새로운 정보기술과 인권

자동화된 프로파일링 등을 통한 개인정보 침해와
인공지능 알고리즘을 이용한 차별을 중심으로

참여위원	• 홍성욱 서울대학교 생명과학부 교수(대표집필 / 위원장) • 윤정로 KAIST 인문사회과학부 교수(위원장) • 이병호 서울대학교 전기정보공학부 교수 • 이은우 정보인권연구소 이사(변호사) • 이진우 KAIST 생명화학공학과 교수 • 이호영 정보통신정책연구원 연구위원
원고검수	• 오요한 RPI 박사과정
간사	• 이지혜 서울대학교 대학원생

발간사

과학기술의 발전은 인류에게 삶의 윤택함과 편리함을 가져다주었다. 하지만 상상 속에서만 가능했던 일들이 과학기술을 통해 현실이 되고, 과학기술이 사회에 미치는 영향력이 점점 더 커짐에 따라 일부 연구결과들은 오히려 개인의 자유와 권리를 위협하는 데 이용될 수 있다는 우려의 목소리가 있다. 인공지능이 창출할 혜택이 소수의 개인이나 기업에 의해 장악되어 불평등을 심화시킬 수도 있고 유전자가위 등 의생명과학기술은 인간존엄성에 양날의 칼이 될 수 있다는 주장이다.

이러한 사회의 문제제기와 국민들의 불안에 대해 우리 과학기술계는 답해야 할 의무가 있지만 그동안의 활동은 미진했다. 몇몇 과학기술인들이 인권 기구에 참여하거나 활동한 적은 있지만 주체로서 이끈 경우는 드물고, 과학기술의 발전이 사회에 미치는 윤리적·법적·사회적 영향(ELSI)에 대해 다방면으로 성찰해야 할 책무를 게을리 했다.

이제 과학기술계가 과학기술을 만드는 자의 권리 뿐 아니라 과학기술을 향유하는 자의 권리까지 함께 생각하고 과학기술의 발전에 따라 위협 받을 수 있는 보편 인권을 보호하기 위해 우리가 할 수 있는 역할이 무엇인지 찾아보고자 한다.

한국과학기술한림원과 대한민국의학한림원은 한림원 회원을 비롯해 다양한 분야의 이 시대 지성인들과 함께 신기술이 인권을 침해하는 현실과 미래에 침해를 유발할 수 있는 가능성을 짚어보고 이를 해결하거나 완화할 수 있는 방안을 깊이 고민해 보았다. △정보기술 △의생명과학 등의 분야에서 발생할 수 있는 인권이슈를 살펴보고 △젊은 과학기술인의 인권 개선 방안 등을 찾았으며, 오픈포럼을 통해 각계각층의 여론을 수렴하였다.

지난 9개월 간 공부하고 수차례 토론한 결과를 정리하여 많은 분들께 공유함으로써 다양한 논의들이 교류되는 장을 만들고자 과학기술과 인권을 주제로 4권의 '이슈페이퍼' 시리즈를 발간한다. 이번 이슈페이퍼가 인권과 과학기술의 조화로운 발전을 만들어 갈 토대를 만드는 데 초석이 되길 바라며, 나아가 과학 기술계, 정책입안자, 언론, 국민과 함께 과학적 진보와 인권 향상이 함께 가는 사회에 대한 공감대를 형성하는 기회가 되길 기대한다.

마지막으로 이슈페이퍼에 수록된 내용은 해당 주제 집필진들이 토론을 통해 도출한 견해이며, 한국과학기술한림원의 공식적인 의견이 아님을 밝힌다.

2018년 11월

한국과학기술한림원 원장 이 명 철

목 차

새로운 정보기술과 인권

자동화된 프로파일링 등을 통한 개인정보 침해와
인공지능 알고리즘을 이용한 차별을 중심으로

새로운 정보기술과 인권

자동화된 프로파일링 등을 통한 개인정보 침해와
인공지능 알고리즘을 이용한 차별을 중심으로

최근 인공지능, 빅데이터, 사물인터넷, 로보틱스 등 새로운 정보기술이 급속하게 발전하고 있다. 이들은 4차 산업혁명의 핵심이 되는 기술로 개인용 컴퓨터, 인터넷 등으로 대변되는 기존 정보통신기술의 고도화에 따라 새롭게 등장한 정보기술이다. 정부와 민간 모두 새로운 정보기술 개발에 많은 예산을 투자하고 있으며, 이는 앞으로 더 늘어날 전망이다.

한편 이전부터 정보기술이 사회에 미치는 영향에 대한 연구는 긍정적인 측면뿐만 아니라 위험과 그에 대한 대처방안을 고민해왔다. 선행연구들은 정보 기술이 거래비용을 감소시켜 신산업의 가능성을 열어주는 등의 순기능이 있는 반면에, 프라이버시 침해문제와 정보격차 심화 등 사회적 불평등을 초래하는 역기능이 있음을 보여주었다. 이러한 이슈들과 더불어 새로운 정보기술은 또 다른 사회문제를 낳고 있으며, 새로운 형태의 인권 문제를 만들고 있다.

따라서 본 이슈페이퍼에서는 새로운 정보기술이 인권에 미치는 영향을 분석하려 한다. 신기술이 인권 신장에 기여하는 측면에 대해 살펴보고 이와는 반대로 인권을 침해하는 현실과 잠재적 위험성을 검토하였다.

이를 바탕으로 국내의 현황과 문제점을 분석한 결과, 정보기술과 관련된 인권 이슈는 크게 ①자동 프로파일링과 얼굴인식 등을 통한 개인정보 수집문제, ②알고리즘을 통한 차별 등이 있다.

연구진은 이러한 문제점을 해결하거나 완화할 수 있는 규제와 지속가능한 거버넌스 구조를 고안해보았으며, 이를 통해 새로운 정보기술에 의한 인권 침해를 방지하고, 이것이 인권 수호와 고양을 위해 사용될 조건을 모색하고자 한다.

