인공지능 시민사회포럼

인권과 안전을 위한
인공지능과 시민사회의 역할

11월 22일(수) 오후 2:00 ~ 6:00
참여연대 2층 아름드리홀

공동주최 건강권 실현을 위한 보건의료단체연합, 미디어기독연대, 민변 디지털정보위원회, 민주언론시민연합, 서울YMCA 시민중계실, 소비자시민모임, 언론개혁시민연대, 언론인권센터, 전국장애인차별철폐연대, (사)정보인권연구소, 진보네트워크센터, 참여연대

후원 진보통신연합 APC (Association for Progressive Communications)

문의 진보넷 사무국 02-774-4551 | jinbo.policy@gmail.com

시민사회포럼 순서

2:00 ~ 2:10	전체사회	오병일 (진보네트워크센터 대표)
세션1	**시민사회를 위한 인공지능 리터러시**	
2:10 ~ 3:00	강좌1	알파고에서 챗GPT까지, 인공지능에 대한 이해 | 이종민 (과학기술연합대학원대학교 교수)
3:00 ~ 3:50	강좌2	인공지능 알고리즘의 불투명성과 편향성 | 권오성 (성신여자대학교 법학부 교수)
3:50 ~ 4:10	**쉬는 시간**	
세션2	**한국의 인공지능 규율, 어떻게 할 것인가** 사회 | 장여경 (정보인권연구소 상임이사)	
4:10 ~ 4:40	발제	인공지능 법안의 시민사회 대안 | 김하나 (민주사회를위한변호사모임 디지털정보위원회, 법무법인 두율 변호사)
4:40 ~ 5:50	토론	박한희 (공익인권변호사모임 희망을만드는법 변호사)
		조아라 (언론인권센터 활동가)
		송경재 (민주언론시민연합 정책위원, 상지대 사회적경제학과 교수)
		윤명 (소비자시민모임 사무총장)
		이장희 (참여연대 공익법센터 운영위원, 창원대 법학과 교수)
	플로어 질의응답 | 전체토론	

<목 차>

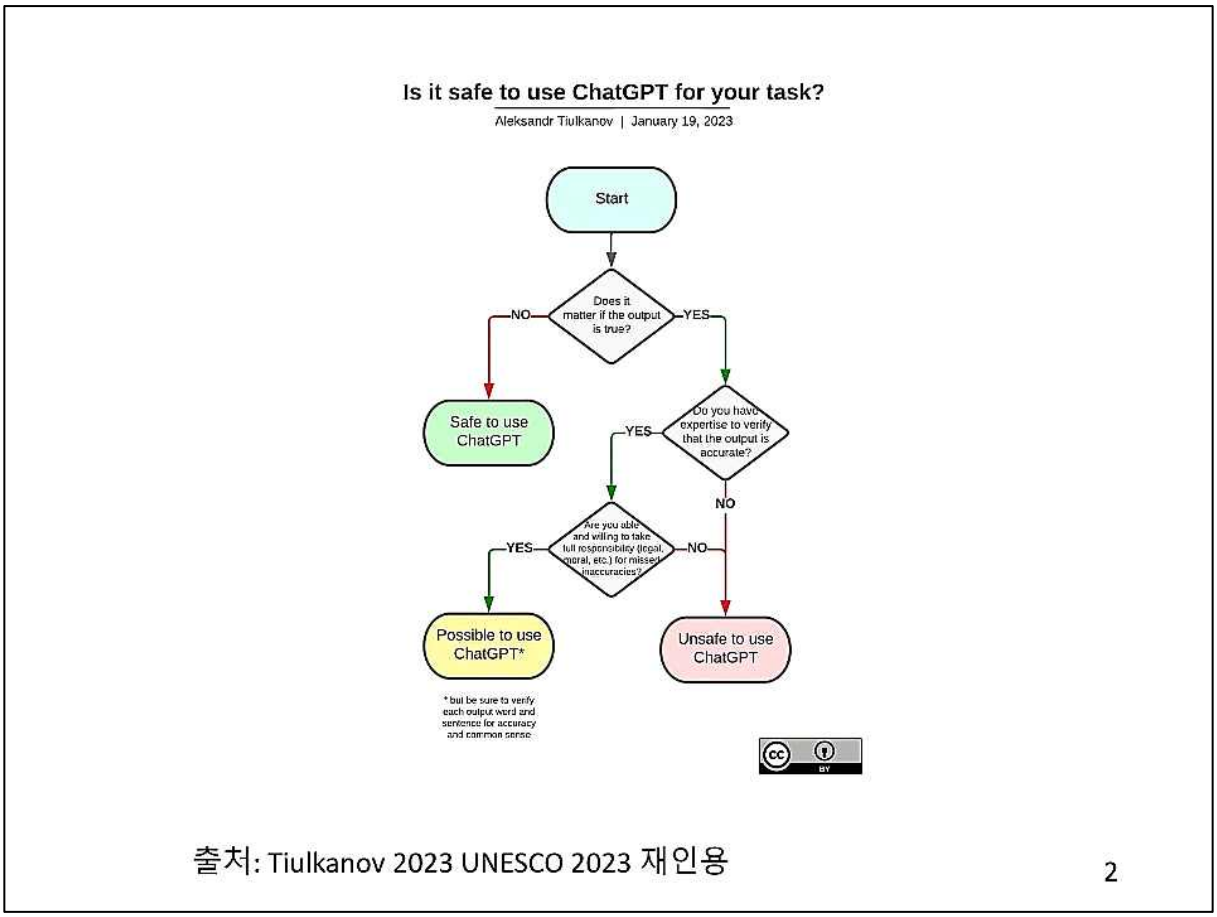
세션 1

시민사회를 위한 인공지능 리터러시

【 강좌 1 】

알파고에서 챗GPT까지, 인공지능에 대한 이해

이종민
(과학기술연합대학원대학교 교수)



인공지능…

- 알파고
- 기계번역
- 챗GPT

- 기술, 알고리즘
- 상품, 프로그램
- 시스템, 의사결정 프로세스
- 주체, 포스트휴먼

3

투명한

신뢰할 수 있는

설명가능한

책임감 있는

공정한

편향 없는

안전한

견고한

AI

검색엔진

내비게이션

로봇

번역

비서

자율주행

챗봇

추천 알고리즘

코딩

4

체스 대결 (인간 대 인공지능)
Garry Kasparov vs. IBM's Deep Blue

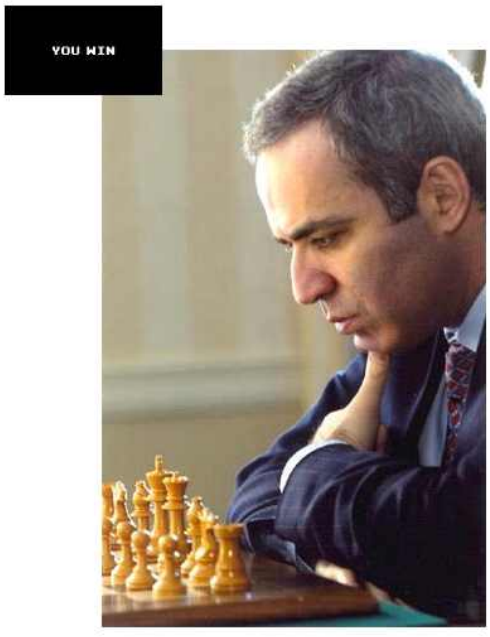
6번째 게임

1996. 2. 17, 미국 필라델피아
출처: AP, H. Rumph, Jr.

2번째 게임

1996. 2. 11, 미국 필라델피아
출처: AP, H. Rumph, Jr.

5



출처: 위키피디아

3 : 2 : 1

4 : 2

6

체스 대결 (인간 대 인공지능) Try again
Garry Kasparov vs. IBM's Deep Blue

2번째 게임, 재대결

1997. 5. 4, 미국 뉴욕
출처: AP, Adam Nadel

Feng-Hsuing Hsu 수 펑슝(오른쪽 아래)
와 Deep Blue 팀

1997. 5. 8
출처: AFP, Stan Honda Getty Images

7

YOU WIN

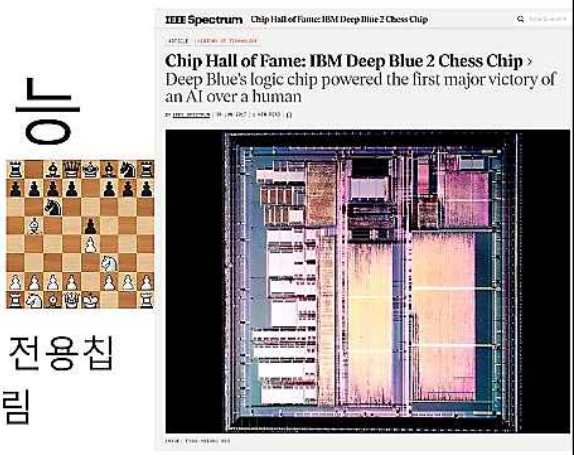
출처: 위키피디아

2 : 3 : 1

3.5 : 2.5

8

딥 블루의 인공지능


IEEE Spectrum Chip Hall of Fame: IBM Deep Blue 2 Chess Chip

Chip Hall of Fame: IBM Deep Blue 2 Chess Chip ›
Deep Blue's logic chip powered the first major victory of an AI over a human

완전 탐색 Brute Force search

"무식하게 푼다"

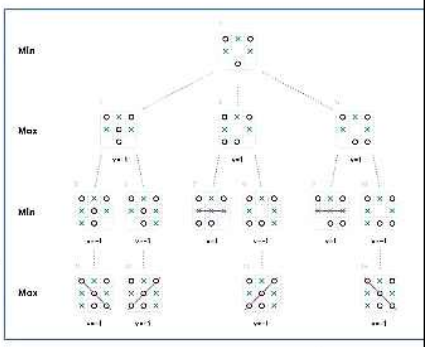
컴퓨터의 빠른 계산능력 활용, 전용칩

숫자가 커지면 시간이 오래 걸림

가지 치기 pruning

막다른 길, "다시는 탐색 안 해."

오프닝, 엔드게임 테이블베이스 활용.



+ 심리 전략 (속도 조절), 체스 전문가

출처: 네이버블로그 whitebearmjf (왼쪽 위)
IEEE Spectrum (오른쪽 위)
Computing.or.kr (오른쪽 아래)

9

nature

Explore content ∨ About the journal ∨ Publish with us ∨ Subscribe

nature › articles › article

Published: 27 January 2016

Mastering the game of Go with deep neural networks and tree search

David Silver ✉, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis ✉

Nature **529**, 484–489 (2016) | Cite this article

461k Accesses | **7900** Citations | **3082** Altmetric | Metrics

Abstract

The game of Go has long been viewed as the most challenging of classic games for artificial intelligence owing to its enormous search space and the difficulty of evaluating board positions and moves. Here we introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play. Without any lookahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play. We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks.

출처: Nature, 2016.1.28.

10

바둑 대결 (인간 대 인공지능)
AlphaGo vs. Lee Sedol

2번째 게임

2016. 3. 10, 한국 서울
출처: AlphaGo – The Movie

Demis Hassabis 데미스 하사비스(왼쪽),
David Silver 데이비드 실버(오른쪽 위),
Aja Huang 아자 황(오른쪽 아래)

출처: 구글 딥마인드 & techtarian

11

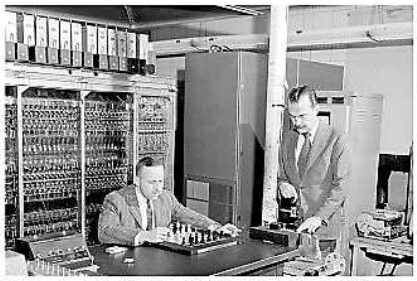
YOU WIN

출처: 나무위키(좌), 연합뉴스(우)

4 : 1

12

알파고의 인공지능(1)



맨해튼 프로젝트에 참여했던 물리학자 니콜라스 메트로폴리스(Nicholas Metropolis)가 1950년대 로스 앨러모스에 만든 컴퓨터 MANIAC과 체스를 변형한 게임을 하고 있다.

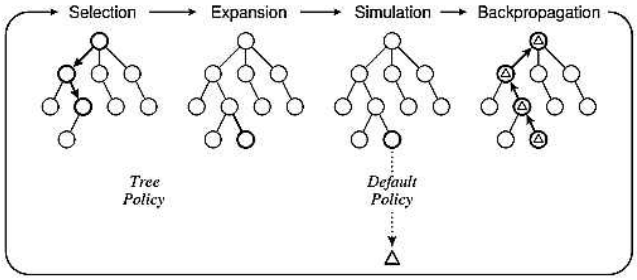
몬테카를로

무작위 표본 추출을 반복
확률적 방법으로 결과 유추

\+

트리 탐색

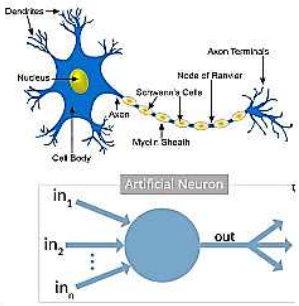
선택,
확장,
시뮬레이션,
역전파 (업데이트)



출처: Atomic Heritage Foundation (위)
Stackoverflow (아래)

13

알파고의 인공지능(2)



인간의 신경세포 뉴런의 작동원리에 착안하여 코넬항공연구소의 프랭크 로젠블랫 Frank Rosenblatt이 해군의 지원을 받아 고안함.

인공신경망 1957년

퍼셉트론 perceptron = perception + neuron
너무 이른 등장, 잊혀져버림.

+ 머신러닝

컨볼루션 convolution, 힌튼 Jeffrey Hinton팀
이미지넷 대회 우승 **2012년**

=딥러닝 2016년

정책망 policy network – 다음번 돌 놓기

가치망 value network – 승자 예측

지도학습 정책망
- 기보 (연습문제 풀이)

롤아웃 정책망
- 정확도 떨어짐, 빠름

강화학습 정책망
- 대국

출처: 이정원 바탕으로 작성

14

알파고 이후

2017. 05. 알파고 마스터 vs. 커제 3 : 0

2017. 10. 알파고 제로 vs. 알파고 마스터 89 : 11
알파고 제로 vs. 알파고 Lee 100 : 0

2018. 10. 알파 제로 Alpha Zero
바둑, 체스, 일본 장기 등에서
기존의 챔피언 프로그램들을 능가함

15

데미스 하사비스

우리는 AI의 리스크를 기후 변화와 같은 다른 주요한 글로벌 도전처럼 심각하게 여겨야 합니다. …
기후변화에 대한 효과적인 지구적 대응을 조율하는데 국제 사회는 너무 늦었고 우리는 그 결과 고초를 겪고 있습니다. AI에는 그렇게 늦어서는 안 됩니다.

출처: Dan Milmo, Guardian(2023.10.24.)

16

번역 대결 (인간 대 인공지능)
구글, 파파고, 시스트란 Vs. 번역사

문학 & 비문학

2017. 02. 21, 한국 서울
출처: 매일경제(2017.2.23)

<전체평가표>

번역사/AI	AI 1	AI 2	AI 3	번역사 합계
K▶E	**(13/30)**	(7/30)	(8/30)	**(24/30)**
E▶K	**(15/30)**	(8/30)	(9/30)	**(25/30)**
총점	**(28/60)**	(15/60)	(17/60)	**(49/60)**

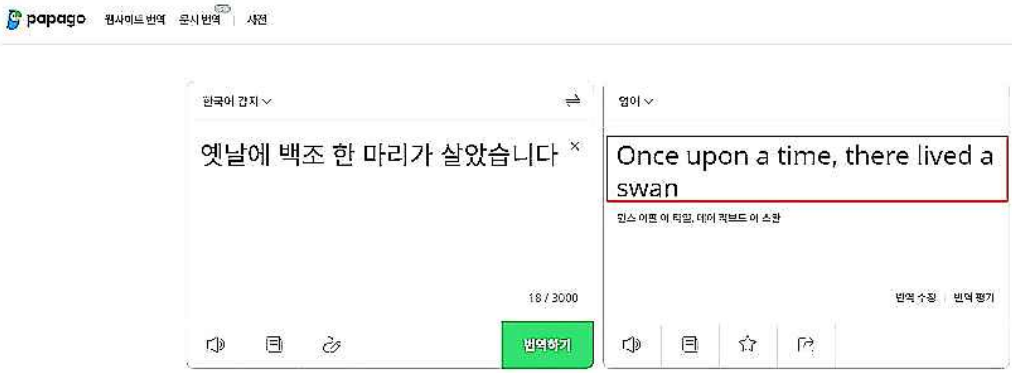
17

기계 번역

규칙 기반	1950	
예시 기반	1980	자연어 처리 Natural Language Processing
통계 기반	1990	트랜스포머 transformer
신경망 기반	2015	

18

Source: google translate, 2023.11.16



Source: papago, 2023.11.16

19

규칙 기반 Rule-based

변화를 따라갈 수 없다.
신조어
여러 의미를 가진 단어

언어는 문화를 반영

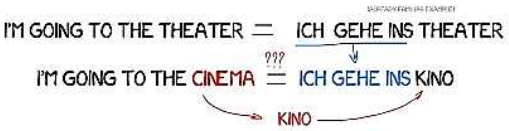
1950
1980
1990
2015

21

1950

예시 기반 Example-based

1980



출처: Pestov, 2018

경험을 통한 모방
유추에 의한 번역

長尾眞 나가오 마코토
출처: 일본 학사원

Ex) 영어와 일본어

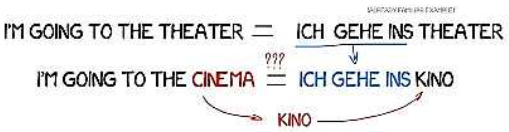
1990

2015

1950

예시 기반 Example-based

1980



출처: Pestov, 2018

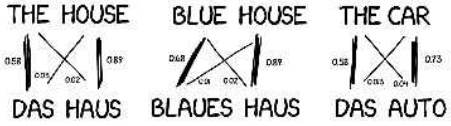
경험을 통한 모방
유추에 의한 번역

長尾眞 나가오 마코토
출처: 일본 학사원

Ex) 영어와 일본어

1990

통계 기반 Statistical



출처: Pestov, 2018

수많은 문장을 분석
단어 word, 어구 phrase, 구문 syntax

Ex) 통째로 외우기

2015

1950

1980

1990

2015

신경망 기반 Neural

양질의 데이터 多

수많은 문장 **통째로** 분석

문장 sentence

인코더 –띄어쓰기 단위로 분석

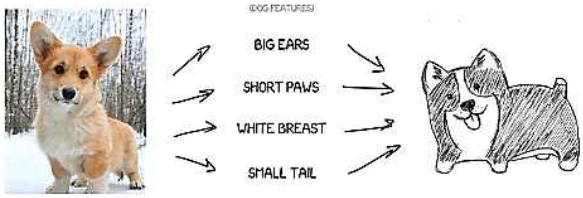
+(계산을 거쳐 압축)

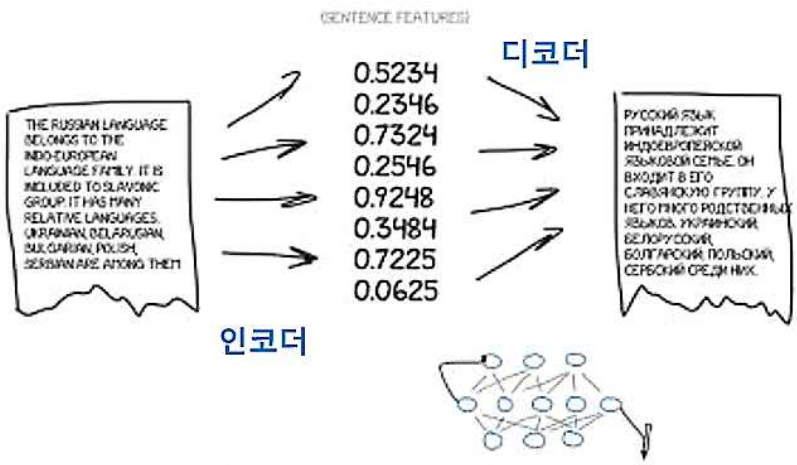
+(벡터 vector 형태로 전달)

디코더 – 압축벡터를 순서대로 품

어텐션 attention – 중요한 단어

24





단어마다 어텐션 바꿔감. 종료시 <끝> 전송

출처: Pestov, 2018

25

출처: Vaswani et al. (2017)

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

1 Introduction

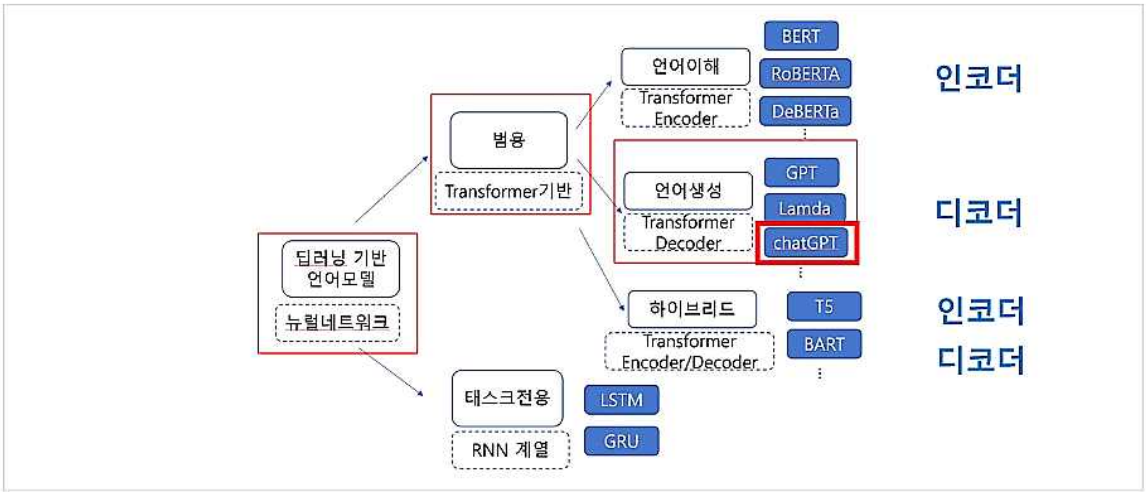
... 우리는 오직 어텐션 메커니즘에만 기반한 새롭고 단순한 네트워크 아키텍쳐, **트랜스포머**를 제안한다. 순환 신경망, 컨볼루션 신경망이 필요가 없다. ...

조경현

10년 넘게 헬싱키, 몬트리올, 뉴욕에서 살며 번역의 중요성을 느꼈어요. 그리고 인터넷 세상에선 번역이 더 중요해요. **온라인** 콘텐츠의 60%가 영어, 나머지 40%가 중국어·아랍어·불어 등으로 돼 있다고 해요. **영어** 편중이 너무 심하죠. … AI 번역이 잘 되면 이런 정보 **비대칭**을 해결하고, 디지털 **장벽**도 확 낮출 수 있어요.

출처: 조경현 인터뷰, 김미리 기자, 조선일보(2021.7.3)

<그림 3> 딥러닝 기반 언어모델



출처: 김은희·이민호·정유나·황명권, 2023

28

챗GPT Chat Generative Pre-trained Transformer

OpenAI

GPT-3.5 아키텍처 기반 대화형 인공지능

2022년 11월

언어 모델
거대 언어 모델
언어 생성 모델
생성형 AI

모델	매개변수	학습데이터	발표시기	주요특징
GPT	1억 1700만	미공개	2018.06	
GPT-2	15억	40GB	2019.02	유기적, 맥락이해 텍스트 생성(주로 영어)
GPT-3	1750억	570GB	2020.06	대규모, 다양한 응용 (여러 언어)
GPT-3.5	1조		2022.11	다목적, 다언어
GPT-4	~100조	미공개	2023.03	다중모달(텍스트, 이미지), 입력 크기 증가

출처: 박상길, 2022 바탕으로 내용 추가

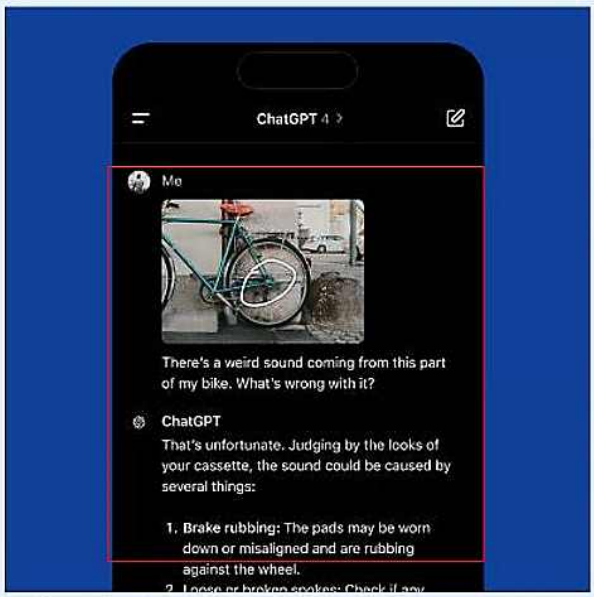
29

Ask me anything

Teach me to negotiate	Help me train for a half marathon	Explain why popcorn pops
Quiz me on vocabulary	Translate this recipe into Greek	Plan an itinerary for teens in Hong Kong
Plan a surf trip to Costa Rica	Rank dog breeds for a small apartment	Design a database schema
Draft a thank-you note	Help me plan a fun dinner party	Recommend an easy potluck dish
Explain this code	Draft a social media content calendar	Help me build a budget
Critique my short story	Write a polite rejection email	Explain nostalgia to a kindergartener
Find gentle lower back stretches	Generate fantasy football team names	Help brainstorm interview questions
Make this recipe vegetarian	Explain airplane turbulence	Summarize my meeting notes
Explain options trading like I'm 5	Write a spreadsheet formula	Write a Python script
Help me pick a halloween costume	Brainstorm domain names	Plan a college tour
Suggest rainy day activities	Write a thank-you note	Suggest photo shoot locations
Write a SQL Query	Help me debug this code	Teach me Mahjong for beginners
Help me with gift ideas for my dad	Create a mobility training workout	Draft a checklist for a dog-sitter
Draft an email for a repair quote	Brainstorm podcast episode ideas	Help me improve this job description
Troubleshoot my printer set-up	Review my argument for a debate	Rank e-bikes for daily commuting

출처: OpenAI, ChatGPT

30



Chat with images

You can now show ChatGPT images and start a chat. Troubleshoot why your grill won't start, explore the contents of your fridge to plan a meal, or analyze a complex graph for work-related data.

Learn more

Create new images

Create images simply by describing them in ChatGPT. Invent new logos, comic strips, and photorealistic scenes right in the chat. You can bring your ideas to life with our most capable image model, DALL·E 3.

Learn more

출처: OpenAI, ChatGPT

31

챗GPT의 인공지능

사전학습 Pre-training

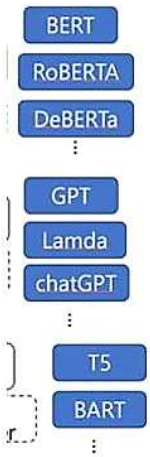
인간에 의한 강화학습 Reinforcement Learning from Human Feedback

미세조정 Fine-Tuning

32

챗GPT 이후 (1)

네이버	하이퍼클로바 HyperCLOVA	구글	바드 Bard
			버트 Bert
카카오	코지피티 KoGPT		미나 Meena
			T5
LG전자	엑사원 EXAONE	Meta	로베르타 RoBERTa
		MS	DialoGPT
SKT	에이닷 A.		
ETRI	엑소브레인		



출처: 고윤미 ·심정민, 2023; 김은희·이민호 · 정유나 ·황명권, 2023

33

챗GPT 이후 (2)

신약 개발, 로봇제어

문제풀이, 연구 보조

코딩

인공지능 창작?

실업, 고스트워크

양극화

디지털 원주민/이민자

표절, 부적절 인용

출처: 김은희·이민호·정유나·황명권, 2023; 손화철, 2023; 김병필, 2023; 오요한, 2023; 정한별·한경희, 2023

34

샘 알트먼

우리는 우리가 현재까지 출시한 이 도구들(챗GPT)이 줄 수 있는 혜택이 리스크보다 훨씬 크다고 믿는다. 그러나 동시에 이 도구들을 안전하게 사용하는 것이 매우 중요하다고 생각한다.

출처: 샘 알트먼, 미국 상원 공청회 (2023.5.16)

35

맺음말 (1)

이미 우리 주변에 와 있는 인공지능

학습 데이터의 유래

리스크에 대한 정의, 합의, 관리

자원, 노동, 그리고 대안들

36

맺음말 (2)

대상&수단: 인공지능에 대한, 인공지능을 이용한

주체&과정: 누구와 함께, 어떻게, 왜

목적&지향: 정확, 안전, 효율, 공정, 평화, 공존

37

참고문헌 (1)

- 고윤미·심정민, 생성형 AI 관련 주요 이슈 및 정책적 시사점, KISTEP 브리프 66, 2023.4.13.
- 김미리, 조경현 인터뷰, 조선일보, 2021.7.3.
- 김병필, 인공지능은 우리로부터 무엇을 추출하는가?, 2022. 12.
- 김은희·이민호 ·정유나 ·황명권, 언어모델, 사람과 소통하다, KISTI 이슈브리프, 53, 2023.2.28.
- 박상길, 비전공자도 이해할 수 있는 AI 지식, 2022.
- 손화철, ChatGPT가 던지는 물음들, 2023. 7.19.
- 오요한, 언어 모델의 크기를 늘리다 보면 예측 불가능한 능력이 창발하는가?, 2023. 9.
- 이정원, 알파고는 어떻게 바둑을 둘까, 2016.
- 정한별 ·한경희, ChatGPT가 한국 공학교육에 던지는 질문, 공학교육연구, 26, 5, 2023.9, 17-28.
- 크로퍼드, 케이트, AI 지도책, 2021/2022.

38

참고문헌 (2)

- Computing.or.kr
- Milmo, Dan, AI risk must be treated as seriously as climate crisis, Guardian (2023.10.24.)
- Pestov, Ilya, A history of machine translation from the Cold War to deep learning, FreeCodeCamp, 2018. 3. 12.
- Sam Altman, U.S. Senate Hearing (2023.5.16) cited by New York Times.
- Silver, David, Huang, Aja et al., Mastering the Game of Go with Deep Neural Networks and Tree Search, Nature, 529 (2016), 484-489.
- Vaswani, et al, Attention is All You Need, 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 2017.

39

이미지 출처

- 구글 번역
- 나무위키
- 네이버 파파고
- 네이버블로그
- 매일경제
- 연합뉴스
- 위키피디아
- 윤대기, 고려대
- 일본학사원
- 한국경제
- AFP
- AlphaGo – The Movie.
- AP
- Atomic Heritage Foundation
- Google Deepmind
- IEEE Spectrum, 2017.
- OpenAI
- Stackoverflow
- Systran
- Techtarian
- Tiulkanov, Aleksandr
- UNESCO, 2023.
- Wikidocs

40

【 강좌 2 】

인공지능 알고리즘의 불투명성과 편향성

– 채용과정에서의 AI 활용을 소재로 –

권오성
(성신여자대학교 법학부 교수)

Ⅰ. 들어가며

OECD는 인공지능 시스템을 주어진 목표 집합에 대하여 환경에 영향을 미치는 (예측, 추천, 결정 등의) 결과를 생성하여 환경에 영향을 미칠 수 있는 기계 기반 시스템이라고 정의한다.[1] 이러한 인공지능 시스템은 기계 및/또는 인간에 기반한 데이터 및 입력값(inputs)을 사용하여 (i) 실제 및/또는 가상 환경을 인식하고, (ii) 자동화된 방식(예컨대, 기계학습) 또는 수동의 분석을 통하여 이러한 인식을 모델로 추상화하며, (iii) 이러한 모델을 사용하여 결과값(outcomes)에 대한 선택을 공식화(formulate)한다.[2] 인공지능 시스템에 관한 OECD의 이러한 정의는 EU를 포함한 다양한 규제 당국에 의해 참조되고 있다.

오늘날 이러한 인공지능이 사용되는 영역은 음성인식, 자율주행차량, 언어의 자동번역, 질병 진단, 콘텐츠 큐레이션, 금융 투자, 맞춤형 학습, 예방적 방범, 사이버 보안까지 확대되고 있다.[3] 한편, 인공지능은 편익과 기술위험(technological risk)이라는 양면성을 수반하는데, 인공지능에 따르는 위험은 판단오류, 알고리즘의 편향성, 비도덕적 판단, 예측 불가

1) "An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives." (Organization for Economic Cooperation and Development (OECD). 2019. "OECD AI Principles Overview." https://oecd.ai/en/ai-principles)
2) Ibid.
3) 최은창(2021), "인공지능 위험인지의 차이와 거버넌스", p.146.

능한 오작동, 제어 불가능성 등 다양하다.[4] 현재 약인공지능은 다양한 분야에서 활용되고 있고, 이에 따라 약인공지능의 현실 세계에 대한 영향력이 커지고 있다. 그런데 인공지능은 대량의 데이터를 학습하여 성능을 향상시키는 머신러닝에 기반하고 있어 불확실성과 불투명성을 갖고 있으며, 노이즈 데이터로 오류를 일으킬 가능성이 상존한다. 인공지능 알고리즘의 파급력을 생각하면 문제가 발생한 이후에 법적 규율을 도모하는 것은 규율의 지연이 아니라 규율 불가로 귀결될 가능성이 크다.[5]

한편, 이러한 인공지능이 노동관계에 활용되는 방식은 다양할 것인바, 최근 인공지능을 활용한 채용면접이 늘어나면서 기존의 전통적 법규범이 인공지능을 활용한 채용면접으로 인한 제반 위험을 규율하지 못한다는 비판이 제기되고 있다.

