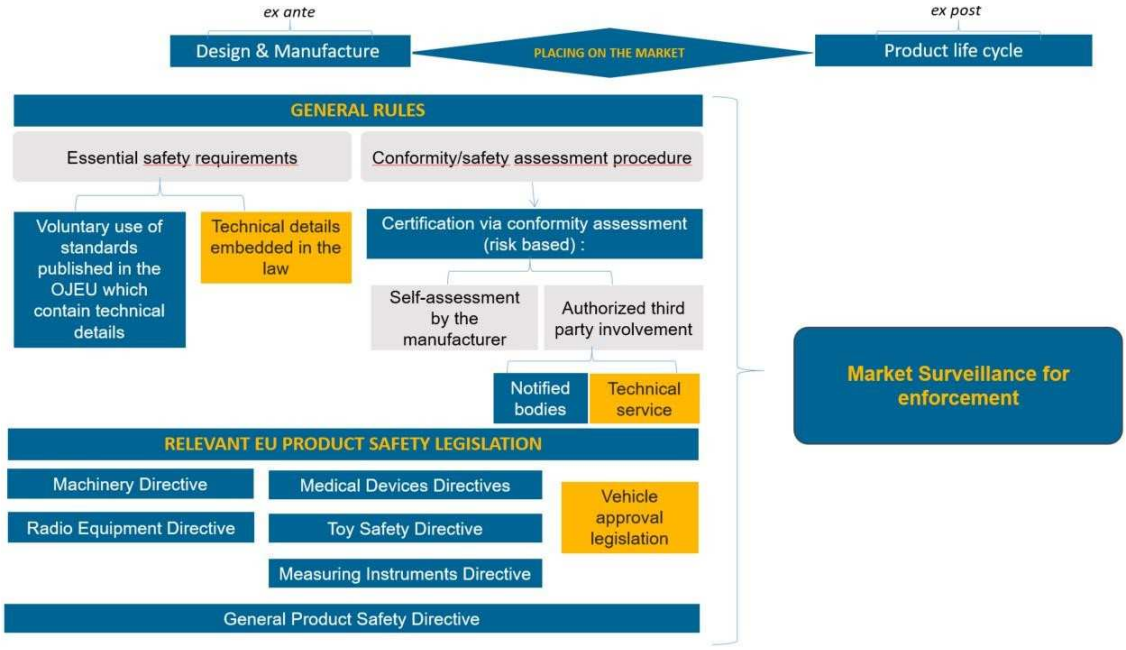
20) 이 스키마에는 유럽연합의 운송 및 자동차 법규는 포함되어 있지 않다.

21) 2001년 12월 3일자 일반 제품 안전에 관한 지침 - Directive 2001/95/EC the European Parliament and of the Council (OJ L 11, 15.1.2002, p. 4-17).

22) 제품의 마케팅에 관한 인정 및 시장 감시에 대한 요구사항을 규정하고, Regulation (EEC) No 339/93, OJ L 218, 13.8.2008, p. 30-47, ELI: http://data.europa.eu/eli/reg /2008/765/oj를 폐지하는 유럽의회 및 2008년 7월 9일자 위원회의 Regulation (EC) No 765/2008, 그리고 2021년부터는 제품의 마케팅에 관한 인정 및 시장 감시에 대한 유럽의회 및 2008년 7월 9일자 위원회의 Regulation (EU) 2019/1020 (Directive 2004/42/EC and Regulations (EC) No 765/2008 및 (EU) No 305/201,1 OJ L 169, 25 .6.2019, p. 1-44, ELI: http://data.europa.eu/eli/reg/2019/1020/oj의 수정).

23) 일반제품안전지침의 article 8 (1) (b) (3)

24) 예를 들면, Directive 2007/46/EC - 자동차 및 트레일러, 그러한 자동차를 위한 시스템/구성요소/별도기술 단위 등의 승인, 자동차 및 트레일러, 그러한 자동차를 위한 시스템/구성요소/별도기술 단위 등의 승인과 시장 감시에 관한 유럽의회 및 2018년 5월 30일자 위원회의 Regulation (EU) 2018/858 (Regulations (EC) No 715/2007과 (EC) No 595/2009의 수정 및 Directive 2007/46/EC의 폐지)



The underlying logic of the current Union product safety legislation[25]

****

(사전) (사후)

설계 및 제조 ---------- 시장 출시 ---------- 제품 수명주기

<일반 규칙>

▣안전에 관한 필수 요구사항

○ 기술적 세부사항을 담고 있는 OJEU에
공표된 표준의 자발적인 사용

○ 법률에 내재되어 있는 기술적 세부사항

▣적합성/안전 평가 절차

○ 적합성 평가를 통한 인증(위험 기반)

- 제조자에 의한 자가 평가
- 승인된 제삼자의 관여
  . 인증 기관
  . 기술 서비스

<집행을 위한 시장감시>

<관련 EU 제품 안전 법규>

.기계 지침 .의료 기기 지침

.라디오 장비 지침 .장난감 안전 지침 .자동차 승인 법규

.측정도구 지침

일반제품안전지침

현행 유럽연합 제품안전 법규의 기반 논리[25)]

****

아래에서는 디지털 신기술이 유럽연합의 제품안전 프레임워크에 야기하는 도전에 대해 설명한다.

연결성(Connectivity)은 지속적으로 증가하고 있는 제품/서비스의 핵심적인 특성이다. 이 특성은 안전에 대한 전통적인 개념에 이의를 제기한다. 왜냐하면, 연결성은 제품의 안전을 직접적으로 해치고, 해킹되었을 때 안전을 간접적으로 해쳐서, 보안 위협을 야기하고, 사용자의 안전에 영향을 미치기 때문이다.

한 예로 아동용 스마트 시계에 관한 Iceland의 EU 급속 경보 시스템(EU Rapid Alert System)을 통한 통보를 들 수 있다.[26)] 이 제품은 시계를 차고 있는 아동에게 직접적인 해를 끼치지는 않지만, 최소 수준의 보안이 미흡하여 아동을 접근할 수 있는 도구로 쉽게 사용될 수 있다. 이 제품의 원래 기능 중의 하나는 위치 확인(localisation)을 통해서 아동을 안전하게 지키는 것이므로, 소비자는 이것이 아동을 추적하거나 누군가에 의한 접촉을 통해서 아동의 안전에 영향을 미칠 수 있는 보안 위협이 제기되리라고는 생각하지 않을 것이다.

또 다른 예로는 승용차에 관련하여 독일이 제출한 보고를 들 수 있다.[27)] 자동차의 라디오는 승인받지 않은 제삼자가 자동차의 상호연결된 통제 시스템을 접근할 수 있는 소프트웨어 보안 상의 갭을 가지고 있을 수 있다. 이러한 소프트웨어의 보안 갭을 제삼자가 악의적인 목적으로 활용하면, 교통 사고가 발생할 수 있다.

산업에서의 적용 또한 이러한 적용에 필요한 수준의 보안이 미흡할 때 많은 사람들의

---

25) 이 그림에는 제품 수명주기에 관한 법규 요구사항(즉, 사용 및 유지보수)이 포함되어 있지 않고, 단지 일반적인 예를 나타내고 있다

26) EU Safety Gate의 웹사이트에 게시된 Iceland의 RAPEX 통보(A12/0157/19).

27) EU Safety Gate의 웹사이트에 게시된 독일의 RAPEX 통보(A12/1671/15).

안전에 영향을 미치는 사이버 위협에 노출될 수 있다. 이것은 예를 들면 폭발을 촉발시켜 많은 사람들의 생명을 앗아갈 수 있는 산업 시설의 핵심 통제 시스템에 대한 사이버 공격의 경우가 될 수 있다.

유럽연합 제품안전에 관한 법규는 사용자들의 안전에 영향을 미치는 사이버 위협에 대해 구체적인 의무적 필수 요구사항을 제시하지 않고 있다. 그러나 의료기기에 대한 규정(Regulation on Medical Devices)[28], 측정도구에 관한 지침(Directive on measuring instruments)[29], 라디오 장비 지침(Radio Equipment Directive)[30], 또는 자동차 유형에 대한 승인 법규 or the vehicle-type approval legislation[31] 등에 보안에 관련된 조항이 존재한다.

사이버 보안법(Cybersecurity Act)[32]은 정보통신기술(ICT) 제품/서비스/프로세스에 대한 자발적인 사이버보안 인증 프레임워크를 제시하고 있고, 관련된 유럽연합 제품안전 법규는 의무 요구사항을 규정하고 있다.

그 뿐만 아니라 새로이 출현하는 기술의 연결성 손실의 위험 또한 안전에 관련된 위험이 수반될 수 있다. 예를 들면, 연결되어 있는 화재 경보기가 연결성을 잃으면, 화재가 발생한 경우에 이를 사용자에게 알리지 못할 수 있다.

현행 유럽연합 제품안전 법규에서 안전은 공공 정책적인 목적이다. 안전의 개념은 제품의 사용, 그리고 제품을 안전하게 만들기 위해 처리해야 할 위험(예: 기계적/전기적 등)과 연결되어 있다. 유럽연합의 제품안전 법규에 따르면, 제품의 사용은 계획된 사용뿐만 아니라 예측 가능한 사용, 그리고 기계지침[33]에서와 같이 어떤 경우에는 을 합리적으로 예측 가능한 오용조차도 커버한다는 점에 유의하기 바란다.

---

28) 의료기기에 관한 Regulation (EU) 2017/745.
29) 측정도구 시장에 관한 Directive 2014/32/EU.
30) Radio Equipment 2014/53/EU Directive.
31) Directive 2007/46/EC — 자동차 및 트레일러, 그러한 자동차를 위한 시스템/구성요소/ 별도기술 단위 등의 승인, 자동차 및 및 트레일러, 그러한 자동차를 위한 시스템/구성요소/ 별도기술 단위 등의 승인. 이 지침은 폐지되고, 자동차 및 트레일러, 그러한 자동차를 위한 시스템/구성요소/별도기술 단위 등의 승인에 관한 Regulation (EU) 2018/858 (Regulations (EC) No 715/2007과 (EC) No 595/2009의 수정 및 2020년 9월 1일자로 Directive 2007/46/EC의 폐지)로 대체될 것이다.
32) Regulation (EU) 2019/881.
33) 기계에 관한 Directive 2006/42/EC.