I 서론

1. PC, 인터넷 등의 정보기술과 인권의 문제

1990년대 이후 PC, 인터넷의 확산과 대중화는 이전에는 존재하지 않았던 새로운 사회 문제를 낳음

정보기술에 의한 프라이버시(Privacy) 침해 문제

- 정보기술은 기존의 개인정보를 연동시키거나 새로운 개인정보를 생성하는 것을 용이하게 하지만, 정보를 쉽게 보관 및 공유하고 전송하게 되면서 정보유출, 변조의 위험이 커져 개인의 프라이버시가 침해될 우려가 있음
- 프라이버시 권리는 「세계인권선언」 12조와 「프라이버시 보호와 개인정보의 국제적 유통에 관한 지침」(OECD, 1980)에 의해 정의됨
- 이에 의하면 ① 개인정보의 수집은 합법적이고 정당한 수단에 의해 정보 주체의 인지와 동의 아래 수집 되어야 하고, ② 수집한 개인정보는 처음에 명시된 목적과 다른 목적으로 사용되어서는 안 되며, ③ 유실이나 해킹에 대한 적절한 안전장치가 마련되어야 함

정보 불평등과 인권

- 정보기술의 발전이 '정보 부자'와 '정보 빈자' 사이의 간극을 넓히고 사회적 불평등으로 이어져 경제적 불평등을 악화시킨다는 비판이 있음(DiMaggio and Hargittai, 2001; Hargittai and Hsieh, 2013)

비트(bit)화에 따른 지적 재산권의 침해

- 디지털 시대로 변화하면서 음반, 책 등과 같은 전통적인 매체가 사라지고 저작물의 형태가 바뀌었음
- 저작물의 콘텐츠가 비트화되면서 복제(copy)와 유통이 쉬워져 지적재산권이 침해될 소지가 커짐

2. 최근 새로운 정보기술의 발달과 인권 문제의 부상

☑ 인터넷의 발전에 따라 등장했던 프라이버시, 정보 불평등, 지적 재산권의 침해 등의 문제가 완전히 해결된 것은 아니나 최근 새로운 정보기술의 발달은 아래와 같은 새로운 문제를 야기함

- 경제적 불평등(economic inequality) 심화
- AI 알고리즘 기반의 사회적 편견 구조화
- 데이터 편중
- 사회안전망에 취약한 집단의 차별 가중
- AI 분야의 다양성 결여와 여성·소수자들의 과소대표성(under-representation)

☑ 강한 인공지능(슈퍼인텔리전스)에 대한 우려

- 먼 미래의 일이지만 인간보다 똑똑한 인공지능이 만들어졌을 때, 이것이 인간을 무력하게 할 우려가 있음(커즈와일, 2007; 보스트롬, 2017)

3. 국제기구에 의한 문제제기

☑ 앰네스티 인터내셔널의 「AI의 선한 사용 극대화를 위한 AI 이니셔티브(2017)」는 인권을 위한 AI의 선용을 위해서 AI를 인권 조사 및 연구, 캠페인에 사용할 것을 권장함(Shetty, 2017; Sherif Elsayed-Ali, 2017)

- 앰네스티 인터내셔널은 차별적인 AI의 부정적 영향을 막기 위해 기업 및 정책전문가들과 함께 연구·조사 사업을 진행할 것을 제안하며 다음 4가지 주제에 대해 주목하였음
  - 치안, 형벌체계, 기본적 경제사회서비스에 대한 접근에서의 차별가능성
  - 자동무기체계
  - 고용의 질 저하 등과 같은 자동화의 사회적 영향(ILO, 2016; Siberman and Irani, 2016; De Stefano, 2016)
    * 플랫폼을 이용한 노동의 원격제어 및 평판시스템의 빅데이터에 의한 인사관리 등 전통적인 표준고용의 틀을 벗어나는 노동의 양산
  - 프라이버시와 정보 신뢰 문제

막스 테그마크가 설립한 Future of Life Institute에서 개최한 '2017년 아실로마 컨퍼런스'에서 제정한 인공지능에 대한 윤리 및 가치 원칙을 간단히 정리하면 다음과 같음(테그마크, 2017)

〈인공지능에 대한 윤리 및 가치 원칙, (테그마크, 2017)〉

- 인공지능 시스템은 안전해야 하고, 안전을 검증할 수 있어야 한다.
- 사법 시스템 등에서 인공지능이 사용된다면 권위 있는 인권기구에게 만족스러운 설명을 제공할 수 있어야 한다.
- 인공지능 시스템의 디자이너와 설계자는 인공지능의 사용, 오용 및 행동의 도덕적 영향에 관해 책임과 기회를 가진다.
- 고도로 자율적인 인공지능 시스템은 작동하는 동안 그의 목표와 행동이 인간의 가치와 일치하도록 설계되어야 하며, 인공지능 시스템은 인간의 존엄성, 권리, 자유 및 문화적 다양성의 이상에 적합하도록 설계되어 운용되어야 한다.
- 인공지능 시스템에서 사람들은 그 자신들이 생산한 데이터를 액세스, 관리 및 통제할 수 있는 권리를 가져야 하며, 개인정보를 사용하는 인공지능이 사람들의 실제 또는 인지된 자유를 부당하게 축소해서는 안 된다.
- 인공지능 기술은 최대한 많은 사람에게 혜택을 주고 힘을 실어주어야 하며, 이를 통해 이루어진 경제적 번영은 인류의 모든 혜택을 위해 널리 공유되어야 한다.
- 치명적인 인공지능 무기의 군비 경쟁은 피해야 한다.

Ⅱ 새로운 정보기술의 발전: 인공지능, 빅데이터, 로보틱스 기술을 중심으로

1. 새로운 정보기술의 발전과 신정보사회의 도래

☑ 새로운 정보기술의 정의

- 사물이 인간의 개입 없이 판단이나 의사결정을 하도록 지원하거나, 사물이 스스로 상황판단 등 사고 능력을 갖도록 하기 위해 사물을 지능화하는 데 필요한 각종 요소기술로 사물인터넷(Internet of Things, IoT), 인공지능, 로보틱스, 빅데이터 등이 포함됨

☑ 이러한 정보기술은 하나의 기술이 아니라, 여러 기술들이 그물망처럼 얽혀 있는 기술 시스템인 것이 특징임

☑ 이들은 생산성과 임금 향상에 기여하는 부분이 있으나, 동시에 임금격차를 벌리고 특정 직종에서 일하는 사람들의 직업 수요를 줄일 수 있음

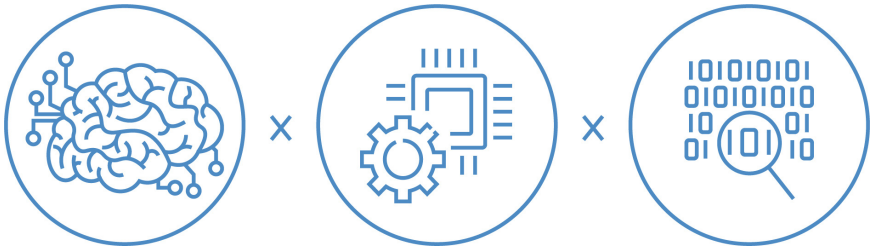
- 자동화에 의한 노동 대체 효과 vs 생산성 효과 간의 긴장이 존재
  - 노동 대체 효과로 노동 수요와 임금을 감소시킴
  - 생산성 효과는 자동화로 인한 비용 절감으로 시장에서 상품의 가격을 떨어뜨리고 새로운 수요를 발생시킨다는 것임
  - 또한 생산성 효과는 부가적인 자본축적과 현존하는 기계의 개선에 의해 보완되며, 이 두 가지는 다시 노동 수요를 증가시킬 수 있음 (Acemoglu and Restrepo, 2018)