종래 채용 시 면접은 사람과 사람 간의 대면으로 행하여졌다. 그러나 점점 더 많은 기업이 채용에 필요한 기간을 단축하고 비용을 줄이기 위하여 대면 면접 대신 비디오 면접을 도입하고 있다. 미국의 HireVue사의 'AI 기반 평가'는 접객업 및 금융업을 포함한 일부 산업에서 널리 활용되고 있으며, 현재 힐튼과 유니레버를 포함한 100개 이상 기업이 이 시스템을 사용하고 있고, 1,000,000명 이상의 구직자가 분석되었다고 한다.[6] 이러한 비디오 면접은 다른 사람이 비디오를 보고 평가하지 않고, 마이다스아이티 등이 제공하는 툴을 사용하여 구직자의 답변을 인공지능에 의해 평가하는 경우가 있다. 이러한 AI 면접 툴은 구직자의 움직임, 단어 선택 및 음성의 분석을 시뮬레이션하여 면접자를 조사한다.[7] 대부분 AI 분석은 1차 면접에만 이루어지지만, 각 구직자에게 고용 가능성에 대한 점수를 부여하는 경우가 있다.[8]

"인공지능이 채용 면접을 진행하는 소위 'AI 면접'의 사례를 살펴보자. 정필모 의원실에 따르면 SW마에스트로 연수생 합격자 150명의 면접 점수를 보면, AI가 A와 B+등급으로 평가

---

4) 최은창(2021)은 AI 기술위험을 작동 시 위험, 보안 위험, 통제 관련 위험, 윤리적 위험, 사회경제적 위험으로 분류한다. 자동화된 판단의 오류, 편향, 불투명성, 설명 불가능성은 작동 시 위험에 해당하며, AI를 이용하여 취약점을 공격하는 사이버 공격, 개인의 데이터 프라이버시 침해 등은 보안 위험이다. 통제 관련한 위험은 갑작스런 오작동, 인간의 통제권 상실, 자동화된 살상 무기 등이다. 상식, 평등, 약자에 대한 배려 등 인간 사회에 통용되는 가치를 코딩화하여 알고리즘 설계에 반영하기 어려운 점, AI와 인간 간의 가치 정렬(value alignment)에 발생하는 간극은 윤리적 위험으로 여겨진다. 실업 발생, 인종과 성별에 따른 차별, 자동화된 판단으로 피해가 발생해도 원인을 찾아내기 어려운 문제 등은 사회경제적 위험으로 분류하고 있다(Ibid).

5) 양종모(2020), "인공지능에 대한 법학의 위험한 해법", p.418.

6) Drew Harwell, *A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job*, The Washington Post (Nov. 6, 2019), <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>

7) *Ibid.*

8) *Ibid.*

했지만 면접관들은 하위 등과 , 2 1등으로 평가했다. 반대로 면접관들이 1등과 2등으로 평가한 합격자는 AI로부터 B0와 B-로 평가받았다. AI 등급 평균과 면접관들의 평균을 비교하면, 어떤 유의미한 값도 나오지 않았다. 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."[9)]

이처럼 AI 면접 툴을 사용하면 각 비디오 인터뷰를 사람이 검토하고 채점해야 하는 사용자의 부담을 줄일 수는 있겠지만, 이에 반대하는 입장에서는 AI 면접이 채용상 차별을 강화할 수 있다는 문제를 제기한다. 또한 대량의 개인 데이터를 인코딩하고 저장하는 시스템과 마찬가지로 개인정보보호 문제도 발생할 수 있다. 알고리즘은 본질적으로 객관적이지 않으며 이를 교육하는 데 사용되는 데이터와 이를 설계하는 사람을 반영한다. 즉, (기존의) 성차별을 포함한 사회적 편견을 학습하고 심지어 증폭시킬 수 있다.

Ⅱ. 인공지능에 대한 기존의 인식

1. 논의의 배경

인공지능은 구인기업이 채용 지원자를 선발하는 데 시간을 절약할 수 있는 도구[10)]이자, 인간의 의사 결정에 내재한 편견을 피할 수 있는 방법이라고 하면서 이를 옹호하는 주장이 있다. 반면, 인공지능 기반 채용 도구가 오히려 편견을 강화할 수 있다는 반론도 있다.[11)] 또한, 인공지능은 사전에 정해진 자격 요건 중 일부만 충족하는 비전통적인 구직자(nontraditional job candidates)에게 불이익을 줄 수 있다고 우려하는 견해도 있다. 또한, 일부 기업은 구성원의 인종적, 민족적 다양성을 높이기 위하여,[12)] 인공지능을 활용하고 있다.[13)]

---

9) 오정미(2021), "공정성, 투명성, 책임성 제고를 위한 인공지능 법제 방향", 「인공지능의 보장을 위한 법제 정비 방안 공정성·투명성·책임성 자료집」, 19-20면.
10) Jack Kelly (2023), How AI-Powered Tech Can Help Recruiters And Hiring Managers Find Candidates Quicker And More Efficiently. Forbes. https://www.forbes.com/sites/jackkelly/2023/03/15/how-ai-powered-tech-can-help-recruiters-and-hiring-managers-find-candidates-quicker-and-more-efficiently/?sh=2085b1f3a3fc
11) Tatiana Walk-Morris (2022), These are the flaws of AI in hiring and how to tackle them. World Economic Forum. https://www.weforum.org/agenda/2022/12/ai-hiring-tackle-algorithms-employment-job/
12) Chika Dunga (2020), How effective is artificial intelligence in removing racial bias in hiring?. Quartz. https://qz.com/work/1923587/can-artificial-intelligence-solve-racism
13) Sascha Brodsky (2022), Why Companies Are Using AI to Increase Diversity, But It May Not Work. Lifewire.

이러한 채용에서 인공지능 사용의 확대는 특히 인종이나 민족에 따른 지원자 대우와 관련하여 채용 과정의 다양성, 차별, 편견에 관한 사회적 논쟁을 불러왔다. 인공지능을 옹호하는 사람들은 인공지능이 무의식적인 편견을 없애고,[14] 일터에서의 다양성을 개선[15]할 수 있다고 주장한다. 그러나 다른 사람들은 AI가 기존의 편견을 고착화하고,[16] 차별적인 결정을 내릴 수 있다는 우려[17]를 제기한다.[18] 이러한 문제는 입법기관[19]과 규제기관[20] 모두에서 활발하게 논의되고 있다. 또한, 기업이 채용 시 알고리즘에 의한 차별 혐의로 소송을 당하기도 한다.[21] 그러나, 인공지능이 채용에서 편견을 없앨 수 있는지,[22] 아니면 오히려 이를 증폭시킬 것인지[23]에 관해서는 여전히 논쟁이 계속되고 있다.

한편, 취업을 희망하는 구직자의 입장에서 이들은 점점 더 사람인 '인사 담당자'뿐만 아니라, 그들을 선별(screen)하는 인공지능 앞에서도 최선을 다해야 한다. 오늘날 채용에 인공지능을 사용하는 것은 일반적이며, 지원서류(입사지원서, ES)의 선별(screening applicants)[24]에서부터 면접 수행(conducting interviews)[25]에 이르기까지 다양한 형태의

---

https://www.lifewire.com/why-companies-are-using-ai-to-increase-diversity-but-it-may-not-work-6754422

14) Frida Polli (2019), Using AI to Eliminate Bias from Hiring, Harvard Business Review. https://hbr.org/2019/10/using-ai-to-eliminate-bias-from-hiring

15) Nick Martindale (2022), Using artificial intelligence to promote diversity & inclusion, Information-age. https://www.information-age.com/using-artificial-intelligence-to-promote-diversity-inclusion-123500943/

16) Madeline Halpert (2022), AI-Powered Job Recruitment Tools May Not Improve Hiring Diversity, Experts Argue, Forbes. https://www.forbes.com/sites/madelinehalpert/2022/10/09/ai-powered-job-recruitment-tools-may-not-improve-hiring-diversity-experts-argue/?sh=6df23a73a743

17) Andrea Hsu (2023), Can bots discriminate? It's a big question as companies use AI for hiring, NPR. https://www.npr.org/2023/01/31/1152652093/ai-artificial-intelligence-bot-hiring-eeoc-discrimination

18) Chris Vallance (2022), AI tools fail to reduce recruitment bias - study, BBC. https://www.bbc.com/news/technology-63228466

19) Nicol Turner Lee, Samantha Lai (2021), Why New York City is cracking down on AI in hiring, Brookings. https://www.brookings.edu/articles/why-new-york-city-is-cracking-down-on-ai-in-hiring/

20) EEOC, Artificial Intelligence and Algorithmic Fairness Initiative. https://www.eeoc.gov/ai

21) Rory Bathgate (2023), Workday hit with claims its AI hiring systems are discriminatory, ITPro. https://www.itpro.com/business/policy-legislation/370133/workday-hit-with-claims-its-ai-hiring-systems-are-discriminatory

22) Liam Barrett (2021), Using Artificial Intelligence to Reimagine Enforcement of Workplace Discrimination Laws. Georgetown Journal on Poverty Law & Policy. https://www.law.georgetown.edu/poverty-journal/blog/using-artificial-intelligence-to-reimagine-enforcement-of-workplace-discrimination-laws/

23) Dawn Zapata (2021), New study finds AI-enabled anti-Black bias in recruiting. THOMSON REUTERS. https://www.thomsonreuters.com/en-us/posts/legal/ai-enabled-anti-black-bias/

인공지능 도구가 사용되고 있다.[26] 전문가,[27] 규제 기관,[28] 인사전문가[29] 모두가 인공지능이 기업의 채용 방식을 변화시키고 있다는 데 동의하고 있다.

채용 과정에서의 인공지능 사용이 증가함에 따라,[30] 일부에서는 이러한 방식으로 이루어지는 채용 결정의 공정성, 품질, 정확성에 의문을 제기하는 반면, 일부에서는 인공지능으로 인해 사람의 개입에 비해 개선되었다고 선전하기도 한다.[31] 인공지능은 이력서의 스캔 및 평가부터 후보자에 대한 점수(scoring) 또는 순위(ranking) 매기기, 면접 진행에 이르기까지 채용 과정에서 다양한 역할을 수행하고 있다.[32] 채용 과정에는 항상 '사람'이 필요하다는 주장도 있지만,[33] 인공지능의 역할을 수용하려는 기업들의 움직임으로 인하여 채용에서 인공지능의 영향력이 어디까지 확대될 것인지에 대한 논의와 토론이 활발해지고 있다.[34]

---

24) Eric Hamilton (2022), From ATS-friendly resumes to interviews, AI is evolving the job search process, UF(University of Florida) News. https://news.ufl.edu/2022/08/ai-resumes-and-interviews/

25) Zahira Jaser, Dimitra Petrakaki (2023), Are You Prepared to Be Interviewed by an AI?, Harvard Business Review. https://hbr.org/2023/02/are-you-prepared-to-be-interviewed-by-an-ai

26) Andrea Hsu (2023), Can bots discriminate? It's a big question as companies use AI for hiring. NPR. https://www.npr.org/2023/01/31/1152652093/ai-artificial-intelligence-bot-hiring-eeoc-discrimination

27) MIT Technology Review (2021), Podcast: Beating the AI hiring machines, https://www.technologyreview.com/2021/08/04/1030513/podcast-beating-the-ai-hiring-machines/

28) EEOC (2022), The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees, https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence

29) John Hilton (2019), Will AI ever replace humans in recruitment?, HRD. https://www.hcamag.com/ca/specialization/hr-technology/will-ai-ever-replace-humans-in-recruitment/171232

30) Courtney Vinopal (2022), A Growing Reliance on AI in Hiring Is Making Regulators and Lawmakers Nervous, Observer. https://observer.com/2022/12/a-growing-reliance-on-ai-in-hiring-is-making-regulators-and-lawmakers-nervous/#:~:text=Nearly%20one%20in%20four%20organizations%20already%20use%20automation%20or%20artificial,with%205%2C000%20or%20more%20employees.

31) Linda Rosencrance (2022), How AI can give companies a DEI boost. Computerworld, https://www.computerworld.com/article/3663053/how-ai-can-improve-diversity-equity-inclusion.html

32) Tomas Chamorro-Premuzic (2023), 5 ways employers use AI to evaluate your career potential, Fastcompany. https://www.fastcompany.com/90838324/ways-employers-use-ai-to-evaluate-your-career-potential

33) Laura McQuillan (2017), Can AI really replace humans in HR?, HRD. https://www.hcamag.com/ca/specialization/hr-technology/can-ai-really-replace-humans-in-hr/129271

34) Rebecca Heilweil (2019), Artificial intelligence will help determine if you get your next job. Vox. https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen

한편, 모든 종류의 인공지능 시스템을 둘러싼 주된 쟁점 중 하나는 인공지능이 인간의 성과를 향상할 수 있는지 여부이다. 채용 과정에서도 비슷한 질문이 제기된다.[35] 예컨대, 인공지능 도구가 인간의 판단에 내재한 잠재적인 결함을 줄이거나 제거할 수 있을까? 인공지능시스템이 입사 지원서에는 바로 나타나지 않는 방식[36]으로 적합한 후보자를 놓치지는(miss) 않을까?[37]

2. 채용에서의 인공지능 활용에 관한 법적 규율의 필요성

오늘날 알고리즘은 기본적으로 은행, 잠재적 파트너, 제안 등 우리를 연결해 주는 역할을 한다. 직장에서도 마찬가지이다. 결원(缺員)을 효율적으로 채우기 위한 여러 소프트웨어 프로그램은, 집계된 정보와 데이터를 기반으로 채용을 위한 실시간 면접 일정을 잡기 훨씬 전에 지원자가 해당 직무에 충분한 자격이 있는가를 판단한다. 이것은 최근까지 인간의 통제 아래 있던 기술의 새로운 기능이 작동하는 곳으로, 종종 그러한 반응을 가져온 원인을 밝히지 않은 채 모호한 결론(ambiguous conclusions)[38]에 도달하기도 한다.

알고리즘은 소프트웨어가 데이터 세트의 확률적 평가를 통해 인간 에이전트가 정의한 대로 원하는 결과를 추구하는 활동을 완료하기 위한, 추상적이고 형식적이며 적응 가능한 지침의 계층 구조(a hierarchy of abstract, formal and adaptable instructions)로 설명할 수 있다.[39] 결과가 무엇인지에 관하여 서로 다른 생각을 가진 인간과 달리 "알고리즘(algocratic)" 시스템은 편견에 의해 방해를 받지 않는다.[40] 알고리즘은 변덕스럽지 않으며 속이거나 뇌물을 받을 수도 없다. 주류적인 견해에 따르면, 알고리즘은 객관성, 기술적 품질(technical quality) 또는 과학의 진정한 승리자이며,[41] 모든 후보자를 동일한 방식으로

35) David Windley (2022), Is AI The Solution To Hiring Bias (Or The Cause Of It)?, Forbes. https://www.forbes.com/sites/forbeshumanresourcescouncil/2022/09/13/is-ai-the-solution-to-hiring-bias-or-the-cause-of-it/?sh=ece98801dab8

36) Meghan McCarty Carino (2013), AI used for hiring and recruitment can be biased. But that's changing. Marketplace. https://www.marketplace.org/shows/marketplace-tech/ai-used-for-hiring-and-recruitment-can-be-biased-but-thats-changing-2/

37) Jodie Zerega (2023), The Perils of AI Resume Screening: Law Firms Are Missing Out on Great Candidates, WFXG, https://www.wfxg.com/story/48405562/the-perils-of-ai-resume-screening-law-firms-are-missing-out-on-great-candidates

38) I Ajunwa. 'The "black box" at work' (2020) 2 Big Data & Society 7. See also P Kim, 'Manipulating Opportunity' (2020) 106 Virginia Law Review 867–935.

39) AE Waldman, 'Power, Process, and Automated Decision-Making' (2019) 88(2) Fordham Law Review 613–32.

40) A Aneesh, 'Global Labor: Algocratic Modes of Organization' (2009) 27(4) Sociological Theory 347–70; J Danaher, 'The threat of algocracy: Reality, resistance and accommodation' (2016) 29(3) Philosophy & Technology 245–68.

41) Algorithm Watch, 'People analytics in the workplace – how to effectively enforce labor rights', https://algorithmwatch.org/en/auto-hr/.

처리하고 편파성을 피하도록 프로그래밍 되어 있다. 기술 낙관주의자(tech-optimists)에게 이러한 시스템은 무의식적 편견(unconscious bias)을 제거함으로써, 인간의 본성에서 가장 오류가 많은 요소를 배제한 선택을 할 수 있다는 점에서 의심할 여지가 없는 장점이 있다. 이는 최적의 후보자 물색(targeting)으로부터 이력서(curriculum vitae, CV) 분류, 지원서(application) 관리, 신원 조회(background screening) 및 원격 인터뷰 실행에 이르기까지 채용의 전체 '깔때기(funnel)'에 적용된다.[42] 수많은 지원자 풀을 처리하는 데 따르는 엄청난 복잡성으로 인하여 고려할 만한 인재를 추리기(narrow down)가 어렵다는 점을 고려할 때, 이는 사실이라고 하기엔 너무 좋은 일이다. 여러 학자가 데이터를 통해 자동화된 채용 및 직장 내 의사 결정 프로세스의 어두운 면을 강조하고 있다. 프로그래머 캐시 오닐(Cathy O'Neil)은 알고리즘은 "수학에 내재된 의견(opinions embedded in mathematics)"일 뿐이라고 말한다.[43] 또한, 이러한 시스템이 머신러닝과 결합할 경우, 통제할 수 없는 규모의 효과가 발생하게 된다.

법학자인 이페오마 아준와(Ifeoma Ajunwa)의 연구에서 설명된 바와 같이, 차별은 근본적인 특징(underlying feature)이 된다.[44] 웹사이트의 드롭다운 메뉴에서 출생 연도를 차단하면 전체 코호트(cohort)를 제외할 수 있으며, 특정한 구인 광고를 회사 내의 다른 구성원과 잘 맞는 어울리는 독자(audience)들이 볼 수 있도록 Facebook 필터를 사용할 수 있다. 심지어 가족에 대한 책임으로 인해 이력서(CV)에 공백이 생길 가능성이 큰 여성에게 다른 자격 요건과 관계없이 간접적으로 불이익을 주는 방식으로 대학 학업이나 경력 발전에 관한 질문을 작성할 수도 있다. 예를 들어, Amazon은 지원서에 기술 업계 전반의 남성 우월주의를 반영하는 "남성적인" 문구가 부족한 여성 지원자에 대한 편견이 드러나자, 그들의 수동적인 지원자를 찾아내는(uncovering passive candidates) 도구를 폐기해야 했다. 아마존의 AI 시스템은 "남성 후보가 더 낫다고 스스로 학습"한 것이다.[45] 경기 침체기에 구직에 어려움을 겪어 정규직 고용에 공백이 생길 수 있는 사람들도 마찬가지이다. 알고리

42) M Bogen and A Rieke, 'Help wanted: An examination of hiring algorithms, equity, and bias' (2018).

43) C O'Neil, Weapons of math destruction: how big data increases inequality and threatens democracy (New York, Crown, 2016). See also AE Waldman, Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power (Cambridge, Cambridge University Press, 2021).

44) I Ajunwa, 'Beware of Automated Hiring' The New York Times (8 October 2019) www.nytimes.com/2019/10/08/opinion/ai-hiring-discrimination.html, I Ajunwa, 'The paradox of automation as anti-bias intervention' (2019) 41(5) Cardozo Law Review 1671–1742. See also P Kim and S Scott, 'Discrimination in Online Employment Recruiting' (2018) 63(1) St. Louis University Law Journal 93–118.

45) J Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women' (Reuters, 11 October 2018) www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G; M Oppenheim, 'Amazon scraps "sexist AI" recruitment tool' Independent (11 October 2018) www.independent.co.uk/life-style/gadgets-and-tech/amazon-ai-sexist-recruitment-tool-algorithm-a8579161.html.

즘의 한계를 꼽으라면, 민감한 데이터에 대용물(proxies)을 사용하여 성별, 인종, 사회적 출신과 관련된 편견을 도입하거나 발전시킬 수 있다는 점이다. 과거의 데이터가 차별적인 관행을 반영하는 경우 ("garbage in, garbage out"이라는 컴퓨터 과학자들의 표현처럼) 결과도 차별적일 수 있다. 따라서 결국 과거의 행동과 시스템 프로그래밍에 정보를 제공한 일련의 값에 따라 예측과 결정을 내리게 된다.[46] 중요한 것은, EU의 경우 법적 구제수단이 새로운 형태의 다차원적 차별(multidimensional discrimination)에 대응할 수 있는 기능을 부분적으로만 갖추고 있다는 점인데, 이는 상관관계(correlation)를 만드는 데 사용되는 알고리즘의 역동적이고 복합적인 특성으로 인한 결과이다.[47]

특정 데이터의 잘못된 해석과 관련된 무의식적이고 통제하기 어려운 프로세스가 있다. 예를 들어, 자동차 사고 발생 빈도에 대한 통계에서, 교통량이 많다는 이유로 도심 지역이 위험한 지역으로 표시된다면, 도심에 거주한다는 사실만으로 거주자의 보험료가 더 높게 책정될 수 있다. 이는 많은 대도시, 특히 비싼 교외에 살 여유가 없는 소수 민족이 도심에 거주하는 경우가 많은 유럽 외 지역의 많은 대도시에서 발생한다. 보험회사의 알고리즘이 소수자 커뮤니티 구성원을 차별하도록 의도적으로 프로그램된 것은 아닐 수 있지만, 겉으로 보기에 중립적인 데이터의 상관관계는 차별적인 결과를 낳는다.[48] 그러므로, 어떤 사용자가 근속 기간을 극대화하기 위해 통근 시간이 이직률의 감소를 결정하는 가장 중요한 변수라는 사실을 알아냈다고 할 때, 이러한 요인이 주거비의 감당 능력(housing affordability)이나 인종과 밀접한 상관관계가 있다는 점이 언급되지 않은 진실(untold truth)이다.[49]

주지하는 바와 같이, 불완전하거나 왜곡된 사회 패턴의 흡수(absorption)는 배제되고 취약한 커뮤니티에 더 많은 불이익과 역사적인 격차를 초래하며, 따라서 여성 혐오(misogyny), 인종 차별, 권위주의의 유산을 강화한다.[50] 특히 실업 수당, 부양 수당, 주택 보조금 또는 예측적 경찰 활동(predictive policing)을 담당하는 정부나 공공 기관에서 이러한 시스템을 채택할 경우, 일종의 빈곤에 대한 과세(tax on poverty)를 조장하게 된다.[51] 정치학자 버지니아 유뱅크스(Virginia Eubanks)는 자동화된 의사 결정 시스템의 결

46) A Agrawal, J Gans and A Goldfarb, Prediction Machines: The Simple Economics of Artificial Intelligence (Brighton MA, Harvard Business Press, 2018); G Resta, 'Governare l'innovazione tecnoϖlogica: decisioni algoritmiche, diritti digitali e principio di uguaglianza' (Rome, Forum Disuguaglianze e Diversità, 2019).

47) R Xenidis, 'Tuning EU equality law to algorithmic discrimination: Three pathways to resilience' (2020) 27(6) Maastricht Journal of European and Comparative Law 736-58.

48) For a similar EU case based on gender, see Case C-236/09 Association belge des Consommateurs Test-Achats ASBL and Others v Conseil des ministres [2011] ECR I-00773.

49) J Walker, 'Meet the New Boss: Big Data' (The Wall Street Journal, 20 September 2012) www.wsj.com/articles/SB10000872396390443890304578006252019616768.

50) See generally R Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code (Cambridge, Polity, 2019); S Skinner-Thompson, Privacy at the Margins (Cambridge, Cambridge University Press, 2020).

51) SU Noble, Algorithms of oppression (New York, New York University Press, 2018). See also

정적인 단점으로 유연성의 부족과 자동화된 프로세스로 인해 부당한 불이익을 받았다고 느낄 때마다 설명을 들을 수 없다는 점을 꼽았다.[52] 이러한 불균형은 고질적인 문제일 가능성이 크며, 그 결과는 카프카적인(Kafkaesque) 것과 디킨스적인 것(Dickensian) 사이의 어딘가에 속할 것이다. 오류에 도전하는 것은 어려운 싸움이다.[53]

채용 전 단계에서의 프로파일링은 점점 더 널리 사용되는 채용 모델이며, 오늘날에는 감정, 성격 특성 및 행동을 추적하기 위해 AI 기반 안면 및 음성 분석도 도입되고 있다. 많은 기업, 특히 대기업에서 이러한 시스템을 사용하고 있으며, 대규모의 신입직원 채용에서 "인재(talent)"를 찾는 데 있어서 이러한 기술이 인간(flesh and blood) 헤드헌터보다 훨씬 더 신뢰할 수 있고 저렴하다는 확신을 갖고 있다. 포춘 500대 기업의 98%가 채용 단계에서 알고리즘 또는 데이터 기반 시스템을 사용한다.[54] 여러 식품 및 위생 제품 브랜드를 소유한 한 대형 다국적 기업은 잠재적 영업 사원에게 휴대폰으로 여러 가지 질문에 답하도록 요청하는 것으로 알려졌다. 그 결과는 과거 성공적인 채용으로 이어진 모든 인터뷰를 기반으로 프로그래밍된 AI 시스템에 의해 분류된다. 평가는 제스처(미소, 눈빛, 떨림) 뿐만 아니라 수동태 동사의 사용, "우리(we)"보다 "나(I)"라는 대명사의 선호도, 단어 선택 및 복잡성, 문장 길이 등을 기준으로 이루어진다. 이렇게 하면 필요한 모델을 구현하지 않는 사람은 배제된다.[55]

필터링은 정보를 수집하고 통계를 처리하며 온라인에 업로드된 이력서에서 찾은 키워드를 기반으로 판단을 내리는 "블랙박스"에 맡겨진다. 전화 통화 횟수와 시간, 근무 시간 동안 검색한 웹사이트 목록, 동료 간 대화의 어조와 내용뿐만 아니라 지리적 위치 및 개인 소셜 미디어의 태그를 통해 추적한 방문 장소 목록 등 직장 내·외부의 다양한 출처에서 데이터를 수집할 수 있다. 여기에 텍스트 문서를 컴퓨터로 분석하는 "자연어 처리(natural language processing)"의 잠재력을 더하면 근로자 자신도 모르는 개인적 특성까지 포함한 매우 상세한 정보를 얻을 수 있다.

B Harcourt, Against Prediction: Profiling, Policing and Punishing in the Actuarial Age (Chicago, University of Chicago Press, 2006).

52) V Eubanks, Automating Inequality. How High-Tech Tools Profile, Police and Punish the Poor (London, St. Martin's Press, 2018).

53) A Murad, 'The computers rejecting your job application' (BBC, 8 February 2021) www.bbc.co.uk/news/business-55932977.

54) J Fuller, M Raman, E Sage-Gavin and K Hines et al, Hidden Workers: Untapped Talent Harvard Business School Project on Managing the Future of Work and Accenture (September 2021); N Lewis and J Marc, 'Want to work for L'Oreal? Get ready to chat with an AI bot' (CNN Business, 29 April 2019) https://edition.cnn.com/2019/04/29/tech/ai-recruitment-loreal/index.html.

55) See the MIT podcast about the automation of everything: https://forms.technologyreview.com/in-machines-we-trust/; B Waber, People Analytics: How Social Sensing Technology Will Transform Business and What it Tells us About the Future of Work (Upper Saddle River, FT Press, 2013).

"Predictim"이라는 상징적인 상호를 사용하는 온라인 서비스는 소셜 미디어에 작성된 모든 문장을 스캔하여 지원자의 성격을 유추하는 시스템을 통하여 프로파일링 활동을 수행할 수 있다고 보장한다. 미국 유타주에 본사를 둔 이 분야의 선도 기업인 "HireVue"는 카메라와 알고리즘을 사용하여 25,000개 이상의 안면 및 언어 정보 샘플(눈썹 올리기부터 미소 짓기, 턱 올리기, 입술 다물기 등)을 데이터베이스와 대조하여 누가 채용에 필요한 자질을 갖추고 있는지 판단함으로써 지원자의 "취업역량(employability)"을 측정할 수 있다고 주장한다. HireVue는 예비 사용자가 선택한 6가지 질문을 사용하여 최대 50만 개의 데이터 포인트와 얼굴 특징을 계산한다. 이것이 검사 대상자 각자에게 할당된 점수의 "구성요소(ingredients)"이다. 성공 지수는 기술, 근성, 학습 태도, 성실성 및 책임감, 가족력, 소비 성향, 개인적 안정성 등 이러한 지표를 기반으로 구성된다. "미소에서 행복을, 찌푸린(scowl) 얼굴에서 분노를, 찡그린(frown) 얼굴에서 슬픔을" 추론하는 것은 기껏해야 터무니없는 일이다.[56] 더 큰 문제는 이 방법이 문화적 차이를 과소평가하고 보편적인 표본으로부터 지나치게 단순화된 방식으로 도출한다는 점이다. 연구자들은 이 시스템이 "과학적 사실에 근거하지 않은 피상적인 측정과 자의적인 숫자 계산이 혼합된 것"이라고 주장한다.[57] 실제로 HireVue는 2021년 1월 "대다수의 직업과 산업에서 시각적 분석은 알고리즘 평가의 다른 요소보다 직무 성과와 상관관계가 훨씬 적다"라는 결론을 내린 후 더는 직무 평가 목적으로 얼굴 분석을 사용하지 않겠다고 발표했다.[58] 그러나 언어 및 구두 데이터는 지원자의 고용 가능성을 평가하는 데 있어 경험에 대한 개방성, 성실성, 외향성, 동의성(agreeableness), 정서적 안정성과 같은 성격적 자질을 추론하는 데 여전히 중요한 요소이다.[59] 졸업생의 경쟁력을 높이기 위해 일부 미국 대학은 이러한 사이비 과학의 기초를 가르치는 과목을 도입할 계획이다. 따라서 전략적인 키워드와 미소를 이용한 게임 시스템이 곧 학업 커리큘럼에 도입될 수 있다. 이러한 시스템이 더 효과적일수록 새로운 직원이 종전 직원의 프로필을 따르는 획일적인 기업 문화, 즉 돌이킬 수 없는 동질화(homogenization) 과정을 만들 위험성이 커진다.

자동화된 채용의 논리는 인사 관리자가 이전의 경험과 진행 중인 프로젝트를 고려하여

56) LF Barrett, R Adolphs, S Marsella, AM Martinez and SD Pollak, 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements' (2019) 20(1) Psychological Science in the Public Interest 1-68.

57) D Harwell, 'A face-scanning algorithm increasingly decides whether you deserve the job' (The Washington Post, 6 November 2019) www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/.

58) L Zuloaga. 'Industry leadership: New audit results and decision on visual analysis' (Hire Vue Blog, 12 January 2021) www.hirevue.com/blog/hiring/industry-leadership-new-audit-results-and-decision-on-visual-analysis.

59) M Murgia, 'Emotion recognition: can AI detect human feelings from a face?' Financial Times (12 May 2021) www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a926109452.

공고된 직책에 적합한 기술, 잠재력, 적성, 성향을 파악함으로써 지원자를 심층적으로 파악할 수 있도록 하는 것이다. 이는 지역(local) 노동시장의 경계와 후보자 풀을 넓힌다는 추가적인 장점이 있다. 소수의 "구매자"와 달리 "공급자"가 너무 많다는 사실은 전자를 유리한 상황에 놓이게 한다. 불확실성은 채용 프로세스의 모든 단계에 영향을 미치며, 장래의 사용자와 구직자 모두의 입장에서 헌신, 좋은 태도, 신뢰성과 같은 자질을 평가하는 것은 거의 불가능하다. 이러한 만성적인 불확실성은 프로필(profiles)과 제안(offers) 사이의 불일치를 초래하고, 이론적으로는 기업을 마비시킬 위험이 있다. 반면에 예비 직원을 꼼꼼하게 선발하면 시간과 비용은 많이 들지만, 안정적 관계와 낮은 이직률이라는 보상을 얻을 수 있으며, 생산성 향상과 성공으로 이어지는 오래 지속되고 유익한 파트너십의 기반을 마련할 수 있다.[60]

Ⅲ. 관련 정책의 유형[61]

1. 개관

회사가 근로자들에 대한 데이터를 수집하고 알고리즘을 활용하여 역량, 생산성을 평가하고 해고, 계약해지 등에 이용하는 상황에서, 회사가 이용하는 데이터와 시스템에 대해 투명하게 알리지 않으면, 근로자들은 이런 시스템으로 인해 발생하는 문제와 피해에 대해 해결할 방법 및 구제수단을 갖지 못하게 된다. 따라서 '통지와 투명성(Notice and Transparency)'은 회사가 근로자에게 데이터처리와 알고리즘에 대해 실질적 정보를 근로자에게 제공하거나, 적어도 제3자 감사기구나 정부기관에 주기적으로 정보를 제공하도록 하여, 위와 같은 상황을 시정하고자 제안되는 정책이다.

그런데, 이러한 투명성만으로는 회사가 데이터처리시스템과 알고리즘을 책임 있게 이용하도록 보장하기에 부족하다. 근로자의 생체정보나 계좌정보 등 중요한 정보가 유출되는 사고로 인해 발생할 수 있는 위협이 존재할 뿐만 아니라, 알고리즘은 사생활 침해, 자기결정권, 존엄성을 위협하고 건강과 안전에 위협이 되는 작동을 할 수 있다. 따라서, 책임성(Accountability)을 강제하는 정책을 통하여 근로자의 데이터를 보호하고, 데이터 수집 및 활용, 알고리즘 이용을 하는데 있어 위험을 방지하고 처리할 수 있게 적절한 조치를 취하도록 강제할 필요가 있다.