유럽연합의 현행 제품안전 법규에서의 안전 개념은 소비자와 사용자를 보호하기 위한 확장된 안전 개념과 일맥상통한다. 따라서 제품 안전의 개념에는 기계적/화학적/전기적 위험뿐만 아니라 사이버 위험과 기기의 연결성 손실에 관련된 위험을 포함하여, 제품으로부터 야기되는 모든 종류의 위험에 대한 보호가 포함된다.

이러한 관점에서 사용자들을 보다 잘 보호하고, 보다 나은 법적인 확실성을 제공할 수 있도록 유럽연합의 관련 법규의 범위에 명시적인 조항을 포함시키는 것을 고려할 수 있다.

자율성(Autonomy)[34]은 AI의 주요한 특성 중의 하나이다. AI 기반의 계획하지 않은 결과는 사용자 및 노출된 사람들에게 해를 끼칠 수 있다.

제품이 출시되기 전에 제조자에 의해 수행된 위험 평가에 의해서 AI 제품의 미래 "행동"이 미리 결정되므로, 유럽연합의 제품안전 프레임워크는 생산자들이 위험 평가에서 제품의 수명주기 전반에 걸쳐서 제품의 "사용"[35]을 고려해야 하는 의무를 이미 제시하고 있다. 또한 이 프레임워크는 제조자들이 사용자들을 위한 설명/안전 정보 또는 경고를 제공해야 한다고 이미 적시하고 있다.[36] 이러한 관점에서, 예를 들면, 라디오장비지침[37]에서는 제조자가 계획된 사용에 따라 라디오 장비를 사용하는 방법에 대한 정보와 함께 설명을 포함하도록 의무화하고 있다.

또한 미래에는 AI 시스템의 결과를 미리 완전하게 파악할 수 없는 상황이 있을 수도 있다. 그러한 상황에서는 제품을 출시하기 전에 수행한 위험 평가는 더 이상 제품의 사용, 기능 또는 행동을 반영하지 못할 수 있다. 이러한 경우, 자율적인 행동으로 인해 제조자들이 최초에 예측했던 계획된 사용이 수정되고[38], 안전 요구사항의 준수가 영향을

---

34) AI 기반 제품들은 자신들의 환경을 인지하고, 미리 정해진 일련의 지시를 따르지 않고 자율적으로 행동할 수 있지만, 이들의 행위는 주어진 목표와 개발자의 설계안에 의해 제약된다.

35) 유럽연합의 제품안전 법규에서 생산자는 제품의 계획된 사용, 예측 가능한 사용 및/혹은 합리적으로 예측 가능한 오용 등을 바탕으로 위험 평가를 수행한다.

36) 유럽의회와 2008년 7월 9일자 위원회의 제품의 마케팅에 대한 공통 프레임워크에 관한 Decision No 768/2008/EC. 이것은 Council Decision 93/465/EEC, OJ L 218, 13.8.2008. p. 82-128. Annex I, Article R2.7을 폐지한다. 폐지된 조항의 내용은 "관련 회원국이 결정한 바와 같이, 제조자는 소비자와 기타 최종 사용자들이 쉽게 이해할 수 있는 언어로 작성된 설명/안전 정보가 제품에 수반되도록 해야 한다"

37) Article 10 (8)은 최종 사용자에 대한 설명을 언급하고, Annex VI는 EU의 적합성 선언(EU Declaration of Conformity)을 언급하고 있다.

38) 현재까지 AI 상황에서 "자율 학습"은 기계는 훈련 과정에서 학습할 수 있다는 것을 나타내는데 사용되고 있다. 즉, AI 기계가 배치된 이후에 지속적으로 학습해야 한다는 것이 아직까지 요구사항은 아니다. 이와 반대로, 특히 의료 서비스 분야에서는 AI 기계는 훈련이 성공적으로 종료된 이후에는 학습을 멈춘다. 따라서 이 단계에서 AI 시스템으로부터 야기된 자율적인 행동은 이 제품이 개발자들이 예측하지 못한 과업을 수행하고 있다는 것을 의미하는 것은 아니다.

받게되면, 자율 학습 제품에 대한 새로운 재평가가 필요한 것으로 생각될 수 있다.[39]

생산자가 자신의 제품이 수명주기 전반에 걸쳐서 안전에 영향을 미치는 위험을 야기하고 있다는 것을 알게 된 경우, 현행 프레임워크 하에서는 생산자는 즉시 권한이 있는 당국에 알리고, 사용자에 대한 위험을 방지하기 위한 조치를 취해야 하는 것이 이미 의무화되어 있다.[40]

> 제품을 출시하기 전에 수행한 위험 평가 이외에도, 제품이 수명주기 동안에 중요한 변경이 되는 경우(최초 위험 평가에서 제조자가 예측하지 못한 제품의 다른 기능), 새로운 위험 평가 절차가 수립될 수 있다. 이것은 제품의 수명주기 전반에 걸쳐서 자율 행동이 안전에 미치는 영향에 초점을 맞추어야 한다. 위험 평가는 적절한 경제 주체에 의해서 수행되어야 한다. 그 뿐만 아니라, 유럽연합의 관련 법규는 제조자에게 사용자에 대한 설명/경고에 대해 강화된 요구사항을 포함시킬 수 있다.
> 이미 운송 분야의 법규에는 이와 유사한 위험 평가가 의무화되어 있다.[41] 예를 들면, 철도 운송 분야의 법규에서는 열차가 인증 이후에 수정되는 경우, 수정의 입안자에게 특정한 절차가 적용되고, 당국이 관여할 필요가 있는지의 여부를 결정하기 위한 명확한 기준이 정의되어 있다.

AI 제품/시스템의 자율 학습 기능은 기계가 생산자가 원래 의도했고, 결과적으로 사용자들이 예상했던 것에서 벗어나는 의사결정을 내리게 할 수도 있다. 이것은 AI 제품/시스템에 의사결정을 위임할지의 여부와 위임 방법을 선택하여, 인간이 선택한 목적을 달성할 수 있도록, 인간의 통제에 대해 의문을 제기한다.[42] 유럽연합의 제품안전에 관한 기존의 법규는 AI 자율 학습 제품/시스템의 상황에서 인간의 감독을 명시적으로 다루지 않고 있다.[43]

---

39) 이것은 EU의 제품 규칙 2016의 실행에 관한 'Blue Guide'의 2.1절과 일치한다.
40) 유럽의회 및 2001sus 12월 3일자, 일반 제품의 안전에 관한 Directive 2001/95/EC 의 Article 5.
41) 안전에 영향을 미칠 수 있는 철도 시스템에 대한 변경의 경우(예: 운영 또는 유지보수 프로세스에 영향을 미칠 수 있는 기술적/운영적 변경 또는 조직적 변경), 따라야 할 프로세스는 COM Implementing 규정 (EU) 2015/1136의 Annex I (OJ L 185, 14.7.2015, p. 6)에 서술되어 있다.
'중대한 변경'의 경우, '평가 기구(국가 안전 당국 또는 기타 기술적으로 권한을 가진 기관)에 의해서 변경 제안자에게 안전 평가 보고서가 제공되어야 한다.
위험 분석 프로세스 이후에, 변경의 제안자는 위험을 경감하기 위한 적절한 조치를 적용할 것이다.
42) Policy and Investment Recommendations for Trustworthy AI, AI고위전문가그룹, 2019년 6월.
43) 그러나 이것이 제품의 출시에 관련된 기존의 보다 일반적인 책임의 결과로서, 주어진 상황에 감독이 필요할 수 있다는 사실을 배제하지는 않는다.

유럽연합의 관련 법규는 AI 제품/시스템의 설계로부터, 그리고 수명주기 전반에 걸쳐서 안전 장치로서 인간의 감독에 대한 구체적인 요구사항이 필요하다는 것을 이미 생각하고 있을 수도 있다.

AI 애플리케이션의 미래 "행동"은 예를 들면, 직장에나 가정에서 인간형(humanoid) AI 로봇/시스템과의 협력으로 인해 사용자의 **정신 건강에 대한 위험**[44]을 생성할 수 있다. 이러한 측면에서 오늘날 안전이란 일반적으로 새로이 출현하는 디지털 기술로부터 야기되는 물리적 피해에 대한 사용자의 인지된 위협을 칭하는데 사용되고 있다. 이와 동시에 안전한 제품이란 유럽연합의 법적 프레임워크에서는 사람의 안전과 건강에 아무런 위험을 주지 않거나 혹은 최소한의 위험만을 주는 제품으로 정의된다. 건강의 정의에는 물리적인 복지와 정신적인 복지가 모두 포함된다는 것은 모두가 동의하는 사실이다. 그러나 정신 건강에 대한 위험은 법규 프레임워크에서 제품 안전의 개념 내에 명시적으로 커버되어야 한다.

예를 들면, 자율성은 장기간에 걸쳐서 과도한 스트레스와 불괘감을 야기해서는 안되고, 정신 건강을 해쳐서도 안된다. 이러한 관점에서 노인들이 안전하다고 느끼는 감정에 긍정적인 영향을 미치는 요인들은 다음과 같은 것들이다[45]. 의료 서비스 요원들과 안전한 관계를 가짐, 평범한 일상을 통제할 수 있음, 일상에 대해 유용한 정보를 받음 등을 들 수 있다.

노인들과 상호작용하는 로봇의 생산자들은 정신 건강에 대한 위험을 방지할 수 있도록 이러한 요인들을 고려해야 한다.