2. 인공지능 · 빅데이터 · 로보틱스 기술의 범용화

- 컴퓨터·인터넷에 이어 사물인터넷(IoT)기술이 범용기술로 부상하였으며, 그 뒤를 이어 인공지능(AI)·로보틱스·빅데이터 등의 기술이 부상

- 특히, 2010년 이후 **인공지능기술**의 급속한 발전이 두드러졌는데, 이는 △딥러닝 △컴퓨팅 파워 △빅데이터 등 세 가지 기술 발전이 수렴되면서 가능해 짐

〈인공지능 발전의 3가지 핵심요소〉



딥러닝 방법론 등장	컴퓨팅 파워 혁신	빅데이터
■ 기존 이론의 한계 극복 및 변화의 시발점 ■ 인공지능 SW 공개로 기술 개발 가속화	■ GPU, 분산처리 환경 진화 ■ 클라우드 접속에 의한 실시간 데이터 실현	■ 모바일, 센서 등의 소형화, 저비용화 등으로 IoT 산업 진전 ■ 수집 가능한 데이터 증대

- 딥러닝 방법론
  - 딥러닝(Deep Learning): 다층구조 형태의 신경망을 기반으로 하는 머신 러닝의 한 분야로, 다량의 데이터로부터 높은 수준의 추상화 모델을 구축하고자 하는 기법
  - 다수의 인공신경(artificial neuron)으로 하나의 층(layer)을 구성하고, 다수의 층이 입력 데이터로부터 출력 결과물을 산출하는 중간 과정에서 여러 개의 은닉층(hidden layer)으로 기능함. 각 층마다 데이터를 표상(representation)하는 수준이 이전 층보다 더 추상화되도록 함으로써, 실세계를 표상하는 개념들을 계층화하여 이해하는 기법임(Goodfellow et al., 2016)
  - 기존 인공신경망(Artificial Neural Network) 이론이 갖는 과대적합(overfitting) 등의 한계를 극복하면서 변화의 시발점이 됨
  - 머신러닝을 훈련시킬 수 있는 인공지능 소프트웨어 라이브러리와 프레임워크의 공개로 기술 개발이 가속화되고 있음

• 예) 구글의 텐서플로우(TensorFlow), 페이스북의 파이토치(PyTorch), 아마존 AWS에서 후원하는 Apache MXNet, 마이크로소프트 리서치의 Microsoft Cognitive Toolkit 등

- 컴퓨팅 파워 혁신
  - GPU(Graphics Processing Unit)의 발달로 분산처리 환경이 개선됨에 따라 컴퓨팅 파워가 급속히 향상됨
  - 1,000개가 넘는 카테고리로 분류된 100만개의 이미지를 인식하는 경진대회인 'ImageNet Large Scale Visual Recognition Challenge'에서 2012년 딥러닝 기반의 AlexNet이 합성곱신경망(Convolution Neural Network)을 고안하여 적용하였고 이 모델을 학습하는 데에 GPU를 활용한 결과, 오류율(error-rate) 15.3%를 보이며 이전까지 최저 오류율인 25%를 획기적으로 개선함. 이는 딥러닝 분야에서 GPU 활용이 보편화되는 초석이 됨(NVIDIA Korea, 2016)
  - 최신 GPU는 5천 여개의 코어가 있기 때문에 단시간에 엄청난 수준의 컴퓨팅 파워를 가지게 되고 클라우드 접속을 통해 실시간 데이터 처리를 가능하게 함
- 빅데이터
  - 2010년대 빅데이터 기술의 발달은 딥러닝모델 훈련을 위한 대량 데이터를 모으는 데에 크게 기여함
    * 딥러닝모델을 훈련시키기 위해서는 막대한 연산 시간을 줄여줄 수 있는 컴퓨팅 기술뿐만 아니라, 훈련에 투입될 대규모의 훈련 데이터가 필요함
  - 빅데이터기술은 모바일 기기의 보급화로 데이터 수집이 용이해 짐에 따라 크게 발전함
    * 애플의 iOS(2007년~), 안드로이드 OS(2008년~) 등으로 구동되는 모바일 기기가 급속히 보급되고 센서 등이 소형화되어 모바일 기기에 다양한 센서 탑재가 가능해짐에 따라 데이터 수집이 용이해짐
  - 많은 양의 데이터는 범세계적인 소셜네트워크시스템(SNS) 등을 활용하여 수집 및 공유가 가능해짐

로봇, 자율주행자동차 등 지능을 갖춘 기계의 확산

인공지능 분야에서 음성인식, 언어인식, 사물인식 기술의 급속한 발전에 따라 의료, 금융, 사법 분야에서 인간의 판단을 대신하거나 보완하는 인공지능 알고리즘 시스템이 확산됨

이러한 정보기술의 발전은 가능성과 함께 새로운 차별 문제를 낳고 있으며, 차별을 통한 인권 침해 문제가 사회적 이슈로 부상함

3. 생산방식 · 산업이나 고용구조, 기술 혁신 등 경제사회 전반에서 변화가 촉발

☑ 기술계에서는 플랫폼과 인공지능 알고리즘의 중요성이 커지고 있으며, 산업계에서는 직업의 변동에 따른 계층 변화가 심화되고 있음

- 사라지는 직장과 새로 생기는 직장의 변동이 빠른 시간 안에 일어나지만(포드, 2016; 브린욜프슨·맥아피, 2014) 산업사회에서 만들어진 교육이나 사회 안전망 등이 이에 충분히 대응하지 못함

- 2017년 맥킨지글로벌 연구소에서 자동화 기술의 발전과 적용이 미래 고용과 생산에 미칠 영향을 분석한 결과, 모든 업무가 자동화 될 직업은 전체의 5% 미만이고, 대부분 경제활동을 차지하는 물리적 활동(81%), 데이터 프로세싱(69%), 데이터 수집(64%) 등은 크게 영향을 받을 전망임

  - 15조 8,000억 달러의 임금을 받는 11억 9,000만 정규직 인력이 자동화 영향을 받을 것으로 예측되며, 강도는 국가별로 상이할 전망이나 중국, 인도, 일본, 미국 등 4개국이 자동화 영향을 크게 받을 것으로 예상됨[1)]

표 Ⅱ-1. 자동화 가능성이 높은 직업·업무

직업·업무	향후 예상되는 자동화 비율(%)
예측 가능한 신체적 업무(기계 작동, 청소, 패스트푸드 조리, 수금·행정보고 등)	81
데이터 처리	69
데이터 수집	64
예측 불가능한 신체적 업무(정원사, 배관공, 아동·노인 돌봄이 등)	26
소통·면담 업무	20
전문직	18
관리직	9

자료: 멕킨지 글로벌연구소(MGI)

- 이에 따라 자동화에 따른 고용변화 대응이 선진국의 주요 정책의제로 떠오름

  - 미국 백악관은 AI기반 자동화에 따른 정책 입안자가 대비해야 할 주요 경제 효과로 △업종, 임금 수준, 교육 수준, 직종 및 지역별로 영향의 고르지 않는 분산과 △다른 일자리가 창출되는 동안 일부 일자리의 소멸에 따른 취업 시장의 혼란 △단기 근로자의 일자리 상실과 정책 대응 수단에 대한 의존 장기화 등의 문제를 지적하였음

1) 미국, 자동화기술이 고용과 생산성에 미치는 영향 분석, S&T GPS, 2017.2.