---

60) L Graham, A Gilbert, J Simons and A Thomas, Artificial intelligence in hiring, Assessing impacts on equality (Institute for the Future of Work) www.ifow.org/publications/artificial-intelligence-in-hiring-assessing-impacts-on-equality.

61) 이 절의 내용은 Emlyn Bottomley, Data and Algorithms in the Workplace: An Overview of Current Public Policy Strategies, Center for Labor Research and Education, UC Berkeley, Working Paper, Technology and Work Program (11.17.20)의 내용을 주로 참고하였다.

한편, 개인에게 자신들의 데이터에 대해 보호권, 즉 개인의 정보권(Individual Data Rights)을 부여하여, 개인이 자기 데이터의 수집·이용·제한 등에 대해 통제를 할 수 있도록 하는 정책이다. 특히, 업무를 하는 기간 전체에 걸쳐 개인정보의 수집·처리·이용에 대해 동의를 받도록 요구하고, 이의제기권의 부여, 부정확한 데이터의 삭제 또는 시정권, 생성된 데이터 접근 및 이동권이 확보하도록 규제해야 한다.

나아가, 데이터처리시스템 및 알고리즘은 확립된 고용 및 노동 법규를 우회하여 근로자에게 피해를 주는 작동을 할 수 있다. 보호해야 하는 계층에 대해 편향성을 드러내는 결과를 생성해 내거나(채용 시 인종과 관련된 요인을 고려하는 채용 알고리즘 사용) 근로자의 건강과 인권에 위협을 가하는 문제적 알고리즘 작동에 대하여 기존의 노동관계법령이 근로자에게 충분한 구제수단을 제공하지 못할 수 있다. 또한, 전자적 모니터링은 단체행동 등 근로자에게 보장되는 권리를 행사하는 것을 위축시키는 효과를 갖는다(월마트의 SNS 활동 감시로 인하여 노동조합의 활동이 제약된 사례). 사업장에서 근로자의 권리(Workplace Rights)를 보장하는 정책은 이러한 문제를 해결하기 위하여 차별금지법이나 고용 및 노동관계법을 보완하는 역할을 한다.

그러나, 위에 언급된 정책들만으로는 알고리즘이 가져오는 총체적, 누적적 피해의 해결이 어려울 수 있다. 개인은 시간 및 전문성의 부족으로 권리 행사에 어려움을 느낄 수 있고, 사용자로부터의 보복이 두려워 권리 행사를 꺼리게 될 수도 있다. 또한, AI 분야의 급속한 발전으로 특정 기술이나 행위 유형으로부터의 보호가 의미 없어질 수도 있다. 따라서, 데이터와 AI 기술이 노동에 미치는 영향에 대해 포괄적으로 관장하는 규제기관의 권한을 확장하거나 FDA와 같이 해당 업계 승인 및 규제 역할을 하는 새로운 기구를 설립하여 관련 표준 확립으로부터 책임 분배 및 결정에 이르기까지 광범위한 규제 권한을 갖도록 해야 한다(Government Oversight and Regulation).

2. 통지 및 투명성(Notice and Transparency)

1) 통지(Notice)

근로자 및 일반에 끼치는 알고리즘과 데이터처리 시스템의 위험 및 혜택을 평가하고 이에 대해 대응하기 위해 투명성 조치 및 충분한 통지는 실질적 정책을 위한 기반으로서 필수적이다.

개인 근로자들은 플랫폼기업의 데이터처리시스템 및 알고리즘의 존재, 활용 정도 및 성격에 대해 모르는 경우가 많으며, 플랫폼 기업은 활용 기술의 의도된 목적에 대해 설명하지 않는 경우가 많다. 통지 정책은 개인들에게 그들의 개인정보가 어떻게 수집, 처리, 이용

되는지 정보를 제공하는 의도로 도입되어 시스템 이용 사실, 목적, 범위 및 데이터 처리 및 알고리즘 작동의 예상되는 결과에 대해 회사가 개인에게 알리도록 강제할 수 있다.

가. 통지와 관련한 기존의 제도

가) 데이터 수집 통지(Notice of data collection)

많은 국가들은 개인의 정보를 직간접적으로 수집 처리하는 자에게 이에 대해 개인에게 통지할 의무를 부과한다. EU의 General Data Protection Regulation(GDPR)에 따르면 개인정보를 수집, 처리, 이용하는 자는 데이터 수집한 자, 데이터 처리의 목적 및 법적 근거와 이 정보에 대해 접근권을 갖는 개인 유형에 대해 알려야 한다. 유사한 법률인 California Consumer Privacy Act(CCPA)는 수집되는 정보 유형 및 이용 목적에 대해 개인에 알리도록 하며, 워싱턴, 뉴멕시코도 유사한 통지 규정을 두고 있다.

나) 알고리즘에 의한 결정 사실의 통지(Notice of decisions made or assisted by algorithms)

알고리즘에 의한 결정의 대상이 된 개인에게 통지를 할 것을 의무화하는 개인정보보호 정책이다. GDPR은 그러한 개인에게 알고리즘에 의한 결정의 존재 및 그러한 시스템이 끼칠 수 있는 영향에 대해 알릴 의무를 부과한다. 어떠한 경우에는 특정 상황에 대해서만 통지의무가 발생하기도 하는데, 예를 들면 Illinois에서는 비디오 녹화 인터뷰를 평가하는 것에 AI를 이용하기 전에 통지를 하고 동의를 받도록 요구된다.

다) 위반의 통지(Breach notifications)

데이터 관련 보안이 침해되었을 때 이에 대해 통지할 의무는 보편적으로 존재하는 정책이다.

라) 생체정보 관련 통지(Biometric notice requirements)

Illinois, Texas, Washington과 같은 주에서는 근로자의 생체식별정보를 수집 또는 처리하는 자에게 사전 통지를 하고 어떻게 이 정보가 이용되는지 설명할 의무를 부과한다.

나. 도입이 제안되는 통지 관련 제도

가) 전자 모니터링 또는 데이터 수집에 대한 통지(Notice of electronic monitoring or data collection)

전자적 방법으로 근로자의 활동이나 소통을 관찰하는 경우 이에 대한 통지를 해야 한다는 제안이다. 카메라, 키보드 사용 감지, 컴퓨터 추적 소프트웨어, 이메일 모니터링 또는 GPS 추적 등을 포함하여 근로자에게 전자적 모니터링의 활용 여부, 어떠한 모니터링 기술이 이용되는지, 수집되는 정보는 무엇이며 그 이용 의도는 무엇인지, 상벌 또는 징계의 대상이 되는 행동에는 어떤 것이 있는지 등을 통지해야 한다.

나) 모니터링 표시(Indication of monitoring)

관찰 대상 근로자에게 감시가 이뤄지고 있다는 시각적, 청각적 표시(감시 카메라에 점등, 메시지 또는 스크린 팝업 창 등)를 제공해야 한다.

다) 공정신용정보법을 개인정보 브로커에 확장 적용 (Expansion of the “Fair Credit Reporting Act” to cover data brokers)

공정신용정보법(The Fair Credit Reporting Act, FCRA)[62]은 신용정보 등 보고서 조회 전에 채용신청자로부터 사전 동의를 얻어야 하며 보고서에 근거하여 부정적인 고용 관련 결정이 있는 경우 통지를 하도록 규정하는데, 데이터 브로커들도 근로자 관련 유사한 정보를 제공하므로 동일하게 공정신용정보법의 적용을 받아야 하며, 미국의 FTC에 상응하는 기관 등이 소비자 보호 기관으로서 법 적용을 집행해야 한다.

라) 소프트웨어 규칙의 공개(Disclosure of “software rules”)

업무 할당, 수행 평가 및 대가 설정 등을 하는 데 있어 근로자를 관리하기 위해 알고리즘을 이용하는 경우, 이들은 근로조건과 관련한 회사의 정책과 마찬가지로 근로조건에 대해 중요한 영향을 미치게 된다. 따라서, 회사는 근로자에게 노동과 관련한 결정을 하거나 그에 영향을 미치는 소프트웨어의 로직에 대해 쉽게 이해할 수 있는 설명을 제공해야 한다. 예를 들어, 업무시간 관리 소프트웨어가 어떻게 작동하는지(공제 시간이나 교대시간의 결정에 사용되는 정보 포함)에 대한 정보를 제공해야 한다는 것이다. 다만, 기계학습 알고리즘의 경우 로직을 문서화하는 것이 용이하자 않다는 현실적인 어려움이 있다.

62) The Fair Credit Reporting Act는 소비자 보고 기관 파일에 포함된 소비자 정보의 정확성, 공정성 및 개인 정보 보호를 촉진하기 위해 제정된 미국의 연방법이다.

2) 투명성(Transparency)

투명성 정책은 데이터처리시스템 및 알고리즘에 대해 제3자, 정부기관이나 공공에 정보를 공개하는 것을 의미한다. 공개대상 정보에는 알고리즘에 대한 기술적 상세사항 및 노동과 관련한 것으로서 업무시간이나 매출데이터와 같은 정보가 포함된다.

투명성은 개인의 권리가 침해되지 않고 알고리즘이 의도된 바대로 작동하는 것을 검증하기 위해 중요하다고 주장되지만, 회사는 이것이 영업비밀 및 지식재산을 잃게 하는 위험을 초래하여 혁신을 저해한다고 주장하고 법원이 이를 인정하곤 한다. 이 문제에 대해 "조건부 투명성" 조치를 통해 균형을 잡고자 하는 시도가 있는데, 이것은 해당 알고리즘의 리스크 레벨 및 알고리즘이 사용된 맥락을 고려하여 그에 비례한 적당한 정보를 제공하는 투명성을 요구하는 것이다. 투명성과 관련하여 제안되는 정책은 다음과 같다.

가. 제3자 공개(Disclosures to third parties)

공인된 제3의 감사기관에 알고리즘 시스템과 관련한 기술 정보를 정기적으로 제출하도록 규정하자는 제안이다. 감사기관이 소스코드 및 트레이닝데이터를 검사하여 편향성 및 차별적 결과가 나올 가능성 판단하여 일정 기준을 만족하면 알고리즘 인증한다. 이 방법을 통해 영업비밀을 보호하며 제한적인 투명성을 보장할 수 있다.

나. 질적 공시(Qualitative public disclosures)

SEC 공시와 유사하게 위험하거나 영향력 큰 알고리즘에 대해 정보를 대중에 공개하도록 하는 것이다. 알고리즘 작동, 유효성, 가능한 오류 등에 대한 정보 정도를 공개하고, 다만 트레이닝 데이터, 코드 등 지식재산 관련 정보는 제외한다.

다. 기술 상세사항의 공시(Public disclosure of technical details)

개인에 중대한 위험을 끼치는 알고리즘의 경우에는 영업비밀에 해당하는 정보까지 포함하여 상당한 기술적 정보를 공개해야 한다는 정책 제안이다.

라. 온라인 플랫폼종사자를 위한 고용정보 공개(Disclose employment data for online labor platform workers)

우버나 TaskRabbit과 같은 온라인 노무제공 플랫폼은 근로자에 대한 방대한 정보를 수집하여 노동력을 조직, 관리하고 있다. 교대 시간, 업무 기록, 업무 일자, 임금, 시간, 위치

등 다양한 정보가 수집되는데, 이를 Internal Revenue Service나 Department of Labor와 같은 규제 당국에 공개해야 한다는 것이다. 규제 당국은 노동 관련 위반을 적발하기 위하여 이러한 정보를 이용하고, 근로자 단체 등은 이들 정보에 접근 분석할 수 있어야 한다.

많은 근로자가 독립계약자로 오분류되어 임금이나 근로시간, 단체교섭(collective bargaining), 차별금지법과 같은 적용에서 제외될 수 있다는 문제가 존재하므로, 이들에 대해서는 별도로 보호받을 수 있도록 규제를 마련하고, Lyft나 Uber와 같은 플랫폼 회사에게 정보공유 요구를 하는 제도를 도입하자는 것이다.

마. 분쟁 시 영업비밀보호에 우선(Preempting trade secrets protections in litigation)

회사가 개인의 법적으로 보호받는 권리, 재정 상태나 고용 기회 등에 중대한 영향을 끼치는 결정을 하는 알고리즘을 이용할 때, 이에 관해 법적 분쟁에서 알고리즘 작동에 대한 정보를 영업비밀 보호를 이유로 공개하지 않는 것은 큰 문제가 된다. 알고리즘이 어떻게 작동하는지에 대한 정보, 고려한 요소들을 모르고는 근로자가 자신의 권리가 알고리즘에 의해 침해되었는지 밝히고 구제를 받는 것은 거의 불가능하므로, 보호 명령 하에서 자신의 클레임과 관련한 중요한 알고리즘 증거를 획득하도록 허락하는 제도가 형사소송 및 근로자 보호 규제 위반과 관련한 상황에서 도입되어야 한다.

Ⅳ. 선도적인 입법 조치와 그 한계[63]

1. 개관

인공지능이 다양한 분야에서 활용되면서 그 영향력이 커지고 있지만, 대량의 데이터를 학습하여 성능을 향상시키는 기계학습에 기반한 인공지능은 그에 수반하는 불확실성과 불투명성을 갖고 있으며, 노이즈 데이터로 오류를 일으킬 가능성도 상존한다.[64] 즉, 인공지능의 활용은 편익과 기술위험(technological risk)의 양면성을 갖는데, 인공지능에 따르는 위험으로는 판단오류, 알고리즘의 편향성, 비도덕적 판단, 예측 불가능한 오작동, 제어 불가능성 등이 제시되고 있다.[65] 따라서 인공지능 알고리즘의 파급력을 생각하면 문제가 발생한

63) 이 부분은 Aloisi, A. and Stefano, V. D. (2022) Your Boss Is an Algorithm. 1st edn. Bloomsbury Publishing.의 내용을 주로 참고하였다.

64) 권오성, 앞의 글, 7쪽.

65) 최은창은 AI 기술위험을 작동 시 위험, 보안 위험, 통제 관련 위험, 윤리적 위험, 사회경제적 위험으로 분류한다. 자동화된 판단의 오류, 편향, 불투명성, 설명 불가능성은 작동 시 위험에 해당하며, AI를 이용하여 취약점을 공격하는 사이버 공격, 개인의 데이터 프라이버시 침해 등은 보안 위험이다. 통제 관련한 위험은 갑작스런 오작동, 인간의 통제권 상실, 자동화된 살상 무기 등이다. 상식, 평등, 약자에 대한

이후에 법적 규율을 도모하는 것은 규율의 지연이 아니라 규율 불가로 귀결될 가능성이 크다.[66]

이러한 배경에서, 각국 정부는 이미 인공지능의 영향에 대한 보다 효과적인 규제를 위해 움직이고 있다. 외국의 경우 일반적인 인공지능 윤리 정립 및 규제와 관련하여 연구와 논의를 진행하고 있으며, 특히 EU가 법제화에 적극적이다. 먼저, 인공지능에 의한 데이터 분석에의 법적 대응의 선진적인 대처로서 EU일반 데이터 보호 규칙(General Data Protection Regulation 이하, "GDPR"이라고 한다)이 있다. GDPR은 빅데이터 분석, 인공지능 및 기계학습에 의한 프로파일링이나 자동처리가 개인의 권리나 자유에 중대한 영향을 미칠 우려가 있다고 하여 개인의 권리나 자유, 특히 프라이버시를 보호하기 위해 프로파일링 등의 개인 데이터의 자동처리에 대해 새로운 규제를 가하고 있다. 이후 EU는 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 ③ 프라이버시 및 데이터 거버넌스 ④ 투명성 ⑤ 다양성, 비차별 및 공정성 ⑥ 사회적·환경적 복지 ⑦ 책임성이다. 이후 EU는 2020년 3월 「인공지능 발전과 신뢰를 위한 백서」를 발표하여 위험 발생 가능성이 높은 분야의 인공지능에 대하여 향후 안전성 요건을 수립하고 사전 적합성 평가를 받도록 하는 방안을 제시하였으며, 2021. 4. 21. "Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)을 발표하였다.

한편, 스페인의 경우 기업이 '플랫폼' 근로자를 관리하는 데 사용하는 알고리즘에 대한 정보를 공개하도록 하는 법(라이더 법으로 불리는 근로자헌장법의 개정법률)이 2021년에 제정되었다. 또한, 스페인에서는 2022년에는 GDPR과 라이더 법의 내용을 포함한 일터에서의 알고리즘에 관한 가이드라인을 제정하였다. 미국의 경우, 2019년 일리노이주에서 2019년에 제정된 법으로, 사용자는 지원자에게 화상 면접에 있어서 AI 시스템이 사용될 것임을 고지할 것을 요구한다. 연방 차원에서는 알고리즘 책임성 법률(Algorithm Accountability Act)이 의회에 수차례 제출되었으나, 아직 입법에 이르지는 못하였다. 그 외에도, 뉴저지주에서도 2022년에 New Jersey Assembly Bill 4909이 제출되었는바, 동 법안은 최근 1년 이내에 감사를 받지 않은 자동화된 의사 결정 도구를 사용하거나 판매하는 것을 불법으로 규정하는 내용이다.[67] 또한, 뉴욕주에도 New York State Assembly Bill A00567이 제출되었는바, 동 법안 또한 2023년 1월 뉴욕주(州)에서 사용되는 자동화된 의사결정 도구에

배려 등 인간 사회에 통용되는 가치를 코딩화하여 알고리즘 설계에 반영하기 어려운 점, AI와 인간 간의 가치 정렬(value alignment)에 발생하는 간극은 윤리적 위험으로 여겨진다. 실업 발생, 인종과 성별에 따른 차별, 자동화된 판단으로 피해가 발생해도 원인을 찾아내기 어려운 문제 등은 사회경제적 위험으로 분류하고 있다(최은창, 앞의 글, 146쪽).

66) 양종모, "인공지능에 대한 법학의 위험한 해법", 「법학에서 위험한 생각들」, 법문사, 2018, 418쪽.

67) https://www.njleg.state.nj.us/bill-search/2022/A4909

대해 매년 영향평가를 받도록 의무화하는 내용이다.[68]

한편, 스페인은 2022년 10월에 인공지능에 관한 시험적인 규제 샌드박스(regulatory sandbox)를 시작했다.[69] 이러한 규제 샌드박스[70]는 관할 당국이 인공지능을 개발하는 기업과 긴밀히 협력하여 향후 유럽위원회의 Artificial Intelligence Act[71]의 시행을 안내할 모범사례를 정의하는 것을 목표로 한다. 스페인 정부가 시작한 이러한 시험적인 규제 샌드박스는 향후 EU AI Act의 요건과 적합성 평가 또는 배포 후 활동과 같은 기타 기능의 운영을 검토할 것입니다. 한편, 미국은 소비자 보호, 고용, 교육, 주택 및 금융, 의료 분야의 기회 형평성 등 다양한 분야를 포괄하는 AI에 대한 '권리장전'의 청사진(Blueprint for an AI Bill of Rights)[72]을 발표했다.

2. GDPR

2018년 EU 일반 데이터 보호 규정(EU General Data Protection Regulation, GDPR)이 시행되기 전, 데이터 보호 분야의 국제적 리더로서 EU의 오랜 입지를 확인한 유럽 데이터 보호 위원회(European Data Protection Board)는 지원자의 개인 프로필에서 데이터를 수집하는 것은 해당 직책의 채용에 필요하고 관련성이 있는 것으로 제한되어야 하며, 지원자에게 정식으로 통지하고 절차가 완료된 후 결과가 어떻든 데이터를 파기해야 한다고 명시했다.[73] EU 데이터 보호 규칙에 대한 공통된 해석을 촉진하기 위해 일반 지침을 발표하는 것을 임무로 하는 이 집행 기관(enforcement agency)은 마우스나 키보드의 움직임이나 화면상의 활동의 추적(tracking) 같은 행위는 프라이버시 및 인간의 존엄성에 대한 국내법의 전통과 상충되는 불균형적인 감시 방법에 해당하므로 EU에서 금지된다고 명시했다.[74]

---

68) https://nyassembly.gov/leg/?default_fld=&leg_video=&bn=A00567&term=2023&Summary=Y&Actions=Y&Text=Y

69) https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented

70) 샌드박스는 혁신기업(innovators)과 규제기관을 연결하고, 이들이 협력할 수 있는 '통제된 환경'을 제공하는 방법이다.

71) https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

72) https://www.whitehouse.gov/ostp/ai-bill-of-rights/

73) Independent EU Advisory Body on Data Protection and Privacy (Art. 29 Working Party, 'WP29'), Opinion 2/2017 on data protection at work, adopted in June 2017 and aimed at complementing the Opinion 08/2001 on the processing of personal data in the employment context. Independent EU Advisory Body on Data Protection and Privacy (Article 29 Working Party, 'WP29'), Working document on the surveillance of electronic communications in the workplace, U.N. Doc. 5401/01/EN/Final; Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01).

74) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection

GDPR은 AI를 활용한 근로자 채용 및 관리 관행에 장벽을 만듦으로써 “자동화된 개별적 의사 결정”과 “프로파일링”으로부터 사람들을 보호한다.[75] 특히 제22조는 강력한 안전장치가 제공되지 않는 한 “법적 효과(legal effects)를 발생시키는 프로파일링을 포함한 자동화된 처리만을 기반으로” 하는 결정 또는 데이터 주체에 “이와 유사하게 중대한 영향을 미치는” 결정의 실행을 금지한다.[76] GDPR 해설서(Recitals of the GDPR) 71항은 GDPR 제22조을 해설하면서 “사람의 개입이 없는 전자 채용 관행(e-recruiting practices)”을 명시적으로 언급하고 있다. 여전히 정보 주체를 쉽게 식별할 수 있는 가명화된 데이터(pseudonymized data)도 이 조항의 적용 범위에 포함된다. 비록 GDPR은 전적으로 자동화된 모델만을 언급하고 있지만, 법원과 다른 해석자들은 인간의 개입이 사소하고 명목상일 뿐이어서 자동화된 결정에 전적으로 종속하는 시스템이 제22조의 범위에서 제외할 수 있는가를 판단할 것이다.[77] 사람의 역할이 다른 선택의 여지 없이 고무인을 찍어주는 것으로 제한된 조치로는 제22조에 규정된 일반적 금지의 적용을 배제할 수 없다고 해석돼야 하고, 이를 통해 데이터 처리에서 “인간의 지휘(human in command)”라는 접근 방식이 채택되도록 유도하여야 한다.[78]

자세히 살펴보면, 자동화된 의사 결정 프로세스의 금지에는 상당한 예외가 존재한다. 프로파일링 또는 알고리즘 관리는 “(a) 정보 주체와 정보처리자 간의 계약 체결 또는 이행을 위해 필요한 경우”, “(b) 정보 주체의 권리, 자유 및 정당한 이익을 보호하기 위한 적절한 조치를 명시한 EU 또는 회원국의 국내법에 의해 승인된 경우”, “(c) 정보 주체의 명시적 동의에 근거한 경우”에는 여전히 허용된다.[79] 자동화된 개별적 의사 결정 프로세스의 금지

Regulation).

75) Council of Europe, The protection of individuals with regard to automatic processing of personal data in the context of profiling Recommendation CM/Rec(2010)13 adopted by the Committee of Ministers of the Council of Europe on 23 November 2010 and explanatory memorandum.

76) Article 4(4) defines ‘profiling’ – a relatively novel concept in European data protection law – as ‘any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements’. A concrete example of this practice would be e-recruiting (Recital 71). See M Hildebrandt, ‘Defining Profiling: A New Type of Knowledge?’ in M Hildebrandt and S Gutwirth (eds), Profiling the European Citizen (Cham, Springer, 2008).

77) ME Kaminski and G Malgieri, ‘Algorithmic Impact Assessments under the GDPR: Producing Multi-layered Explanations’ (2021) 11(2) International Data Privacy Law 125–44.

78) M Veale and L Edwards, ‘Clarity, surprises, and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling’ (2018) 34(2) Computer Law & Security Review 398–404.

79) D Kamarinou, C Millard and J Singh, ‘Machine Learning with Personal Data’ in R Leenes et al (eds), Data Protection and Privacy: The Age of Intelligent Machines (Oxford, Hart Publishing, 2020) 89–114.

는 고용의 맥락에서 곧바로 적용되지는 않는다. 대량의 지원서에 대한 전자적 심사와 같은 인사 부서의 의사 결정은 "계약 체결 또는 이행을 위해 필요한" 경우에는 제22조 제2항에서 허용하는 예외의 범위에 속할 수 있다. 그러나 사용자는 "(노동자의) 권리와 자유 및 정당한 이익을 보호하기 위해 적절한 조치를 취해야 하며, 최소한 관리자의 인간에 의한 개입(human intervention)을 얻고, 자신의 관점을 표현하고, 결정에 대해 이의를 제기할 수 있는 권리"와 나아가 "도출된 결정에 대한 설명을 들을 권리"를 보장해야 한다.[80] 명시적인 동의가 있는 경우에만 일반적인 금지 규정에서 벗어날 가능성이 있다. 그러나, 유럽 데이터 보호 위원회(European Data Protection Board, EDPB)에 따르면 노동자가 "동의를 자유롭게 제공, 거부 또는 철회할 수 있는 위치에 있는 경우가 드물기 때문에" 교섭력의 격차로 인하여 "(업무상 데이터 처리에 대한) 법적 근거가 동의가 될 수는 없고, 동의가 되어서도 안 된다."라고 한다.[81]

또한, GDPR은 당사자가 동의하거나 공익상의 이유로 처리가 필요한 경우를 제외하고는 건강 상태, 성적 지향, 정치적, 이념적, 노동조합에 대한 견해(trade union opinions) 또는 민족적 출신과 같은 민감한 데이터를 기반으로 알고리즘에 따른 결정을 내릴 수 없다는 거래할 수 없는(non-negotiable) 한계를 설정했다. 제22조의 거창한 목적이 서류상으로만 남는 것을 방지하기 위해,[82] 지역 수준(local level)에서 입법과 단체교섭은 자동화된 의사 결정에 대한 가장 설득력 있는 대응책이며, 투명한 기준을 준수하고 근로자에게 영향을 미치는 모든 결정에 대하여 인간 에이전트(human agents)가 최종 통제권과 책임을 보유하도록 보장하고, 집단적 감독 모델을 확립한다.[83] GDPR은 회원국이 법률 또는 단체협약에 따라 "고용(employment)의 맥락에서, 특히 채용(recruitment), 고용 계약의 이행, […] 업무의 관리, 계획 및 조직, 직장 내 평등 및 다양성, 직장 내 보건 및 안전, […] 고용 관계 종료의 목적과 관련한 근로자의 개인 데이터 처리와 관련하여 권리와 자유의 보호를 보장하는 구체적인 규칙"을 도입할 수 있다고 명시하여, 위와 같은 해결 방안을 뒷받침한다. 이러

80) AD Selbst and J Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) International Data Privacy Law 233-42. Interestingly, the draft directive on platform work seems to respond to these uncertainties in a purposive and systematic way, by providing for the right to obtain an explanation 'for any decision taken or supported by an automated decision-making system that significantly affects the platform worker's working conditions' (Article 8).

81) EDPB, Guidelines 05/2020 on consent under Regulation 2016/679.

82) See also Article 9 of the revised Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data concerning the right not to be subject to automated decision-making without human intervention.

83) A model of 'municipal sovereignty' is that of the Coalition of Cities for Digital Rights promoted by Amsterdam, Barcelona and New York City with the aim of protecting, promoting and monitoring the personal data of citizens and visitors by 'opening up' the shared management of information generated by public services. L Dencik, 'Towards Data Justice Unionism? A Labour Perspective on AI Governance' in P Verdegem (ed), AI for Everyone? Critical Perspectives (London, University of Westminster Press, 2021) 267-84.

한 규칙은 “정보 주체의 인간 존엄성, 정당한 이익 및 기본권을 보호하기 위한 적절하고 구체적인 조치를 포함”해야 한다(GDPR 제88조).[84]

GDPR의 프레임워크의 가장 중요한 점은 이것이 외부의 영향 없이 작동하는 것이 아니라, 매우 복잡하고 다양한 소스로 구성된 규제 구조(regulatory architecture)의 기반이 된다는 점이다. 회원국은 근로자의 모니터링 및 업무상 데이터 처리와 관련하여 구체적인 내부 조치를 도입할 수 있다. 또한, 근로자의 활동을 모니터링할 때 그들의 프라이버시가 침해되지 않도록 하기 위해서는 데이터 보호 당국의 역할도 매우 중요하다. 개인의 동의만으로는 충분하지 않으며, 이러한 권리는 포기할 수 없다. 확고한 법적 전통에 따라 많은 EU 관할권에서 근로자 대표와의 사전 협의 단계 또는 근로자 대표의 승인은 채용 과정에서 사용되는 감시 장비를 포함한 감시 장비의 도입을 위한 필수적인 전제 요건이다. 이러한 절차적 규칙은 개인정보 수집 및 처리에 대한 합법적인 근거가 되며, 이러한 활동의 배후에 있는 회사의 정당한 이익에 대한 증거가 뒷받침되어야 한다. 국내법과 판례는 근로자 대표가 정보, 토론 및 공동 결정을 통해 참여하도록 보장한다. 이러한 요건을 준수하지 않을 경우, 불법적으로 수집된 데이터와 정보를 사용할 수 없으며 제재를 받을 수도 있다.

그러나, 특정한 기술의 전체적인 범위를 과소평가하는 논평자가 너무나 많다. 의심할 여지가 없는 보호 가치에도 불구하고, EU의 GDPR은 개인 데이터와 관련하여 AI 및 머신러닝 기술이 제기하는 모든 새로운 도전에 효과적으로 대처하지 못하고, 빠르게 구식이 될 위험이 있다. 옥스퍼드 인터넷 연구소(Oxford Internet Institute)의 변호사 산드라 왓쳐(Sandra Wachter)는 이 EU 규정이 데이터의 수집에만 초점을 맞추고 데이터 처리에 대해서는 소홀히 하고 있다고 비판한다. 그녀의 의견에 따르면, 일단 합법적으로 데이터를 수집하기만 하면, 추론에 이한 분석(inferential analysis), 즉 알고리즘이 대량의 데이터에서 반복되는 패턴을 추출하여 이를 실시간 예측으로 개발하는 과정에 대해서는 아무런 제한이 없다. 이 모든 것은 여전히 위험지역(no-man’s-land)으로 남게 될 것이다.[85] 하지만, 이에 반대하는 좋은 논거가 있다. 구체적인 사실(예컨대, 어떤 사람이 정크푸드를 많이 구매한다

84) In June 2020, the European social partners signed a landmark framework agreement. While acknowledging the significant contribution in terms of security, health and safety and efficiency, the agreement stresses the risk of deterioration of working conditions and well-being of workers and calls for ‘data minimisation and transparency along with clear rules on the processing of personal data limits the risk of intrusive monitoring and misuse of personal data’. Interestingly, it advocates for worker representative’s involvement to address issues related to consent, privacy protection and surveillance. Available at www.etuc.org/en/document/eu-social-partners-agreement-digitalisation.

85) S Wachter, B Mittelstadt and C Russell, ‘Counterfactual explanations without opening the black box: Automated decisions and the GDPR’ (2018) 31(2) Harvard Journal of Law & Technology 841-87. See also S Wachter, Sandra, B Mittelstadt and L Floridi, ‘Why a right to explanation of automated decision-making does not exist in the general data protection regulation’ (2017) 7(2) International Data Privacy Law 76-99.

는 사실)에 기반한 데이터 수집을 제한하면서 알고리즘이 이 데이터로부터 도출하는 추론(예컨대, 같은 개인의 특정 건강 상태 예측)을 제한하지 않는 것은 불합리하다는 것이다. EDPB는 추론도 GDPR의 적용을 받는다고 주장한다.

이론적 관점에서 보면, 또 다른 근본적인 딜레마가 존재한다. 알고리즘은 "현명(wise)"하라고 프로그래밍되어 있지 않다. 오히려 어떤 업무를 지시받으면, 가능한 한 가장 효과적인 방법으로 이를 수행하기 때문에, 인간에게 특유의 중요한 유연성(critical flexibility)을 손상한다. 그렇다면, 이러한 점에서 가치(values)와 원칙(principles)을 인정하지 않는 시스템과 어떻게 상호작용할 수 있을까? AI와 디지털 기술의 윤리에 대한 일반적인 논의에 갇히지 말고, 인권 보호의 수단에 의존해야 한다. 유럽에서는 사생활과 가족생활, 가정 및 통신에 대한 권리를 보호하는 유럽인권협약(ECHR) 제8조가 대표적인 예이다.

스트라스부르에 위치한 유럽 인권재판소(The Court of Human Rights)는 이 조항을 근로자에게 과도하고 불균형적인 원격 모니터링을 금지하는 것으로 해석했다. 2018년 유럽평의회(Council of Europe)는 1981년의 "개인정보의 자동 처리에 관한 개인 보호 협약(Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data)"을 현대화하기 위한 의정서(Protocol)도 채택했다. 개정된 협약은 모든 개인이 "자신의 의견을 고려하지 않고 오로지 자동화된 데이터 처리에 근거하여 자신에게 중대한 영향을 미치는 결정의 대상이 되지 않을 권리"를 명시하고 있습니다(제9조). 이러한 보호는 GDPR이 규정하고 있는 보호와 크게 다르지 않지만, 그러나 그 효력은 비교할 수 없으며, 또한 그 시장 지향(market-oriented)적 목표에 비추어 볼 때도 그러하다. 유럽인권재판소의 판례에 따라 기업은 알고리즘 채용 및 프로파일링 시스템이 차별적이지 않고 다른 유럽 규범과 원칙을 준수한다는 것을 입증해야 한다.