AI 인간형 로봇의 생산자들이 자신들의 제품이 사용자, 그 중에서도 특히 돌봄을 받고 있는 노인들과 같은 취약한 사용자들에게 야기할 수 있는 무형의 피해를 명시적으로 고려해야 하는 책임은 관련 EU 법규의 범위에서 고려될 수 있다.

AI 기반 제품/시스템의 또 다른 근본적인 특성은 **데이터 의존성**(data dependency)이다. 데이터의 정확성과 관련성은 AI 기반 시스템/제품이 생산자가 의도한 바와 같이 의사결정을 내리도록 하는데 필수적이다.

---

44) WHO Constitution, 첫째 항목: "건강은 단순히 질병이나 질환의 부재가 아니라, 완전한 물리적/정신적/사회적 복지의 상태이다". (https://www.who.int/about/who-we-are /constitution).

45) Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction, pp.237-264, Research, Neziha Akalin, Annica Kristoff ersson and Amy Loutfi, 2019년 7월.

유럽연합의 제품안전 법규는 잘못된 데이터로부터 야기되는 안전에 대한 위험을 명시적으로 다루지 않고 있다. 그러나 제품의 "사용"에 따르면, 생산자들은 설계 및 테스트 단계 동안에 안전 기능에 대한 데이터의 정확성과 관련성을 예측해야 한다.

예를 들면, 특정한 물건을 탐지하도록 제작된 AI 기반의 시스템이 조명이 미흡한 조건에서는 물건을 인식하는데 어려움을 겪을 수 있다. 따라서 설계자들은 일반적인 조명 환경과 미흡한 조명 환경 모두에서 제품을 테스트한 결과로부터 도출된 데이터를 포함시켜야 한다.

또 다른 예로 나무나 땅에서, 잘 익은 과일을 탐지하여 찾아내는 과일따기 로봇과 같은 농업용 로봇을 들 수 있다. 관련 알고리즘은 이미 90% 이상의 분류 성공률을 나타내고 있지만, 그러한 앨고리즘에 연료를 제공하는 데이터 세트의 결점은 그러한 로봇들이 미흡한 결정을 내리게 하고, 이에 따라 동물이나 사람을 다치게 할 수 있다.

> 유럽연합 제품안전에 관한 법규가 잘못된 데이터의 안전에 관한 위험을 다루는 특정한 요구사항을 설계 단계에 포함해야 하는지, 그리고 데이터의 품질이 AI 제품/시스템의 사용 기간 내내 유지되도록 하는 메커니즘을 포함해야 하는지에 대해 의문이 제기되고 있다.

불투명성(Opacity)은 일부 AI 기반 제품/시스템의 또 다른 주요한 특성인데, 이러한 불투명성은 경험으로부터 학습하여 자신들의 성능을 향상시킬 수 있는 능력으로부터 야기된다. 방법론적 접근방법에 따라 AI 기반의 제품/시스템은 투명성의 정도로 구분될 수 있다. 이것은 추적하기 어려운 시스템의 의사결정 프로세스를 야기할 수 있다('블랙박스 효과'). 사람들이 의사결정 프로세스의 모든 단계를 이해할 필요는 없겠지만, AI 알고리즘이 더욱 고도화되고 핵심 영역에 배치됨에 따라, 사람들이 시스템의 알고리즘에 의한 결정이 내려진 방법을 이해할 수 있는 것이 필요하다. 이것은 사후 집행 메커니즘에서는 특히 중요할 것이다. 왜냐하면 이러한 메커니즘은 집행기관들이 AI 시스템의 행동과 선택의 책임을 추적할 수 있도록 해 주기 때문이다. 또한 이것은 '인간 중심적 AI에서의 신뢰 구축에 관한 위원회 공보'에서도 인정하고 있다.[46)]

46) https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

유럽연합의 제품안전 법규는 알고리즘을 기반으로 하고 있는 시스템의 불투명성으로부터 야기되는 위험을 명시적으로 다루지 않고 있다. 따라서 알고리즘의 투명성, 그리고 사후 집행 메커니즘에 특히 중요한 견고함/최종 책임/인간의 감독/편향되지 않은 결과[47] 등에 대한 요구사항을 고려하고, 그러한 기술의 사용에 대한 신뢰를 구축하는 것이 필요하다. 이러한 도전에 대응하는 한 가지 방법은 사고가 발생한 경우, 알고리즘의 개발자들에게 설계 파라미터와 데이터 세트의 메타데이터를 공개하는 책임을 부과하는 것이다.

다양한 구성요소, 기기 및 제품들이 통합되고, 서로의 기능(예: 스마트 홈 생태계의 제품 부품)에 영향을 미침에 따라, 안전에 영향을 미칠 수 있는 추가적인 위험은 **제품/시스템의 복잡성**으로부터 야기되는 위험이다.

이러한 복잡성은 본 절의 초반부에 설명한 유럽연합의 안전 법규 프레임워크에 의해서 이미 다루어지고 있다.[48] 특히, 생산자가 제품의 위험 평가를 수행할 때, 계획된 사용, 예측 가능한 사용, 그리고 적용되는 경우 합리적으로 예측 가능한 오용을 고려해야 한다.

이러한 상황에서 **생산자가 자신들의 기기가 상호연결되고, 다른 기기들과 상호작용할 것이라고 예상하면, 이것은 위험 평가 동안에 고려되어야 한다.** 사용 혹은 오용은 예를 들면, 동일한 유형의 제품의 과거 사용 경험, 사고 조사 또는 사람의 행동 등을 기반으로 결정된다.

또한 시스템의 복잡성은 의료기기 규정 등과 같은 산업별 안전 법규, 그리고 어느 정도는 일반제품안전 법규[49]에 의해서 보다 구체적으로 다루어지고 있다. 예를 들면, 스마트 홈 생태계의 일부로 계획된, 연결되어 있는 기기의 생산자는 자신들의 제품이 다른 제품의 안전에 영향을 미칠 것이라는 것을 합리적으로 예측할 수 있어야 한다.

그 뿐만 아니라, 운송 분야의 법규는 이러한 복잡성을 시스템 수준에서 다루고 있다. 자동차, 기차, 비행기에 대해서 승인과 인증은 전체 차량이나 항공기에 대해서 뿐만 아니라 각 구성요소에 대해서도 이루어지고 있다. 도로주행 적성성, 감항성

47) 신뢰성 있는 AI를 위한 윤리 가이드라인에서 고위전문가그룹이 제안한 핵심 요구사항을 바탕으로. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines

48) Regulation (EC) No. 2008/765, Decision (EC) No. 2008/768, 그리고 일치화된 산업별 제품안전 법규(예: 기계지침 2006/42/EC).

49) 일반제품안전지침의 Article 2는 다른 제품과 함께 사용될 것이라는 것이 합리적으로 예측되는 경우, 안전한 제품은 "다른 제품에 대한 영향"을 고려해야 한다고 규정하고 있다.

(air-worthiness), 철도 상호운용성 등은 안전 평가의 일부이다. 운송 분야에서 "시스템"은 명확한 기술적 요구사항에 대비한 적합성을 제삼자가 평가한 결과를 바탕으로, 혹은 위험을 처리하고 있는 방법을 시연한 후에 당국에 의해서 "승인"되어야 한다. 해결방안은 일반적으로 "제품"과 "시스템" 수준의 결합이다.

> 운송 분야의 법규를 포함하여, 유럽연합의 제품안전에 관한 법규는 사용자들의 안전에 영향을 미칠 수 있는 위험에 대응할 수 있도록, 이미 어느 정도 제품이나 시스템의 복잡성을 고려하고 있다.

복잡한 시스템에는 흔히 **소프트웨어**가 관여되는데, 소프트웨어는 AI 기반 시스템의 필수적인 구성요소이다. 일반적으로 봤을 때, 최초 위험 평가의 일부로서, 최종 제품의 제조자는 제품을 출시할 때 해당 제품에 통합되어 있는 소프트웨어의 위험을 예측해야하는 책임이 있다.

유럽연합의 제품안전 법규의 일부는 제품에 통합되어 있는 소프트웨어를 명시적으로 언급하고 있다. 예를 들면, 기계지침[50]은 통제 시스템의 소프트웨어 결함이 위험한 상황을 야기하지 않도록 의무화하고 있다.

유럽연합의 제품안전 법규에서는 이미 출시되어 있는 제품을 크게 수정하지 않고, 최초 위험 평가에서 예측하지 못한 새로운 위험이 도입되지 않는다면, 소프트웨어 업데이터는 안전을 확보하기 위한 유지보수 운영과 비교될 수 있다. 그러나 소프트웨어 업데이트가 이것이 다운로드된 제품을 크게 수정하면, 전체 제품은 새로운 제품으로 간주되고, 수정이 수행될 때에 관련된 제품안전에 관한 법규의 준수가 재평가되어야 한다.[51]

시장에 그대로 출시되었거나, 제품이 출시된 이후에 업로드된 독립형 소프트웨어의 경우, 유럽연합의 산업 고유의 일치화된 제품안전 법규는 일반적으로 구체적인 조항을 가지고 있지 않다. 그러나 유럽연합 법규의 일부(예: 의료 기기에 관한 규정)는 독립형 소프트웨어를 다루고 있다. 그 뿐만 아니라 특정한 라디오 모듈[52]을 통해서 통신하는 연결되어 있는 제품에 업로드된 독립형 소프트웨어는 위임받은 법률을 통해서 라디오장비지침에 의해 규제될 수도 있다. 이 지침은 특정한 등급이나 범주의 라디오 장비는 소프트웨어가 업로드될 때 해당 장비의 준수가 손상되지 않도록 하는 기능을 지원하는 것을 의무화하고 있다.[53]

---

50) 기계지침의 Annex 1 Section 1.2.1
51) The Blue Guide on the implementation of EU product rules, 2016
52) 라디오 모듈은 두 기기들 간에 전파(WiFi, 블루투스)를 전송 및/혹은 수신하는 전자 기기이다.