- 백악관은 2016년 보고서를 통해 AI시대의 고용 없는 성장에 대한 대응방안을 발표함(The White House, 2016)
- 유럽의회에 2016년 발의된 권고안에서는 △가장 진화된 자율 로봇에게 '특수한 권리와 의무를 가진 전자인간'(electronic persons with specific rights and obligations)이란 법적 지위를 부여할 것과 △로봇 및 AI로 인해 일자리를 위협받을 사람들을 지원하기 위해 로봇이 수행한 일에 세금을 매기거나 로봇을 사용하고 유지·보수하는 데에 비용을 청구하는 방안을 검토할 것을 권고하였음. 2017년 2월 유럽 의회는 해당 권고안을 통과시키며 전자는 채택하였으나, 후자는 채택하지 않았음(European Parliament, 2017)[2)]

● MIT의 Acemoglu와 보스턴 대학의 Restrepo 교수의 연구에 따르면 예상보다 로봇과 자동화를 통한 공학 및 금융업 등에서의 고용 창출 효과는 적었으며, 제조업에서의 일자리 상실을 상쇄할 정도로 큰 고용 창출이 일어나지 않았음(Acemoglu & Restrepo, 2017)

- 1990년에서 2007년 사이 미국 내에서 로봇기술과 자동화로 인해 사라진 제조업 일자리는 총 67만개에 달하며, 이러한 경향은 앞으로 로봇 도입이 본격화됨에 따라 더욱 강해질 것으로 전망됨
- 근로자 1,000명당 로봇 1기 도입은 지역 사회 내 평균 6.2개의 일자리와 0.7%의 임금 하락을 야기하는 것으로 분석됨
- 국가 전체로 살펴볼 경우 로봇기술과 자동화가 미치는 부정적 효과는 다소 줄어들지만, 여전히 피해를 입는 근로자가 발생할 것임

▣ 따라서 우리나라도 산업이나 고용구조 등의 변화를 고려하여 신정보사회에 선제적으로 대응할 수 있는 정책을 마련해야 함

● 인공지능 기술과 소프트웨어 교육 지원, 미래의 직업 변화를 위한 교육 도입, 재교육 프로그램 확충 등이 필요함

● 직장을 잃을 것으로 예상되는 분야의 근로자들을 지원하는 사회 안전망을 확충하여 정보기술의 혜택이 특정 계층에게 국한되는 것을 방지해야 함

2) Reuters (2017) "European parliament calls for robot law, rejects robot tax" (2017.2.16.)

4. 신경망 인공지능의 의사결정에 내재한 불확실성과 신뢰의 문제

☑ 신경망 인공지능은 인공지능이 왜 그러한 결정을 내렸는지에 대해 설계한 사람도 알 수 없는 경우가 있다는 문제가 있음

- 뉴욕대의 심리학자이자 인공지능 학자인 게리 마쿠스 (Gary Marcus)는 현 시점 까지의 딥러닝의 10가지 인지적 한계를 다음과 같이 정리함 (Marcus, 2018)[3)]

① 데이터를 너무 많이 필요로 한다
② 지식을 얕게 배우고, 훈련된 도메인 이외의 다른 도메인에 적용하는 데에 한계가 있다
③ 계층적 구조를 다루는 자연스러운 방법을 갖고 있지 않다
④ 개방형(open-ended) 문제 추론에서 난항을 겪고 있다
⑤ 충분히 투명하지 못하다
⑥ 선 지식(prior knowledge)과 잘 융합되어 있지 않다
⑦ 인과관계와 상관관계를 선천적으로 분간하지 못한다
⑧ 문제를 일으킬 수 있는 방식으로 대체적으로 정적인 환경을 가정한다
⑨ 추정으로서 잘 동작하나, 추정 결과를 전적으로 신뢰할 수 없는 일이 있다
⑩ 엔지니어링 하기 어렵다

- MIT 컴퓨터과학 및 인공지능 연구소(MIT Computer Science & Artificial Intelligence Lab)의 연구는 사람이 AI를 속일 수 있음을 보여줌
  - 해당 연구는 인공지능의 오용 가능성을 지적함(Athalye et al., 2018; MIT CSAIL, 2017)
  - 연구진은 시각인식 인공지능이 총의 사진을 총이 아니라고 분류하게 속이거나, 질감을 미묘하게 바꾼 사물(예: 3D 프린터로 인쇄한 장난감 거북, 야구공)을 전혀 다른 종류로(예: 장총, 에스프레소) 인식하도록 속이는 것이 가능함을 증명함
- MIT 미디어랩의 연구는 사물의 부정적인 면을 학습한 AI는 질문에 대해서 부정적인 답을 하는 것을 보여주었는데, 이는 데이터의 편향이 AI의 편향을 가져올 수 있다는 것을 시사함

3) MIT Technology Reivew (2017) "The Dark Secret at the Heart of AI" (2017.4.11.) by Will Knight. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

■ 이에 따라 인공지능 연구자, 실리콘밸리 기업 등에서 인공지능의 투명성(transparency), 책임성(accountability), 설명가능성(explainability)을 높이려는 시도가 있음

- 미 국방부: 설명 가능한 인공지능 프로젝트(DARPA, 2016)
  - 머신러닝 모델이 내린 의사결정이 설명가능해지도록 함
  - 최신 기법의 인간-컴퓨터 인터페이스를 활용하여 대화하는 형태로 최종 인간 사용자에게 제공하는 것을 목표로 함
- NVIDIA: 화면 내에서 자율주행차량의 딥러닝 시스템에 가장 크게 기여하는 영역을 시각적으로 강조하는 방법을 개발(Bojarski et al., 2017)[4]
  - 연구 결과, 딥러닝 신경망은 인간이 관심을 기울이는 대상과 동일하게 도로의 경계선, 차선 표시, 주차된 차량 등에 대해 가장 크게 집중하는 것으로 밝혀짐