증명책임의 전환을 요건은 명백히 더 완화되어야 한다. 이러한 기대는 마치 Nutella의 비밀 재료처럼 영업 비밀과 지적 재산을 방패막이로 삼아 프로세스를 설명가능(explainable)하고 (설명)책임성 있게(accountable) 만들기를 거부하는 많은 "알고리즘 남작(algorithm barons)"에 대항하는 역할을 하지만, 초콜릿 스프레드는 우리의 삶(기껏해야 허리둘레)에 대한 결정을 내리는 것은 아니라는 점에서 큰 차이가 있다. 통계적 의사 결정 시스템의 기능을 공개하지 않으려는 시도는 기각되어야 한다.[86] 무엇보다도, 이러한 거부는 다른 모든 사람의 프라이버시를 침해하면서 자신의 기밀은 보호하라고 요구하는 매우 모순적인 행위이다. 부당하고 책임감 없는 기술로 인한 잠재적 피해는 사용 가능한 모든 법적 수단을 동원하여 억제해야 한다. 특히 정보 및 접근 권한에 관한 GDPR 제13조부터 제15

86) G Buttarelli, Privacy 2030, available at https://iapp.org/media/pdf/resource_center/giovanni_manifesto.pdf.

조까지 규정에 비추어 볼 때, 데이터 주체가 특정 행위의 결과를 이해할 수 있도록 운영 논리를 설명할 수 있는 능력의 중요성은 아무리 강조해도 지나치지 않다.[87] 이러한 정보의 집합은 차별에 관한 일응의 증명(prima facie)을 하는 차별 관련 소송을 제기하는 데 핵심이 될 수 있으므로, 고용 주체가 그 의미에 대한 기술적 이해 없이 제3자가 제공하는 소프트웨어에 의존하지 못하도록 억제할 수 있다. 자동화된 의사 결정 프로세스에 사용된 논리를 검증하고 그 기준과 결과를 평가할 수 있는 능력은 특히 차별 금지 규정의 적용에 있어 매우 중요하다. 예를 들어, EU 판례에 따르면 사용자는 근로자에 대한 고객의 편향된 행동에 대해서는 물론,[88] 대용물 차별(proxy discrimination)에 대해서도 책임을 지게 될 수 있다.[89] 그러므로, 자동 의사 결정의 사슬을 출처까지 추적하는 것이 중요하다.

3. 스페인 노동부의 "직장에서의 알고리즘 정보" 가이드라인(2022. 5)

라이더법(Ley Riders) 시행 이후 스페인 노동부는 2022년 5월 "직장에서의 알고리즘 정보(Información algorítmica en el ámbito laboral)" 가이드라인을 발표하였는다. 이 가이드라인 자체의 규범력은 없지만, 동 가이드라인은 스페인의 근로자헌장법과 EU의 GDPR의 내용을 구체화한 것이므로 이러한 범위에서는 사실상 구속력을 가질 것으로 예상된다.

이 가이드라인에는 현재 업무 관계에서 알고리즘이나 자동화된 의사결정 시스템의 사용에 관하여 GDPR §22에 따른 개별적 차원과 근로자헌장법 §64에 의해 규제되는 집단적 차원의 권리의 양자를 규정하고 있다.

4. 미국

1) 뉴욕시(New York City)

뉴욕시 의회는 채용 시 인공지능 사용을 규제하기 위해 사용자 및 직업소개기관이 자동

87) T Bucher, If ··· then: Algorithmic power and politics (Oxford, Oxford University Press, 2018).
88) Judgment of 14 March 2017, Case C-188/15 Asma Bougnaoui et Association de défense des droits de l'homme (ADDH) contre Micropole SA [2018] ICR 139. See R Ducato, M Kullmann and M Rocca. 'European Legal Perspectives on Customer Ratings and Discrimination' in T Addabbo, E Ales, Y Curzi, T Fabbri, O Rymkevich and I Senatori (eds), Performance Appraisal in Modern Employment Relations (Palgrave Macmillan, Cham, 2020) 225–51.
89) R Xenidis and L Senden, 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination' in U Bernitz et al (eds), General Principles of EU law and the EU Digital Order (Alphen aan den Rijn, Kluwer Law International, 2020) 151–82; L Grozdanovski, 'In search of effectiveness and fairness in proving algorithmic discrimination in EU law' (2021) 58(1) Common Market Law Review 99–136.

화된 고용상 결정 도구(automated employment decision tools, AEDTs)를 사용하여 채용 후보자 또는 승진 후보자를 평가하기 전에 독립적이고 공정한 편향성 감사를 의뢰하도록 하는 법안(Local Law 144)을 의결했다. 편향성 감사가 적용되는 도구의 배포 일자 및 편향성 감사의 결과 요약도 공개적으로 제공해야 한다.

편향성 감사 이외에도 Local Law 144에 따라 사용자는 채용 지원자 또는 근로자에게 자동화된 도구가 자신을 평가하는 데 사용될 것이라는 사실 및 평가 과정에서 사용될 특성, 그리고 대안적인 선발 절차 또는 편의 제공을 요청할 수 있다는 사실을 도구가 사용되기 최소 영업일 10일 전에 통지하여야 한다. 수집된 데이터의 유형, 데이터의 출처 및 회사의 데이터 보존 정책에 대한 정보가 사용자 또는 직업소개기관의 웹사이트에 공개되어 있지 않은 경우, 서면 요청 시 사용자 또는 직업소개기관은 이러한 정보를 제공해야 한다.[90]

※ 참고. 미국의 주 및 지방의 입법권

1. 미국의 연방제

현재 미연방은 50개 주와 Washing D.C.로 구성되어 있으며, 연방정부와 주(州) 정부, 지방(Local) 정부(카운티, 일반 시, 타운 및 타운십, 학교구, 특별구) 구조로 이루어진 정부 형태를 취하고 있다.

정부, 입법부, 사법부로 구성된 연방정부는 권력분립의 원칙에 따라 상호 경제와 균형을 원칙으로 한다. 한편, 연방정부는 주 정부의 사무에 관여할 수 없고, 주 정부 아래 단위인 '지방정부'의 사무에도 관여할 수 없지만, 연방 정책의 수행을 위한 상호 협조 등을 위해서 연방과 주, 지방 상호 간의 정부 간 관계 담당 부서가 있다.

연방정부와 주 정부의 권한은 연방헌법으로부터 부여받고 있다. 연방 수정헌법은 최고성(Supremacy) 원칙에 따라 모든 주와 지방정부의 법률 및 정책 등이 연방헌법과 법률에 우선적으로 구속되고, 이를 위반할 경우 사법(司法) 통제의 대상이 된다. 다만 이러한 연방헌법 및 법률의 우선성은 연방 수정헌법으로부터 수권(授權) 받은 범위 내에서만 인정된다.

연방 수정헌법 제1조 제8절에 명시된 '연방정부 권한'으로는 외교, 국방 관련 권한, 공동 방위와 일반 복지를 위한 조세권, 외국과의 통상 규제, 관세, 화폐 도량형 설정, 주간(inter-state) 통상 규제권, 증권 및 화폐 위조 처벌권, 하급법원 설치, 전쟁 선언 등이 포함된다. 한편 연방수정헌법 제1조 8절 제18항 'Necessary and Proper Clause'에 따라 연방정부는 헌법으로부터 부여받은 권한을 행사하기 위해 필요하고 적절한 법률을 제정할 수 있는 권한을 주고 있다.

2. 연방정부의 주(State, 州)정부의 관계

연방 수정헌법에 의해 연방에 위임되지 않았거나, 각 주의 권한으로 금지되지 않은 권

90) https://portal.311.nyc.gov/article/?kanumber=KA-03552

한은 각 주가 보유한다. 다만, 연방수정헌법의 최고성의 원칙에 따라 연방헌법 혹은 법률은 주와 지방정부의 법률 및 정책에 우선한다. 연방 선점(Federal preemption)의 원칙에 따라 연방의회가 연방법으로 미리 규제하고 있는 경우, 연방법에 반하는 주법은 위헌이다.

3. 연방정부와 지방정부 관계

연방제를 취하고 있는 미국에서 지방자치는 주에 대한 지방의 자치권을 의미한다. 연방수정헌법에는 지방자치나 지방정부에 대한 규정이 없으며 기타 연방 법률도 지방자치에 대하여 언급하고 있지 않다. 따라서, 지방정부는 헌법적 지위를 갖고 있지는 못하다. 따라서 이론적으로는 연방정부와 지방정부 간의 관계에 대한 헌법적 근거는 없다.

4. 주정부와 지방정부 관계

미국의 지방계층구조는 주(州)마다 상이하지만, 대체로 카운티(county), 시(municipalities), 타운 및 타운쉽(town or township), 교육구(school districts) 및 특별구(special districts)로 나눌 수 있다. 연방주의의 특성상 지방정부는 주 정부를 매개하여 자치가 가능하며, 주에 따라 다르지만 대부분의 주에서 주 헌법에 헌장을 통한 지방정부의 법적 지위를 부여하고 있다. 이러한 면에서 미국의 주정부와 지방정부의 관계는 단방제(unitary system)에서의 중앙정부와 지방정부 사이의 관계와 유사하다. 요컨대, 지방정부와의 관계에서 주 정부는 모든 법적 권한을 가지고 있으며, 지방정부의 권한은 주 헌법 또는 법에 명기된 것에 한한다. 이러한 면에서 지방정부는 전적으로 주 정부의 창조물이라고 할 수 있다.

미국의 지방정부는 County, Parish, City, Borough, Town, Township, Village, School Districts, Special Districts 등 다양하게 불리고 있지만, 연방정부에서 발행하는 인구센서스에서는 카운티(Counties), 일반시(Municipalities), 타운(Towns and Townships), 교육구(School Districts), 그리고 특별구(Special Districts)의 다섯 가지 유형으로 분류하고 있다. 이 중 카운티, 일반시, 타운 및 타운십은 일반 목적의 지방정부(general-purpose governments)로 분류되고, 학교구와 특별구는 특별 목적의 지방정부(special-purpose governments)로 분류된다.

이 중 일반 시(municipality)는 시(city)를 의미한다. 역사적으로 시는 지방정부의 기본 단위로, 카운티와는 달리 처음부터 주 정부의 행정보조기관으로 설립된 것이 아니라 주민자치를 위하여 설립되었다. 일반 시는 카운티와 마찬가지로 일반 목적의 지방정부 단위이지만, 카운티보다 의사결정 권한과 재량권을 더 많이 가지고 있다. 일반적으로 시는 인구가 밀집한 지역을 중심으로 주 정부가 부여한 자치헌장(home rule charter)에 따라 운영된다. 거의 모든 시정부는 시장-위원회 형(mayor-council form), 위원회-관리자 형(council-manager form), 시위원회 형(city commision form)라는 세 가지 구조 중에서 하나로 운영되며, 각각의 구조에서 전형적으로 시 위원회(city council)로 불리는 선출된 통치체가 의사결정 권한을 가진다.

※ 참고. 뉴욕시 행정법전(Administrative Code of the City of New York)과 뉴욕시 규

정(Rules of the City of New York)의 관계

1. 뉴욕시 행정법전(Administrative Code of the City of New York) - 뉴욕시 행정법전은 뉴욕시 의회의 의원 또는 시장에 의해 제안된 법안에 대하여 뉴욕시 의회가 법안에 대한 검토를 거쳐 투표로 법안을 승인하면, 뉴욕 시장이 해당 법안에 서명하여 공포한다. 법률로 승인된 법안의 유효일자가 결정되며, 일반적으로 유효일자는 일정한 기간 후로 정해진다. 행정법전(Code)은 주로 범죄, 건축, 주택, 건강, 교통 등 다양한 분야에 대한 법규를 포함하고 있다. **2. 뉴욕시 규정(Rules of the City of New York)** - 뉴욕시 규정은 뉴욕시의 행정규칙을 규정하는 법률문서이다. 이 규정은 뉴욕시 행정법전에 근거하여 제정되며, 뉴욕시의 행정부 기관 및 기타 단체에서 구체적인 규칙과 절차를 규정한다. **3. 양자의 관계** - Administrative Code of the City of New York은 뉴욕시 의회의 의결을 거쳐 성립하는 뉴욕시의 법률(우리나라의 조례와 유사)에 해당하고, Rules of the City of New York은 the Code의 세부적인 집행을 위하여 뉴욕시 행정기관이 제정한 규정(우리나라의 규칙과 유사)한 법률문서이다.

가. 뉴욕시 행정법전(Administrative Code of the City of New York)

Subchapter 25: Automated Employment Decision Tools	
§ 20-870 Definitions. For the purposes of this subchapter, the following terms have the following meanings: **Automated employment decision tool.** The term "automated employment decision tool" means any computational process, derived from machine learning, statistical modeling, data analytics, or artificial intelligence, that issues simplified output, including a score, classification, or recommendation, that is used to substantially assist or replace discretionary decision making for making employment decisions that	§ 20-870 **정의.** 이 절(subchapter)에서 사용되는 아래 용어의 의미는 다음과 같다: **자동화된 고용상 결정 도구.** “자동화된 고용상 결정 도구”라는 용어는, 기계 학습, 통계적 모델링, 데이터 분석 또는 인공지능에서 파생되어(derived from), 자연인에게 영향을 미치는 고용상 결정에 관한 재량적인(discretionary) 의사 결정을 실질적(substantially)으로 보조하거나 대체하는 데 사용되는 점수, 분류 또는 추천을 포함한 단순화된 결과값(simplified output)을 발행하는 모든 계산 프로세스(computational process)를 의미한다. 재량적 의사 결정 프로세스를 자동화하거나, 지원하거나, 실질적

impact natural persons. The term "automated employment decision tool" does not include a tool that does not automate, support, substantially assist or replace discretionary decision-making processes and that does not materially impact natural persons, including, but not limited to, a junk email filter, firewall, antivirus software, calculator, spreadsheet, database, data set, or other compilation of data.	으로 보조하거나, 대체하지 않고, 자연인에게 중대한 영향을 미치지 않는, 정크 이메일 필터, 방화벽, 바이러스 백신 소프트웨어, 계산기, 스프레드시트, 데이터베이스, 데이터 세트 또는 기타 데이터 편집(compilation)을 포함하되 이에 제한되지 않는 도구는 “자동화된 고용상 결정 도구”라는 용어에 포함되지 않는다.
Bias audit. The term "bias audit" means an impartial evaluation by an independent auditor. Such bias audit shall include but not be limited to the testing of an automated employment decision tool to assess the tool's disparate impact on persons of any component 1 category required to be reported by employers pursuant to subsection (c) of section 2000e-8 of title 42 of the United States code as specified in part 1602.7 of title 29 of the code of federal regulations.	**편향성 감사**. “편향성 감사”라는 용어는 독립적인 감사인(independent auditor)에 의한 공정한 평가를 의미한다. 이러한 편향성 감사에는 29 CFR § 1602.7에 기재된 42 USC § 2000e-8 (c)에 따라 사용자가 보고해야 하는 ‘component 1’[91] 범주에 해당하는 사람에게 대한 해당 도구의 차별적 영향(disparate impact)을 평가하기 위한 자동화된 고용 결정 도구에 대한 테스트가 포함되나, 이에 제한되지 않는다.
Employment decision. The term "employment decision" means to screen candidates for employment or employees for promotion within the city. (L.L. 2021/144, 12/11/2021, eff. 1/1/2023)	**고용상 결정**. “고용상 결정”이란 뉴욕시 내에서 채용 후보자 또는 승진 후보자의 적격성 심사(screen)를 하는 것을 의미한다.
§ 20-871 Requirements for automated employment decision tools. a. In the city, it shall be unlawful for an employer or an employment agency to use an automated employment decision tool to screen a candidate or employee for an employment decision unless:	**§ 20-871 자동회된 고용상 결정 도구의 요건.** a. 뉴욕시 내에서, 사용자 또는 직업소개기관이 자동화된 고용상 결정 도구를 사용하여 고용상 결정의 대상이 되는 후보자 또는 근로자를 선별(screen)하는 것은, 다음의 요건을 충족한 경우를 제외하고 위법(unlawful)이다:

1. Such tool has been the subject of a bias audit conducted no more than one year prior to the use of such tool; and 2. A summary of the results of the most recent bias audit of such tool as well as the distribution date of the tool to which such audit applies has been made publicly available on the website of the employer or employment agency prior to the use of such tool.	1. 해당 도구가 그러한 도구의 사용 전 1년 이내에 편향성 감사를 받은 경우; 그리고 2. 사용자 또는 직업소개기관이 해당 도구를 사용하기 전에, 그러한 도구에 대한 가장 최근의 편향성 감사 결과의 요약(summary of the results)과 해당 감사가 적용되는 도구의 배포일자(distribution date)이 사용자 또는 직업소개기관의 웹사이트에 공개적으로 게시된 경우.
b. **Notices required.** In the city, any employer or employment agency that uses an automated employment decision tool to screen an employee or a candidate who has applied for a position for an employment decision shall notify each such employee or candidate who resides in the city of the following: 1. That an automated employment decision tool will be used in connection with the assessment or evaluation of such employee or candidate that resides in the city. Such notice shall be made no less than ten business days before such use and allow a candidate to request an alternative selection process or accommodation; 2. The job qualifications and characteristics that such automated employment decision tool will use in the assessment of such candidate or employee. Such notice shall be made no less than 10 business days before such use; and	b. **통지 의무**. 뉴욕시 내에서, 자동화된 고용상 결정 도구를 사용하여 고용상 결정을 위해 지원한 근로자나 후보자를 선별하는 사용자 또는 직업소개기관은, 뉴욕시 내에 거주하는 각 근로자 또는 후보자에게 다음 사항을 통지해야 한다: 1. 뉴욕시 내에 거주하는 근로자 또는 후보자의 평가(assessment) 또는 사정(evaluation)과 관련하여 자동화된 고용상 결정 도구가 사용될 것이라는 사실. 이러한 통지는 해당 도구가 사용되기 최소 10영업일 전에 이루어져야 하며, 지원자가 대안적 선발 절차(alternative selection process) 또는 편의 제공(accommodation)을 요청할 수 있도록 해야 한다; 2. 해당 자동화된 고용상 결정 도구가 해당 후보자 또는 근로자를 평가할 때 사용하는 직업상 적격(job qualifications) 및 특성(characteristics). 이러한 통지는 해당 도구가 사용되기 최소 10영업일 전에 이루어져야 한다; 그리고 3. 자동화된 고용상 결정 도구를 위하여 수집된 데이터의 유형, 해당 데이터의 출처 및 사용자 또는 직업소개기관의 데이터 보존 정책에 대한 정보가 사용자 또는 직업소개기관의 웹사이트에 게시되

3. If not disclosed on the employer or employment agency's website, information about the type of data collected for the automated employment decision tool, the source of such data and the employer or employment agency's data retention policy shall be available upon written request by a candidate or employee. Such information shall be provided within 30 days of the written request. Information pursuant to this section shall not be disclosed where such disclosure would violate local, state, or federal law, or interfere with a law enforcement investigation. (L.L. 2021/144, 12/11/2021, eff. 1/1/2023)	지 않은 경우, 후보자 또는 근로자의 서면 요청 시 이러한 정보에 접근이 허용되어야 한다. 이러한 정보는 서면 요청일로부터 30일 이내에 제공되어야 한다. 본조에 따른 정보는 그 공개가 지방, 주 또는 연방 법을 위반하거나 법집행기관의 조사(investigation)를 방해할 수 있는 경우에는 공개되어서는 안 된다.
§ 20-872 Penalties. a. Any person that violates any provision of this subchapter or any rule promulgated pursuant to this subchapter is liable for a civil penalty of not more than $500 for a first violation and each additional violation occurring on the same day as the first violation, and not less than $500 nor more than $1,500 for each subsequent violation. b. Each day on which an automated employment decision tool is used in violation of this section shall give rise to a separate violation of subdivision a of section 20-871. c. Failure to provide any notice to a candidate or an employee in violation of paragraphs 1, 2 or 3 of subdivision b of section 20-871 shall constitute a	**§ 20-872 벌칙.** a. 이 절(subchapter)의 어느 조항 또는 이 절에 따라 공포된 규칙(rule)을 위반하는 자는, 최초 위반 및 최초 위반과 같은 날에 이루어진 각 추가적인 위반에 대해 500달러 이하의 민사벌(civil penalty)에 처하고, 이후의 각각 위반에 대해서는 500달러 이상 1,500달러 이하의 민사벌에 처한다. b. 본조를 위반하여 자동화된 고용상 결정 도구를 사용하는 각각의 날마다 § 20-871 (a)에 대한 별도의 위반에 해당한다. c. § 20-871 (b)의 (1), (2) 또는 (3)을 위반하여 후보자 또는 근로자에게 통지를 제공하지 않는 것은 각각 별도의 위반을 구성한다. d. 이 절에 의해 부과된(authorized) 민사벌의 집행에 관한 소송(proceeding to recover)은 행정 소송 및 청문 사무소(Office of Administrative Trials and Hearings) 또는 그러한 소송을 수행하도록

separate violation. d. A proceeding to recover any civil penalty authorized by this subchapter is returnable to any tribunal established within the office of administrative trials and hearings or within any agency of the city designated to conduct such proceedings. (L.L. 2021/144, 12/11/2021, eff. 1/1/2023)	지정된 뉴욕시 기관 내에 설치된 심판소(tribunal)에 제기할 수 있다.
§ 20-873 Enforcement. The corporation counsel or such other persons designated by the corporation counsel on behalf of the department may initiate in any court of competent jurisdiction any action or proceeding that may be appropriate or necessary for correction of any violation issued pursuant this subchapter, including mandating compliance with the provisions of this chapter or such other relief as may be appropriate. (L.L. 2021/144, 12/11/2021, eff. 1/1/2023)	**§ 20-873 집행.** (뉴욕시 법무부의) 기업 법무관(corporation counsel) 또는 기업 법무관이 법무부(department)를 대표하여 지정한 기타 사람은, 관할 법원에 이 장의 규정의 준수의 의무화 기타 적절한 구제수단을 포함하여 이 절에 따라 발행된 위반(violation) 사항의 시정하는 데 적절하거나 필요할 수 있는 모든 소송 또는 절차를 개시할 수 있다.
§ 20-874 Construction. The provisions of this subchapter shall not be construed to limit any right of any candidate or employee for an employment decision to bring a civil action in any court of competent jurisdiction, or to limit the authority of the commission on human rights to enforce the provisions of title 8, in accordance with law. (L.L. 2021/144, 12/11/2021, eff. 1/1/2023)	**§ 20-874 해석.** 이 절의 조항은 고용상 결정에 대한 후보자 또는 근로자가 관할 법원에 민사 소송을 제기할 수 있는 권리를 제한하거나, 또는 법률에 따라 제8장의 조항을 집행하는 뉴욕시 인권위원회(commission on human rights)의 권한을 제한하는 것으로 해석되어서는 안 된다.

나. 뉴욕시 규정 제6편(Title 6 of the Rules of the City of New York)

91) EEO-1 'component 1' Report는 100명 이상의 근로자를 보유한 모든 민간 부문 사용자와 특정 기준을 충족하는 50명 이상의 근로자를 보유한 연방 계약자가 직업 범주, 성별, 인종 또는 민족별 데이터를 포함한 인력 인구 통계 데이터를 EEOC에 제출하도록 요구하는 필수적인 연간 보고서이다.

Subchapter T: Automated Employment Decision Tools	
§ 5-300 Definitions. As used in this subchapter, the following terms have the following meanings:	**§ 5-300 정의.** 이 절(subchapter)에서 사용되는 아래 용어의 의미는 다음과 같다:
Automated Employment Decision Tool. "Automated employment decision tool" or "AEDT" means "Automated employment decision tool" as defined by § 20-870 of the Code where the phrase "to substantially assist or replace discretionary decision making" means: i. to rely solely on a simplified output (score, tag, classification, ranking, etc.), with no other factors considered; or ii. to use a simplified output as one of a set of criteria where the simplified output is weighted more than any other criterion in the set; or iii. to use a simplified output to overrule conclusions derived from other factors including human decision-making.	**자동화된 고용상 결정 도구.** "자동화된 고용상 결정 도구" 또는 "AEDT"는 뉴욕시 행정법전(the Code) § 20-870에서 정의된 "자동화된 고용상 결정 도구"를 의미하며, 여기서 "재량적 의사 결정을 실질적으로 지원하거나 대체하는(to substantially assist or replace discretionary decision making)"이라는 문구는 다음과 같은 의미를 갖는다: i. 다른 요인을 고려하지 않고, 단순화된 결과값(점수, 태그, 분류, 순위 등)에만 의존하는 것; 또는 ii. 단순화된 결과값(simplified output)을 일련의 기준 중 하나로 사용하는 경우에, 세트의 다른 기준보다 단순화된 산출물의 가중치가 더 높은 경우; 또는 iii. 단순화된 결과값을 사람의 의사 결정을 포함한 다른 요소에서 도출된 결론을 번복(overrule)하는데 사용하는 경우.
Bias Audit. "Bias audit" means "Bias audit" as defined by § 20-870 of the Code.	**편향성 감사.** "편향성 감사"란 뉴욕시 행정법전(the Code) § 20-870에 정의된 "편향성 감사"를 의미한다.
Candidate for Employment. "Candidate for employment" means a person who has applied for a specific employment position by submitting the necessary information or items in the format required by the employer or employment agency.	**채용 후보자.** "채용 지원자"란 사용자 또는 직업소개기관(employment agency)에서 요구하는 양식에 따라 필요한 정보 또는 항목을 제출하여, 특정한 직위(specific employment position)에 지원한 사람을 말한다.
Category. "Category" means any component 1 category required to be reported by employers pursuant to subsection (c) of section 2000e-8 of title 42 of the United States Code as specified in part 1602.7 of title 29 of	**범주.** "범주"란 29 CFR § 1602.7에 기재된 42 USC § 2000e-8 (c)에 따라 EEOC의 정보 보고서 'EEO-1'에 의해 사용자가 보고해야 하는 모든 'component 1' 범주를 말한다.

the Code of Federal Regulations, as designated on the Equal Employment Opportunity Commission Employer Information Report EEO-1.	
Code. "Code" means the Administrative Code of the City of New York.	**법전** “법전”은 뉴욕시 행정법전을 말한다.
Distribution Date. "Distribution date" means the date the employer or employment agency began using a specific AEDT.	**배포일자.** “배포일자”란 사용자 또는 직업소개기관이 특정 AEDT의 사용을 개시한 날을 말한다.
Employment Decision. "Employment decision" means "Employment decision" as defined by § 20-870 of the Code.	**고용상 결정.** “고용상 결정”이란 뉴욕시 행정법전 § 20-870에 정의된 “고용상 결정”을 말한다.
Employment Agency. "Employment agency" means "Employment agency" as defined by 6 RCNY § 5-249.	**직업소개기관.** “직업소개기관”는 6 RCNY(뉴욕시 규정) § 5-249에 정의된 “직업소개기관”를 말한다.
Historical data. "Historical data" means data collected during an employer or employment agency's use of an AEDT to assess candidates for employment or employees for promotion.	**이력 데이터.** “이력(履歷) 데이터”란 사용자 또는 직업소개기관이 채용 후보자 또는 승진 대상자를 평가하기 위해 AEDT를 사용하는 동안 수집된 데이터를 말한다.
Independent Auditor. "Independent auditor" means a person or group that is capable of exercising objective and impartial judgment on all issues within the scope of a bias audit of an AEDT. An auditor is not an independent auditor of an AEDT if the auditor: i. is or was involved in using, developing, or distributing the AEDT; ii. at any point during the bias audit, has an employment relationship with an employer or employment agency that seeks to use or continue to use the AEDT or with a vendor that developed or distributes the AEDT; or iii. at any point during the bias audit,	**독립 감사인.** “독립 감사인”은 AEDT의 편향성 감사의 범위 속하는 모든 사항에 대해 객관적이고 공정한 판단을 내릴 수 있는 개인 또는 그룹을 말한다. 감사인이 다음에 해당하는 때에는 독립 감사인이 아니다: i. AEDT의 사용, 개발 또는 유통에 데 관여하고 있거나, 관여했던 경우; ii. 편향성 감사 중 어느 시점에서든, AEDT를 사용하거나 계속 사용하려는 사용자·직업소개기관 또는 AEDT를 개발 또는 유통한 공급업체(vendor)와 고용관계를 맺은 경우; 또는 iii. 편향성 감사 중 어느 시점에서든, AEDT를 사용하거나 계속 사용하려는 사용자·직업소개기관 또는 AEDT를 개발 또는 유통한 공급업체(vendor)와 직접적인 재정적 이해관계 또는 중대한(material) 간접적 재정적 이해관계를 맺

has a direct financial interest or a material indirect financial interest in an employer or employment agency that seeks to use or continue to use the AEDT or in a vendor that developed or distributed the AEDT.	은 경우.
Impact Ratio. "Impact ratio" means either (1) the selection rate for a category divided by the selection rate of the most selected category or (2) the scoring rate for a category divided by the scoring rate for the highest scoring category. Impact Ratio= selection rate for a category ÷ selection rate of the most selected category OR Impact Ratio= scoring rate for a category ÷ scoring rate of the highest scoring category	**영향률**. "영향률"이란 (1) 어떤 범주(a category)의 선발률(selection rate)을 가장 많이 선택된 범주의 선택률로 나눈 값 또는 (2) 어떤 범주의 득점률(scoring rate)을 가장 높은 점수를 받은 범주의 득점률로 나눈 값 중 하나를 말한다. 영향률 = 어떤 범주의 선발률 ÷ 가장 많이 선택된 범주의 선발률 또는 영향률 = 어떤 범주의 득점률 ÷ 가장 높은 점수를 받은 범주의 득점률
Machine learning, statistical modeling, data analytics, or artificial intelligence. "Machine learning, statistical modeling, data analytics, or artificial intelligence" means a group of mathematical, computer-based techniques: i. that generate a prediction, meaning an expected outcome for an observation, such as an assessment of a candidate's fit or likelihood of success, or that generate a classification, meaning an assignment of an observation to a group, such as categorizations	**기계 학습, 통계적 모델링, 데이터 분석 또는 인공지능.** "기계 학습, 통계적 모델링, 데이터 분석 또는 인공지능"이란 수학적, 컴퓨터 기반의 일련의 기술로 다음의 것을 의미한다: i. 지원자의 적합성(fit) 또는 성공 가능성(likelihood of success)에 대한 평가(assessment)와 같은 관찰에 의한 예상 결과(expected outcome)라는 의미에서의 예측(prediction)을 생성하거나, 또는 관찰에 따라 기능(skill sets) 또는 적성(aptitude)에 기반한 분류와 같은 그룹에 할당한다는 의미에서의 분류(classification)를 생성하는 경우; 및 ii. 컴퓨터가 적어도 부분적으로 입력값

based on skill sets or aptitude; and ii. for which a computer at least in part identifies the inputs, the relative importance placed on those inputs, and, if applicable, other parameters for the models in order to improve the accuracy of the prediction or classification.	(inputs), 해당 입력값에 부여된 상대적 중요도 및 적용 가능할 경우 예측(prediction) 또는 분류(classification)의 정확성을 개선하기 위해 모델의 다른 매개변수(parameters)를 식별하는 경우.
Scoring Rate. "Scoring Rate" means the rate at which individuals in a category receive a score above the sample's median score, where the score has been calculated by an AEDT.	**득점률.** "득점률"이란 어떤 범주에 속하는 개인이 AEDT에 의해 부여받은 점수 중에서 표본의 중앙값(median score)보다 높은 점수를 받은 비율을 의미한다.
Screen. "Screen" means to make a determination about whether a candidate for employment or employee being considered for promotion should be selected or advanced in the hiring or promotion process.	**적격심사.** "적격심사"란 채용 후보자 또는 승진 후보자를 채용 또는 승진 절차에서 선발 또는 승진시킬 것인지에 대한 결정을 하는 것을 말한다.
Selection Rate. "Selection rate" means the rate at which individuals in a category are either selected to move forward in the hiring process or assigned a classification by an AEDT. Such rate may be calculated by dividing the number of individuals in the category moving forward or assigned a classification by the total number of individuals in the category who applied for a position or were considered for promotion. **Example.** If 100 Hispanic women apply for a position and 40 are selected for an interview after use of an AEDT, the selection rate for Hispanic women is 40/100 or 40%.	**선발률.** "선발률"이란 어떤 범주에 속하는 개인이 AEDT에 의해 채용 절차에서 다음 단계로 진행(move forward)하도록 선발되거나, 직급(classification)을 부여받는 비율을 말한다. 이러한 비율은 해당 범주에서 채용되거나 직급을 부여받은 개인의 수를 해당 범주에서 해당 직위에 구직을 신청하거나 승진 후보로 고려된 개인의 총수로 나누어 계산할 수 있다. **예시.** 히스패닉계 여성 100명이 어떤 직책(position)에 지원하여 40명이 AEDT에 의해 면접 대상자로 선발된 경우, 히스패닉계 여성의 선발률은 40/100 또는 40%이다.
Simplified output. "Simplified output" means a prediction or classification as	**단순화된 결과값.** "단순화된 결과값"이란 "기계 학습, 통계적 모델링, 데이터 분석 또는