유럽연합 제품안전 법규는 제품을 출시할 때, 그리고 제조업체가 예상한 차후의 업데이트 시에 제품에 통합되어 있는 소프트웨어로부터 기인하는 안전 위험을 감안하고 있지만, 독립형 소프트웨어에 대한 구체적이고 명시적인 요구사항이 필요할 수 있다 (예: 다운로드될 '앱'). 독립형 소프트웨어가 AI 제품/시스템에서 안전 기능을 확보하도록 하는데 특별한 주의를 기울여야 한다.
제조자들이 AI 제품의 수명주기 동안에 안전에 영향을 미치는 소프트웨어의 업로드를 방지할 수 있는 기능을 제공할 수 있도록 제조자에게 추가적인 책임이 필요할 수도 있다.

마지막으로, 새로이 출현하는 기술들은 **복잡한 가치 체인**의 영향을 받고 있다. 그러나 이러한 복잡성은 새로운 것이 아니고, AI나 IoT와 같은 새로운 디지털 기술만이 야기하는 이슈도 아니다. 이것은 예를 들면, 컴퓨터, 서비스 로봇 또는 운송 시스템과 같은 제품의 경우에도 마찬가지이다.

유럽연합의 제품안전 프레임워크 하에서는 가치 체인이 얼마나 복잡하냐에 상관없이, 제품의 안전에 대한 책임은 해당 제품을 시장에 출시하는 생산자에게 귀속된다. 생산자들은 제품에 통합된 부품(예: 컴퓨터의 소프트웨어)을 포함하여, 최종 제품의 안전에 대한 책임을 진다.

유럽연합 제품안전 법규는 가치 체인의 복잡성을 감안하여, 여러 경제 주체에게 "공유 책임"의 원칙을 따라야 할 책임을 부과한다.
최종 제품의 안전에 대한 생산자의 책임은 현재의 복잡한 가치 체인에 적합한 것으로 입증되고 있지만, 공급망의 경제 주체들과 사용자들 간의 협력을 구체적으로 요구하는 명시적인 조항은 보다 복잡한 가치 체인에서조차도 법적인 확실성을 제공할 수 있다. 특히, 제품의 안전에 영향을 미치는 가치 체인 상의 각 주체(예: 소프트웨어 생산자)와 사용자(제품을 수정함으로써)들은 자신들의 책임을 맡고, 체인 상의 다음 주체에게 필요한 정보와 조치를 제공할 것이다.

유럽연합의 제품안전 법규의 일부는 해당 제품이 시장에 출시되기 전에 여러 경제 주체들이 주어진 제품에 개입하는 상황을 명시적으로 언급하는 조항을 이미 포함하고 있다. 예를 들면, 승강기지침(Lifts Directive)[54]은 승강기를 설계하고 제조하는 경제 주체는 "*설치자들이 정확하고 안전한 설치 및 승강기의 테스트를 수행할 수 있도록 필요한*

53) 라디오장비지침의 Article 3 (3) (i)
54) Directive 2014/33/EU의 Article 16(2)에 따라

모든 필요한 문서와 정보"를 설치자들[55]에게 제공하도록 의무화하고 있다. 기계지침은 장비의 제조자들이 운영자에게 해당 장비와 다른 기계를 조립하는 방법에 대한 정보를 제공하도록 의무화하고 있다.[56]

3. 책임

유럽연합 차원에서 제품의 안전과 제품의 책임에 관한 조항들은 두 개의 상호보완적인 메커니즘으로서, 높은 수준의 안전을 보장(즉, 사용자에게 미칠 수 있는 피해의 위험을 최소화)하고, 결함을 가진 제품으로부터 야기되는 피해에 대해서는 보상을 제공하는 제품의 단일 시장을 작동시킨다는 동일한 정책 목표를 추구한다.

국가적인 차원에서 일치화되지 않은 민사 책임 프레임워크는 다양한 원인(제품과 서비스 등과 같은)으로 인해서 발생하는 피해에 대해 보상이 이루어지도록 하고, 책임이 있는 서로 다른 사람들(소유자, 운영자 또는 서비스 제공자와 같은)을 언급함으로써 이러한 유럽연합의 규칙을 보완한다.

유럽연합의 AI에 대한 규칙들을 최적화하는 것은 사고를 방지하는 것을 도와줄 수 있지만, 그럼에도 불구하고 사고는 발생할 수 있다. 이 때에 민사 책임이 개입되게 된다. 민사 책임 규칙들은 우리 사회에서 이중의 역할을 수행한다. 즉, 한편으로는 다른 사람에 의해서 야기된 피해의 희생자들이 보상을 받게 해 주는 것이고, 또 다른 한편으로는 책임이 있는 당사자가 그러한 피해를 야기하는 것을 방지할 경제적인 동기를 제공하는 것이다. 책임에 관한 규칙들은 항상 시민들을 피해로부터 보호해주는 것과 기업의 혁신을 가능하게 해 주는 것 간의 균형을 맞추어야 한다.

유럽연합에서 책임에 관한 프레임워크들은 잘 작동되어 오고 있다. 이러한 프레임워크들은 결함이 있는 제품의 생산자에 대한 책임을 일치화시킨 제품책임지침(Directive 85/374/EEC)과 기타 일치화되지 않은 국가별 책임 체계를 병행해서 적용하고 있다. 제품책임지침은 국가별 결함 기반 책임만으로는 제공되지 않는 보호 계층을 제공한다. 이것은 생산자의 제품의 결함에 의해 야기된 피해에 대해 무과실 책임(strict liability) 체계를 도입하고 있다. 물리적 또는 유형의 피해의 경우, 부상을 당한 당사자는 피해, 제품의 결함(즉, 대중이 기대할 권리가 있는 안전을 제공하지 않음), 결함이 있는 제품과

55) Lifts Directive 2014/33/EU에서 설치자는 제조자에 해당하고, 승강기의 설계/제조/설치/ 출시에 대한 책임을 진다.

56) Machinery Directive, Annex I, Article 1.7.4.2에 의하면, "각 사용설명 매뉴얼은 적어도 다음과 같은 정보를 담고 있어야 한다." (i) "그림, 도표, 기계를 탑재하는 새시나 설비의 부착/지정 수단 등을 포함하여 조립/설치/연결에 관한 설명"

피해 간의 인과관계를 입증하면, 보상을 받을 권리가 있다.

국가별로 일치화되지 않은 체계는 결함 기반의 책임 규칙을 제공하고 있다. 여기에 따르면, 책임 소송에서 승소하기 위해서는 피해자는 책임을 져야 할 사람의 과실, 피해, 결함과 피해 간의 인과관계를 입증해야 한다. 이것은 또한 무과실 책임 체계를 제공하는데, 이 경우, 입법자들은 피해자가 결함/과실이나 결함/과실과 피해 간의 인과관계를 입증할 필요 없이 특정인에게 위험에 대한 책임을 귀속시켰다.

국가별 책임 체계는 제품/서비스로 인해서 야기된 피해의 희생자들에게 과실이나 무과실 책임을 기반으로 여러 가지 병렬적인 보상 청구를 제공한다. 이러한 청구들은 흔히 청구의 유형에 따라 책임을 지울 대상과 조건이 다르다.

예를 들면, 자동차 사고의 피해자는 국가별 민사법, 그리고 자동차가 결함을 가진 경우에는 생산자에 대해 제품책임지침 하에서, 자동차 소유자(자동차의 책임 보험을 가지고 있는 사람)에게 무과실 책임(strict-liability)을, 운전자에게는 과실 기반 책임(fault-based liability)을 청구한다.

자동차 보험에 관한 일치화된 규칙에 따르면, 자동차의 사용은 보험에 가입해야 하고[57], 보험회사는 항상 개인의 부상이나 물질적인 피해에 대한 최초의 보상 청구점이다. 이러한 규칙에 따르면, 의무(obligatory) 보험은 피해자에게 보상을 제공하고, 국가별 민사법 규칙 하에서 자동차 사고로 인한 재무적인 피해를 지불할 책임이 있는 보험 가입자를 보호한다.[58]

생산자들은 제품책임지침 하에서 의무 보험의 적용을 받지 않는다. 자율 자동차는 유럽 연합의 법규에서는 자동차 보험에 대해 비자율 자동차와 달리 취급되지 않는다. 모든 자동차와 마찬가지로 자율 자동차들도 제삼자 자동차 책임 보험에 의해 커버되어야 하고, 이것은 부상을 당한 당사자가 보상을 받을 수 있는 가장 손쉬운 방법이다.

적절한 보험에 드는 것은 피해자에게 원활한 보상을 제공함으로써 사고의 부정적인 결과를 경감시킬 수 있다. 책임에 관한 명확한 규칙들은 보험회사들이 자신들의 위험을 계산하고, 피해에 대해 궁극적으로 책임이 있는 당사자로부터 구상권을 청구하는 것을 도와준다. 예를 들면, 사고가 결함에 의해 야기되었다면, 자동차 보험회사는 피해자에게

57) 자동차의 사용에 대한 민사 책임에 대비한 보험, 그리고 그러한 책임에 대해 보험에 가입해야 할 의무의 집행에 관련된 Directive 2009/103/EC에 의해 일치화됨.
58) 대부분의 회원국에서는 자신의 이름으로 자동차를 등록한 사람에게 무과실 책임이 적용되고 있다.

보상한 후에 제조업체에게 보상을 청구할 수 있다.

그러나 AI, IoT, 로보틱스와 같은 새로이 출현하는 디지털 기술들의 특성 때문에 유럽연합과 국가별 책임 프레임워크는 도전을 받고 있고, 이것들의 효과성이 줄어들 수 있다. 이러한 특성들의 일부는 피해로부터 이를 야기한 사람의 행동으로 역추적하는 것을 어렵게 만들 수 있고, 이것은 국가 규칙에 따라 과실 기반 청구의 근거를 제공할 수 있다. 이것은 국가별 불법행위법(tort law)을 기반으로 한 책임 청구가 어렵거나, 입증하는데 비용이 많이 소요되어, 결과적으로 피해자들이 적절한 보상을 받지 못할 수도 있다는 것을 의미한다. AI와 같이 새로이 출현하는 디지털 기술을 포함하여 제품/서비스로 인한 사고의 피해자들이 국가별 불법행위법 하에서 보상을 받을 다른 유사한 제품/서비스에 비해서 낮은 수준의 보호를 받지 않도록 하는 것이 중요하다. 이것은 이러한 신기술의 사회적인 수용도를 낮추어, 이러한 기술을 사용하는 것을 주저하게 만들 수 있다.