■ 국내에서도 이러한 논의가 시작되었으며, 정부에서 **'설명 가능한 인공지능의 개발'**을 대규모로 지원하고 있음

- 2017년 9월 울산과학기술원(UNIST)은 설명가능 인공지능 연구센터를 개소하였고 과기부에서 2021년까지 150억 원의 예산을 지원할 예정임
  - 최재식 교수 연구팀은 "학습, 추론이 가능한 AI" 세부 과제를 수행하기 위해 데이터의 인과관계를 분석해 의사결정에 적절한 이유를 제공함으로써 '사람이 이해할 수 있는 형태로 설명하는 소프트웨어를 개발'하고 있음[5]

4) MIT Technology Review (2017) "Nvidia Lets You Peer Inside the Black Box of Its Self-Driving AI" (2017.5.3.)
https://www.technologyreview.com/s/604324/nvidia-lets-you-peer-inside-the-black-box-of-its-self-driving-ai/

5) 울산과기대 (2017) "의사결정 설명·사람과 상호작용하는 인공지능 연구 본격화" (2017.9.25.)
http://news.unist.ac.kr/kor/newsletter/20170925/

5. 인공지능의 무기화

- 인공지능의 무기화는 인공지능을 이용해 재래식 무기를 제어하는 것과 음성인식, 영상인식을 이용한 감청 등 인공지능이 새로운 형태의 군사 기술로 사용되는 두 가지 유형이 있음
  - 2018년 4월, 해외 학계에서 카이스트와 한화시스템의 국방 인공지능 융합연구센터의 연구의도에 대한 우려를 표명하면서 카이스트와의 관계를 끊겠다고 함(boycott). 이에 카이스트는 "인공지능 분야 관련 연구에서 대량 살상 무기나 공격용 무기 개발 계획이 없으며, 통제력이 결여된 자율무기를 포함해 인간 존엄성에 어긋나는 연구 활동을 수행하지 않을 것"이라고 해명함으로써 보이콧이 철회됨[6)]
  - 얼굴인식 인공지능은 공공 CCTV 카메라 영상에서 개별 시민을 식별하는 것이 가능해 대테러 방지 등 군사기술에 응용될 수 있음
    - 중국은 얼굴인식 CCTV 시스템으로 수배중인 범인을 검거하고 무단횡단을 단속하는 용도로 활용하고 있음
    - 미국 캘리포니아 주 올랜도 경찰국은 아마존의 얼굴인식 소프트웨어 "Rekognition Service"를 실시간으로 제공받는 파일럿 프로그램을 실행하기로 하였음
    - 이 프로그램은 용의자의 얼굴을 특정하면 과거 행적을 재구성할 수 있는 기능이 있기 때문에 향후 CCTV 카메라 등에 찍히는 불특정 다수의 시민의 얼굴을 식별해 용의자를 탐색·추적하는 기법으로 발전할 수 있음[7)]

- 우리나라는 안보분야의 특수성이 있기 때문에 인공지능을 군사기술에 활용하는 것을 완전히 막기는 어려우므로 이에 대한 사회적 논의와 토론을 통해 합의를 얻는 과정이 중요함

---

6) CIO Korea (2018) "살상 로봇 보이콧 '철회'… 카이스트, 치명적인 자율 무기 개발 않겠다" (2018.04.11.) http://www.ciokorea.com/news/37881

7) 연합뉴스 (2018) "미 올랜도 경찰, 아마존 리얼타임 얼굴인식 시스템 도입" (2018.5.23.) http://www.yonhapnews.co.kr/bulletin/2018/05/23/0200000000AKR20180523010200075.HTML

Ⅲ 정보기술이 낳는 인권의 문제

1. 알고리즘에 의한 차별가능성의 문제

▣ 인공지능 알고리즘 응용 범위 확대와 부작용의 발생

- 산업, 의료 분야에서 금융, 치안, 사법, 행정까지로 응용범위가 확대
- 미국이나 유럽의 몇몇 국가에서는 프라이버시 침해, 개인정보남용을 넘어 소수자에 대한 차별문제가 이슈가 되고 있음
- 캐시 오닐(Cathy O'Neil)은 이러한 차별가능성 문제를 분석해 『대량살상 수학무기』를 출간, 아마존 베스트셀러에 오름
- 2017년 뉴욕은 시에서 사용되는 모든 알고리즘이 책임성을 가진(혹은 해명 가능한, accountable)것이 되어야 한다는 법안 발의하였으며, 미국에서 최초의 선례가 됨[8)]

▣ 알고리즘에 의한 차별(Gillispie 2014)은 왜 발생하는가

- 대부분 인공지능 알고리즘은 인간의 판단을 보완하거나 대신하며, 초기에 알고리즘은 인간의 판단에 개입하는 여러 가지 선입견, 편견 등을 극복하기 위해 도입되었음(Caplan et al., 2018)
- 알고리즘 도입이유에는 소수자에 대한 편견극복도 포함되어 있음
- 그러나 빅데이터의 처리 과정에서 데이터 편향 등 차별적인 요소가 포함될 수 있는데, 이는 다음의 4가지 유형이 있음(Barocas & Selbst, 2016; Kroll, Barocas, Felten, Reidenberg, Robinson, & Yu, 2017)

- 목표 변수를 정의하는 과정에서의 편향
- 훈련 데이터를 레이블링하는 과정에서의 편향
- 특징 선택의 사용에서의 편향
- 특정 문제를 풀기위한 판단지표 자체가 특정한 계층을 분간하는 대리지표가 되는 경우

8) https://www.businessinsider.com/algorithmic-bias-accountability-bill-passes-in-new-york-city-2017-12

- 사용자는 이러한 문제를 해결하기 위해 데이터를 더욱 세부적으로 수집하는데, 이것이 다시 특정 계층에 편향적인 데이터를 생산하기도 함

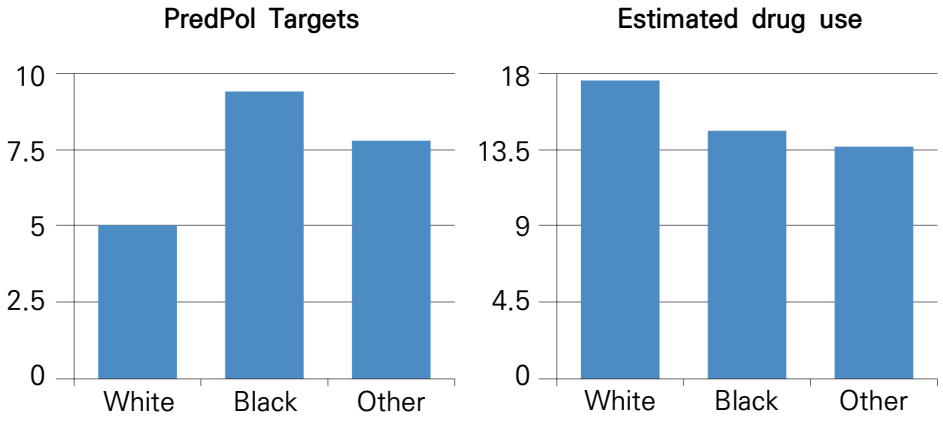
☑ 미국 경찰이 사용하는 '예측적 순찰(predictive policing)' 프로그램 PredPol 알고리즘의 차별사례(Perry et al., 2013; Moses and Chan, 2016)