specified in the definition for "machine learning, statistical modelling, data analytics, or artificial intelligence." A simplified output may take the form of a score (e.g., rating a candidate's estimated technical skills), tag or categorization (e.g., categorizing a candidate's resume based on key words, assigning a skill or trait to a candidate), recommendation (e.g., whether a candidate should be given an interview), or ranking (e.g., arranging a list of candidates based on how well their cover letters match the job description). It does not refer to the output from analytical tools that translate or transcribe existing text, e.g., convert a resume from a PDF or transcribe a video or audio interview.	인공지능"에 대한 정의에 명시된 예측(prediction) 또는 분류(classification)를 말한다. 단순화된 결과값은 점수(예컨대, 후보자의 예상 기술력 평가), 태그 또는 분류(예컨대, 키워드를 기반으로 한 지원자의 이력서(resume) 분류, 후보자에게 기술 또는 특성 할당), 추천(예컨대, 후보자에게 면접 기회를 부여할 것인지 여부) 또는 순위(예컨대, 후보자의 자기소개서(cover letters)와 직무기술서(job description)가 얼마나 부합하는지에 따른 후보자 목록의 정렬)의 형태를 취할 수 있다. 예컨대, PDF에서 이력서를 전환하거나 비디오 또는 오디오 인터뷰를 전사(轉寫)하는 것과 같이, 기존의 텍스트를 번역하거나 전사한 산출물을 이에 해당하지 않는다.
Test data. "Test data" means data used to conduct a bias audit that is not historical data.	**시험 데이터.** "시험 데이터"란 편향성 감사를 수행하는 데 사용되는 데이터로, 이력 데이터(historical data)가 아닌 것을 말한다.
§ 5-301 Bias Audit. (a) An employer or employment agency may not use or continue to use an AEDT if more than one year has passed since the most recent bias audit of the AEDT.	**§ 5-301 편향성 감사.** (a) 사용자 또는 직업소개기관은 AEDT에 대한 가장 최근의 편향성 감사 이후 1년 이상 경과한 경우, 해당 AEDT를 사용하거나 사용을 계속할 수 없다.
(b) Where an AEDT selects candidates for employment or employees being considered for promotion to move forward in the hiring process or classifies them into groups, a bias audit must, at a minimum: (1) Calculate the selection rate for each category; (2) Calculate the impact ratio for each category; (3) Ensure that the calculations	(b) AEDT가 채용 후보자 또는 승진 후보자를 선정하여 채용 절차를 진행하거나 이들을 그룹으로 분류하는 경우, 최소한 다음 항목을 포함한 편향성 감사를 실시해야 한다: (1) 각 범주에 대한 선발률(selection rate)의 산정; (2) 각 범주에 대한 영향률(impact ratio)의 산정; (3) 본 항의 (1)호 및 (2)호에서 요구하는 산정에 있어 AEDT가 아래의 범주에 미치는 영향을 각각(separately) 산정할 것:

required in paragraphs (1) and (2) of this subdivision separately calculate the impact of the AEDT on: i. Sex categories (e.g., impact ratio for selection of male candidates vs female candidates), ii. Race/Ethnicity categories (e.g., impact ratio for selection of Hispanic or Latino candidates vs Black or African American [Not Hispanic or Latino] candidates), and iii. intersectional categories of sex, ethnicity, and race (e.g., impact ratio for selection of Hispanic or Latino male candidates vs. Not Hispanic or Latino Black or African American female candidates). (4) Ensure that the calculations in paragraphs (1), (2), and (3) of this subdivision are performed for each group, if an AEDT classifies candidates for employment or employees being considered for promotion into specified groups (e.g., leadership styles); and (5) Indicate the number of individuals the AEDT assessed that are not included in the required calculations because they fall within an unknown category.	i. 성별 범주 (예컨대, 남성 지원자 대 여성 지원자의 선발에 관한 영향률); ii. 인종/민족 범주 (예컨대, 히스패닉 또는 라틴계 후보자 대 [히스패닉 또는 라틴계가 아닌] 흑인 또는 아프리카계 미국인 후보자의 선정에 관한 영향률), 및 iii. 성별, 민족 및 인종의 교차(intersectional) 범주(예컨대, 히스패닉 또는 라틴계 남성 후보자 대 히스패닉 또는 라틴계가 아닌 흑인 또는 아프리카계 미국인 여성 후보자의 선정에 관한 영향률). (4) AEDT가 채용 후보자 또는 승진 후보자를 특정한 그룹(예컨대, 리더십 스타일)으로 분류하는 경우, 이러한 각각의 그룹에 대하여 이 본항 (1), (2) 및 (3)호에 따른 산정을 수행할 것. (5) AEDT에 의해 평가되었지만, 범주가 불분명하여(unknown) 요구되는 산정에 포함되지 않은 개인의 수를 표시할 것.
Example: An employer wants to use an AEDT to screen resumes and schedule interviews for a job posting. To do so, the employer must ensure that a bias audit of the AEDT was conducted no	**예시:** 사용자가 채용 공고에 대한 이력서를 적격심사(screen)하고 면접 일정을 잡기 위해 AEDT의 사용을 희망한다. 이를 위해서는, 사용자는 AEDT의 예정된 사용일의 1년 이내에 AEDT에 대한 편향성 감사가 수행되

more than a year before the planned use of the AEDT. This bias audit is necessary even though the employer is not using the AEDT to make the final hiring decision, but only to screen at an early point in the application process. The employer asks the vendor for a bias audit. The vendor provides historical data regarding applicant selection that the vendor has collected from multiple employers to an independent auditor who will conduct a bias audit as follows:	도록 해야 한다. 사용자가 AEDT를 최종 채용 결정(final hiring decision)을 내리는 데는 사용하지 않고, 단지 지원 절차의 초기 단계에서 적격심사를 하는 데만 사용하는 경우에도 이러한 편향성 감사는 필요하다. 사용자는 공급업체에 편향성 감사를 요청한다. 공급업체는 여러 사용자로부터 수집한 후보자 선택에 관한 이력 데이터(historical data)를 편향성 감사를 수행할 독립 감사인에게 다음과 같이 제공한다.

Sex Categories				
	# of Applicants	# Selected	Selection Rate	Impact Ratio
Male	1390	667	48%	1.00
Female	1181	555	47%	0.979

Race/Ethnicity Categories				
	# of Applicants	# Selected	Selection Rate	Impact Ratio
Hispanic or Latino	408	204	50%	0.97
White (Not Hispanic or Latino)	797	412	52%	1.00
Black or African American (Not Hispanic or Latino)	390	170	44%	0.84
Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	119	52	44%	0.85
Asian (Not Hispanic or Latino)	616	302	49%	0.95
Native American or Alaska Native (Not Hispanic or Latino)	41	18	44%	0.85
Two or More Races (Not Hispanic or Latino)	213	96	45%	0.87

Intersectional Categories						
			# of Applicants	# Selected	Selection Rate	Impact Ratio
Hispanic or Latino		Male	205	90	43.9%	0.841
		Female	190	82	43.2%	0.827
Non/Hispanic or Latino	Male	White	412	215	52.2%	1.000
		Black or African American	226	95	42.0%	0.806
		Native Hawaiian or Pacific	87	37	42.5%	0.815

		Islander				
		Asian	321	167	52.0%	0.997
		Native American or Alaska Native	24	11	45.8%	0.878
		Two or More Races	115	52	45.2%	0.866
	Female	White	385	197	51.2%	0.981
		Black or African American	164	75	45.7%	0.876
		Native Hawaiian or Pacific Islander	32	15	46.9%	0.898
		Asian	295	135	45.8%	0.877
		Native American or Alaska Native	17	7	41.2%	0.789
		Two or More Races	98	44	44.9%	0.860

Note: The AEDT was also used to assess 250 individuals with an unknown sex or race/ethnicity category. Data on those individuals was not included in the calculations above.	참고: 성별 또는 인종/민족이 불분명한 250명의 개인을 평가하는 데에도 AEDT를 사용했다. 이러한 개인에 대한 데이터는 위의 산정에 포함되지 않았다.
(c) Where an AEDT scores candidates for employment or employees being considered for promotion, a bias audit must, at a minimum: (1) Calculate the median score for the full sample of applicants; (2) Calculate the scoring rate for individuals in each category; (3) Calculate the impact ratio for each category; (4) Ensure that the calculations required in paragraphs (1), (2), and (3) of this subdivision separately calculate the impact of the AEDT on: i. Sex categories (i.e., impact ratio for selection of male candidates vs	(c) AEDT가 채용 후보자 또는 승진 후보자에게 점수를 매기는(scores) 경우, 최소한 다음 항목을 포함한 편향성 감사를 실시해야 한다: (1) 전체 지원자 표본의 점수의 중앙값(median score)의 산정; (2) 각각의 범주에 속하는 개인의 득점률(scoring rate)의 산정; (3) 각 범주에 관한 영향률(impact ratio)의 산정; (4) 본항의 (1), (2) 및 (3)호에서 요구하는 산정에 있어 AEDT가 아래의 범주에 미치는 영향을 각각(separately) 산정할 것: i. 성별 범주 (예컨대, 남성 지원자 대 여성 지원자의 선발에 관한 영향률); ii. 인종/민족 범주 (예컨대, 히스패닉 또는

female candidates), ii. Race/Ethnicity categories (e.g., impact ratio for selection of Hispanic or Latino candidates vs Black or African American [Not Hispanic or Latino] candidates), and iii. intersectional categories of sex, ethnicity, and race (e.g., impact ratio for selection of Hispanic or Latino male candidates vs. Not Hispanic or Latino Black or African American female candidates); and (5) Indicate the number of individuals the AEDT assessed that are not included in the required calculations because they fall within an unknown category.	라틴계 후보자 대 [히스패닉 또는 라틴계가 아닌] 흑인 또는 아프리카계 미국인 후보자의 선정에 관한 영향률), 및 iii. 성별, 민족 및 인종의 교차(intersectional) 범주(예컨대, 히스패닉 또는 라틴계 남성 후보자 대 히스패닉 또는 라틴계가 아닌 흑인 또는 아프리카계 미국인 여성 후보자의 선정에 관한 영향률); 그리고 (5) AEDT에 의해 평가되었지만, 범주가 불분명하여(unknown) 요구되는 산정에 포함되지 않은 개인의 수를 표시할 것.
(d) Notwithstanding the requirements of paragraphs (2) and (3) of subdivision (b) and paragraphs (3) and (4) of subdivision (c), an independent auditor may exclude a category that represents less than 2% of the data being used for the bias audit from the required calculations for impact ratio. Where such a category is excluded, the summary of results must include the independent auditor's justification for the exclusion, as well as the number of applicants and scoring rate or selection rate for the excluded category.	(d) (b)항의 (2) 및 (3)호와 (c)항의 (3) 및 (4)호의 요건에도 불구하고, 독립 감사인은 편향성 감사에 사용되는 데이터의 2% 미만을 차지하는 범주를 영향률 산정에서 제외할 수 있다. 이러한 범주가 제외되는 경우, 제외된 범주의 지원자 수와 득점률(scoring rate) 또는 선택률(selection rate)은 물론 이러한 제외에 관한 독립 감사인의 정당성 증명(justification)이 결과 요약(summary of results)에 포함되어야 한다.
Example: An employer uses an AEDT to score applicants for "culture fit." To do so, the employer must ensure that a bias audit of the AEDT was conducted no more than a year before the use of the AEDT. The employer provides	**예시**: 사용자가 AEDT를 사용하여 지원자의 “기업 문화 적합성(culture fit)”에 대한 점수를 부여한다. 이를 위해서는, 사용자는 AEDT 사용의 1년 이내에 AEDT에 대한 편향성 감사가 수행되도록 해야 한다. 사용자는 다음과 같이 편향성 감사를 수행할 수 있도

historical data on "culture fit" score of applicants for each category to an independent auditor to conduct a bias audit as follows:	록, 각 범주의 지원자의 “기업 문화 적합성” 점수에 관한 이력 데이터(historical data)를 독립 감사인에게 제공한다.

Sex Categories			
	# of Applicants	Scoring Rate	Impact Ratio
Male	92	54.3%	1.00
Female	76	44.7%	0.82

Race/Ethnicity Categories			
	# of Applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	28	64.2%	1.00
White (Not Hispanic or Latino)	40	37.5%	0.58
Black or African American (Not Hispanic or Latino)	32	50.0%	0.78
Native Hawaiian or Pacific Islander (Not Hispanic or Latino)	8	62.5%	0.97
Asian (Not Hispanic or Latino)	24	41.7%	0.65
Native American or Alaska Native (Not Hispanic or Latino)	16	62.5%	0.97
Two or More Races (Not Hispanic or Latino)	20	50.0%	0.78

Intersectional Categories					
			# of Applicants	Scoring Rate	Impact Ratio
Hispanic or Latino		Male	16	75%	1.00
		Female	12	50%	0.67
Non/ Hispanic or Latino	Male	White	20	35%	0.47
		Black or African American	20	50%	0.67
		Native Hawaiian or Pacific Islander	4	75%	1.00
		Asian	12	58.3%	0.78
		Native American or Alaska Native	8	62.5%	0.83
		Two or More Races	12	50%	0.67
	Female	White	20	40%	0.53
		Black or African American	12	50%	0.67
		Native Hawaiian or Pacific Islander	4	50%	0.67
		Asian	12	25%	0.33
		Native American or Alaska Native	8	62.5%	0.83
		Two or More Races	8	50%	0.67

Note: The AEDT was used to assess 15 individuals with an unknown sex or	참고: 성별 또는 인종/민족이 불분명한 15명의 개인을 평가하기 위해 AEDT를

race/ ethnicity category. Data on these individuals was not included in the calculations above.	사용했다. 이러한 개인에 대한 데이터는 위의 산정에 포함되지 않았다.
§ 5-302 Data Requirements. (a) **Historical Data.** A bias audit conducted pursuant to 6 RCNY § 5-301 must use historical data of the AEDT. The historical data used to conduct a bias audit may be from one or more employers or employment agencies that use the AEDT. However, an individual employer or employment agency may rely on a bias audit of an AEDT that uses the historical data of other employers or employment agencies only in the following circumstances: if such employer or employment agency provided historical data from its own use of the AEDT to the independent auditor conducting the bias audit or if such employer or employment agency has never used the AEDT.	**§ 5-302 데이터 요건.** (a) **이력 데이터**. 6 RCNY § 5-301에 따라 수행되는 편향성 감사는 해당 AEDT의 이력 데이터를 사용해야 한다. 편향성 감사를 수행하는 데 사용되는 이력 데이터는 AEDT를 사용하는 하나 또는 그 이상의 사용자 또는 직업소개기관으로부터 제공받을 수 있다. 그러나, 개별 사용자 또는 직업소개기관은 해당 사용자 또는 직업소개기관이 편향성 감사를 수행하는 독립 감사인에게 AEDT의 사용에 따른 자신의 이력 데이터를 제공하였거나, 또는 해당 사용자 또는 직업소개기관이 AEDT를 사용한 적이 없는 경우에만 다른 사용자 또는 직업소개기관의 이력 데이터를 사용한 편향성 감사에 의존할 수 있다.
(b) **Test Data**. Notwithstanding the requirements of subdivision (a) of this section, an employer or employment agency may rely on a bias audit that uses test data if insufficient historical data is available to conduct a statistically significant bias audit. If a bias audit uses test data, the summary of results of the bias audit must explain why historical data was not used and describe how the test data used was generated and obtained.	(b) **시험 데이터.** 본조 (a)항의 요건에도 불구하고, 통계적으로 유의미한 편향성 감사를 수행하기에 이력 데이터가 충분하지 않은 경우, 사용자 또는 직업소개기관은 시험 데이터(test data)를 사용한 편향성 감사에 의존할 수 있다. 편향성 감사에서 시험 데이터를 사용하는 경우, 편향성 감사의 결과 요약(summary of results)에는 이력 데이터를 사용하지 않은 이유가 설명되고, 사용된 시험 데이터가 어떻게 생성되고 취득되었는지가 기술되어야만 한다.
Example 1: An employer is planning to use an AEDT for the first time. The employer may rely on a bias audit conducted using the historical data of other employers or employment	**예시 1**: 사용자가 최초로 AEDT를 사용할 계획이다. 사용자는 다른 사용자 또는 직업소개기관의 이력 데이터를 사용하여 수행한 편향성 감사 또는 시험 데이터를 사용하여 수행한 편향성 감사에 의존할 수 있다.

agencies, or on a bias audit conducted using test data.	
Example 2: An employment agency has been using an AEDT for 6 months. The bias audit the employment agency relied on before its first use of the AEDT was conducted 10 months ago using test data. The employment agency will need an updated bias audit if it will continue to use the AEDT once 12 months have passed since the bias audit it first relied on was conducted. The employment agency's data from 6 months of use of the AEDT is not sufficient on its own to conduct a statistically significant bias audit. The employment agency may rely on a bias audit using the historical data of other employers and employment agencies if it provides its 6 months of historical data to the independent auditor for use and consideration. The employment agency may also rely on a bias audit that uses test data.	**예시 2**: 어떤 직업소개기관에서 6개월 동안 AEDT를 사용해 왔다. 직업소개기관이 AEDT를 최초로 사용하기 전에 의존한 편향성 감사는 10개월 전에 시험 데이터를 사용하여 수행되었다. 직업소개기관이 최초로 편향성 감사를 실시한 때로부터 12개월이 경과한 후에도 AEDT를 계속 사용하기 위해서는 업데이트된 편향성 감사를 받아야 한다. 직업소개기관이 6개월 동안 AEDT를 사용한 데이터만으로는 통계적으로 유의미한 편향성 감사를 수행하기에 충분하지 않다. 직업소개기관이 독립 감사인에게 사용 및 검토를 위해 6개월 간의 이력 데이터를 제공한 경우, 다른 사용자 및 직업소개기관의 이력 데이터를 사용한 편향성 감사에 의존할 수 있다. 직업소개소는 또한 시험 데이터를 사용한 편향성 감사에 의존할 수도 있다.
Example 3: An employer has been using an AEDT for 3 years and will soon need an updated bias audit. The employer has statistically significant data from its 3 years of use of the AEDT. The employer may rely on a bias audit conducted using historical data from multiple employers if it provides its 3 years of historical data to the independent auditor for use and consideration. The employer may also rely on a bias audit conducted using historical data from its own use of the AEDT, without any data from other	**예시 3**: 사용자가 3년 동안 AEDT를 사용해 왔으며, 곧 업데이트된 편향성 감사를 받아야 한다. 사용자는 3년간 AEDT 사용을 통하여 통계적으로 유의미한 데이터를 보유하고 있다. 사용자가 독립 감사인에게 사용 및 검토를 위해 3년간의 이력 데이터를 제공하는 경우, 다수 사용자의 이력 데이터를 사용하여 수행된 편향성 감사에 의존할 수 있다. 또한, 사용자는 다른 사용자나 직업소개기관의 데이터 없이, 자신이 AEDT를 사용한 이력 데이터만을 사용하여 수행한 편향성 감사에 의존할 수도 있다. 사용자는 시험 데이터를 사용하여 수행된 편향성 감사에 의존해서는 안 된다.

employers or employment agencies. The employer may not rely on a bias audit conducted using test data.	
§ 5-303 Published Results. (a) Before the use of an AEDT, an employer or employment agency in the city must make the following publicly available on the employment section of their website in a clear and conspicuous manner: (1) The date of the most recent bias audit of the AEDT and a summary of the results, which shall include the source and explanation of the data used to conduct the bias audit, the number of individuals the AEDT assessed that fall within an unknown category, and the number of applicants or candidates, the selection or scoring rates, as applicable, and the impact ratios for all categories; and, (2) The distribution date of the AEDT.	**§ 5-303 결과의 게시.** (a) AEDT의 사용에 앞서, 뉴욕시 내의 사용자와 직업소개기관은 자신의 웹사이트 내 고용 섹션(employment section)에 다음 사항을 명확하고 눈에 잘 띄는 방식으로 공개해야 한다: (1) 가장 최근의 AEDT 편향성 감사 실시일과 편향성 감사 수행에 사용된 데이터의 출처와 설명, AEDT에 의해 평가되었지만 범주가 불분명한 개인의 수, 지원자 또는 후보자의 수, 해당하는 경우의 선발률 또는 득점률 및 모든 범주에 대한 영향률이 포함된 결과의 요약; 그리고, (2) AEDT의 배포일자(distribution date).
(b) The requirements of subdivision (a) of this section may be met with an active hyperlink to a website containing the required summary of results and distribution date, provided that the link is clearly identified as a link to results of the bias audit.	(b) 해당 링크가 편향성 감사의 결과에 대한 링크로 명확하게 식별되는 한, 요구되는 결과의 요약 및 배포일자(distribution date)이 포함된 웹사이트에 대한 활성 하이퍼링크(active hyperlink)로 본조 (a)항의 요건을 충족할 수 있다.
(c) An employer or employment agency must keep the summary of results and distribution date posted for at least 6 months after its latest use of the AEDT for an employment decision.	(c) 사용자 또는 직업소개기관은 고용상 결정을 위해 AEDT를 마지막으로 사용한 때로부터 최소 6개월 동안 결과의 요약과 배포일자를 게시해야 한다.
§ 5-304 Notice to Candidates and Employees. (a) The notice required by § 20-871(b)(1) of the Code must include	**§ 5-304 후보자 및 근로자에 대한 통지.** (a) 뉴욕시 행정법전 § 20-871(b)(1)에 의해 요구되는 통지에는, 해당하는 경우, 개인이 대안적 선발 절차(alternative selection

instructions for how an individual can request an alternative selection process or a reasonable accommodation under other laws, if available. Nothing in this subchapter requires an employer or employment agency to provide an alternative selection process.	process) 또는 다른 법률에 따라 합리적인 편의 제공(reasonable accommodation)을 요청할 방법에 대한 지침이 포함되어야 한다. 본조의 어떠한 규정도 사용자 또는 직업소개기관에 대하여 대안적 선발 절차를 제공하도록 요구하지 않는다.
(b) To comply with § 20-871(b)(1) and (2) of the Code, an employer or employment agency may provide notice to a candidate for employment who resides in the city by doing any of the following: (1) Provide notice on the employment section of its website in a clear and conspicuous manner at least 10 business days before use of an AEDT; (2) Provide notice in a job posting at least 10 business days before use of an AEDT; or, (3) Provide notice to candidates for employment via U.S. mail or e-mail at least 10 business days before use of an AEDT.	(b) 뉴욕시 행정법전 § 20-871(b)(1) 및 (2)를 준수하기 위하여, 사용자 또는 직업소개기관은 뉴욕시에 거주하는 채용 후보자에게 다음 중 어느 하나에 해당하는 방법으로 통지할 수 있다: (1) AEDT 사용의 영업일 기준 최소 10일 전에 자신의 웹사이트 고용 섹션에 명확하고 눈에 잘 띄는 방식으로 통지할 것; (2) AEDT 사용의 영업일 기준 최소 10일 전에 채용 공고(job posting)를 통해 통지할 것; 또는 (3) AEDT 사용의 영업일 기준 최소 10일 전에 미국 우편(U.S. mail) 또는 이메일을 통해 채용 후보자에게 통지할 것.
(c) To comply with § 20-871(b)(1) and (2) of the Code, an employer or employment agency may provide notice to an employee being considered for promotion who resides in the city by doing any of the following: (1) Provide notice in a written policy or procedure that is provided to employees at least 10 business days before use of an AEDT; (2) Provide notice in a job posting at least 10 business days before use of an AEDT; or,	(c) 뉴욕시 행정법전 § 20-871(b)(1) 및 (2)를 준수하기 위하여, 사용자 또는 직업소개기관은 뉴욕시에 거주하는 승진 후보자에게 다음 중 어느 하나에 해당하는 방법으로 통지할 수 있습니다: (1) AEDT 사용의 영업일 기준 최소 10일 전에 근로자에게 제공되는 서면 정책(written policy)이나 절차를 통해 통지할 것; (2) AEDT 사용의 영업일 기준 최소 10일 전에 사내 공모(job posting)를 통해 통지할 것; 또는

(3) Provide notice via U.S. mail or e-mail at least 10 business days before use of an AEDT.	(3) AEDT 사용의 영업일 기준 최소 10일 전에 미국 우편(U.S. mail) 또는 이메일을 통해 채용 후보자에게 통지할 것.
(d) To comply with § 20-871(b)(3) of the Code, an employer or employment agency must: (1) Provide information on the employment section of its website in a clear and conspicuous manner about its AEDT data retention policy, the type of data collected for the AEDT, and the source of the data; (2) Post instructions on the employment section of its website in a clear and conspicuous manner for how to make a written request for such information, and if a written request is received, provide such information within 30 days; and (3) Provide an explanation to a candidate for employment or employee being considered for promotion why disclosure of such information would violate local, state, or federal law, or interfere with a law enforcement investigation.	(d) 뉴욕시 행정법전 § 20-871(b)(3)을 준수하기 위하여, 사용자 또는 직업소개기관은 다음을 수행해야 한다: (1) 자신의 웹사이트 고용 섹션에 명확하고 눈에 잘 띄는 방식으로, AEDT 데이터 보존 정책, AEDT를 위해 수집하는 데이터의 유형 및 데이터 출처에 대한 정보를 제공해야 한다; (2) 자신의 웹사이트 고용 섹션에 명확하고 눈에 잘 띄는 방식으로, 이러한 정보를 서면으로 요청(written request)하는 방법에 대한 지침을 게시하고, 서면 요청이 접수될 경우 30일 이내에 해당 정보를 제공해야 한다; 그리고 (3) 채용 후보자 또는 승진 후보자에게, 해당 정보의 공개(disclosure)가 지방(local), 주 또는 연방 법을 위반하거나, 또는 법집행기관의 조사(investigation)에 방해가 되는 이유에 관한 설명을 제공해야 한다.

2) 캘리포니아주

캘리포니아주는 고용과 관련된 결정을 포함한 중대한 결정에 있어 인공지능, 기계 학습 및 기타 데이터 기반 통계 프로세스 사용을 규제하기 위한 조치를 취하고 있다.[92] 캘리포니아 민권위원회(Civil Rights Council, CRC)는 2023년 2월 10일에 “자동화된 결정 시스

92) https://www.littler.com/publication-press/publication/update-californias-efforts-regulate-use-ai-employment-decision-making

템에 관한 고용 규정에 대한 개정(안)(Proposed Modifications to Employment Regulations Regarding Automated-Decision Systems)"을 발표했다. 이 외에 고용에 관하여 인공지능을 추가로 규제하기 위한 몇몇 법안이 캘리포니아주 상원[93]과 하원[94]에 각각 제출되었다. 이중 인공지능의 규제와 관련하여 의미가 있는 것은 하원에 제출된 AB-1651 Worker rights: Workplace Technology Accountability Act과 Assembly Bill No. 331 이다.[95]

가. Proposed Modifications to Employment Regulations Regarding Automated-Decision Systems

Civil Rights Council(CRC)는 2023년 2월 10일에 자동화된 결정 시스템에 관한 고용 규정(Employment Regulations Regarding Automated-Decision Systems)에 대한 수정안을 제안했다. 종전에는 '캘리포니아 공정 고용 및 주택 위원회(California Fair Employment and Housing Council)'라고 불렸던 CRC는 캘리포니아 민권부(California Civil Rights Department)에 소속되어 있으며, 캘리포니아 공정 고용 및 주택법(California Fair Employment and Housing Act, "FEHA")을 비롯한 캘리포니아주의 민권법을 집행업무를 담당한다.

한편, 동 수정안은 캘리포니아주의 고용 및 채용 관행을 규율하는 기존 규정에 자동 의사 결정 시스템(ADS)의 사용을 통합하기 위한 전면적인 변경을 제안한 2022년 3월의 CRC 수정안의 개정판이다.

동 개정안은 "자동화된 의사 결정 시스템"을 "기계 학습, 통계 및 기타 데이터 처리 또는 인공지능 기술에서 파생된 것을 포함하여, 선별, 평가, 분류, 추천 또는 기타 방식으로 근로자 또는 구직자에게 영향을 미치는 결정을 내리거나 인간의 의사 결정을 촉진하는 계산 프로세스"라고 광범위하게 정의하고 있다. 한편, 동 개정안은 자동화된 의사 결정 시스템의 예시를 다음과 같이 제시한다.

- 특정 용어 또는 패턴에 대해 이력서를 선별하는 알고리즘;

---

93) Senate Bill No. 721, "California Interagency AI Working Group," to add and repeal Section 11546.47 of the Government Code, relating to artificial intelligence.

94) AB-1651 Worker rights: Workplace Technology Accountability Act.(2021-2022) (2022. 11. 30. 현재 From committee without further action 상태로, 하원의 회기(2년) 만료로 폐기되었음); AB-331 Automated decision tools.(2023-2024) (2023. 5. 18. 현재 보류(held under submission) 상태임).

95) https://www.dwt.com/blogs/employment-labor-and-benefits/2023/05/california-ai-employment-law-regulations#page=1

- 안면 및/또는 음성인식을 사용하여 얼굴표정, 단어 선택, 음성을 분석하는 알고리즘;
- 근로자 또는 지원자에 대한 예측 평가를 하거나 손재주(dexterity), 반응 시간 또는 기타 신체적 또는 정신적 능력이나 특성을 포함하되 이에 국한되지 않는 특성을 측정하는 데 사용하는 게임화된 테스트(gamified testing)를 사용하는 알고리즘; 및
- 성격 특성, 적성, 인지능력 및/또는 문화적 적합성(cultural fit)을 측정하기 위한 온라인 테스트를 사용하는 알고리즘.

또한, 동 개정안은 "알고리즘"을 "일반적으로 컴퓨터에서 계산, 문제 해결 또는 의사 결정을 내리는 데 사용되는 일련의 규칙 또는 명령어"로 광범위하게 정의하고 있다. 알고리즘에 대한 이러한 정의의 범위는 매우 광범위하여 고용상 결정과 밀접한 관련이 있을 수 있는 특정 애플리케이션이나 시스템은 이에 해당할 가능성이 크다.

한편, 동 개정안에 따르면, 적용 대상 기관은 직무와 관련이 있고 업무상 필요성에 부합하는 것으로 입증되지 않는 한, 보호 대상 특성(protected characteristic)에 근거하여 지원자 또는 근로자를 "선별(screen out)하거나 선별하는 경향이 있는" 자동 결정 시스템을 사용하는 것은 FEHA에 따라 불법이 된다. 동 개정안은 채용 결정은 물론 급여, 승진, 징계, 해고 등의 고용관계 전반에 걸쳐 사용자 및 적용대상인 제3자의 의사 결정에 적용된다. 즉, 동 개정안은 사용자뿐만 아니라 고용상 결정과 관련하여 고용주에게 인공지능 및 기계학습 기술을 제공하는 공급업체(vendor)가 포함될 수 있는 "고용 대행사(employment agencies)"에도 적용된다.

나. AB-1651 Worker rights: Workplace Technology Accountability Act(2021-2022)[96]

제1장. 일반 조항(General Provisions)

- 이 법안은 근로자(employee), 독립계약자(independent contractors) 및 구직자를 포함한 모든 노동자(worker)에게 적용된다.
- 모든 보호와 권리는 노동자 대표(worker representatives)에게도 적용된다.
- 노무도급의 수급인(labor subcontractors)를 포함한 모든 사용자가 적용 대상이다.
- 사용자에게 기술 역무(technology services)를 제공하는 공급업체(vender)도 규제를 받

96) 이 법안에 대한 상세한 논의는 Airlie Hilliard, Emre Kazim, Tom Kemp & Kelvin Bageire (2023) Overview and commentary of the California Workplace Technology Accountability Act, International Review of Law, Computers & Technology, 37:1, 91-109, DOI: 10.1080/13600869.2022.2115749 및 Tom Kemp(2022), Overview of AB 1651: The California Workplace Technology Accountability Act, Medium. at https://medium.com/golden-data/overview-of-ab-1651-the-california-workplace-technology-accountability-act-1f887b88d346 참조.

는다.

제2장. 노동자 정보(Worker Data)

- 사용자는 노동자에게 어떤 정보를 수집할 것인지, 노동자에게 어떤 권리가 있는지, 고용상 결정을 하는 데 정보가 어떻게 사용될 것인지에 관한 자세한 사전통지를 해야 한다.
- 노동자는 자신의 정보에 접근 및 수정(access and correct)할 권리가 있다.
- 사용자는 노동자가 업무를 수행하거나 사용자가 임금 및 복리후생을 관리하기 위하여 필요한 경우에만 노동자 정보를 수집할 수 있다.
- 사용자는 노동자 정보를 제3자에게 판매하거나 라이선스를 부여할 수 없다.

제3장. 전자 감시(Electronic Monitoring)

- 사용자는 노동자에게 최소한의 영향을 미치는 특정한 목적을 위해서만 전자 감시를 사용할 수 있다.
- 사용자는 모든 전자 감시에 대해 사전통지를 해야 한다.
- 사용자는 노동관계법(labor and employment laws)의 위반을 초래하거나, 비번(off-duty)인 근로자 또는 민감한 영역(sensitive areas)의 정보를 기록하거나, 안면 인식과 같은 고위험 기술을 사용하거나, 노동관계법에 따라 권리를 행사하는 노동자를 식별하는 전자 감시를 사용할 수 없다.
- 생산성 감시시스템(productivity monitoring systems)은 시행 전에 문서화되어야 하고, 사업장 보건 및 안전을 감독하는 규제 기관의 검토를 받아야 한다.
- 사용자는 채용, 해고, 징계, 승진 등의 결정을 내릴 때 전자 감시 데이터에만 의존하는 것이 금지된다. 대신, 사용자는 데이터를 독립적으로 검증하고 노동자에게 전체 문서를 제공해야 한다.