신기술의 기존 프레임워크에 대한 도전이 기존의 법률이 적용되는 방법(즉, 과실의 개념이 AI에 의해 야기된 피해에 적용되는 방법) 측면에서 법적인 불확실성을 야기할지의 여부를 평가할 필요가 있을 것이다. 이것은 결국 투자 의욕을 저하시키고, 공급망에서 생산자와 기타 기업, 그 중에서도 특히 중소기업들의 정보 및 보험 비용을 증가시킬 것이다. 그 뿐만 아니라, 회원국들이 국가별 책임 프레임워크에 대한 도천을 궁극적으로 해결하려고 하면, 이것은 분열을 더욱 야기할 수 있고, 이에 따라 혁신적인 AI 솔루션의 출시 비용을 증가시키고, 단일 시장에서의 국가간 교역을 줄일 수 있다. 기업들이 가치 체인 전반에 걸쳐서 자신들의 책임에 대한 위험을 알고, 이러한 위험을 줄이거나 예방하고, 이러한 위험에 대비하여 효과적으로 보험을 드는 것이 중요하다.

본 장에서는 신기술이 기존의 프레임워크에 도전하는 방법, 그리고 이러한 도전에 대처할 수 있는 방법을 설명한다. 그 뿐만 아니라, 예를 들면 의료 서비스와 같은 일부 산업의 특이성에 대해서는 추가적인 고려를 할 필요가 있다.

제품/서비스/가치 체인의 복잡성: 지난 10여년 동안 기술과 산업은 크게 발전해 오고 있다. 특히 제품과 서비스의 경계선이 이제 더 이상 과거처럼 명확하지 않을 수 있다. 제품과 서비스의 제공은 점차 서로 뒤얽히고 있다. 복잡한 제품과 가치 체인은 유럽 산업이나 규제 모델에 새로운 것이지만, 소프트웨어와 AI 또한 제품 책임 측면에서 특별한 주의를 기울일 필요가 있다. 소프트웨어는 많은 수의 제품을 작동시키는데 필수적이고, 이러한 제품들의 안전에 영향을 미칠 수 있다. 소프트웨어는 제품에 통합되지만, 의도한 바와 같이 제품을 사용하는 것이 가능하도록 별도로 공급될 수도 있다. 소프트웨어 없이는 컴퓨터나 스마트폰은 크게 소용이 없을 것이다. 이것은 소프트웨어는 유형의

제품이 결함이 있도록 만들 수 있고, 물리적인 손상을 야기할 수 있다는 것을 의미한다. 이것은 궁극적으로 제품책임지침 하에서 제품 생산자의 책임으로 귀결될 수 있다.

> 제품책임지침에서의 제품의 정의는 광범위하지만, 그 범위를 더욱 명확히 하여, 새로이 출현하는 기술의 복잡성을 더 잘 반영하고, 소프트웨어나 기타 디지털 기능 때문에 결함을 가지게 된 제품에 의해 야기된 피해에 대해 항시 보상이 가능하도록 할 수 있다, 이것은 소프트웨어 개발자들과 같은 경제 주체들이 제품책임지침에 따라 자신들이 생산자로 간주될 수 있는지의 여부를 보다 잘 평가할 수 있도록 해 줄 것이다.

그러나 소프트웨어는 많은 유형 및 형식을 띄므로, 서비스로서의 소프트웨어 또는 제품으로서의 소프트웨어로 분류해야 하는지에 대한 대답이 항시 간단한 것은 아니다. 따라서 유형 제품의 운영을 조정하는 소프트웨어는 해당 제품의 부품이나 구성요소로 간주될 수 있지만, 특정한 유형의 독립형 소프트웨어는 분류하기 더욱 어려울 수 있다.

AI 애플리케이션들은 흔히 많은 서로 다른 연결된 기기와 서비스들이 상호작용하는 복잡한 IoT 환경에 통합되어 있다. 복잡한 생태계에서 서로 다른 디지털 구성요소들을 결합하고, 여기에는 여러 주체들이 관여하므로, 잠재적인 피해가 어디에서 비롯되었고, 누가 여기에 대한 책임이 있는지를 평가하는 것은 어려울 수 있다. 이러한 기술들의 복잡성 때문에 피해자들이 책임자를 식별하고, 국가의 법률이 요구하는 바와 같이 성공적인 청구의 모든 필요조건들을 입증하는 것이 매우 어려울 수 있다. 이러한 전문성에 대한 비용은 경제적으로 엄두도 못낼 정도로 높고, 피해자가 보상을 청구하는 것을 포기하게 만들 수도 있다.

그 뿐만 아니라 AI에 의존하는 제품/서비스들은 전통적인 기술들과 상호작용할 것이고, 이에 따라 책임 측면에서 복잡성이 더욱 높아지게 된다, 예를 들면, 자율주행 자동차는 일정 기간 동안 전통적인 자동차와 도로를 공유할 것이다. 부분적으로 자동화된 AI 시스템이 인간의 의사결정을 지원할 일부 서비스 부문(예: 교통 관리, 의료 서비스)에서도 상호작용하는 주체들 간에 이와 유사한 복잡성이 발생할 것이다.

책임및신기술 전문가그룹의 신기술 분과 보고서[59]에 따르면, AI 관련 피해자에 대한 입증 책임을 촉진하기 위해 국가별 법률의 수정이 고려될 수 있다. 예를 들면, 입증 책임은 특정한 사이버 보안이나 법률에 규정된 기타 안전 의무의 준수(관련 주체에 의한)

59) AI 및 기타 신기술에 대한 책임 보고서(Liability for Artificial Intelligence and other emerging technologies' Report, https://ec.europa.eu/newsroom/dae/document.cfm ?doc_id=63199

와 연결될 수 있다. 즉, 일방이 이러한 규칙들을 준수하지 않으면, 과실과 인과관계에 대한 입증 책임에 대한 변경이 적용될 수 있다.

> 본 위원회는 AI 애플리케이션의 운영에 의해서 야기되는 피해에 대한 국가별 책임 규칙들에서 요구하는 입증 책임을 경감/반전시킴으로써 복잡성의 결과를 경감시킬 필요가 있는지의 여부와 그 정도에 대한 견해를 적절한 EU 이니셔티브를 통해서 구하고 있다.
> 유럽연합 법규의 경우, 제품책임지침에 따르면 의무적인 안전 규칙을 충족시키지 못하는 제품은 생산자의 과실 여부에 상관없이 결함을 가진 것으로 간주될 것이다, 그러나 이 지침 하에서 피해자의 입증 책임을 촉진하는 방법들에 대해 심사숙고해야 할 이유가 있을 수 있다. 즉, 이 지침은 증거 및 인과관계의 수립에 대한 국가별 규칙들에 의존하고 있다.

연결성 및 개방성: 제품의 사이버 보안 위반으로 인해서 야기되는 피해에 대해 어떤 수준의 안전을 기대하는지, 그리고 그러한 피해가 제품책임지침 하에서 적절하게 보상될 것인지의 여부는 현재 분명하지 않다.

사이버 보안의 취약점은 제품이 유통되기 시작하는 처음부터 존재할 수 있지만, 제품이 유통된지 한참 지난 이후의 단계에서 나타날 수도 있다.

과실 기반의 책임 프레임워크에서 사이버 보안에 대한 명확한 의무를 수립하게 되면, 운영 주체들은 책임의 결과를 피하기 위해서 자신들이 무엇을 해야 하는지를 파악할 수 있게 된다.

제품책임지침 하에서 생산자가 합리적으로 예측 가능한 제품의 사용을 감안하여 특정한 변경을 예측할 수 있었다면, 이 문제는 더욱 중요해진다. 예를 들면, '이후 결함 방어(later defect defence)'의 사용 혹은 '개발 위험 방어가 증가하는 것을 볼 수 있을 것이다. 이후 결함 방어에 따르면, 제품이 유통되기 시작할 때 결함이 존재하지 않았다면, 생산자는 책임을 지지 않는다. 개발 위험 방어는 그 시점에 최첨단의 지식도 결함을 예측할 수 없다는 것이다. 그 뿐만 아니라, 부상을 입은 당사자가 안전 관련 업데이트를 수행하지 않은 경우, 책임은 줄어들 수 있다. 이것은 부상을 입은 사람에 의한 과실 기여(contributory negligence)로 간주될 수 있고, 이에 따라 생산자의 책임이 줄어들 수 있다. 예측 가능한 합리적 사용의 개념과 과실 기여(안전 업데이터를 다운로드 하지 않은 것)의 문제가 확산됨에 따라, 부상을 입은 사람은 제품의 결함으로 인해 야기된 피해에 대해 보상을 받는 것이 더 어려워졌다는 것을 알게 될 수 있다.

자율성 및 불투명성: AI 애플리케이션이 자율적으로 행동할 수 있는 경우, 이들은 모든 단계를 미리 정할 필요도 없고, 즉각적인 사람의 통제나 감독이 덜하거나 궁극적으로는 전혀 없이 과업을 수행한다. 기계학습을 기반으로 한 알고리즘은 이해하기 불가능한 것은 아니지만, 매우 어려울 수 있다('블랙박스 효과').