- 이 알고리즘은 경범죄를 데이터에 포함시켰기 때문에 경범죄가 많이 발생하는 지역인 빈민가, 흑인 밀집 거주지가 집중 순찰의 대상이 되었으며, 자동차 절도 등과 같은 재산범죄(property crime)에 집중하고 있음
- 경찰이 알고리즘이 선택한 지역을 위주로 순찰하다보면 더 많은 흑인을 검문하고 대마초 소지자, 흉기 소지자 등을 검거하게 되는 결과를 초래함
- 중산층 이상의 백인들이 주로 사는 지역은 순찰지역에서 제외되어 이들이 저지르는 화이트칼라 범죄는 순찰의 표적에서 제외될 수 있음

그림 Ⅲ-1. 예측적 순찰 프로그램 PredPol 알고리즘의 차별사례

Predictive policing in Oakland vs. actual drug use

The chart on the left shows the demographic breakdown of people targeted for policing based on a simulation of PredPol in Oakland. The chart on the right shows actual estimate use of illicit drugs.



source: National Survey on Drug Use and Health, Human Rights Data Analysis Group

미국 사법부에서 사용하는 COMPAS 알고리즘의 공정성에 대한 논란[9)]

- 미국은 법원에서 교도소에 수감된 죄수의 보석과 가석방을 결정하고자 하는 경우나, 판사가 형량을 결정할 때 광범위하게 노스포인트(Northpointe)사의 COMPAS 알고리즘을 사용함
- 뉴욕의 탐사매체인 ProPublica는 플로리다 주 Broward 카운티에 거주하고 있는 시민 7000명 이상을 대상으로 알고리즘의 예측결과와 실제 재범여부를 조사하였음
- 그 결과, 재범률이 높다고 예측되었지만 실제로 2년간 범죄를 저지르지 않은 흑인이 45%로 백인(23%)의 두 배에 달한다는 점, 그리고 재범률이 낮은 것으로 예측되었지만 실제로 2년간 범죄를 저지른 확률이 백인이 48%로 흑인(28%)보다 훨씬 높다는 점이 밝혀져 논란이 되었음(Angwin et al., 2016)
- 이에 대해 Northpointe사는 다음과 같이 반론함(Northpointe Inc. Research Department 2016)
  - 조사의 목적은 재범자를 예측하는 것이며, 이에 있어 백인과 흑인에 차별적인 결과는 나타나지 않았다고 반론함
  - 알고리즘 분석결과, 위험점수가 높다고 분류되고 실제로 재범한 사람의 확률은 백인 59%, 흑인 63%로 유의미한 차이가 나지 않았음
  - 반면에 위험 점수가 낮다고 분류되고 실제로 2년 안에 재범하지 않은 사람의 확률 역시 백인 71%, 흑인 65%로 크게 차이를 보이지는 않았음
  - 일반적으로 흑인의 일반범죄 재범률은 51%로 백인의 39%에 비해 높으며, 흑인의 강력범죄 재범률 역시 14%로 백인의 9%에 비해 높게 나타남
  - 이러한 기본 구성비율(base rate)에서의 흑인과 백인의 차이가 ProPublica 측이 제시한 결과에 영향을 끼친 것이라 볼 수 있음
- 이런 논쟁은 '공정성'(fairness)이란 무엇인가에 대한 사회적 논쟁을 유발함 (Chouldechova, 2017; Grgić-Hlača et al., 2018)

9) https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

아마존 얼굴인식 시스템의 오인식 문제(ACLU, 2018)[10]

- 미국시민자유연맹(ACLU: American Civil Liberties Union)이 아마존 얼굴인식시스템 '레코그니션(Rekognition)API'의 오인식 문제를 지적하였음
- 연맹은 이 시스템에게 범죄자 2만5천명의 공개 머그샷 데이터를 훈련시킨 뒤, 미국 의회의원 535명의 머그샷 중 범죄자와 일치하는 인물을 찾도록 하였음
- 그 결과 이 시스템은 의원 중 28명을 범죄자로 인식하였으며, 범죄자로 오인식된 28명 중 39%가 유색인종인 것으로 나타나 인종차별적인 결과가 큰 이슈가 되었음(※ 실제 연방의원 중 유색인종 비율은 20%)
- 아마존은 이에 대해 해당 테스트가 아마존이 법 집행 기관에 권장하는 신뢰도 95% 이상이 아닌 핫도그, 의자, 동물 등의 사진에 허용되는 80% 신뢰도 수준으로 진행되었기 때문이라고 해명함
- 이에 다시 ACLU는 레코그니션시스템 설치 과정에서 해당 권장 내용이 나오지 않는다는 점과 법 집행기관이 기본 설정 값 80%로 설정해 사용하는 것을 막을 수 있는 방법도 없다는 점을 지적하였음
- 이후 미국 연방 의회 의원들은 제프 베조스 아마존 최고경영자에게 얼굴인식 기술의 효능 및 영향을 설명하라는 서한을 보내는 등 즉각적으로 대응하였음

금융 분야에서 사용하는 알고리즘의 문제

- 도입영역 : 신용, 보험, 로보 어드바이저(Robo-advisor) 서비스
  - 신용 평가회사(Equifax, CallCredit)의 경우에는 신용점수 평가에 알고리즘을 사용할 수 있으며, 대부자들도 돈을 빌려줄 사람과 이자율 선택시에 활용 가능함
  - 보험계는 잠재 고객에 대한 평가에 알고리즘을 사용할 수 있음
  - 로보 어드바이저(robo-advisor)는 낮은 최저 투자액, 싼 수수료 때문에 인기가 있는데, 실제 이들은 투자 포트폴리오에 대해 조언하지만 소비자의 망각, 잘못된 결정 등에 대한 대비가 없음
  - 또한 로보 어드바이저를 이용하는 소비자는 자신이 받는 서비스에 대해 잘 모르고 있는 경우가 많음

10) ZDNet Korea (2018) "황당한 아마존 얼굴인식시스템… '美의원 28명 범죄자로 인식'" (2018.7.27.) http://www.zdnet.co.kr/news/news_view.asp?artice_id=20180727101046