제4장. 알고리즘(Algorithms)

- 사용자는 알고리즘 결정 시스템(algorithmic decision systems)을 사용하기 전에 노동자에게 상세한 사전통지를 제공해야 한다. 또한, 사용자는 관련 영향평가(impact assessments)를 공유해야 한다.
- 사용자는 노동관계법 위반을 초래하는 알고리즘을 사용하거나, 직무와 무관한 노동자의 행동을 예측하거나, 노동관계법에 따른 법적 권리를 행사하는 노동자를 식별하거나, 안면 또는 감정 인식(facial or emotion recognition) 기술과 같은 고위험 기술을 사용할 수 없다.
- 생산성 알고리즘(productivity algorithms)은 구동(implementation)하기 전에 문서화되어야 하고, 사업장 보건 및 안전을 감독하는 규제 기관의 검토를 받아야 한다.
- 사용자는 채용, 해고, 징계, 승진 등의 결정을 내릴 때 알고리즘의 결과에만 의존하는

것이 금지된다. 대신, 사용자는 의미 있는 사람에 의한 감독(meaningful human oversight)을 통해 알고리즘의 결과를 독립적으로 검증하고, 노동자에게 전체 문서를 제공해야 한다.

제5장. 영향평가(Impact Assessments)

- 사용자는 노동자에게 영향을 미치는 모든 알고리즘 및 데이터 수집 시스템에 대해 영향평가를 실시하고, 이를 주 노동관서(state labor agency)에 제출해야 한다.
- 영향평가는 차별, 건강 및 안전, 개인정보 보호의 위험을 포함하여, 노동자에게 미칠 수 있는 위험을 충분히 평가해야 한다.
- 사용자는 영향평가 전반에 관하여 노동자와 협의하고, 노동자의 피드백을 반영해야 한다.
- 잠재적 피해(potential harms)가 확인되면, 사용자는 완화 조치(mitigation measures)를 시행해야 한다.
- 주 노동관서는 추가 정보를 요구하거나, 유해한 알고리즘의 사용을 금지하는 등 영향평가에 관한 조치를 취할 수 있는 권한이 있다.

제6장. 집행(Enforcement)

- 노동자는 본 법안의 위반에 관하여, 민사 소송(civil action)을 제기하거나 또는 행정구제의 청구(administrative claim)를 통하여 벌금(penalties) 및 금지명령(injunctive relief)을 청구할 권리가 있다.

다. AB-331 Automated decision tools(2023-2024)[97]

2023년 1월에 발의된 AB-331은 자동화된 의사 결정 도구의 개발자와 배포자가 해당 도구에 대한 "영향평가(impact assessment)"를 수행하도록 의무화한다.[98]

동 법안에서 자동화된 의사 결정 도구는 "인공지능을 사용하며 결과적 결정(consequential decisions)을 내리거나 결정에 영향을 미치기 위해 특별히 개발 및 판매되거나 특별히 수정된 시스템 또는 서비스"라고 정의된다. 한편, 동 법안은 "인공지능"에 대한 정의조항을 두고 있지는 않지만, "결과적 결정(consequential decisions)"을 급여, 승진, 채용, 해고, 자동화된 업무 할당 등 고용에 대한 접근성이나 고용조건 또는 가용성에 "법적 또는 이와 유사하게 중용한 영향(legal or similarly significant effects)을 미치는" 결정 또는 판단이라고 정의한다.

---

97) https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB331&search_keywords=artificial+intelligence

98) https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB331

한편, 동 법안은 자동화된 의사 결정 도구에 대해 위험 기반 평가로 정의되는 영향평가를 요구하며, 여기에는 해당 도구가 수집하고 처리한 데이터의 요약, 성별, 인종 또는 민족에 따른 부정적 영향 분석, 설명 등 여러 요소가 포함된다. 자동화된 의사 결정 도구의 중요한 업데이트에도 영향평가가 필요하다.

동 법안에 따르면 영향평가의 결과는 2년간 유지되어야 하며, 뉴욕시 “Local law 144”와 유사하게 성별, 인종 또는 민족에 따른 부정적 영향 분석을 요구한다. 그러나 동 법안은 결과를 공개적으로 게시하는 대신, 영향평가가 완료된 후 60일 이내에 주 민권부서에 제출하도록 규정하며, 이를 이행하지 않을 경우 과태료(administrative fine)가 부과된다. <끝>

세션 2

한국의 인공지능 규율, 어떻게 할 것인가

사회 : 장여경

(정보인권연구소 상임이사)

【 발제 1 】

인공지능 법안에 대한 시민사회 대안 입법안

김하나
(민주사회를위한변호사모임 디지털정보위원회, 법무법인 두율 변호사)

01

논의 배경

인공지능 기술 개발과 활용 관련 환경이 급변하고 있다.

AI 개발자들조차 이해, 예측하기 어려운 인공지능 기술에 대한 위험을 경고하며 안전한 프로토콜의 필요성을 강조하고 있고, 세계 주요 국가의 정상이 모여 AI의 '안전한' 활용과 거버넌스 구축을 위한 협력방안을 논의하였다.

전세계 AI 관련 논의의 화두는 AI 기술을 어떻게 안전하게 개발 및 활용할 것인가로 모아진다.

하지만 정작 대한민국에는 AI 개발과 활용을 실질적으로 규제할 수 있는 법안이 존재하지 않는다.

1. 논의 배경

[AI 설계 및 개발과정에서 안전한 규범의 필요성]

← All Open Letters

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures
33709

Add your signature

Published
March 22, 2023

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research[1] and acknowledged by top AI labs.[2] As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.* Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.

Contemporary AI systems are now becoming human-competitive at general tasks,[3] and we must ask ourselves: *Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should* we risk loss of control of our civilization? Such decisions must not be delegated to unelected tech leaders. **Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable.** This confidence must be well justified and increase with the magnitude of a system's potential effects. OpenAI's recent statement regarding artificial general intelligence, states that *"At some point, it may be important to get independent review before starting to train future systems, and for the most advanced efforts to agree to limit the rate of growth of compute used for creating new models."* We agree. That point

이미지 출처 : Future of Life Institute

FLI 공개서한의 함의

- 현지시각으로 2023. 3. 29. 미국의 비영리단체인 'Future of Life Institute(미래생명연구소)'가 "6개월간 AI시스템의 개발을 일시 중단하자."는 공개서한을 냈다. 일론 머스크 테슬라 최고경영자(CEO), 「사피엔스」의 저자 유발 하라리, 스티브 워즈니악 애플 공동창업자를 비롯해 유명인들의 서한에 서명했다.

- 공개서한은 현대 AI 시스템이 이제 일반적인 작업에서 인간과 경쟁하게 된 상황을 지적하며 강력한 AI 시스템은 그 효과가 긍정적이고, 위험을 관리할 수 있을 때만 개발되어야 한다고 언급하였다.

- 공개서한은 AI 연구소와 독립 전문가는 일시 중지를 활용하여 독립적인 외부 전문가가 엄격하게 감사하고 감독하는 고급 AI 설계 및 개발을 위한 일련의 공유 안전 프로토콜을 공동으로 개발하고 구현하고, 이러한 프로토콜을 준수하는 프로그램이 의심의 여지없이 안전하다는 것을 보장하여야 한다고 지적하였다.

1. 논의 배경

[안전한 AI 기술의 중요성]

제1회 AI 안전 정상회담

- 2023. 11. 1.(현지시간) 미국, 중국, 한국 등 28개국과 유럽연합(EU)은 영국에서 개막한 '제1차 AI 안전 정상회의'에서 '블레츨리 선언'을 채택했다. '블레츨리 선언'은 고도의 능력을 갖춘 프런티어 AI가 파국적 피해를 초래할 수도 있다는 점을 확인하고, 위험 대처에 관한 이해를 넓히는 한편 국가 간 협력이 필요하다는 점을 명확히 하였다.

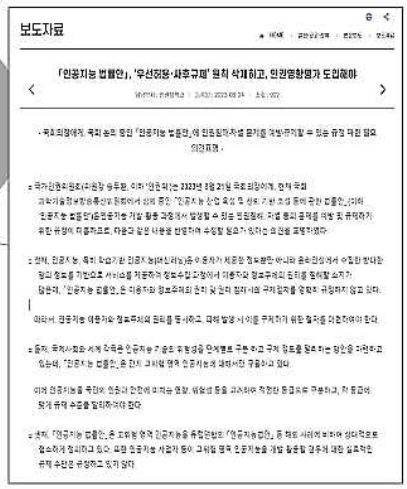
- 윤석열 대통령은 이 회의에서 화상 연설을 통해 인공지능의 안전한 활용'을 위한 협력 방안을 논의하였는데, 글로벌 차원의 디지털 국제 규범 정립을 위한 국제사회의 연대와 국제기구 설립 추진의 필요성을 다시 한번 강조하는 한편, 우리 정부가 발표한 '디지털 권리장전'의 의미를 각국의 정상들과 공유하였다.

1. 논의 배경

[안전한 AI 설계 및 개발에 대한 규범 부존재]

❖ 2023. 3. 7개 안을 통합한 인공지능 제정법안이 국회 과방위 법안심사소위 통과
(인공지능 육성 및 신뢰기반 조성 등에 관한 법률안)

2023. 8. 24. 국가인권위원회 비판적인 의견표명


보도자료

「인공지능 법률안」, '무선허용·사후규제' 원칙 삭제하고, 인권영향평가 도입해야

2023. 3~8. 시민사회 등 비판

❖ 2023. 11.
시민사회 인공지능 책임법안 전달

❖ 2023. 11. 정부 과학기술정보통신부 수정안 전달

02

과방위 법안심사소위 통과 법안의 한계

국회 과학기술방송통신위원회 법안심사소위를 통과한 법안은 그동안 국회에 계류 중이었던 7개 법안을 통합한 안이다.

그런데, '안전한' AI기술 개발과 활용을 위하여 전세계가 뜻을 모으고 있는 추세와 달리 해당 법안은 여전히 규제를 배제한 채 인공지능 기술의 활용에 방점을 두고 있다는 비판을 많이 받았다.

그중 가장 많은 비판을 받았던 몇가지 요소를 살펴본다.

2-1. 과방위 법안심사소위 통과 법안의 한계 [AI 기술 개발 및 활용 중심]

01 소 위 안

대원칙으로 우선 허용, 사후 규제 원칙 명문화

제11조(우선허용 · 사후규제 원칙) ① 누구든지 인공지능기술 및 알고리즘의 **연구 · 개발** 및 인공지능제품 또는 인공지능서비스의 **출시 등과 관련된 행위를 할 수 있다.** 이 경우 인공지능기술, 인공지능제품 또는 인공지능서비스가 **국민의 생명 · 안전 · 권익에 위해가 되거나 공공의 안전 보장, 질서 유지 및 복리 증진을 현저히 저해할 우려가 있는 경우**가 아니라면 이를 **제한하여서는 아니 된다.**

② 국가 및 지방자치단체는 인공지능기술, 인공지능제품 또는 인공지능서비스와 관련된 **법령 및 제도가 제1항의 원칙에 부합하게 정비**되도록 노력하여야 한다.

01 반대 의견

- ✓ 소위안은 "인공지능산업 기반 조성" 제하의 절에서 산업 진흥을 위한 **우선허용, 사후규제 원칙**을 명시하고 있고, 단순히 우선허용 사후규제 원칙을 천명한 것에 그치지 않고, **관련 법령 및 제도를 원칙에 부합하도록 정비**할 것을 있어 파급력이 상당함
- ✓ 국내외에서 권장하는 인공지능 입법 기준 뿐 아니라 주요 국가의 인공지능법안의 내용은, 국가가 인공지능 제품의 안전과 인권 기준 준수 여부를 감독하고 피해를 구제하도록 요구하고 있음. 특히 유럽연합 인공지능법안은 **제품 적합성 평가 기관 또는 시장 감시와 관련된 기존의 규제기관들이 이 법의 규제를 우선 집행**하고 **인증이나 기술 부처는 그 기준을 반영하도록 규정**함. '우선 허용, 사후 규제' 원칙을 천명하는 것은 실질적으로 감독기능을 하지 않겠다는 의미로 볼 수밖에 없음
- ✓ 소위안은 동시에 효과적인 감독과 피해구제 절차를 규정하지 않았을 뿐만 아니라 원칙적으로 '현저한 우려가 있는 경우'가 아니면 제한할 수 없도록 함.
- ✓ 우선허용·사후규제 원칙으로 위해가 있을 수 있는 인공지능을 시장에 우선적으로 출시할 수 있도록 함.

2-2. 과방위 법안심사소위 통과 법안의 한계 [AI 기술 개발 및 활용 중심]

02
소 위 안

고위험인공지능 관련 규정

제2조(정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.

3. "고위험영역인공지능"이란 다음 각 목의 어느 하나에 해당하는 인공지능으로서 사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 영역에서 활용되는 인공지능을 말한다.

가. 「에너지법」 제2조제1호에 따른 에너지, 「먹는물관리법」 제3조제1호에 따른 먹는물 등의 공급을 위하여 사용되는 인공지능

나. 「보건의료기본법」 제3조제1호에 따른 보건의료의 제공 및 이용체계 등에 사용되는 인공지능

다. 「의료기기법」 제2조제1항에 따른 의료기기에 사용되는 인공지능

라. 「원자력시설 등의 방호 및 방사능 방재 대책법」 제2조제1항에 따른 핵물질과 원자력시설의 안전한 관리 및 운영을 위하여 사용되는 인공지능

마. 범죄 수사나 체포 업무에 있어 생체정보(얼굴 · 지문 · 홍채 및 손바닥 정맥 등 개인을 식별할 수 있는 신체적 · 생리적 · 행동적 특징에 관한 개인정보를 말한다)를 분석 · 활용하는 데 사용되는 인공지능

바. 채용, 대출 심사 등 개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가 목적의 인공지능

사. 「교통안전법」 제2조제1호부터 제3호까지에 따른 교통수단, 교통시설, 교통체계의 주요한 작동 및 운영에 사용되는 인공지능

아. 국가, 지방자치단체, 「공공기관의 운영에 관한 법률」에 따른 공공기관 등(이하 "국가기관등"이라 한다)이 사용하는 인공지능으로서 국민에게 영향을 미치는 의사결정을 위하여 사용되는 인공지능

자. 그 밖에 국민의 안전 · 건강 및 기본권 보호에 중대한 영향을 미치는 인공지능으로서 대통령령으로 정하는 인공지능

제26조(고위험영역 인공지능의 확인) ① 제2조제3호 각 목에 따른 인공지능 또는 그 인공지능을 이용한 제품 또는 서비스를 개발·활용·제공하려는 자는 해당 인공지능이 고위험영역 인공지능에 해당하는지에 대한 확인을 과학기술정보통신부장관에게 요청할 수 있다.

2-2. 과방위 법안심사소위 통과 법안의 한계 [AI 기술 개발 및 활용 중심]

02
반대 의견

고위험 인공지능의 정의가 자의적이며 부분적임

✓ 안전과 인권에 미치는 위험이 방지될 수 없는 수준에 달해 **금지해야 하는 인공지능**에 대해서는 전혀 규정하고 있는 바가 없음

✓ 유럽연합 인공지능법안은 인권에 미치는 위험이 높은 것으로 알려진 인공지능을 대체로 모두 법률에 규정하였으며, 그 위험도가 완화되기 어려운 분야의 경우 '금지'하는 대상으로 우선적으로 규정하고, 그 위험도가 관리될 수 있는 분야의 경우 '고위험'으로 규정함. 그런데 소위안은 위험도 존재하는 영역 중에서도 일부만 정함

① 소위안은 생체정보의 분석·활용의 경우 "범죄 수사나 체포 업무라는 일부 업무"에 한정하여 고위험으로 지정함. 그러나 유럽연합 인공지능법안은 공공장소 원격 실시간 생체 인식을 우선 원칙적으로 금지하였으며, 나머지 자연인에 대한 '실시간' 및 '사후' 원격 생체 인식을 모두 고위험으로 지정함. 특히 공공장소 원격 실시간 생체 인식은 이동의 자유, 집회시위의 자유 침해 우려 등으로 유엔인권최고대표를 비롯하여 국제인권기구 및 우리나라 국가인권위원회에서 원칙적으로 금지를 권고한 바 있음. 그러나 소위안은 한국에서도 현안이 되고 있는 공공장소 얼굴인식 등 원격 생체 인식과 추적에 대하여 별다른 제한을 하지 않았으며, 범죄 수사나 체포 업무와 무관한 생체 인식을 고위험에서 제외함.

② 소위안은 "채용, 대출 심사 등 개인의 권리·의무 관계에 중대한 영향을 미치는 판단 또는 평가 목적의 인공지능"을 고위험으로 규정하였으나, '중대한 영향'이 무엇인지 명확히 규정하지 않았으며, '채용, 대출 심사' 등 매우 자의적이고 제한적인 예시 규정을 두고 있음. 다른 나라에서 인권 침해와 차별 위험이 높은 인공지능 분야를 자세하게 규정하고 있는 것과 다르며 많은 인권 침해 분야를 고위험에서 제외함

✓ 소위안이 열거 방식으로 규정한 고위험 인공지능의 정의는 **주요 국가에서 추진하는 고위험 인공지능 규제 대상과 내용에 비하여 중요한 분야**를 누락하고 있음.

✓ 소위안은 고위험영역 인공지능 여부를 판단할 때 국가나 지방자치단체에서 제품안전 등에 대하여 판단하고 규제해 온 기관이 아니라 과학기술정보통신부가 우선적으로 확인(판단)하도록 함.

2-3. 과방위 법안심사소위 통과 법안의 한계 [AI 기술 개발 및 활용 중심]

03

소 위 안

고위험인공지능 관련 규정

제7조(위원회의 기능) 위원회는 다음 각 호의 사항을 심의 · 의결한다.
9. 고위험영역 인공지능 규율에 관한 사항
10. 고위험영역 인공지능과 관련된 사회적 변화 양상과 정책적 대응에 관한 사항

제27조(고위험영역 인공지능 고지 의무) ① 고위험영역 인공지능을 이용하여 제품 또는 서비스를 제공하려는 자는 해당 제품 또는 서비스가 고위험영역 인공지능에 기반하여 운용된다는 사실을 이용자에게 사전에 고지하여야 한다.

제28조(고위험영역 인공지능과 관련한 사업자의 책무)
① 고위험영역 인공지능을 개발하는 자 또는 고위험영역 인공지능을 사용하여 제품 또는 서비스를 제공하는 자(이하 "고위험영역 인공지능과 관련한 사업자"라 한다)는 인공지능의 신뢰성과 안전성을 확보하기 위한 조치(이하 "신뢰성 확보조치"라 한다)를 하여야 한다.
② 과학기술정보통신부장관은 제1항에 따른 신뢰성 확보조치의 구체적 내용을 정하여 위원회의 심의·의결을 거쳐 고시하고, 고위험영역 인공지능과 관련한 사업자에게 이를 준수하도록 권고할 수 있다.
③ 제2항에 따른 고시에는 다음 각 호의 사항을 포함하여야 한다.
1. 위험관리방안의 수립·운영에 관한 사항
2. 신뢰성 확보 조치 내용을 확인할 수 있는 문서의 작성과 보관에 관한 사항
3. 기술적으로 가능한 범위 내에서의 인공지능이 도출한 최종결과, 인공지능의 최종결과 도출에 활용된 주요 기준, 인공지능의 개발·활용을 위해 학습된 데이터 개요 등에 대한 설명 방안
4. 이용자 보호 방안
5. 고위험영역 인공지능에 대한 사람의 관리 · 감독에 관한 사항
6. 기타 고위험영역 인공지능의 신뢰성과 안전성 확보를 위해 필요한 사항

2-3. 과방위 법안심사소위 통과 법안의 한계 [AI 기술 개발 및 활용 중심]

03

반대 의견

고위험 인공지능 규제가 형식적이기 때문에 실질적인 위험 방지 역할을 하기 어려움

✓ 인공지능위원회와 과학기술정보통신부는 고위험 인공지능에 대한 의사결정에 있어 독립적인 조사와 조치 권한을 가지고 있는 기존 규제기관의 판단에 대하여 중복적이거나 충돌되는 의사결정을 내릴 위험이 있음

✓ 소위안은 고위험 인공지능 사업자의 책무를 일부 규정하였으나, 고위험이라는 사실을 이용자에게 고지하도록 하고(제27조) 대부분 내부적인 '방안'을 마련하는데 그침(제28조). 게다가 이에 대한 **준수 여부는 아무런 벌칙이나 조치 사항 없이 기업 자율에 맡기고 있음**. 문제가 확인된 고위험 인공지능의 시판 및 이용을 방지하기 위하여 향후 **구체적인 위험 방지 조치의 내용이나 준수 사항을 규정하는 바가 없음**

✓ 인공지능의 **투명성과 관련한 설명의무의 범위에 대하여 "기술적으로 가능한 범위 내에서" 내부적인 '설명방안'을 마련하면 족하도록 규정하였고, 인공지능 위험성에서 가장 논란이 되고 있는 데이터 편향이나 오류를 방지할 의무를 규정하고 있지 않음**

✓ 고위험 인공지능 시스템의 개발이나 활용하는 사업자에 대해서는 위험 영향을 평가하거나 완화할 의무, 출시 전 검사하거나 사후에 모니터링할 의무, 개발이나 운영 중 문서화하거나 기록할 의무, 데이터 편향이나 오류를 방지할 의무, 작동에 대해 투명하게 설명할 의무, 인간이 관리감독할 의무, 시스템의 견고성·정확성·보안성, 인증·등록·보고 의무 등이 세세하게 규정되어야 마땅함

2-4. 과방위 법안심사소위 통과 법안의 한계 [AI 기술 개발 및 활용 중심]

04 기타 반대 의견

과학기술 정보통신부가 인공지능 관련 사회 정책 일반을 소관하는 것에 반대함

✓ 소위안이 규정하는 인공지능윤리, 신뢰성 및 인공지능 사회에 대한 정의가 매우 모호하고 광범위하여, 이 법을 소관하는 과학기술정보통신부가 인공지능과 사회 정책 일반에 대하여 폭넓은 권한을 부여하는 것으로 귀결됨

✓ 하지만 소위안이 정의하고 소관하는 인공지능 '윤리'와 '신뢰'에 대한 정책은, 현재 공정거래위원회, 고용노동부, 국가인권위원회, 개인정보보호위원회, 산업통상자원부, 행정안전부 등 국가 및 지방자치단체의 다른 규제기관에서 소관하는 제품 안전, 소비자 보호, 인권, 차별, 개인정보 보호 분야까지 모두 포괄할 수 있는 매우 광범위한 분야를 다루고 있음

✓ 과학기술정보통신부가 '윤리', '신뢰성', '인공지능사회와관련된 제도 개선 및 권고' 관련 역할을 실질적으로 수행하기에 한계가 있고, 정책 관련 논의를 할 때는 반드시 이를 수행할 수 있는 정부부처가 참여할 필요가 있음

기술의 개발과 활용에 중점을 둔 법안의 중복 제정에 반대함

✓ 의도와 무관하게 소위안은 기술의 개발과 활용을 촉진하기 위한 내용이 주를 이룸 그런데 인공지능 기술과 산업에 대한 기본법으로는 같은 부처 소관으로 「지능정보화기본법」이 이미 시행 중이고, 이미 시행 중인 지능정보화 기본법의 목적과 소관하는 내용 대부분이 유사하거나 중복됨.

✓ 국민은 '안전한 인공지능'을 요구하는 반면, 국회는 기술의 개발과 활용을 중복으로 보장하는 법률만 반복적으로 양산하는 모양새임. 이는 '국민의 권익과 존엄을 보호' 한다는 법의 제정 취지에도 맞지 않음

03

시민사회 인공지능책임법안

진보네트워크, 참여연대, 민변 등이 '안전한 인공지능'을 위하여 법안에 필수적으로 포함되어야 하는 부분을 중심으로 '인공지능책임법(안)을 구상하여 보았다.

안전한 인공지능을 위해 필요한 요소가 무엇인지와 시민사회 입법안 중요 내용을 살펴본다.

3. 시민사회 인공지능 책임법(안)

[필수적인 요소]

금지되는 인공지능

- 국민의 **생명, 안전 인권에 미치는 해악이 커 인공지능사회의 책임기반을 해치는 인공지능**을 규정하고 그 제공과 활용을 금지할 필요성이 있음

고위험 영역 인공지능규제

- 금지에는 이르지 않았지만 국민의 안전과 인권이 미치는 **영향력이 커 위험성을 통제할 필요성이 큰 인공지능을 고위험 영역 인공지능**으로 정의함
- 고위험 영역 인공지능의 경우 이를 제공하는 자와 활용하는 자에게 의무를 부과하여 위험성을 통제할 수 있도록 제도를 정비할 필요가 있음

- 인공지능을 활용하려는 공공기관과 고위험 영역 인공지능 활용자는 활용하려는 **인공지능이 인권에 미칠 인권위험을 사전에 평가**하여 대책을 마련할 필요가 있음

인권영향평가

- 안전성과 신뢰성이 확보된 인공지능 기술을 기반으로 한 인공지능 산업의 진흥 및 책임확보에 관한 **사무를 독립적으로 수행**하기 위하여 국무총리 소속 인공지능위원회를 둘 필요가 있음

인공지능위원회

3-1. 시민사회 인공지능 책임법(안)

[금지되는 인공지능 정의 등]

금지되는 인공지능 활용 금지

제29조 (인공지능 개발과 활용의 금지)

① 인간의 존엄성을 침해하여 인간을 수단이나 도구로 사용하기 위한 목적이나 헌법에서 정한 기본권의 본질적 내용을 침해할 목적으로 인공지능을 개발하거나 활용할 수 없다. 또한 누구든지 다음 각 호의 사항에 해당하는 인공지능을 개발하거나 활용할 수 없다.

1. 성별·연령·장애·지역·인종·종교·국가 등 개인의 특성에 따른 차별을 목적으로 하는 인공지능
2. 사회적 약자 및 취약 계층에 속한 사람의 행동을 중대하게 왜곡하려는 의도로 이들의 생명, 신체, 재산 등을 침해하는 인공지능
3. 인간의 심리, 사고, 행동 등을 왜곡하기 위한 의도로 개인의 잠재의식에 영향을 미치는 인공지능
4. 특정된 개인 또는 집단을 차별하거나 불이익을 줄 의도로 인간의 사회적 행동, 인격적 특성에 기초하여 개인의 신뢰도를 평가하거나 분류하는 인공지능
5. 공공장소에서 실시간 원격 신원확인을 위한 목적의 인공지능
6. 인간의 감정을 인식하거나 예측하여 개인의 권리와 의무에 중대한 영향을 미치기 위한 목적의 인공지능
7. 인간의 실질적 통제 없이 무기를 운용할 목적의 인공지능
8. 인간의 실질적 통제 없이 운용되는 예측 치안 및 경찰직무집행을 목적으로 하는 인공지능
9. 기타 인간의 존엄을 침해할 우려가 있는 인공지능으로서 대통령령으로 정하는 인공지능

② 인공지능 제공자는 제1항 각호에 해당하지 않는 인공지능을 개발 또는 제공하는 과정에서 해당 인공지능이 제1항 각호에 해당하게 되는 경우 즉시 해당 인공지능의 개발 또는 제공을 중단하여야 한다.

③ 인공지능 활용자는 인공지능 시스템이 제1항 각호에 해당하는 인공지능에 해당함을 알게 된 경우에는 즉시 활용을 중단하고 관계기관에 신고하여야 한다

3-1. 시민사회 인공지능 책임법(안) [금지되는 인공지능 정의 등]

금지되는 인공지능 활용금지

제14조(인공지능위원회의 심의 · 의결 사항 등) ① 인공지능위원회는 다음 각 호의 사항을 심의 · 의결한다.
7. 제29조 제1항에 따른 금지되는 인공지능의 확인에 관한 사항

제46조(벌칙) 다음 각 호의 어느 하나에 해당하는 자는 10년 이하의 징역 또는 1억 원 이하의 벌금에 처한다.
1. 제29조 제1항을 위반하여 개발과 활용이 금지된 인공지능을 개발 및 활용한 자

제47조(벌칙) 다음 각 호의 어느 하나에 해당하는 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처한다.
4. 제29조 제2항을 위반한 자

시민사회 의견

- 인간의 존엄성을 침해하여 인간을 수단이나 도구로 사용하기 위한 목적이나, 헌법에서 정한 기본권의 본질적 내용을 침해할 목적으로 인공지능을 개발하거나 활용할 수 없음을 확인함
- 인공지능을 개발 또는 제공하는 과정에서 금지된 인공지능이라는 점이 확인된 경우 개발 또는 제공을 즉시 중단하고, 인공지능 활용자는 사후 인공지능이 금지된 인공지능에 해당한다는 사실을 알게 된 경우 즉시 활용을 중단하고 관계기관에 신고하여야 함을 명시함
- 금지된 인공지능을 개발 및 활용한 자는 10년 이하의 징역 또는 1억원 이하의 벌금에 처하고, 개발 또는 제공 과정에서 금지된 인공지능임을 알았음에도 개발 또는 제공을 중단하지 않은 경우 3년 이하의 징역 또는 3천만원 이하의 벌금에 처함
- 금지된 인공지능에 해당하는지 여부는 인공지능위원회 심의·의결을 통해 확인 받을 수 있음

3-2. 시민사회 인공지능 책임법(안) [고위험 인공지능 정의 및 규제]

정의

제2조 (정의) 이 법에서 사용하는 용어의 뜻은 다음과 같다.
2. **"고위험영역 인공지능"**이란 다음 각 목의 어느 하나에 해당하는 인공지능으로서 사람의 생명, 신체의 안전 및 기본권을 침해할 우려가 있는 영역에서 활용되는 인공지능을 말한다.
가.「에너지법」 제2조제 1호에 따른 에너지, 「먹는물관리법」 제3조 제1호에 따른 먹는물 등의 공급을 위하여 사용되는 인공지능
나.「보건의료기본법」 제3조제1호에 따른 보건의료의 제공 및 이용체계 등에 사용되는 인공지능
다.「의료기기법」 제2조제1항에 따른 의료기기에 사용되는 인공지능
라.「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제2조 제1호에 따른 정보통신망 내에서 사용되는 인공지능
마. 생체정보(얼굴·지문·홍채 및 손바닥 정맥 등 개인을 식별할 수 있는 신체적·생리적·행동적 특징에 관한 개인정보를 말한다)를 분석, 활용하는 데 사용되는 인공지능
바. 수사, 기소, 형집행 업무에 사용되는 인공지능
사. 사법기관, 행정기관(위탁기관까지 포함)의 소송 및 조정, 중재, 화해 등 업무처리에 사용되는 인공지능
아. 근로자 및 노무제공자(산업재해보상보험법 제91조의 15)에 관하여, 채용 과정에서의 지원자 평가, 승진과 해고의 결정, 인사평가, 직무배치, 업무 할당의 결정 등에 사용되는 인공지능
자. 대통령령으로 정하는 유해하거나 위험한 기계, 기구, 설비에 사용되는 인공지능
차. 어린이제품 안전 특별법」 제2조 제9호에 따른 안전인증 대상 어린이 제품의 전체 또는 부분이 인공지능인 경우
카. 사회보험, 공공부조, 사회서비스 등 혜택의 수급 자격의 평가, 부여 등에 사용하는 인공지능
타. 이주, 난민, 출입국 관리에 사용하는 인공지능
파. 군 또는 정보기관에서 첩보, 방첩 업무에 사용하는 인공지능
하. 교육 및 직업훈련과 관련된 기회 제공 및 자원 할당업무에 사용하는 인공지능
거. 금융 및 보험 분야에 사용하는 인공지능
너. 그 밖에 국민의 안전·건강 및 기본권을 침해할 우려가 있는 인공지능으로서 대통령령으로 정하는 인공지능

3-2. 시민사회 인공지능 책임법(안)

[고위험 인공지능 정의 및 규제]

규제

제33조(고위험영역 인공지능 기술문서 작성 및 유지 의무)

① 고위험영역 인공지능 제공자는 인공지능 시스템의 주요 요소와 개발과정의 상세, 시스템 변경사항, 인공지능 개발에 사용된 데이터 및 데이터의 위험성 완화를 위한 조치 등에 대하여 주기적으로 문건을 기록하여 작성하고, 보관하여야 한다.

② 고위험영역 인공지능 제공자는 시스템이 자동으로 생성하는 로그를 보존하여야 한다. (이하 생략)

제34조(데이터 평가 및 위험 관리)

① 고위험영역 인공지능 제공자는 제품 개발 또는 출시 이전과 이후 시점에 주기적으로 인공지능의 학습데이터, 검증데이터, 테스트 데이터 및 그 결과물로 인한 차별 및 기본권 침해 위험에 대하여 평가하고, 그 위험을 제거 또는 완화하기 위한 조치(이하 '위험성 완화조치'라 한다)를 하여야 한다.

② 인공지능위원회는 제1항에 따른 위험성 완화 조치의 구체적 내용을 정하여 고시하고, 고위험영역 인공지능 제공자에게 이를 준수하도록 권고할 수 있다. (이하 생략)

제35조(적합성 인증기관의 지정 등)

① 인공지능위원회는 고위험영역 인공지능 제품 또는 서비스의 적합성 인증업무를 수행하는 기관을 지정할 수 있다.