위에서 설명한 복잡성뿐만 아니라, 일부 AI에서의 블랙박스 효과 때문에, 자율 AI 애플리케이션에 의해서 야기된 피해에 대한 보상을 받는 것이 어려워질 수 있다. AI가 사용하는 알고리즘과 데이터를 이해하기 위해서는 분석 역량과 기술적 전문성이 필요하고, 이것은 피해자가 감당하지 못할 정도로 비용이 많이 소요된다. 그 뿐만 아니라, 알고리즘과 데이터에 대한 접근이 잠재적으로 책임이 있는 당사자의 협조 없이는 불가능할 수 있다. 그 뿐만 아니라, 자율적으로 행동하는 AI의 과실을 어떻게 입증할 것인지, 또는 AI의 사용에 의존하고 있는 사람의 어떤 것을 과실로 간주해야 할 것인지 등은 불분명할 것이다.

국가별 법률은 이와 유사한 상황에서 피해자의 입증 책임을 줄이기 위한 몇 가지 해결방안을 이미 수립하고 있다.

유럽연합의 제품 안전과 제품 책임에 대한 지침이 되는 원칙은 출시되는 모든 제품은 이들의 수명주기 전반에 걸쳐서, 그리고 합리적으로 기대할 수 있는 제품의 사용에 대해서 안전하도록 하는 책임은 생산자에게 있다는 것이다. 이것은 제조자는 AI를 사용하는 제품이 특정한 안전 파라미터를 존중하도록 해야 한다는 것을 의미한다. 이러한 AI의 특성 때문에 자동 잔디깍는 기계인지 수술 로봇인지에 상관없이, 제품의 안전에 대한 기대를 배제하는 것은 아니다.

자율성은 제품의 안전에 영향을 미칠 수 있다. 왜냐하면, 자율성은 안전 기능을 포함하여 제품의 특성을 크게 변화시킬 수 있기 때문이다. 어떤 조건 하에서 자율 학습 기능이 생산자의 책임을 연장시키는지, 그리고 생산자가 특정한 변경을 어느 정도나 예측했어야 했는지 등이 문제이다.

유럽연합의 안전 프레임워크의 변경사항과 긴밀하게 조정하여, 변경되고 수정되어야 할 제품들을 고려할 수 있도록, 제품책임지침에서 현재 사용하고 있는 "유통을 시작하는 시점"의 개념을 다시 논의할 수 있다. 또한 이것은 제품에 가해진 변경에 대해 누가 책임을 져야 하는지를 명확화하는 것을 도울 수 있다.

책임및신기술 전문가그룹의 신기술 분과 보고서[60]에 따르면, 일부 자율 AI 기기/서비스들의 운영은 책임 측면에서 특정한 위험 프로파일을 가지고 있다. 왜냐하면, 이들은 생명, 건강, 재산과 같은 중요한 법적 이익에 큰 해를 끼칠 수 있고, 많은 대중들을 위험에 노출시키기 때문이다. 이것은 주로 공공 장소에서 이동하는 AI 기기들(예: 완전 자율 자동차, 드론[61], 택배 로봇) 또는 유사한 위험을 가진 AI 기반의 서비스(예: 자동차를 가이드하거나 통제하는 교통 관리 서비스, 전력 배포 관리)와 관련되어 있다. 자율성과 불투명성이 국가별 불법행위법에 대한 도전은 위험 기반의 접근방법을 따름으로써 해결될 수 있다. 무과실 책임 체계는 위험이 현실화될 때마다 피해자는 과실에 상관없이 보상을 받도록 할 것이다. AI의 개발 및 수용에서 그러한 운영에 대한 과실 책임을 누가 져야하는지를 선택하는 것이 미치는 파급효과를 주의깊게 평가하고, 위험 기반의 접근방법을 고려해야 할 것이다

특정한 위험 프로파일을 가진 AI 애플리케이션의 운영을 위해, 본 위원회는 대중들이 노출되어 있는 유사한 위험(예를 들면, 자동차, 비행기 또는 원자력 발전소의 운영)에 대한 국가별 법률에 존재하는 것과 같이, 피해자에게 효과적인 보상을 제공할 수 있도록 무과실 책임이 필요한지의 여부와 필요하다면 어느 정도나 필요하지에 대한 의견을 구하고 있다. 또한 본 위원회는 책임을 져야할 사람의 지불 능력에 상관없이 보상이 이루어지도록 하고, 피해의 비용을 줄이는 것을 도와줄 수 있도록, 자동차보험지침의 예를 따라, 과실 책임과 가용한 보험의 가입 의무를 결합하는데 대한 견해를 구하고 있다.

기타 모든 AI 애플리케이션의 운영에 대해 본 위원회는 인과관계 및 과실에 대한 입증 책임이 수정되어야 할 필요성이 있는지의 여부에 대해 생각하고 있다. 이러한 측면에서 책임및신기술 전문가그룹의 신기술 분과 보고서[62]에서 제기한 이슈 중의 하나는 잠재적으로 책임을 져야 할 당사자가 책임 평가에 관련된 데이터를 로깅하지 않았거나, 이러한 데이터를 피해자와 공유할 의사가 없는 상황이다.

4. 결론

AI, IoT, 로보틱스 등과 같은 새로운 디지털 기술의 등장은 연결성, 자율성, 데이터 의존성, 불투명성, 제품/시스템의 복잡성, 소프트웨어 업데이트, 보다 복잡한 안전 관리

60) AI 및 기타 신기술에 대한 책임 보고서(Liability for Artificial Intelligence and other emerging technologies' Report, https://ec.europa.eu/newsroom/dae/document.cfm ?doc_id=63199

61) 무인 항공기의 운영에 대한 규칙과 절차에 대한 2019년 5월 24일자 Commission Implementing Regulation (EU) 2019/947에서 언급된 무인 항공기 시스템

62) AI 및 기타 신기술에 대한 책임 보고서(Liability for Artificial Intelligence and other emerging technologies' Report, https://ec.europa.eu/newsroom/dae/document.cfm ?doc_id=63199

및 가치 체인 등과 같은 제품의 안전과 책임 측면에서 새로운 해결과제를 야기하고 있다.

제품 안전에 관한 현행 법규, 그 중에서도 특히 특히 일반제품안전지침, 기계지침, 라이오장비지침, 새로운 법규 프레임워크 등은 해결되어야 할 몇 가지 갭을 포함하고 있다. 이 프레임워크에서 서로 다른 법규의 수정을 위한 향후 작업은 일관성 있고, 조화로운 방법으로 수행될 것이다.

안전 측면에서 새로운 도전은 책임 측면에서도 새로운 도전을 창출할 것이다. 책임에 관련된 도전은 기술 혁신의 니즈와 균형을 유지하면서, 전통적인 기술의 피해자와 비교하여 동일한 수준을 보호를 받을 수 있도록 해결되어야 한다. 이것은 이러한 새로운 디지털 기술에 대한 신뢰를 창출하고, 투자 안정성을 창출하는 것을 도울 것이다.

유럽연합 및 국가별 기존 법률은 새로이 출현하는 기술에 원칙적으로는 대응할 수 있지만, AI의 가진 특성, 그리고 AI 관련 해결과제들의 결합된 효과 때문에 피해자들에게 정당성이 있는 모든 경우에 보상을 제공하는 것이 보다 어려울 수 있다.[63] 이에 따라 피해가 발생했을 때, 비용의 분배가 현재의 규칙 하에서는 불공평하거나 비효율적일 수 있다. 이러한 문제를 바로잡고, 기존 프레임워크의 잠재적인 불확실성을 해결하기 위해서는, 적절한 EU 이니셔티브를 통해서, 제품책임지침과 국가별 책임 체계에 대한 조정이 목표 지향적이고 위험 기반의 접근방법(즉, AI 애플리케이션의 종류에 따라 위험이 달라질 수 있다는 것을 감안)으로 고려될 수 있다. ▢

63) 신기술 분과의 보고서 p3, AI고위전문가그룹의 정책 권고사항 27.2를 참조하시오.

자료

FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS

미국 연방거래위원회

인공 지능 및 알고리즘의 활용

번역제공 : 한국과학기술정보연구원(KISTI)

앤드류 스미스(Andrew Smith), 연방거래위원회 소비자보호국 국장

2020년 4월 8일

각종 매스 미디어가 인공지능 기술의 급속한 발전에 찬사를 보내고 있다. 예측하고, 추천하고, 의사 결정하는 기계-알고리즘의 조합인 인공지능 기술은 삶의 질 개선 및 생산성 향상을 위한 엄청난 잠재력을 지니고 있다. 그러나 여기에는 위험도 따른다. 불공정하고 차별적인 결과를 낳을 수 있으며 지금까지 이어져 온 사회 경제적 격차를 영구화시킬 가능성도 있다. 의료용 인공지능이 좋은 예이다. 사이언스 지에 보고된 최근 연구에 따르면 병세가 가장 심한 중증 환자에 대한 의료 개입을 목적으로 고안된 알고리즘, 즉 선한 의도로 고안된 알고리즘이 실제로는 건강한 백인 환자들에게 가용한 리소스가 편중되도록 하여 그보다 더 아픈 흑인 환자들이 피해를 보게 만들고 있다.

그나마 다행인 것은 인공지능 및 기계 학습의 고도화와 복잡성이 우리에게 낯선 경험이기는 해도, 자동화된 의사결정은 새로운 개념이 아니며, 또한 우리 연방거래위원회가 소비자에 대한 의사 결정을 위한 데이터 및 알고리즘 사용에 수반되는 여러 문제를 오랜 세월 다루며 경험을 축적해왔다는 점이다. 연방거래위원회는 수년간 집행 법령에 대한 위반 혐의가 있는 인공지능 및 자동화된 의사결정과 관련된 다수의 사건에 대해 소송을 제기했으며 수많은 관련 기업들을 조사해왔다. 예를 들어, 1970년에 제정된 공정신용보고법(Fair Credit Reporting Act, FCRA)과 1974년에 제정된 평등신용기회법(Equal Credit Opportunity Act, ECOA) 모두 자동화된 의사결정을 다루고 있으며, 금융 서비스 기업들은 이미 수십 년간 이러한 법들을 기계 기반 신용 인수 모델에 적용해왔다. 연방거래위원회 또한 인공지능 및 자동화된 의사결정의 적용으로 인해 발생하는 소비자 피해를 해결하기 위해 연방거래위원회법에 따른 권한을 기반으로 불공정하고 기만적인 관행들을 금지해왔다.