그 동안 금융 분야에서 제기된 알고리즘의 위험성[11]

- 가격차별(Mcsweeny and O'dea, 2017)
- 신용도 평가에서 금융 배제로 이어질 위험
- 알고리즘의 예측 오류: 상업적 목적과 인간의 본래적 편향에 의해 알고리즘이 교육됨
- 로보 어드바이저의 경우에는 법률 변경 등에 의한 시장 전체의 변동에 따른 체계적 위험(Systemic risk)이 존재[12]

문제 해결을 위한 제언(Andrews et al., 2017)

- 인공지능 알고리즘의 특성상 결정을 어떻게 설명할 수 있을지 불명확하기 때문에 감독(Monitoring)이 매우 중요함
- 알고리즘 윤리위원회를 감독기관과 서비스제공자로 구성하여, 알고리즘을 통한 결정에 대해 논의하고, 설명할 수 있도록 하는 것이 바람직함
- 감독기관의 조사가 가능하도록 알고리즘을 열어 놓는 것이 중요

11) 공공 영역과 민간 경제 영역에서 의사결정시 알고리즘 사용에 대한 영국 의회 Science and Technology Committee의 의견 요청에 대한 Financial Services Consumer Panel의 의견서 https://www.fs-cp.org.uk/sites/default/files/fscp_response_stc_algorithms_inquiry.pdf

12) SCM Direct (2016). Fintech Folly: the sense and sensibilities of UK robo-advice https://scmdirect.com/press-and-videos#block-views-resourcesscm-research-tab; Storey, A. (2016). How to monitor robo-advice https://www.linkedin.com/pulse/how-monitor-robo-advice-andrew-storey

2. 인공지능을 이용한 자동화의 문제

◪ 자율주행자동차의 문제

- 맥킨지 보고서(McKinsey & Company, 2016)에 따르면 자율주행자동차가 사고를 90% 감소시킬 수 있다고 보고되었으나 소위 "트롤리 문제"(Trolley Problem)가 남아 있음
- 이는 보행자나 운전자의 생명이 달린 상황에서 자율주행 자동차의 주행을 어떻게 프로그래밍 할 것인가의 문제임
  - 인공지능에 가치를 입력하는 문제
  - 자율주행자동차 확산에 매우 중요한 문제 (※많은 응답자가 사망확률이 같을 때에는 운전자를 희생하는 방향으로 프로그래밍되어야 한다고 생각하지만, 이럴 경우에 차를 사겠다는 비율이 매우 낮아짐)
- MIT에서 투표 기반의 윤리 학습 시스템인 Moral Machine 사이트를 운영하고 각국의 반응의 차이를 연구 중이며, 이 연구는 사회적 합의의 중요성에 근거하고 있음
- 2017년 독일 정부는 자율주행자동차와 관련된 인권 이슈에 대해 입장을 내어 놓았는데 (Lüge, 2017), 크게 두 가지 원칙으로 구성됨

  ① 인간의 생명을 보호해야 함

  ② 인종, 직위, 성별, 나이 등을 차별해서는 안 됨(모든 인간 생명은 동등하게 존엄)
  - 해당원칙은 차별을 하지 않는다는 원칙을 지켰지만, 사회구성원의 평균적인 인식과 같지는 않음

◪ 자동화와 직업의 감소 문제

- 기존 직장의 감소 경향과 새로운 직장의 생성
  - 변화의 속도가 얼마나 빠를 것인가에 대해서는 학자들 간 이견이 있으나, 멀지 않은 미래에 직장 생태계에 큰 변화가 있을 것임은 분명함(고용정보원, 2017)
- 고용정보원은 자동화로 인해 사무직, 생산직 직장의 감소 가능성을 예견하였으며, 이는 중산층의 붕괴, 민주주의의 근간 와해 가능성과 연관이 있기 때문에 사회적으로 매우 심각한 문제임
  - 자동화로 인해 직장을 잃을 수 있는 직업에 종사하는 사람들의 재교육의 문제가 사회적 이슈로 대두됨

3. 프라이버시

☑ 인공지능, 빅데이터 시대의 프라이버시 문제는 개인이 아닌 전체 사회구조적, 계층적 문제임

- 그 이유는 정보제공자와 수집자가 불분명하고 정보주체의 권한이 모호해 기존의 동의모델 적용이 어렵거나 불가능하기 때문임
- 최근 많은 이들이 개인이 적어도 자신에게 악영향을 미칠 수 있는 데이터의 정확성은 확인할 수 있어야 한다고 목소리를 높였고, 이에 개인정보 보호를 위한 새로운 법률과 가이드라인 등이 만들어지고 있음

표 Ⅲ-1. 2017년 국내외 프라이버시 유출 이슈

국내이슈 TOP 5	해외이슈 TOP 5
1. 가상화폐 거래소 해킹, 개인정보 유출, 거래소 폐쇄	1. GDPR, 핵심은 개인정보의 보호와 활용의 균형
2. 해커 한명의 PC에서 3천 3백만건의 개인정보 발견	2. 중국 네트워크보안법 시행 (중국에서 생산, 수집된 개인정보, 데이터는 중국 내 저장)
3. KISA 갱니정보 유출 사고 분석 보고: 해킹과 내부자의 고의유출	3. NYCRR 500 뉴욕주의 모든 금융기관에 적용되는 새로운 사이버보안 규정
4. 소상공인과 대리점의 개인정보 관리 문제 지속	4. 야후 개인정보 30억명 유출
5. 개인정보를 불모로 하는 랜섬웨어, 북한발 해킹의 지속	5. 미국 신용평가기관 에퀴팩스 개인신용정보 유출

☑ 프라이버시 문제와 관련하여 고객의 정보를 제3자(third party)에게 제공하는 현재의 구조는 개인정보유출에 취약한 구조임

- 개인의 중요 데이터를 보유한 소수의 기술업체가 제3의 업체에게 API 등을 통해 고객 데이터 활용 서비스를 제공하는 구조는 해킹에 의해 보안 취약점이 뚫리게 되면 개인정보가 쉽게 유출될 수 있음
- 2017년 9월, 미국 내 3대 개인 신용정보 관리업체이자 8억 2000만명의 개인 신용정보를 보유한 업체인 에퀴팩스(Equifax)로부터 1억 4300만명의 고객들의 개인 정보가 유출된 것으로 밝혀짐[13]

13) 뉴스1 (2017) "에퀴팩스 '미국내 1억4300만명 개인정보 해킹됐다'" (2017.9.8.)