② 인공지능위원회는 제1항에 따라 지정받은 기관(이하 '인증기관'이라 한다)에 적합성 인증업무를 수행하는 데 필요한 지원을 할 수 있다. (이하 생략)

제36조(적합성 인증)

① 고위험영역 인공지능 제공자는 인공지능 제품 또는 서비스의 출시 전에 제34조 제1항의 위험성 완화조치를 포함한 일련의 조치가 이 법에 부합하는지 등에 관하여 적합성 인증을 받아야 한다.

② 제1항에 따른 적합성 인증의 유효기간은 2년으로 한다.

③ (중략)

⑦ 고위험영역 인공지능 제공자는 적합성 인증을 받은 후 대상 제품 또는 서비스에 대하여 대통령령으로 정하는 바에 따라 기본권 침해 또는 차별 가능성이 없는지 자체 검사를 하여야 한다. (이하 생략)

3-2. 시민사회 인공지능 책임법(안)

[고위험 인공지능 정의 및 규제]

규제

제44조(과징금의 부과)

② 인공지능위원회는 다음 각호의 어느 하나에 해당하는 경우 해당 고위험영역 인공지능 제공자에게 직전 연도 재무상태표상 전체 매출액의 100분의3를 초과하지 아니하는 범위에서 과징금을 부과할 수 있다. 다만, 매출액이 없거나, 매출액의 산정이 곤란한 경우로서 대통령령으로 정하는 경우에는 20억원을 초과하지 아니하는 범위에서 과징금을 부과할 수 있다.

1. 제36조 제1항을 위반하여 적합성 인증을 받지 아니하고 고위험영역 인공지능 제품 또는 서비스를 출시한 자

제47조(벌칙) 다음 각 호의 어느 하나에 해당하는 자는 3년 이하의 징역 또는 3천만원 이하의 벌금에 처한다.

5. 제36조 제1항을 위반하여 적합성 인증을 받지 아니하고 고위험영역 인공지능 제품 또는 서비스를 출시한 자
6. 제36조 제1항을 위반하여 거짓 또는 그밖의 부정한 방법으로 적합성 인증을 받은 자
7. 제36조 제4항을 위반하여 적합성 인증의 기준에 적합하지 않음에도 적합성 인증을 한 자(이하 생략)

시민사회

고위험 영역 인공지능은 그 내재된 위험성에 비추어 **① 고위험 영역 인공지능이 포함되었다는 사실을 활용자에게 고지하고(정보제공의무), ② 개발에 사용된 데이터, 데이터의 위험성 완화 조치에 관한 문서, 로그 기록 등 기술문서를 작성하고 유지할 의무가 있으며, ③ 제품 개발 및 출시 전후 위험성 완화조치를 하고 적합성 인증을 받아야 함. 또한, 이하 살펴볼 ④ 인공지능 활용자는 사용 전 인권영향평가를 받아야 함**

만약, 적합성 인증을 받지 않고 고위험 영역 인공지능 제품과 서비스를 출시한 경우 인공지능위원회는 직전 연도 재무상태표상 전체 매출액의 100분의 3의 범위에서 과징금을 부과할 수 있고, 3년 이하의 징역 또는 3천만원 이하에 벌금에 처해짐
거짓 또는 부정한 방법으로 적합성 인증을 받고, 적합성 인증 기준에 적합하지 않음에도 인증을 한 경우에도 3년 이하의 징역 또는 3천만원 이하에 벌금에 처해짐

3-3. 시민사회 인공지능 책임법(안) [인권 영향평가]

규정

제38조 인공지능 인권영향평가

① 인공지능을 활용하려는 공공기관과 고위험영역 인공지능 활용자는 그 사용 전에 인권에 미치는 부정적 영향의 분석 · 평가(이하 "인공지능 인권영향평가"라 한다)와 개선 사항 도출을 위한 평가를 하여야 하며, 다음 사항을 고려하여 현존하거나 잠재하는 위험(이하 '인권위험'이라 한다)에 대한 식별을 포함하여야 한다.

가. 인공지능을 사용하는 목적, 적용 범위 및 그 사용이 예상 또는 의도된 상황

나. 학습·테스트·검증에 사용된 데이터의 특성과 양

다. 알고리즘 실행 결과물의 목적에 부합하는 신뢰성, 안전성 및 정확성, 특히 영향을 받는 인구집단별 편향적인 결과물의 생성 여부

라. 인공지능의 데이터, 알고리즘, 결과물의 투명성, 설명가능성 및 문서화 정도

마. 인공지능의 자동화된 의사결정과 인간 개입의 수준

바. 인공지능 사용으로 인해 예상되는 인권 침해와 차별의 심각성과 발생 개연성

사. 인공지능 사용으로 인해 영향을 받는 개인 또는 집단의 규모, 특히 소외되거나 취약한 개인 또는 집단에 미치는 부정적 영향 또는 위험

아. 인권침해 가능성과 차별적 영향이 확인되었을 경우 그에 대한 방지 및 완화 방안

자. 영향을 받는 개인 및 단체가 이용할 수 있는 이의제기 및 구제 방안

차. 기타 대통령령으로 정하는 사항

3-3. 시민사회 인공지능 책임법(안) [인권 영향평가]

규정

제38조 인공지능 인권영향평가

② 인공지능 인권영향평가는 데이터·알고리즘 등 해당 인공지능으로부터 직무상 독립적이고 전문적인 자격을 갖춘 사람 또는 기관이 실시하여야 한다.

③ 인공지능 인권영향평가는 영향을 받을 가능성이 있는 개인 또는 집단 및 그 대리인의 의견을 수렴하여야 한다.

④ 인공지능을 활용하려는 공공기관과 고위험영역 인공지능 활용자는 인공지능 인권영향평가에서 현존하거나 잠재하는 인권위험을 식별한 경우에는 해당 활용자는 지체 없이 위험성을 방지, 완화하고 영향을 받은 사람들의 권리를 보호, 구제할 수 있는 대책을 수립하고 실행하여야 한다.

⑤ 인공지능을 활용하려는 공공기관과 고위험영역 인공지능 활용자는 인공지능 인권영향평가에서 식별한 현존하거나 잠재하는 인권위험에 대한 대책을 수립하고 실행할 것이 어려운 경우에는 그 사용을 중단하고 이 사실을 제공자와 인공지능위원회에 알려야 한다.

⑥ 인공지능을 활용하려는 공공기관과 고위험영역 인공지능 활용자는 인공지능 인권영향평가를 실시한 때로부터 인공지능의 기능 또는 활용 범위 및 대상이 중대하게 변경되었을 경우 해당 평가를 갱신하여야 한다.

⑦ 인공지능을 활용하려는 공공기관과 고위험영역 인공지능 활용자는 인권영향평가와 병행하여 개인정보보호법 제33조에 따라 개인정보 영향평가를 실시한다. 이때 해당 조항에 따른 평가의무자가 아니더라도 개인정보 영향평가를 실시하여야 한다.

⑧ 인공지능을 활용하려는 공공기관과 고위험영역 인공지능 활용자는 인공지능 인권영향평가와 개인정보 영향평가가 완료된 날로부터 3개월 이내에 제4항에 따른 대책의 수립 및 실행 결과가 포함된 평가의 요약 보고서를 일반에 공개하여야 한다. 또한 평가가 완료된 날로부터 3개월 이내에 완전한 보고서를 인권영향평가의 경우 인공지능위원회에, 개인정보 영향평가의 경우 개인정보보호인공지능위원회에 제출하여야 한다.

⑨ 인공지능위원회는 인공지능 인권영향평가의 지침 · 표준 마련, 평가기준 및 평가지표 설정 등을 위하여 국가인권위원회의 의견을 듣고, 협의하여야 한다. (이하 생략)

3-3. 시민사회 인공지능 책임법(안) [인권 영향평가]

시민 사회 의견

✓ 인공지능 인권영향평가는 인공지능이 인권에 미치는 부정적인 영향을 분석·평가하는 도구임

✓ 주된 의무주체는 인공지능을 활용하려는 공공기관, 고위험영역 인공지능 활용자인데,
① 직무상 독립적이고 전문적인 자격을 갖춘 사람 또는 기관을 통해 인공지능 사용 이전에 인권영향평가를 하여야 하고,
② 과정에서 영향을 받을 가능성이 있는 개인 또는 집단 및 대리인의 의견을 수렴해야 하며,
③ 인권위험을 식별한 경우 지체없이 대책을 수립·실행하고, 그러한 대책 수립·실행이 어려운 경우에는 인공지능 사용을 중단하고 인공지능 위원회에 그 사실을 알려야 함
④ 또한, 인공지능 기능 또는 활용 범위 및 대상이 중대하게 변경되었을 경우 해당 평가를 갱신하고, 인권영향평가를 완료한 주체는 요약 보고서를 일반에 공개하고 인공지능위원회와 개인정보보호위원회 제출하여야 함

✓ 인공지능 인권영향평가에 관한 업무는 기본적으로 인공지능위원회가 수행하되, 인권 영향평가 기준 지표 설정·수정을 할 때는 국가인권위원회의 의견을 듣도록 하여 자의적인 해석을 방지하고, 전문성을 확보하도록 함

3-4. 시민사회 인공지능 책임법(안) [인공지능위원회]

규정

제6조(인공지능위원회) ① 안전성과 신뢰성이 확보된 인공지능기술을 기반으로 한 인공지능산업의 진흥 및 책임 확보에 관한 사무를 독립적으로 수행하기 위하여 국무총리 소속으로 인공지능위원회를 둔다.

제7조(인공지능위원회의 구성 등) ① 인공지능위원회는 상임위원 2명(위원장 1명, 부위원장 1명)을 포함한 **9명의 위원**으로 구성한다.
② 인공지능위원회의 위원은 인공지능에 관한 경력과 전문지식이 풍부한 다음 각 호의 사람 중에서 위원장과 부위원장은 국무총리의 제청으로, 그 외 위원 중 2명은 위원장의 제청으로, 2명은 대통령이 소속되거나 소속되었던 정당의 교섭단체 추천으로, 3명은 그 외의 교섭단체 추천으로 대통령이 임명 또는 위촉한다.
1. 판사 · 검사 · 변호사의 직에 10년 이상 있거나 있었던 사람
2. 공공기관 또는 단체(사업자단체, 소비자단체 및 인권옹호단체를 포함한다)에 3년 이상 임원으로 재직하였거나 이들 기관 또는 단체로부터 추천받은 사람으로서 안전, 인권 및 차별에 관한 업무를 3년 이상 담당하였던 사람
3. 인공지능기술, 안전, 인권 및 차별 관련 분야에 전문지식이 있고 「고등교육법」 제2조제1호에 따른 학교에서 부교수 이상으로 5년 이상 재직하고 있거나 재직하였던 사람

제13조(인공지능위원회의 소관 사무) ① 인공지능위원회는 다음 각 호의 소관 사무를 수행한다.
1. 인공지능과 관련된 법령의 개선에 관한 사항
2. 인공지능과 관련된 정책 · 제도 · 계획 수립 · 집행에 관한 사항
3. 인공지능 인권영향평가의 지침 · 표준 마련, 평가기준 및 평가지표 설정 등에 관한 사항
4. 인공지능 관련 국제기구 및 외국 기관와의 교류 · 협력
5. 인공지능에 관한 법령 · 정책 · 제도 · 실태 등의 조사 · 연구, 교육 및 홍보에 관한 사항
6. 인공지능 기술 개발의 지원 · 보급, 기술의 표준화 및 전문인력의 양성에 관한 사항
7. 이 법 및 다른 법령에 따라 인공지능위원회의 사무로 규정된 사항

② 인공지능위원회가 인지하거나 접수한 다음 각 호 사무의 경우 다른 국가기관이 권한을 가진 경우 해당 사안을 지체없이 해당 기관으로 이송하여야 한다. 다만 제14조 제1항 제5호 내지 제6호에 따른 심의 · 의결을 거치는 경우 인공지능위원회 소관 사무로서 수행한다.
1. 인공지능의 중대한 안전 위해, 인권 침해 및 차별에 대한 조사 및 이에 따른 처분에 관한 사항
2. 인공지능의 중대한 안전 위해, 인권 침해 및 차별과 관련한 진정처리 및 권리구제에 관한 사항

3-4. 시민사회 인공지능 책임법(안)

[인공지능위원회]

규정

제14조(인공지능위원회의 심의 · 의결 사항 등) ① 인공지능위원회는 다음 각 호의 사항을 심의 · 의결한다.

1. 제23조에 따른 기본계획에 관한 사항
2. 인공지능과 관련된 정책, 제도 및 법령의 개선에 관한 사항
3. 인공지능에 관한 공공기관 간의 의견조정에 관한 사항
4. 인공지능에 관한 법령의 해석 · 운용에 관한 사항
5. 인공지능의 중대한 안전 위해, 인권 침해 및 차별에 관한 조사 및 이에 따른 처분에 관한 사항
6. 인공지능의 중대한 안전 위해, 인권 침해 및 차별에 관한 진정처리 및 권리구제에 관한 사항
7. 제29조 제1항에 따른 금지되는 인공지능의 확인에 관한 사항
8. 제2조 제2호 각 목에 따른 고위험영역 인공지능의 확인 및 제30조 면제에 관한 사항
9. 제34조에 따른 위험성 완화 조치에 관한 사항
10. 제38조에 따른 인권영향평가에 관한 사항
11. 소관 법령 및 인공지능위원회 규칙의 제정 · 개정 및 폐지에 관한 사항
12. 인공지능과 관련하여 인공지능위원회의 위원장 또는 위원 2명 이상이 회의에 부치는 사항
13. 그 밖에 이 법 또는 다른 법령에 따라 인공지능위원회가 심의 · 의결하는 사항

② 인공지능위원회는 제1항 각 호의 사항을 심의 · 의결하기 위하여 필요한 경우 다음 **각 호의 조치**를 할 수 있다.

1. 관계 공무원, 인공지능 기술에 관한 전문 지식이 있는 사람이나 시민사회단체 및 관련 사업자로부터의 의견 청취
2. 관계 기관 등에 대한 자료제출이나 사실조회 요구

③ 제2항제2호에 따른 요구를 받은 관계 기관 등은 특별한 사정이 없으면 이에 따라야 한다.

④ 인공지능위원회는 제1항제2호의 사항을 심의 · 의결한 경우에는 관계 기관에 그 개선을 권고할 수 있다.

⑤ 인공지능위원회는 제4항에 따른 권고 내용의 이행 여부를 점검할 수 있다.

시민사회 의견

- ✓ 과학기술정보통신부 산하 인공지능위원회를 둘 경우 '인공지능 인권영향평가', ' 인공지능으로 인한 인권 침해 및 차별에 관한 조사 및 진정처리' 등 업무를 수행할 수 없거나 형식적으로 이루어질 염려가 있음
- ✓ 이에 인공지능과 관련된 업무를 독립적·유기적으로 처리할 수 있도록 독립기관을 두되, 독립성을 확보할 수 있도록 조직을 국무 총리 산하 기관으로 상정함. 안전, 인권 및 차별에 관한 업무를 3년 이상 담당하였던 사람, 인공지능기술, 안전, 인권 및 차별 관련 분야에 전문지식는 사람과 같이 자격을 다양화하여 인공지능 위원회가 실질적으로 역할을 할 수 있도록 정함

04

정부 수정안의 한계

정부 과학기술정보통신부는 20233. 11. 기존 인공지능법(안)을 수정한 안을 국회에 제출하였다.

하지만 해당 법(안)도 기존 법안심사소위를 통과한 법안과 동일하게 한계가 존재한다.

핵심적인 한계점을 살펴본다.

4. 수정안의 한계 [비판점]

규제 공백

✓ 금지된 인공지능 규제 부존재

EU 인공지능법도 '금지된 인공지능' 관련 규정을 두어 인간에게 돌이킬 수 없는 악영향을 미치는 특정 영역에서 AI 기술 개발 및 활용을 금지

기술 개발을 무조건적으로 허용하는 것이 기술 개발을 장려하고 촉진하는 일인지, 근본적인 성찰이 필요

✓ 인권영향평가에 대한 규정 부존재

소위안은 인권영향평가에 대한 규정이 부존재하였고, 이에 대하여 국가인권위원회가 도입을 권고하였으나, 여전히 관련 내용 부존재

01

고위험 인공지능 규제 한계

✓ 시민사회는 기존 소위안의 협소한 고위험 인공지능 범위에 대하여 강력하게 비판하였으나, 수정안은 기존 소위안에 대한 비판의견을 전혀 수용하지 않음

✓ 국가인권위원회는 "*유럽연합은 인공지능 법안에서 형사사법 분야에서 재범 위험 또는 범죄의 잠재적 피해 위험을 평가하기 위하여 사용하는 인공지능, 거짓말 탐지기 및 유사한 도구로 사용하거나 감정 상태를 감지하기 위하여 사용하는 인공지능, 딥페이크 감지를 위하여 사용하는 인공지능, 증거의 신뢰성을 평가하기 위하여 사용하는 인공지능, 프로파일링을 기반으로 범죄행위의 발생 또는 재발을 예측하거나 과거의 범죄 행위를 평가하는데 사용하는 인공지능 등을 소위험 인공지능으로 규정하고 있다.*"라며 소위안이 이미 사회 전반에 문제가 된 인공지능조차 규제하지 못하고 있다는 취지의 지적을 하였음

02

피해자 구제제도 부존재

✓ 소위안은 피해자에 대한 규제제도를 명문화하지 않음

✓ 국가인권위원회는 "*인공지능 기술은 다양한 분야에서 활용되므로, 권리 침해 역시 다양한 형태로 발생할 수 있다. 따라서 인공지능으로 인해 발생할 수 있는 권리 침해 사안을 유형별로 분석·분류하고, 이미 제도화되어 있는 구제 절차 중 각 유형별로 이를 처리하기 적합한 구제 절차를 이행할 수 있도록 인공지능 법률안에 근거 규정을 마련하여야 하며 관련 법률 역시 이에 맞게 정비하여야 한다.*"라고 권고함

✓ 2023. 6. 유럽의회안에도 '국가감독기구 ' 를 강화하였고 '권리구제 ' 장을 신설하여 피해자가 감독기관에 진정을 접수할 수 있도록 하였음

03

실효성 없는 고지의무

✓ 소위안은 '고위험 인공지능 여부 ' 를 고지하도록 규정한 반면 수정안은 '생성형인공지능'과 '고위험 인공지능 ' 으로 범위를 확대함

✓ 그러나 고지하지 않았을 경우 이에 대한 과태료, 벌칙 규정이 존재하지 않아 실효성이 없다는 평가를 받을 수밖에 없음

✓ 또한, 이번에 새롭게 규정한 '생성형 인공지능 ' 의 경우에도 유럽연합 의회안은 '생성형 인공지능에 국한되지 않고, 챗봇, 감정인식, 생체분류까지 포괄하고 있음. 즉, 생성형 고지의무 대상을 확대할 필요가 있음

04

4. 수정안의 한계 [비판점]

인공지능 관련 논의는 과학기술정보통신부가 아닌 새로운 기관이 신설되어야 함

✓ 과학기술정보통신부는 오랜 기간 과학기술 관련 산업의 진흥을 목적으로 각종 정책을 수행하여 왔기에 AI기술과 관련해서도 규제에 매우 소극적인 입장으로 일관하고 있음. 그러나 이번 AI '안전' 정상회담에서 확인한 것과 같이 AI기술은 이제 더 이상 기술 개발과 활용이 아니라 어떻게 규제해야 하는지에 대한 논의를 하여야 하고 같은 취지에서 시민사회도 과학기술정보통신부 중심의 정책 입안에 한계를 여러 차례 지적하였음

✓ 국가인권위원회도 "*인공지능으로 인한 인권침해, 차별 등의 문제를 실효성 있게 예방하고 감독하기 위해서는 인공지능 감독·규제 업무는 인공지능 산업 진흥에 관한 사항을 소관하는 기관이 아닌 제3의 기관이 독립적으로 수행하여야 한다.*"라고 지적한바 있음

✓ 유럽연합 인공지능 법안은 '국가감독기관 ' 을 별도로 설립하되 각 분야별로 기존에 존재하는 기구들과 협력하도록 정하고 있음

✓ 수정안은 이러한 비판과 대안을 전혀 반영하지 않은 채 소위안의 내용을 그대로 확인하고 있으며, 이러한 한계점도 과학기술정보통신부 중심으로 관련 법안 제정을 추진하고 있기 때문으로 분석됨

<끝>

【 토론 1 】

박한희
(공익인권변호사모임 희망을만드는법, 변호사)

차별금지와 평등실현을 위한 인공지능 규율 필요성

1. 인공지능과 차별

2020년 12월 스캐터랩의 챗봇 이루다는 공개된 직후에 인공지능의 차별 문제에 대한 사회적인 논쟁을 불러왔다. 지하철 임산부석, 미투에 대한 질문을 하면 '싫다, 혐오스럽다'는 답변, 여성전용헬스장에 대한 질문에 '시러, 거기 여자들 다 줘패고 싶을 듯'이라는 답변, 성소수자를 싫어하냐는 질문에 '싫다, 혐오한다'고 답변, 장애인처럼 한다'는 이야기를 하거나, 만일 장애인이면 어떠냐는 질문에 죽어야지라고 답변 등, 이루다가 보여준 차별과 혐오의 모습은 여러 사람들에게 충격을 안겨줬다.[99] 이후 스캐터랩은 2022년 10월 이루다 2.0을 재출시했고, 이루다 2.0에서는 아직까지 소수자에 대한 혐오발언이 드러나지는 않고 있다.[100] 그럼에도 이루다가 남긴 과제인 인공지능이 차별과 혐오를 학습하고 확산하는 것을 어떻게 방지할 수 있는지에 대해서는 여전히 더 많은 논의가 필요하다. 이와 관련하여 지금까지 알려진 인공지능에 의한 차별사례들을 몇 가지 살펴보면 다음과 같다.

【사례 ①】 형사 재범 위험성 프로그램의 인종차별[101]

- 미국 위스콘신주 대법원은 2016년 피고인의 재범 위험성을 평가할 때 참고하는 콤파스(COMPAS) 알고리즘의 평가지수가 법원 결정의 유일한 요소가 되었다면 위법이지만, 보조적인 수단으로 사용되는 경우 적법절차 위반이 아니라고 판결함
- 그러나 언론사 프로퍼블리카에서 2013년부터 2014년까지 콤파스 알고리즘에 의해 법원의 결정이 이루어진 피고인 1200명의 기록을 검증한 결과, 재범률이 높은 것으로 예

99) 성희롱·혐오논란에 3주만에 멈춘 '이루다'…AI윤리 숙제 남기다, 연합뉴스(2021.1.11.) https://www.yna.co.kr/view/AKR20210111155153017
100) "카톡도 제쳤다"…'혐오 논란' 이루다 어떻게 바뀌었길래 [조아라의 IT's fun], 한국경제(2022.11.11.) https://www.hankyung.com/it/article/202211113349g

측되었지만 실제로 2년간 범죄를 저지르지 않은 경우가 흑인의 경우 45%, 백인의 경우는 23.5%이었던 반면, 재범률이 낮은 것으로 예측되었지만 실제로 2년간 범죄를 저지른 경우가 백인이 48%로 흑인 28%보다 훨씬 높았던 것으로 드러남

【사례 ②】 채용 인공지능의 학습 데이터와 성차별

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

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

【사례 ③】 여성으로 설정된 AI 비서와 성별, 성적지향 편견 발언[102)]

■ 국내의 KT, SK텔레콤 등 통신사에서는 알고리즘 비서 서비스를 제공하고 있는데, 이들 AI 비서가 성차별적 편견을 강화하는 문제가 제기됨

■ KT의 「기가지니」와 SK 텔레콤의 「누구」는 스스로를 '여성'으로 규정하고, 기가지니는 '보시다시피 아리따운 여자랍니다', '전 샤방샤방한 핑크색이 좋아요', '제가 여자라서 그런지 자동차에 관심이 없어요' 라며 성차별적 인식을 드러냄. '나는 레즈비언이야'라는 이야기에도 '그런 말씀을 하시면 너무 슬퍼요'라고 대답함

【사례 ④】 편향된 데이터로 인한 흑인에 대한 의료 차별[103)]

■ 미국의 민영의료보험사인 유나이티드헬스그룹은 과거의 병력 및 진단 결과와 잠재적인 건강 위험을 예측하여 질병 가능성이 높은 사람에게 우선적으로 의료 서비스를 제공하기 위한 알고리즘으로 임팩트 프로를 사용함. 그런데 실제 의료기록에 따르면 당뇨병, 빈혈, 신부전, 고혈압 등의 질병 위험은 흑인이 비흑인보다 높았지만, 건강위험 점수는 흑인이 비흑인에 비해 10점 낮게 나타남.

■ 임팩트 프로는 인종, 피부색 변수는 사용하지 않았으나 대리 변수로 의료비를 사용함. 그리고 흑인이 의료 서비스를 받을 때 사용한 연간 의료비는 다른 인종에 비해 평균 1,800달러 적었기 때문에 결과적으로 고액 치료를 받지 못하는 흑인환자에 대한 차별이 발생함

---

101) Machine Bias, There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, 2016. 5. 23. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

102) "여자라서 자동차에 관심없다"...성차별 부추기는 AI 비서, 블로터(2019.6.21.) https://www.bloter.net/news/articleView.html?idxno=29428

103) 이윤아;윤상오(2022), 인공지능 알고리즘이 유발하는 차별 방지방안에 관한 연구, 한국거버넌스학회

이러한 인공지능에 의한 차별은 크게 의도적인 경우와 비의도적인 경우로 나눠볼 수 있다. 의도적인 차별은 인공지능 개발 당시부터 편견을 가진 개발자가 차별을 하고자 하는 목적에서 알고리즘이 왜곡된 데이터를 수집하도록 설계하는 것이다. 가령 신입사원을 선발하는 알고리즘 설계에 있어 남녀별 가중치를 다르게 설정하는 경우가 여기에 해당한다.[104] 비의도적인 차별은 인공지능 알고리즘 자체는 공정하게 설계되었다 하더라고 결과적으로는 특정 집단이나 개인에게 차별을 가져오는 것을 말한다. 가령 컴퓨터 프로그래머 전공자를 대상으로 한 타깃형 광고 알고리즘이 있을 때 해당 알고리즘은 개발자에서 상대적으로 높은 비율을 차지하는 남성에게 광고를 노출시킬 것이고 여성 개발자의 취업기회를 박탈하는 결과를 낳을 수 있다.[105] 문제는 인공지능의 경우 알고리즘이 명확히 공개되지 않아 차별에 개발자의 의도가 있었는지 아니면 알고리즘이나 수집된 데이터의 문제인지를 명확히 알 수가 없다는 것이다. 그렇기에 사람이 하는 것에 비해 인공지능에 의한 차별은 더 큰 피해를 가져오고 구제도 쉽지 않을 수 있다. 따라서 이를 막기 위한 구체적인 대안이 요구된다.

2. 인공지능 차별에 대한 대응 방안

1) 국가인권위원회 인공지능 개발과 활용에 관한 인권 가이드라인[106]

국가인권위원회는 2022년 5월 11일 인공지능 개발과 활용 과정에서 발생할 수 있는 인권 침해와 차별을 방지하기 위하여 <인공지능 개발과 활용에 관한 인권 가이드라인)>을 마련하고, 국무총리와 관련 부처 장관에게 권고를 하였다. 해당 가이드라인에서 국가인권위는 차별 금지를 위한 방안을 다음과 같이 이야기하고 있다.

인공지능을 개발하고 활용할 때는 인공지능으로 인해 영향을 받는 사람의 다양성과 대표성을 반영하기 위해 노력해야 하고, 성별, 종교, 장애, 나이, 출신 지역, 신체조건, 피부색, 성적 지향, 사회적 신분 등 개인과 집단의 특성에 따라 편향적이고 차별적인 결과가 나오지 않아야 한다.

데이터의 수집 및 시스템 설계, 활용 등 인공지능 개발 전반에 걸쳐 편향이나 차별을 배제해야 하고, 데이터 요소를 검사하고 차별적인 데이터를 조정하는 등의 조치를 해야 한다.

개발한 인공지능에 대해 주기적인 모니터링을 거쳐 데이터 품질과 위험을 관리하고, 차별적

보 제29권 제2호, 185쪽.

104) 정보통신정책연구원(2018). 지능정보화 이용자행태 조사방법론 개발 및 실증, 9쪽. 여기서는 의도적인 차별을 '차별적 처우', 비의도적인 차별을 '차별적 효과'라고 부르고 있다.

105) 위 보고서, 10쪽.

106) https://www.humanrights.go.kr/base/board/read?boardManagementNo=24&boardNo=7609439&menuLevel=3&menuNo=91

결과나 의도하지 않은 결과에 대해 개선의 조치를 주기적으로 수행하여야 한다.

2) 유럽연합 인공지능법안[107)]

유럽연합 집행위원회는 2021년 4월 21일 인공지능법안(이하 “위원회 안”)을 발표하였으며, 유럽연합 의회는 2023년 6월 14일 이에 대한 수정안(이하 “의회 수정안”)을 채택하였다. 인공지능 기술에 관한 포괄적인 규제를 담은 최초의 법안으로서, 인공지능과 차별에 관하여 다음과 같이 이야기하고 있다.

알려진 또는 추정되는 민감 내지 보호되는 특성에 따라 개인을 특정 범주로 할당하고 분류하는 인공지능 시스템은 특히 침해적이며 인간의 존엄성을 해하고 차별을 일으킬 큰 위험이 있다. 이러한 특성은 성별, 성별정체성, 인종, 민족적 기원, 이주민 또는 시민권, 정치적 견해, 성적지향, 종교, 장애, 그리고 유럽 기본권 헌장 제21조, EU 규칙 2016/769에 따른 기타 차별금지사유가 모두 포함된다. 이러한 시스템은 따라서 금지되어야 한다.

고용, 근로자 관리 및 자영업 분야, 특히 채용, 승진 및 해고에 대한 의사 결정 또는 이에 실질적인 영향을 미치는 사람의 모집 및 선택, 개인의 행동, 개인의 특성 또는 생체 정보에 기반한 개인화된 업무 할당, 업무 관련 계약 관계에 있는 사람들의 모니터링 또는 평가에 사용되는 인공지능 시스템 역시 고위험으로 분류되어야 한다. 채용 과정 전반 및 업무와 관련된 계약 관계에 있는 사람들의 평가, 승진 또는 업무 유지에 있어서, 이러한 시스템은 여성, 특정 연령대, 장애인 또는 특정 인종, 민족적 출신, 성적 성향을 가진 사람들에 대한 역사적인 차별을 지속시킬 수 있다. 또한 이러한 사람들의 성과와 행동을 감시하기 위해 사용되는 인공지능 시스템은 데이터 보호 및 사생활 보호에 대한 그들의 기본권의 본질을 훼손할 수 있다.

3) EEOC 고용주의 인공지능, 그 밖의 알고리즘 의사결정 도구 활용에 있어 Title 7 지침[108)]

미국 고용평등기회위원회(EEOC)는 2022년 5월 인공지능 및 알고리즘을 통한 채용절차가 장애인 에 대한 차별을 일으킬 가능성을 우려하여 가이드라인을 발표하였다. 이후 2023년 7월에는 고용주의 AI 사용에 관한 두 번째 지침을 발표했다. 해당 지침은 민권법 제7장의 차별금지 조항 중의 한 측면, 이질적 또는 부정적인 영향을 주는 차별을 금지하는 것에 초점을 두고 있다. 해당 지침의 주요 내용은 다음과 같다.

107) https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf
108) https://www.mayerbrown.com/en/perspectives-events/publications/2023/07/eeoc-issues-title-vii-guidance-on-employer-use-of-ai-other-algorithmic-decisionmaking-tools

선발절차 : 고용주가 고용에 있어 알고리즘 의사 결정을 이용하는 것은 Title 7에 의한 '선발절차'이며 해당 절차는 민권법 Title 7의 적용을 받는다. 따라서 고용주는 알고리즘을 이용한 선발절차가 Title 7이 금지하는 이질적인 부정적인 차별을 주지 않도록 해야 한다.

4/5 규칙 : 고용주는 알고리즘 의사 결정 도구와 관련하여 부정적인 영향을 분석할 때 일반적으로 적용되는 경험칙인 4/5 법칙을 사용할 수 있다. 여기에 따르면 일반적으로 보호 특성에 대한 "선택 비율"이 가장 높은 선택 비율을 가진 그룹 비율의 80%(4/5) 미만인 경우 부정적인 영향을 준 것으로 분석된다. 다만 이는 일반적 경험칙이며 구체적 사안에서는 다르게 판단할 수 있다.

제3자 보호 불가 : 고용주는 알고리즘 의사 결정 도구를 이용하기 위해 소프트웨어 공급업체 등 제3자의 도움을 받을 수는 있지만, 그러한 도구의 사용으로 발생하는 부정적 영향에 대한 책임은 고용주가 궁극적으로 책임져야 한다.

3. 평등실현을 위한 인공지능 규율 과제

1) 정부의 인공지능법안에 대한 수정 방향

발제문에서도 지적하듯이 과방위 법안소위를 통과한 법안은 여러 가지 한계를 담고 있고, 인공지능으로 인하여 발생할 수 있는 차별을 방지하기에 부족한 지점이 많다. 이에 다음과 같은 방향으로 법안 수정이 필요하다.

첫째, 고위험 인공지능의 정의에 있어 '성별, 종교, 인종, 장애, 성적지향 등 개인이나 집단의 특성을 근거로 한 차별을 할 목적으로 설계된 인공지능'과 '성별, 종교, 인종, 장애, 성적지향 등 특성을 근거로 개인이나 집단을 분류하고 할당하는 인공지능'을 포함시킬 필요가 있다.