2016년, 연방거래위원회는 "빅데이터: 포용을 위한 도구인가 배제를 위한 도구인가?"라는 제목의 보고서를 발간한 바 있다. 이 보고서는 빅데이터 분석도구와 기계 학습을 사용하는 기업이 편향 가능성을 줄이도록 권고하였다. 가장 최근인 2018년 11월에는 인공지능, 알고리즘, 예측 분석에 관한 공청회를 개최한 바 있다.

연방거래위원회의 모든 법 집행 활동, 연구 및 지침이 강조하는 바는 인공지능 도구의 활용이 투명하고, 설명 가능하고, 공정하고, 실증적으로 건전한 방식으로 이루어져야 한다는 것이며, 이를 통해 책임 의식을 고취하고 있다. 우리의 이러한 경험과 기존 법령을 바탕으로, 기업이 소비자 보호와 관련하여 인공지능 및 알고리즘이 수반하는 여러 리스크를 해결하는 데 필요한 중요한 교훈을 제공해줄 수 있을 것으로 생각한다.

투명성을 확보해야 한다.

자동화 도구의 사용과 관련하여 소비자를 기만해서는 안 된다. 인공지능은 백그라운드에서 동작하는 경우가 많아서 소비자 경험과는 어느 정도 동떨어져 있는 듯하다. 그러나 인공지능을 소비자와의 상호작용(챗봇 등과 같이)에 활용하는 경우에는 상호작용의 본질에 대한 소비자의 오해를 불러일으키지 않도록 주의가 요구된다. 에슐리 메디슨(Ashley Madison) 사례는 불륜 조장 사이트인 에슐리 메디슨이 잠재적 고객의 데이트 서비스 가입을 유도하기 위해 매력적인 남녀의 거짓 "참여자 프로필"을 만들어 고객을 기만한 혐의로 제소된 사건이다. 데부미(Devumi) 사례는 소셜 미디어에서의 존재감을 강화하고자 하는 기업이나 개인에게 가짜 팔로워, 거짓 구독자, 허위 "좋아요"를 판매한 혐의로 데부미 사가 제소된 사건이다. 결론은? 가짜 데이트 프로필, 거짓 팔로워, 딥페이크, 인공지능 챗봇과 같은 "도플갱어"의 사용으로 소비자를 오도하는 기업은 연방거래위원회의 법 집행 대상이 될 수 있다.

민감한 데이터의 수집에 있어 투명성 확보해야 한다. 데이터 세트의 규모가 커질수록, 알고리즘의 성능이 향상되고, 그에 따라 소비자 제품의 품질이 개선된다. 더 이상 무슨 말이 필요하겠는가. 하지만 과연 그럴까? 너무 성급한 결론일 수 있다. 그 데이터 세트를 취득하는 방식에 대해서도 주의를 기울일 필요가 있다. 음성이나 시각 데이터 또는 민감한 데이터를 비밀리에 수집하여 알고리즘에 적용하는 행위 또한 연방거래위원회의 조치 대상이 될 수 있다. 바로 작년에만 해도 연방거래위원회는 페이스북이 얼굴 인식 기능이 이미 기본 설정이었음에도 불구하고 사용자에게 해당 옵션을 선택할 수 있다고 안내함으로써 사용자를 오도했다는 혐의를 제기했다. 페이스북 사례에서 알 수 있듯, 데이터 수집 방식은 매우 중요한 사안이 될 수 있다.

제3의 판매인으로부터 제공된 정보에 기반하여 자동화된 의사결정을 내릴 경우, 해당 소비자에게 "불이익 조치(adverse action)" 통지를 제공할 의무를 지게 될 수 있다. 공정신용보고법(FCRA)에 의거하여, 신용 자격, 고용, 보험, 주거나 유사한 혜택 및 거래에 대한 의사 결정의 자동화를 위해 소비자 정보를 수집하는 판매인은 "소비자 보고기관(consumer reporting agency)"이 될 수 있다. 이는 결국 해당 정보의 사용자에 대한 의무를 발생시킨다. 구체적으로 말하면, 공정신용보고법에 따라 소비자에게 특정 통지를 제공해야 한다는 것이다. 예를 들어, 여러분이 인공지능 도구를 기반으로 소비자에 대한 점수를 책정하여 해당 소비자가 좋은 세입자일지를 예측하는 배경조사회사(background check company)로부터 보고서나 소비자 점수를 구입했다고 가정해보자. 이 인공지능 모델은 공공 기록 정보, 범죄 기록, 신용 기록뿐만 아니라 소셜 미디어 사용, 쇼핑 기록, 공개된 사진, 비디오 등 소비자에 대한 광범위한 데이터를 활용하게 된다. 만약 여러분이 이러한 보고서나 점수를 토대로 누군가에게 아파트를 임대하지 않거나 더 높은 임대료를 책정하려 한다면, 반드시 해당 소비자에게 불이익 조치 통지를 제

공해야만 한다. 불이익 조치 통지는 해당 소비자가 자신에 대해 보고된 정보를 열람하고 부정확한 정보를 수정할 권리를 가짐을 명시한다.

의사결정에 대해 소비자에게 설명한다.

알고리즘 의사결정에 기반하여 소비자에게 가치 있는 무언가를 제공하지 않게 되는 경우에는 그 이유를 설명해야 한다. 알고리즘 의사결정에 영향을 미칠 수 있는 다수의 요인들이 너무 복잡해서 설명이 불가능하다고 말하는 사람도 있다. 하지만, 신용 공여에 있어 기업들은 신용이 거부된 소비자에게 해당 의사 결정의 주요 근거를 설명해야 할 의무를 진다. "점수가 너무 낮다"거나 "기준이 충족되지 않는다" 정도의 설명으로는 충분하지 않다. 구체적으로 설명할 필요가 있다(예를 들어, "신용 채무를 이행하지 않았다"라거나 "신용 조회수가 부족하다"와 같은 설명) 이것이 의미하는 바는 당신이 적용한 모델에 **어떤** 데이터가 사용되고 있으며 그 데이터가 의사결정에 있어 **어떤 방식으로** 활용되고 있는지를 반드시 알고 있어야 한다는 것이다. 또한 소비자에게 그러한 사안에 관해 설명할 수 있어야만 한다. 어떤 맥락에서든 인공지능을 활용해 소비자에 대한 의사결정을 하는 경우, 소비자가 원한다면 소비자에게 당신의 의사결정 과정을 설명하도록 한다.

알고리즘을 이용해 소비자 위험 지수를 평가하는 경우에도 점수 및 등급에 영향을 미치는 주요 요인을 중요도 순으로 정리하여 공개한다. 다른 알고리즘 의사결정과 마찬가지로 점수 평가 또한 여러 다양한 요인에 기반하여 이루어지며, 이들 중 일부는 소비자에게 설명하기가 쉽지 않다. 예를 들어, 신용 점수가 누군가에게 신용을 거부하거나 불리한 조건으로 신용을 제공하는 근거로 사용된다면, 소비자는 법에 따라 사전 통지, 해당 점수에 대한 설명(출처 및 적용된 신용 모델의 점수 범위 등), 그리고 영향도에 따라 중요도 순서로 정리된 신용 점수에 부정적인 영향을 미친 적어도 네 개 이상의 주요 요인을 제공받을 수 있어야 한다.

자동화 도구에 기반하여 거래 조건을 변경하는 경우, 반드시 소비자에게 알리도록 한다. 10여 년 전 연방거래위원회는 서브프라임 신용 마케터인 CompuCredit이 소비자의 신용한도액을 축소하는 행동 기반 스코어링 모델을 사용했다는 사실을 고의로 공개하지 않음으로써 연방거래위원회법을 위반하였다는 혐의를 제기하였다. 예를 들어, 신용카드 사용자가 술집, 나이트클럽, 안마 시술소 등의 특정 장소에서 신용카드 현금 서비스를 이용하거나 신용카드로 결제를 하는 경우, 그들의 신용한도액이 줄어들게 되는 것이다. 해당 기업은 이러한 카드 사용으로 인해 신용한도액이 축소될 수 있다는 사실을 소비자에게 알리지 않았다. 카드 발급 시에도, 신용한도액이 축소되는 시점에서도 이를 알리지 않은 것이다. 10년 전의 일이지만 오늘날에도 여전히 중요한 문제이다. 알고리즘을 사용해 거래 조건을 변경하는 경우, 반드시 소비자에게 알리도록 한다.

의사결정은 반드시 공정해야 한다.

보호 계층을 차별해서는 안 된다. 인공지능의 무신경한 사용은 보호 계층에 대한 차별로 이어질 수 있다. 평등신용기회법(ECOA), 1964년 공민권법의 VII 편 등 여러 연방 기회균등 법이 이러한 행위와 관련되어 있다. 연방거래위원회는 인종, 피부색, 종교, 국적, 성별, 결혼여부, 나이, 공적 부조의 수령 여부를 기반으로 한 신용 차별을 금지하는 평등신용기회법을 집행한다. 예를 들어, 어떤 기업이 소비자의 우편번호를 기반으로 신용 의사결정을 내려 특정 인종 집단이 "차별적 영향"을 받게 되는 경우, 연방거래위원회는 평등신용기회법에 의거하여 그러한 관행에 이의를 제기할 수 있다. 알고리즘을 적용하기에 앞서, 그리고 그 이후에도 주기적으로 해당 알고리즘이 보호 계층에 대한 차별적 영향을 야기하지 않는지를 철저히 검사함으로써 여러 복잡한 문제의 발생을 미리 방지할 수 있다.