둘째, 차별적인 의도가 없더라고 데이터의 편향, 데이터의 부족, 알고리즘의 결함 등으로 인하여 차별적인 결과가 발생하지 않도록 사전, 사후로 평가하기 위한 절차가 필요하다. 시민사회 법안에서 규정한 인공지능 인권영향평가가 대표적인 절차라 할 것이다.

셋째, 인공지능으로 인한 차별의 발생은 알고리즘 자체의 문제에서 발생하기도 하지만 이를 운용하고 평가하는 사람들의 문제이기도 하다. 그런데 현재 정부법안에서는 인공지능위원회의 구성과 기능에 대해서 세세히 규정하고 있는데 그 구성에 있어 다양성과 인권 전문가의

포함에 대한 규정이 없다. 따라서 인공지능위원회의 구성에 있어 인권과 차별에 대한 경험과 지식이 있는 사람이 포함되어야 하고, 「국가인권위원회법」 제5조 4호와 같이 '다양한 사회계층의 대표성이 반영될 수 있도록 위원을 선출·지명하여야 한다' 등의 규정을 두어야 한다.

넷째, 국가인권위원회는 2023. 7. 13.「인공지능산업 육성 및 신뢰 기반 조성 등에 관한 법률안」에 대한 의견표명에서 "인공지능으로 인해 발생할 수 있는 권리 침해 사안을 유형별로 분석·분류하고, 이미 제도화되어 있는 구제 절차 중 각 유형별로 이를 처리하기 적합한 구제 절차를 이행할 수 있도록 인공지능 법률안에 근거 규정을 마련하여야 한다"는 권고를 하였다. 국가인권위원회의 권고처럼 인공지능으로 인해 차별을 받은 이들이 피해를 진정하고 구제받을 수 있는 절차가 별도로 규정되어야 한다. 현재의 국가인권위원회법은 인공지능과 차별에 대한 규정이 없어 현재의 규정만으로는 차별 피해에 대한 규제가 쉽지 않기 때문이다.

2) 포괄적 차별금지법 제정

한편 인공지능에 의해 발생하는 차별을 예방하고 대처하기 위해서는 인공지능법안에 구체적인 규정을 넣는 것도 중요하지만, 궁극적으로는 포괄적 차별금지법의 제정이 요구된다. 인공지능 차별에 있어 차별금지법이 필요한 이유는 다음과 같다.

첫째, 인공지능에 의해 이루어지는 혐오와 차별은 결국은 사회의 혐오와 차별적 구조를 투영한 결과물인 경우가 많다. 따라서 아무리 데이터 수집, 활용 알고리즘을 정교하게 개발하더라도 사회 전반의 차별적 구조가 변경되지 않는 한 이 사건과 같은 문제는 계속될 수밖에 없고, 이를 해소하기 위해서는 차별금지법이 역시 필요하다. 이와 관련하여 영국 공직생활윤리위원회의 <인공지능과 공공규범보고서(Artificial Intelligence and Public Standards: report)> 는 "평등법은 특정사유를 이유로 한 차별을 금지하고 있기에 데이터 편향을 방지하는 핵심적인 법적안전장치"라고 이야기하고 있다.[109)]

둘째, 인공지능의 데이터 수집, 활용 과정에서 발생할 수 있는 편향, 차별, 혐오를 예방하고 제거하는 알고리즘을 개발하기 위해서는 우선 개발자들이 무엇이 차별이고 왜 그러한 차별이 문제가 되는 것인지를 알아야 한다. 문제는 현재 한국사회에는 구체적으로 차별의

109) The Committee on Standards in Public Life, Artificial Intelligence and Public Standards : Report, 2020, 39p https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868284/Web_Version_AI_and_Public_Standards.PDF

개념과 판단기준을 통일적으로 규정한 법률은 없다는 것이다. 그렇기에 직접차별, 간접차별, 복합차별 등 차별에 대한 개념과 구체적인 판단기준을 제시하는 법규범으로서 차별금지법 제정이 필요하다.

셋째, 인공지능에 의한 차별 문제에서 특히 중요한 것이 간접차별의 법리이다. 인공지능 자체가 차별적 의도를 갖고 설계된 경우도 있지만 많은 차별은 의도하지 않았지만 데이터의 편향 등으로 결과적 차별이 발생하는 경우가 많기 때문이다. 유럽평의회 민주주의총국(Directorate General of Democracy)에서 발간한 <차별, 인공지능과 알고리즘 의사결정(Discrimination, artificial intelligence, and algorithmic decision-making)> 보고서 역시 "가령 특정한 인종적 배경을 지닌 사람들에게 더 많은 재화 용역 비용을 감당하도록 하는 인공지능의 결정은 간접차별 위반이 될 수 있다"면서, "차별금지법은 차별적인 인공지능 결정에 맞서는 데 이용될 수 있다"고 이야기하고 있다.[110] <끝>

110) Prof. Frederik Zuiderveen Borgesius et al, Discrimination, artificial intelligence, and algorithmic decision-making, Directorate General of Democracy, Council of Europe, 2018, 19p.

【 토론 2 】

조아라
(언론인권센터 활동가)

더 나은 인공지능을 위한 시민사회의 또 다른 역할: 디지털 시민성 교육

올해 초 챗GPT가 2개월만에 월 사용자 1억 명을 돌파하면서 AI에 대한 관심이 한껏 높아졌습니다. 그와 동시에 이와 같은 AI가 야기할 수 있을 위험 또는 폐해에 대해서도 꾸준히 비판적인 논의가 진행되어 왔습니다. 가령 2020년 챗봇 '이루다'에서 이미 문제가 제기된 바 있었던 혐오의 문제들은 물론, 대체할 수 없으리라고 예측되어 왔던 활동에 AI가 활용되기 시작하면서 나타난 노동 문제들, AI 개발 과정에서 확인된 각종 노동 착취와 저작권 침해, 인권 침해 등, 그 편리함만큼이나 많은 문제가 나타나고 있습니다. 그러나 여전히 챗GPT를 비롯한 AI는 다양한 분야에서 활용되기 시작했고, 이 흐름은 이미 멈출 수 없는 것이 된 것 같습니다. 그렇기에 이번 인공지능 법안에 대한 이번 비판적 분석이 의미가 있으며, 해당 법안과 수정안의 허점을 분석하여 보충한 시민사회 인공지능책임법(안)에 대한 제안 역시 매우 소중한 것입니다.

금지해야 하는 인공지능에 대한 보다 상세한 언급과 함께, 실질적인 규제 방안을 제시하고 다양한 정부 부처가 인공지능 관련 정책에 참여할 것을 요구하는 이번 시민사회 인공지능책임법(안) 제안의 필요성에 대해서는 두말할 나위가 없다고 생각합니다. 이에 저는 이와 같은 법안 및 정책에 힘을 더하고 더 나은 인공지능을 만들어나가기 위하여, 시민사회가 할 수 있을 또 다른 역할을 이번 시민사회포럼에서 함께 이야기할 수 있었으면 합니다. 바로 AI 시대의 디지털 시민성 교육입니다.

디지털 시민성이란 '인터넷이 제공하는 기회를 활용한 정보통신기술 사용 기술, 협업 기술, 시민 참여, 창의적 생산, 타인과의 존중과 같은 영역에서 긍정적인 성장을 이룰 수 있는 인

지적, 사회-정서적 역량과 위협을 최소화하고 해결하기 위한 적절한 조치를 취할 수 있는 역량'을 포괄하는 개념입니다[111]. 이와 같은 디지털 시민성에 대한 정의와 이를 함양하고자 하는 교육안에 대한 논의는 디지털 기술의 발달과 함께 계속되어 온 것이지만, AI 기술이 대두되면서 이에 대한 논의 역시 이에 포함시켜야 한다는 제안이 나타나고 있는 것 같습니다.

"(…)그러한 목적을 달성하기 위한 정책 제언은 ▲교육 목표 설정 ▲교육 내용 ▲교육방식, 접근법, 실천 ▲정책행동의 관점에서 나누어 제시했다. 첫째, 교육 목표는 ①지속가능하고 포용적인 사회를 만드는데 기여하는 디지털 시민성 ②'모두를 위한, 모두에 의한' 디지털 시민성 ③인공지능 기술 개발의 방향성에 개입하는 디지털 시민성으로 설정할 필요가 있다. 둘째, 교육 내용 측면에서는 ①생성AI를 포함해 인공지능 기술이 어떻게 작동하는지, 기술의 이점과 한계는 무엇인지, 구체적인 적용 사례와 콘텐츠 생성 및 소비에 미치는 영향, 사회 전반에 미치는 영향 등에 관한 교육 ②생성AI가 가져온 정보폭증 대응 능력을 기르는 교육 ③AI 생성물로 인해 발생하는 저작권 침해, 개인정보 침해에 대한 교육과 더불어 패러다임 자체의 변화에 대해 논의하는 교육 ④AI의 잠재적 영향과 윤리적 고려사항에 대한 교육 등이 새롭게 추가되어야 한다."[112]

유네스코는 이미 "2023 세계 교육 현황 보고서"에서 교육 분야에서의 기술을 중점적으로 다루면서, 교육 현장에서의 기술적 적용과 그 한계, 그리고 더 나은 적용을 위한 목표와 원칙 설정의 중요성을 설파한 바 있습니다. 그리고 기술이 이끌어 낼 교육의 질적 변화는 적절한 기술 인프라를 제공할 수 있느냐의 여부 뿐만 아니라 기술 활용에 대해 가르칠 수 있는 교사의 역량, 디지털 리터러시 통합 교육 등 교육 콘텐츠, 해당 교육의 사회적 성과 등을 모두 포함하고 있다고 언급하고 있습니다.[113] 이와 같은 내용에서 역시 AI에 대한 디지털 역량을 기르는 것이 빠질 수는 없을 것입니다.

최근 AI의 안정성과 윤리성에 대한 요구가 높아지면서, 다양한 국가 및 기업에서 "신뢰 가능한 AI(trustworthy AI)"에 대한 가이드라인을 마련하고 있습니다. 발제에서 제시된 시민사회 인공지능 책입법안에서도 인공지능의 인권영향평가를 필수 요소로 제안한 바, 이를 바탕으로 시민사회의 구성원들이 나서서 AI 시대의 디지털 시민성 교육안을 마련, 함께 진행한다면, 해당 법안과 함께 보다 나은 AI를 만들어나가고 이를 긍정적인 방향으로 활용하는데 큰 도움이 되리라 생각합니다. <끝>

---

111) 김민정, 「생성형 AI 시대의 디지털 시민성 함양을 위한 세계시민교육의 과제 검토 - '생성형 AI'의 낯선 도전, 우리는 무엇을 준비해야 하나」, 『이슈 브리프』, 유네스코, 2023.11., pp.16~17.
112) 위 글.
113) 유네스코, 『2023 세계 교육 현황 보고서 요약본 - 교육 분야에서의 기술: 누구를 위한 도구인가?』, 2023.

【 토론 3 】

송경재
(민주언론시민연합 정책위원, 상지대 사회적경제학과 교수)

인공지능 법안에 대한 시민사회 입법안 토론문

Ⅰ. 인공지능(AI)의 위험성

○ 인공지능(AI)의 등장
- 인공지능(artificial intelligence, AI)은 인간의 학습, 추론, 지각능력을 컴퓨터 과학을 이용하여 인공적으로 구현하려는 기술로 4차 산업혁명의 핵심기술로 주목
- 국내에서는 2016년 3월 이세돌 9단과 구글의 딥마인드(DeepMind)가 개발한 바둑 AI인 알파고(AlphaGo)와의 대국에서 1승 4패로 이세돌 9단이 패하면서 이른바 '알파고 쇼'크로 불리며 AI에 관한 관심이 집중
- AI는 빠르게 인간을 대체해 나갈 것이며 AI를 기반으로 IoT, 클라우드 컴퓨팅, 빅데이터 등이 융합되면서 새로운 사회변화를 주도할 것으로 예측(송경재 2023)

○ 생성형 AI(Generative AI)
- AI기술은 현재 머신러닝과 딥러닝을 중심으로 발전하고 있으며 '약한(Weak) AI'와 '강한(Strong) AI'으로 구분 가능
- 생성형 AI(Generative AI)는 이용자의 특정 요구에 따라 결과를 생성해내는 AI로 정보 데이터 학습을 통해 작문, 이미지 생성, 비디오 생성, 음악, 미술 등 콘텐츠 생성에 이용
- 생성형 AI는 2022년 말 오픈AI(OpenAI)의 대형언어모델 기반의 GPT-3기술을 기반으로 하는 챗GPT(ChatGPT)가 일상화되면서 주목을 받고 있음

○ 생성형 AI의 사회·정치·경제적 영향에 관한 관심 증가
- 세계경제포럼(WEF)과 경제협력개발기구(OECD)의 자료를 인용한 정보통신정책연구원(KISDI) 『생성형 AI의 등장과 AI의 일자리 영향에 대한 소고』에 따르면(2023), 생성형 AI의 등장은 기술 발전이 미래 일자리와 인력 수요 변화 등의 사회 불확실성을 점점 더 높이는 방향으로

진행되고 있다고 예측
- 2023년 5월 G7정상회담의 히로시마 프로세스
→ ‘인간 중심의 신뢰성 있는 AI’를 구축하기 위해서도 '신뢰성 있는 자유로운 정보 유통'(DFFT·Data Free Flow with Trust)을 구체화
- 언론의 관심, 대중의 호기심이 결합하여 세간의 주목을 받으며 생성형 AI의 사회·정치·경제적 영향에 관한 관심도 고조

○ AI의 위험성 주목
- AI는 근본적으로 학습을 위한 데이터의 안전성과 학습의 본질적인 한계에서 파생하는 정치적 편향성, 가치 판단의 왜곡이 시민들의 생활에 악영향을 미칠 수 있다는 우려감도 증가(하라리 2023)
- 실제 AI의 활용에 따른 정치적 위험성은 다양하지만 가장 많이 논의되는 것은 첫째, 이미지와 동영상을 제작한 가짜뉴스의 조작, 둘째, 편향성과 확증 편향의 강화, 셋째, 사이버 공격 등
- AI의 정치적 악용에서 가장 많이 거론되는 것은 2023년 5월 22일 미국 국방부 청사(펜타곤)에서 검은 연기가 피어오르는 사진이 급속도로 온라인에서 확산한 사건은 생성형 AI가 만든 것으로 추정되는 가짜 이미지였으나 몇 시간 만에 이 사진은 소셜미디어로 확산되었고 금융시장에까지 영향
- 이에 미국과 유럽연합 국가 내의 주요 선거에서 러시아발 챗봇과 생성형 AI를 이용한 가짜뉴스의 확산은 이미 공공연한 비밀로 간주(송태은 2021)

○ 생성형 AI 등 AI 발전에 따른 정치사회적 대응에 대한 논의 필요
- AI를 악용할 경우 미국의 펜타곤 화재 사진과 같은 사회 혼란을 부추길 수 있는 여지가 강해짐에 따라 AI의 현명한 이용과 사회적 활용에 대한 국제적인 관심도가 제고되고 있는 상황
- 다만 현실적으로 AI기술발전을 막을 방법도 없고 역으로 한국과 같은 후발 AI연구국가에서는 지나친 규제로 인한 기술개발의 지연 등 경제적인 효과에 대한 우려도 제기
- 그러나 전 세계적으로 AI의 위험성에 대한 논의가 이미 진행되고 있어 기술 발전과 시민의 이익 또는 자유권을 조화롭게 발전시키기 위한 논의가 필요함

Ⅱ. 발표문에 대한 보완 토론

○ 김하나 변호사님이 제시한 내용에 대해서는 이견보다는 논의를 풍부하게 하기 위한 보완적 고려사항이 있어 이에 대한 고민이 필요
- 발제문은 시민사회 입장에서 국가가 주도하는 방식의 진흥중심 또는 규제중심적 인공지능 법제화의 문제점을 잘 지적하였음
- 단기적으로 AI의 위험성에 대한 경고가 제기되고 있으나 여전히 국가 중심주의적인 규제 또는 진흥 논의가 가지는 위험성이 또 반복되고 있음

○ 고위험 인공지능의 모호함
- 현재 정부가 추진 중인 법안의 핵심은 고위험 인공지능의 규정의 모호성

- 법안에서 제시하고 있는 “사람의 생명, 신체의 안전 및 기본권의 보호에 중대한 영향을 미칠 우려가 있는 영역”에 대한 사회적 합의의 부재

○ 관련 인공지능위원회 구성과 역할
- 위원회를 위한 위원회 기능에 머무를 가능성 높음
- 개념규정이 모호한 상태에서 자의적인 판단 근거도 높고, 위원회의 업무가 기존 대통령실, 총리실 산하의 위원회와 중복될 가능성 높음

○ AI의 부작용에 대한 실질적 규제조항 부재
- 인공지능위원회가 준수여부에 대한 벌칙조항이 없다는 것은 친기업적인 정책결정의 가능성 높음
- 세계적으로도 히로시마 프로세스 이후 AI 개발 및 운영사의 사회적 책무를 강화하는 시대적인 조료와 전혀 부합하지 않는 법안

○ AI의 투명성에 대한 모호한 규정
- 법안의 기술적으로 가능한 범위라는 조항은 있으나 마나한 조항으로 책무성과 투명성이 향후 AI의 핵심적인 의제가 되고 있는 점을 감안하면 시대적인 흐름에 한참 미흡
- 엄격한 책무성과 투명성을 바탕으로 AI의 인간친화적 사용을 위한 다양한 논의가 마련되지 않음

○ 섣부른 시민사회 인공지능 책임법(안)의 오용 가능성
- 시민사회 인공지능 책임법(안)이 필요한지에 대한 고민 필요
- 현재 개별 국가의 흐름을 보면 진흥과 규제가 다양하게 나타나고 있어 이에 대한 핵심적인 내용을 중심으로 법제화하는 것은 필요함
- 그러나 디지털 인권과 관련한 기본적인 조항들이 아직 국내에서 미비한 상황에서 인공지능 책임법은 자칫 기업의 기술발전을 저해할 소지가 있어, 법안의 세밀한 고려사항을 반영하기 어려움
- 오히려 과도한 입법안의 마련보다 국제기준을 면밀하 타산하고 종합적인 법안은 준비하는 것도 고려해야 할 것임
- 국제적으로 논의되고 있는 AI관련 정책이 시민권, 책무성, 투명성, 안전 등의 가치를 잘 살리고 있다는 점을 고려해 너무 근시안적인 법안 마련은 오히려 정부의 규제권한을 강화하는 문제점이 나타날 수도 있음을 시민사회가 경계해야 할 것 <끝>

【 토론 4 】

윤명
(소비자시민모임 사무총장)

이용자보호 관점의 인공지능 책임법 제정 필요성

인공지능이라는 단어가 우리의 일상 곳곳에서 보편적으로 사용되고 있지만, 국민 개개인들은 인공지능이 활용된 것인지도 모른 채 이용하고 있을 수 있다.

우리는 그동안 인공지능에 있어 이용자 보호를 이야기 할 때 데이터 활용과 관련한 이슈들을 중심으로 고민해 왔다. 그러나 이제 인공지능의 수준과 활용의 범위, 발전수준 등을 고려해 볼 때 인공지능(AI)이 가져올 이익 못지않게 인간의 기본권까지 위협할 수 있다는 두려움을 갖게 된다.

아직은 인공지능의 개발이 초기단계라고 하고 있지만 그 발전 속도를 볼 때 다른 어떤 분야보다 인공지능 책임과 관련한 법안 제정의 필요성에 공감하게 된다.

현재까지 국내의 인공지능 법안 제정과 관련한 논의를 보면, 진흥의 관점이 우선시 되고 있다. 그러나 인공지능 법안과 관련하여서 가장 우선 시 되어야 할 것은 이용자의 안전과 헌법에서 보장하고 있는 최소한의 기본권을 보장하는 차원뿐만 아니라 전 세계적인 공통의 규범과 문화적 측면에서의 피해나 불이익 등이 발생하지 않는 차원의 사회적 합의를 통한 규제가 먼저 정립되어야 한다.

발제에서 제안하고 있는 안의 내용에 전반적으로 동의를 하면서, 특히 금지되는 인공지능과 관련한 조항의 필요성에 특히 더 공감하게 된다.

일부에서는 선언적이고 모호한 부분이 있다고 할 수 있으나, 인간의 존엄성을 침해하지 않

는 방향에서 인공지능을 개발하거나 활용하는 것은 변함없는 대원칙의 합의로 이를 위한 법적인 근거 마련은 매우 중요하다고 본다.

국제기구나 해외 소비자단체에서도 지난 몇 년간 디지털 경제, 사물인터넷, 인공지능과 관련된 소비자 관점에서의 연구와 활동을 해오고 있다.

전세계 이용자들 중 인공지능에 대한 경험을 담은 이용자의 의견을 예를 들면, (1) 인도에서는 인공지능기술이 여성과 낮은 카스트 계급에 대한 차별에 적용될 것을 우려하였고, (2) 일본에서는 일본인의 정신이나 문화에 맞는 방법으로 인공지능이 도입되기를 희망한다는 의견을 표명했다. 따라서 한 나라에 적용된 인공지능 기술이 그대로 다른 나라에 적용되었을 경우에, 인종, 문화, 국민적인 인식의 차이 등으로 인해 수입된 인공지능 기술이 수입된 국가의 사정에는 맞지 않을 수 있어 이에 대한 우려를 나타내고 있다.

또한 이번 책임법(안)에서 논의하고 있는 고위험 영역 인공지능 규제와 관련해서국제소비자기구 역시 인공지능이 적용되는 분야 중 소비자에 영향을 미치는 분야로는 보건의료 분야, 자율주행자동차 또는 위치 정보 지도서비스, 금융서비스, 온라인 상거래, 가정내 사물인터넷, 공공서비스 등 고위험이 발생할 수 있는 분야에 대해서는 사회나 인권에 미치는 영향이 클 수 있으므로 통제를 위한 제도적 장치가 필요하다는 데 의견을 내고 있다. 인간에게 치명적인 영향을 미칠 수 있는 분야에 대한 서비스 제공 시 이용자의 권익에 직접적인 영향을 미칠 수 있는 것으로, “고위험 인공지능군”으로 정하고 이에 대한 통제를 위한 제도적 장치가 필요하다. 발제의 제안에서는 이러한 통제의 역할을 위한 인공지능위원회 설치를 제안하고 있는데, 특히 이용자의 관점에서는 인공지능과 관련하여 문제가 발생했을 경우 이에 대한 대응을 개인의 지식으로는 해결하기 어려운 상황에 놓일 수 있어, 다른 어떤 분야보다 사회 각계의 논의 구조를 통한 문제해결이 필요할 것으로 보인다. 다만, 인공지능위원회의 역할이 사안의 심의 의결보다는 좀 더 인류보편적인 관점에서 정의와 제도적 장치를 마련하기 위한 논의구조가 될 수 있도록 구성원과 운영방식 등이 기존의 법적위원회와는 다르게 구성 운영되어야 하지 않을까 생각한다.

이용자 관점에서 인공지능과 관련하여 가장 관심이 되는 분야는 투명성과 통제가능성에 관한 것이다. 인공지능을 활용했을 때, 이용자의 대표적인 우려는 정보의 수집 및 이용방법에 대한 우려가 컸고, 이용자 입장에서의 선택가능성의 결여 등으로 이용자가 무엇을 모르는지 모를 때 발생하는 리스크가 가장 큰 리스크라고 생각한다.

이용자의 입장에서는 인공지능으로 인한 어떠한 손해가 발생되었을 때 누구를 상대로 어떠한 책임을 불어야 할 지 모르는 경우가 많다. 민간의 영역에서 인공지능기술을 도입/적용

하였을 때 귀책관계를 따질 때, 공공기관 또는 민간업체의 책임이 불분명할 수 있어 이에 대한 우려가 크다.

실제로 국제소비자기구에서도 이용자가 인공지능 기술을 감독하거나 감시할 기술적인 능력이 없다는 것에 문제를 두고, 이를 해결하기 위한 법적 제도적 장치가 필요함을 주장하고 있다.

(1) 책무성의 측면에서 살펴보면, 민간업체의 책무성을 강화할 수 있는 법적 제도적인 장치가 있어야 할 것이고,
(2) 투명성 및 설명가능성에 대해, 투명성은 민간기업/인공지능 기술을 만든 기업이 국가의 감독기관의 감시 감독의 대상이 되어야 하고, 이용자에 대한 정보비대칭 해소를 위한 투명성의 제고도 필요하다고 할 것입니다. 여기서 투명성은 이용자에게 유의미한 정보의 제공이여야 하고, 가령 십만 단어 이상의 어떤 긴 문서, 다량의 정보를 이용자에게 제공하는 것만으로 투명성이 확보된다고 보기는 어렵다고 하고 있다.
(3) 또한 윤리적인 지침이나, 자율규제만으로 인공지능의 책무성을 확보하기는 어렵기 때문에 법규적인 수준에서 관련 규정(법안 제정)을 마련해야 한다는 의견을 제시하고 있다.

인공지능규율하는 법을 만들더라도 세부적인 분야에서는 각 영역별 규제 틀 내에서 규율될 것으로 예상된다. 따라서 인공지능을 위한 법안에서는 특히 이용자의 인권과 사회적 규범을 헤치지 않기 위한 인공지능 책임법 특히, 개발자가, 활용 이용자에게 일정부분의 역할과 책임을 명확하게 하는 법안이 우선되어야 한다.

물론 기술개발이 이루어지고 있는 시점에서 많은 혼선과 견해의 대립 등 사회적 논의와 합의의 과정이 발생할 수밖에 없다. 그러나 기술을 개발해 가는 단계에서부터 이용자의 권익과 이용자보호의 관점이 배제되거나 등안시 된다면, 오히려 추후에 이를 돌이킬 수 없는 사회적 혼란을 가져올 수 있다고 생각한다.

산업육성, 기술혁신 지원 등도 물론 중요하겠지만 그보다 더 이를 이용하려는 목적이 인간에게 도움이 되고자 하는 것인 만큼 개발단계에서부터 이용자의 권익보호 차원의 법안의 필요성이 더욱 강조된다. <끝>

【 토론 5 】

이장희
(참여연대 공익법센터 운영위원, 창원대 법학과 교수)

“인공지능 법안에 대한 시민사회 대안 입법안”에 대한 토론문

인공지능 시민사회포럼을 준비해 주신 관계자 여러분들께 감사드리며, 소중한 자리에 토론자로 초대되어 매우 기쁘게 생각합니다. 그리고 토론에 앞서 인공지능의 위험성을 잘 지적하고 안정성 확보를 위한 다각적인 고민을 담은 훌륭한 발표를 해 주신 민변 디지털정보위원회의 김하나 변호사님의 수고에도 깊이 감사드립니다.

이제 인공지능 사회에 진입하는 길목에서 인공지능으로부터의 안전과 인간적 삶의 보존은 매우 중요한 현안이 되고 있습니다. 인공지능은 크게 보아 약한 인공지능과 강한 인공지능으로 분류되지만, 인공지능 기술은 지속적인 발전을 거듭하고 있고 또 머지않아 인공지능이 고도화된 새로운 사회로 변화할 것으로 예상합니다. 그런 점에서 여전히 인공지능에 관한 기본법의 입법은 매우 중요하고 시급한 과제라는 점에 이견(異見)이 없을 것 같습니다.

하지만 문제는 과연 지금의 인공지능 기본법안이 올바른 방향으로 마련되고 있는지에 있습니다. 유독 인공지능에 대해서만은 관대한 입장에서 인공지능 산업의 진흥에만 초점을 맞추고 있다고 보이기 때문입니다. 물론 현재 인공지능 관련 기술 패권을 다투는 중국이나 미국 등의 인공지능 선두국가들에 비해 우리나라의 인공지능 수준은 상당히 부족한 것도 사실입니다. 인공지능의 기술 패권과 우위를 선점하기 위한 글로벌 경쟁이 격화될 조짐까지 있는 상황에서 정부당국자의 고민과 대응에도 이해가 안가는 것은 아닙니다만, 그렇다고 인공지능에 내재한 위험성까지 도외시하는 법제도를 마련하는 것은 자칫 소잃고 외양간 고치는 격의 우를 범할 우려가 있습니다. 특히 인공지능 관련 산업의 글로벌 선도를 위해 “원칙적 허용, 사후적 규제”라는 원칙을 세우는 것은 인공지능에 내재된 위험성을 그대로 방치하는 결과가 될 수 있다는 점에서 매우 잘못된 입법태도가 아닐 수 없습니다.

우리가 생각하는 바람직한 법제 방향은 역시 인공지능의 기술발전 가능성과 인공지능의 안전성을 형량하여 함께 추구하고 균형을 도모하는 것이라 할 수 있습니다. 이것은 단순히 인공지능에만 특히 요구되는 것은 아니며, 기본적으로 모든 국가 역할에서 우리 헌법이 요구하는 기본 원칙이기 때문입니다. 우리 헌법은 전문(前文)에서 “우리들과 우리들 자손의 안전, 자유, 행복을 영원히 확보”하자는 대한민국의 국가적 목표를 세우고 있습니다. 달리 말하자면 우리가 행복하려면 자유로워야 하고, 자유로우려면 먼저 안전이 확보되어야 한다는 의미일 것입니다. 당대 뿐만 아니라 미래 세대에게도 그것이 가능해야 한다는 것입니다. 그런데 과연 지금 인공지능의 기본 법제가 ‘안전’을 확보하는데 얼마나 가치를 두고 있는지 알기 어려우며, 인공지능의 위험성이 무방비로 방치되는 상황에서 과연 인간의 자유, 무엇보다 인간의 존엄과 다양한 인권의 가치가 얼마나 존중되고 보호될 수 있는지 의문이 제기되고 있습니다.

인공지능 분야에서 ‘안전’의 핵심은 역시 인간의 존엄을 침해할 우려가 있는 인공지능의 분야를 원칙적으로 금지하는 입법이라고 봅니다. 이것은 우리뿐만 아니라 유럽 등 주요 선진국의 인공지능 발전 방향과도 궤를 같이 하는 부분입니다. 그러한 점에서 김하나 변호사님의 발표 및 인공지능 시민사회 대안에서 제시한 금지항목은 매우 적절해 보입니다.

또한 고위험 인공지능의 분야도 세부적으로는 조정의 여지가 있을 순 있지만, 기본적으로 이 부분에 대해서는 상당한 통제 가능성을 확보하면서 개발과 활용을 허용하는 균형이 도모되고 있다고 보입니다. 사실 계속적인 기술 발전이 예상되는 인공지능에서 언제나 위험이 없을 수는 없습니다. 무엇보다 인간의 지능을 모방하는 인공지능에게 인간의 삶을 모두 맡기는 것이 가능하지 않다는 점에서 인공지능에게는 ‘인간성 확보를 위한 비상브레이크’로서 인간의 개입과 통제가 반드시 필요하다고 생각합니다.

또한 과기정통부의 그늘에서 벗어나, 국무총리 소속으로 ‘독립된’ 인공지능위원회를 구성하고 인공지능 관련 계획, 법과 정책을 주관하려는 것은 상당히 중요한 대안으로 생각됩니다. 특히 과기정통부가 주도하는 행정부 중심의 인공지능 사회는 근본적으로 ‘빅브라더’의 출현이라는 점에서 매우 우려스러운 점이 많은데, 인공지능위원회는 민주적이고 인권친화적인 인공지능 거버넌스의 구축에도 좋은 대안이라고 볼 수 있습니다. 특히 인공지능의 개발 이용에 앞서 ‘인권영향평가’를 실시하는 것은 새 법안에서 반드시 필요하고 중요한 내용이라 보입니다.

다만 몇 가지 점에서 의문점이 있어 이를 지적하는 것으로 저의 토론을 마무리 하겠습니다.

먼저, 인공지능위원회와 다른 국가기관, 특히 과기정통부나 행안부, 개인정보위원회, 국방부, 국정원 등과의 소관 업무 분장 및 관계 설정이 중요할 거 같은데, 이 부분이 명확하지 않아 발표자의 의견을 듣고 싶습니다.
둘째로, 방첩 분야를 넘어서, 군사, 국방분야의 인공지능은 어떻게 봐야 하는지도 보충이 가능하시면 의견을 듣고 싶습니다.
셋째로, 고도화된 인공지능에 일종의 독립된 법인격을 부여하자는 의견이 사회 일각에서 제시되는 경우가 있는데, 토론자는 인공지능은 인간과 대등한 목적이 될 수 없다는 점에서 인공지능 자체가 결코 법인격을 가질 수 없다는 입장을 가지고 있는터라 발표자의 생각이 궁금합니다.
넷째로, 인공지능으로부터 초래되는 피해, 위험의 책임은 결국 인공지능 자체가 아니라 그 개발자, 사업자, 사용자 등이 져야 하는데, 이 부분에 대한 책임 부분에 대해서는 법안에 어떤 내용으로 수용될 수 있는지 궁금합니다.
다섯째로, 과징금의 한도를 전체 매출의 3%로 한다면 상당히 경제적으로 남는 장사가 아닐까 싶은데, 이 정도의 과징금 액수로 과징금 부과의 목적 내지 취지를 살릴 수 있을까요?
여섯째로, 과태료 부과의 주체는 인공지능위원회인지요?
마지막으로, 인공지능위원회 위원으로 국립대 교수가 추천될 경우 교육공무원이라서 법률상 불가능하게 되는 부분은 어떻게 생각하시는지 여쭤봅니다.

이번 인공지능 시민사회포럼이 인공지능 기본법안의 입법과 방향성을 확인하는데 중요한 역할을 하고 최고의 대안이 모색되는 기회가 되길 바라며, 이상으로 저의 토론을 마칩니다. 감사합니다. <끝>