입력 데이터뿐만 아니라 그 결과물에도 주의를 기울여야 한다. 불법적인 차별과 관련하여 알고리즘이나 기타 인공지능 기반 도구를 평가할 때, 연방거래위원회는 해당 모델에 적용되는 입력 데이터를 살펴본다. 즉, 해당 모델이 인종에 기반한 요인, 그러한 요인을 대변할 수 있는 다른 요인, 또는 인구 조사 표준지역 등의 데이터를 포함하는지를 조사한다. 그러나 입력 데이터와는 별개로, 그에 따른 결과물 또한 조사한다. 예를 들어, 해당 모델이 허용되지 않는 근거를 기반으로 실제로 차별적 결과를 내놓는가? 겉으로는 중립적으로 보이는 모델이지만 보호 계층에 대해 불법적인 차별적 영향을 미치고 있지 않은가? 우리의 경제 분석은 소비자가 신용에 지불하는 비용 등과 같은 실제 결과물을 조사함으로써 어떤 모델이 보호 계층에 대한 차별적 영향을 미치는지의 여부를 평가하는 방식으로 이루어진다. 차별적 모델을 사용하는 기업에 대해서는 그러한 모델 사용에 대한 타당한 근거가 있는지를 조사하고, 덜 차별적인 대안을 통해 동일한 결과를 얻을 수 있을지를 검토한다. 인공지능 및 알고리즘 도구를 사용하는 기업들은 그러한 모델 사용에 수반되는 소비자 보호와 관련된 고유 리스크를 해결하기 위해 인공지능 기반 결과물에 대한 자체적인 평가 및 검사의 수행을 고려해야 한다.

소비자가 본인에 대한 의사결정에 사용된 정보를 수정할 수 있는 권한과 기회를 가져야 한다. 공정신용보고법(FCRA)은 소비자에 대한 의사결정에 사용되는 데이터를 규제한다. 여기에는 취업, 신용, 보험 가입, 아파트 임대 등과 관련된 모든 정보가 포함된다. 공정신용보고법에 따라 소비자는 자신에 대해 작성된 정보를 획득하고 부정확한 것으로 여겨지는 정보에 대해 이의를 제기할 법적 자격이 있다. 또한, 해당 정보가 소비자에게 불리한 의사결정에 사용되는 경우, 소비자는 반드시 불이익 조치 통보를 제공받을 수 있어야 한다. 이 통지에는 해당 의사결정에 사용된 정보의 출처가 포함되어야 하고 소비자가 정보 접근 및 이의 제기에 대한 권리를 가진다는 사실이 표시되어야 한다.

소비자에 대한 중요한 의사결정을 내리는 과정에서 타인으로부터 획득한 자료를 사용하거나 또는 해당 소비자 본인으로부터 직접 취득한 정보를 사용하는 경우에도, 해당 소비자에게 사용된 정보의 사본을 제공하고 사용된 정보의 정확성에 대해 이의를 제기할 수 있도록 허용해야 한다.

확실하고 실증적으로 건전한 데이터와 모델을 사용하도록 한다.

소비자에 대한 정보를 신용 접근성, 고용, 보험, 주거, 정부 보조, 수표의 현금화 등의 거래에 대한 의사결정을 원하는 타인에게 제공할 경우, 여러분은 공정신용보고법(FCRA)의 적용을 받는 소비자 보고기관으로 취급될 수 있으며, 이에 따라 해당 데이터가 정확하고 최신의 데이터가 맞는지 등을 확인해야 할 의무가 있다. 물론 이렇게 생각할 수도 있을 것이다: 우리는 하는 일은 인공지능이지 소비자 보고가 아니야. 그러니 우린 공정신용보고법의 적용을 받지 않아. 과연 그럴까? 다시 생각해보는 것이 좋겠다. 만약 당신이 신용, 고용, 보험, 주거 등과 같이 특정 혜택이나 거래에 대한 소비자의 적격성을 결정하는데 활용되는 또는 활용될 것으로 여겨지는 소비자 정보를 취합해서 판매한다면, 당신은 실제로 공정신용보고법의 적용 대상이 될 수 있다. 이것이 의미하는 바는 무엇일까? 무엇보다, 가능한 가장 정확한 소비자 보고를 제공하고, 소비자가 원할 경우 본인의 정보를 열람할 수 있을 뿐만 아니라 잘못된 부분은 수정할 수 있도록 담보하는 타당한 절차를 수행해야 할 의무를 지게 된다는 점이다. 주택 지원자의 범죄 기록을 실시간으로 또는 거의 실시간에 가깝게 조회하는 소프트웨어 도구를 배포했던 RealPage, Inc라는 기업은 이와 관련하여 매우 값비싼 교훈을 얻었다. 즉, 임대주나 부동산 관리자에게 제공되는 정보의 정확성을 담보하는데 필요한 타당한 조치를 취하지 않아 공정신용보고법을 위반하였으며, 이로 인해 3백만 달러의 벌금을 물어야 했다.

소비자 보고기관이 아닐지라도 자동화된 의사결정에 사용될 소비자 정보를 타인에게 제공하는 경우, 해당 데이터가 정확한지 확인해야 할 의무를 갖게 된다. 소비자 보고기관에 소비자 정보를 제공하는 기업은 공정신용보고법에 따라 “공급자”로 지칭한다. “공급자”는 정확하지 않은 것으로 여길만한 타당한 이유가 있는 데이터는 제공해서는 안 된다. 또한, 제공하는 데이터의 정확성과 무결성을 담보하는 데 필요한 서면 정책 및 절차를 마련해야만 한다. 공급자들은 또한 소비자뿐만 아니라 소비자 보고기관으로부터 입수된 분쟁을 조사할 의무를 갖는다. 이러한 요건은 인공지능 모델에 가능한 한 가장 정확한, 최신의 정보가 적용될 수 있게 하는 데 있어 중요하다. 소비자 보고기관에 정보를 제공하는 기업 중 해당 보고의 정확성을 보장하는 데 필요한 서면 정책 및 절차를 마련해야 할 의무를 이행하지 않은 기업에 대해 연방거래위원회는 법적 조치를 취하고 많은 벌금을 부과해 왔다.

인공지능 모델의 타당성을 검토하고, 해당 모델이 의도한 대로 작동하고 불법적인 차별

을 야기하지 않는지 재검증해야 한다. 소비자 금융 대출 분야로부터의 교훈을 좀 더 살펴보자. 신용 공여자들은 신용 인수 과정의 자동화를 위해 수십 년간 여러 가지 데이터와 알고리즘을 사용해 왔다. 대출법은 "경험에 기반하여 유래되고 입증 가능하며 통계적으로 건전한" 인공지능 도구의 사용을 권장한다. 무엇보다 중요한 포인트는, 해당 도구가 과거 상당 기간에 걸쳐 수집된 신용 지원자의 샘플 집단이나 개체군 내의 신용 가능한 지원자와 신용 불가능한 지원자에 대한 경험적 비교로부터 얻어진 데이터에 기반한다는 점이다. 널리 인정받는 통계적 원칙과 방법론에 따라 개발되고 검증되며, 적절한 통계적 원칙과 방법론을 기반으로 주기적으로 재검증되고, 예측 성능을 유지할 수 있도록 조정을 거치게 된다.

준수, 윤리, 공정, 차별 철폐에 대해 책임 의식을 가져야 한다.

알고리즘을 사용하기에 앞서 다음의 질문에 답해 본다. 2016년 빅데이터 보고서로 다시 돌아가 보자. 위원회는 해당 보고서를 통해 빅데이터 기반 분석이 편견이나 소비자 피해로 이어질 수 있음을 기업에 경고한 바 있다. 알고리즘 운영자는 이를 방지하기 위해 다음의 네 가지 핵심 질문에 답해야 한다.

당신의 데이터 세트는 대표성을 갖는가?
당신의 데이터 모델은 편견을 야기하는가?
빅데이터에 기반한 예측이 얼마나 정확한가?
빅데이터에 대한 의존으로 인해 윤리성 또는 공정성에 관한 우려가 발생하는가?

알고리즘이 인가되지 않은 방식으로 사용되지 않도록 한다. 다른 기업에 판매할 목적으로 인공지능을 개발하는 사업자는 그러한 도구들이 오용될 가능성에 대해 생각해봐야 하고 액세스 제어 등과 같은 기술을 활용해 오용을 방지할 수 있을지 따져봐야 한다. 바로 지난 달, 연방거래위원회는 음성 복제 기술에 대한 워크숍을 개최했다. 기계 학습 기반으로 한 이러한 음성 복제 기술을 통해 기업들은 누군가의 목소리가 담긴 5초짜리 영상만 있어도 실제와 같은 목소리로 원하는 다른 말을 하는 프로그램을 구현할 수 있다. 이러한 기술은 무엇보다 말하는 능력을 상실한 사람들에게 큰 도움이 될 것으로 생각되지만, 혹시라도 사기꾼의 손에 들어가게 되면 오용되기 십상이다. 이 음성 복제 기술을 소개한 한 기업은 사용자에 대한 심사를 수행하고 내부 서버를 통해 해당 기술을 운용함으로써 어떠한 오남용도 발생하지 않도록 주의를 기울이고 있다.

책임 메커니즘의 도입을 고려한다. 어떻게 하면 스스로 책임 의식을 가질 수 있을지 그리고 독립적인 기준의 적용이나 외부 전문가를 통해 자신의 인공지능을 한발 물러서 꼼꼼히 살펴보는 것이 타당한지를 따져 볼 필요가 있다. 예를 들어, 흑인 환자에게 불리한 차별적인 의사결정을 내리는 알고리즘 이야기로 돌아가 보자. 선의를 가진 직원들

은 알고리즘이 병세가 가장 심한 중증 환자에게 의료 개입이 집중되는 방향으로 사용될 수 있게 하려고 노력하고 있었다. 그 문제를 처음 발견한 건 외부에서 독립적으로 알고리즘을 점검했던 객관적 관찰자들이었다. 인공지능의 사용이 빈번해지면서 이러한 외부 도구 및 서비스도 점차 늘고 있다. 기업들은 이러한 도구의 적용을 고려해 볼 필요가 있다. ▢

메모