- 2017년 5월부터 7월까지 어느 단체가 웹사이트의 취약 부분을 공격하여 해킹함으로써, 사회보장번호(Social Security Number), 생일, 주소, 일부 자동차 면허 번호와 신용카드 번호 등 민감한 개인 정보를 편취함
- 해당 단체가 편취한 정보의 규모는 미국 인구의 약 절반에 달하는 수치이며, 피해자들은 신분도용, 금융 계좌 해킹, 신용도하락 등의 위험에 처하게 되었음[14]
- 회사는 7월 29일에 해킹 사실을 인지했지만, 이를 공개적으로 시인한 9월 7일까지 6주 동안 고객에게 경고하기 보다는 오히려 피해자들에게 유료 서비스인 '신용감시 서비스'에 가입하도록 유도한 정황이 드러남[15]
- 이후 2018년 2월에는 유출가능성이 있는 데이터의 범주가 기존에 알려진 것 보다 훨씬 광범위하다는 것이 추가로 밝혀지면서, 에퀴팩스의 고객 프라이버시 관리에 대한 불투명하고 폐쇄적인 태도가 문제가 됨[16]
  - 공격자들이 접근할 수 있었던 정보에는 세금 식별번호, 이메일 주소, 여권 번호 등이 포함되어 있었음이 뒤늦게 밝혀짐
- 많은 양의 주요 개인정보를 얻게 되면 개인을 사칭하는 것이 더욱 쉬워지게 됨
- 그러나 이와 같은 사건의 재발 방지를 위해 구조적인 변화를 가하는 것이 쉽지 않고, 해당 사건이 에퀴팩스 등 신용평가사의 향후 사업에 큰 지장을 주기 어렵다는 것이 문제임
  - 식품, 의약품, 완구, 다른 소비재와 달리 고객 데이터를 부주의하게 다룬 기업을 처벌하는 형법이나 민법 조항은 많지 않음
  - 신용평가사 입장에서는 고객 정보 보호에 대한 동기부여가 없음
    * 신용평가사의 주 사업 모델은 정보수집의 대상인 일반 소비자에게 서비스를 파는 것보다는 은행, 주택 담보 대출자, 그리고 마케팅 업체들에게 신용평가 데이터를 판매하는 것임(2017년 매출 31억 달러 중 거의 3분의 2가 이들을 통해 발생)
  - 더불어 소비자로서는 신용점수가 필요하기 때문에 신용평가사 서비스 이용을 포기할 수도 없는 실정임
    * 현대사회에서 신용평가 점수가 없다는 것은 크게는 주택이나 차량 구매에서, 보다 일상적으로는 신용카드 사용에 이르는 광범위하고 필수적인 금융 서비스를 포기하는 것과 마찬가지이기 때문임

14) 포춘코리아 (2018) "일반 고객 vs 신용업계 지배자들" by Jen Wieczner, John Roberts (2018년 1월)
15) 포춘코리아 (2018).
16) ITWorld Korea (2018) "'에퀴팩스 해커, 알려진 것보다 더 많은 데이터를 훔쳤다'…미 상원의원" (2018.2.14.)

- 2018년 9월 28일, 페이스북은 해커 조직이 2900만 명 사용자의 이름, 전화번호, 이메일 주소에 접근한 것을 확인했다고 밝힘.[17)]
  - 해킹은 9월 14일부터 25일까지 11일 동안 이루어졌으며, 40만 명의 계정 통제권을 덮어쓴 뒤 이를 기반으로 정보를 해킹함
  - 페이스북이 이를 인지한 이후 이틀 간 자체 조사를 진행한 결과, 약 절반인 1400만 명의 경우 상기의 3가지 정보 외에 더 민감한 정보(연락처 정보, 성별, 구사하는 언어, 종교, 친구 관계·지위, 최근 로그인 정보와 검색기록, 사용하는 디바이스 유형 등)가 해커들에게 노출된 것으로 밝혀짐

■ GDPR(General Data Protection Regulation)[18)]: 개인정보 보호와 관련된 유럽연합의 단일 법률

- GDPR은 삭제권, 개인정보 이동권, 자동화된 결정(프로파일링 포함) 관련 권리 등을 새롭게 도입해서 정보주체의 권리를 강화함(표 Ⅲ-2 참조)
- GDPR의 프로파일링등 인공지능에 의해 자동화된 의사결정에 대한 규율은 GDPR 제22조에 명시되어 있음

표 Ⅲ-2. GDPR의 강화된 정보주체 권리

No	정보주체의 권리	관련 조문
1	정보를 제공받을 권리(Right to be informed)	제13조, 제14조
2	정보주체의 열람권(Right of access by the data subject)	제15조
3	정정권(Right of rectification)	제16조
4	삭제권(Right of erasure. 잊힐 권리)	제17조
5	처리에 대한 제한권(Right of restriction of processing)	제18조
6	개인정보 이동권(Right to data portability)	제20조
7	반대할 권리(Right to object)	제21조
8	자동화된 결정 및 프로파일링 관련 권리 (Right to related to automated decision making and profiling)	제22조

자료: Europeen union(2016)

17) 전자신문 (2018) "페이스북 '사용자 2900만명 개인정보, 해커들에 뚫렸다'" (2018.10.14.)

18) 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 Regulation)

- 자동화된 의사결정만을 근거로 한 결정 금지: “법적 효과(legal effects)나 중대한 영향(significantly affects)을 미치는 결정을 자동화된 처리로만 내리는 것(a decision based solely on automated processing, including profiling)은 금지(22조 제1항)”
- 자동화된 결정이 허용되는 경우에도, 특별한 범주의 개인정보 (민감정보, special category)에 근거한 결정은 명백한 동의 또는 법률에 근거하여 상당한 공익상의 이유로 필요한 경우로 보호조치가 마련되어지지 않는 한 허용되지 않음(22조 4항)

제22조 프로파일링 등 자동화된 개별 의사결정

1. 개인정보주체는 프로파일링 등, 본인에 관한 법적 효력을 초래하거나 이와 유사하게 본인에게 중대한 영향을 미치는 자동화된 처리에만 의존하는 결정의 적용을 받지 않을 권리를 가진다.
2. 결정이 다음 각 호에 해당하는 경우에는 제1항이 적용되지 않는다.
   (a) 개인정보주체와 개인정보처리자 간의 계약을 체결 또는 이행하는 데 필요한 경우
   (b) 개인정보처리자에 적용되며, 개인정보주체의 권리와 자유 및 정당한 이익을 보호하기 위한 적절한 조치를 규정하는 유럽연합 또는 회원국 법률이 허용하는 경우
   (c) 개인정보주체의 명백한 동의에 근거하는 경우
3. 제2항 (a)호 및 (c)호의 사례의 경우, 개인정보처리자는 개인정보주체의 권리와 자유 및 정당한 이익, 최소한 개인정보처리자의 인적 개입을 확보하고 본인의 관점을 피력하며 결정에 대해 이의를 제기할 수 있는 권리를 보호하는 데 적절한 조치를 시행해야 한다.
4. 제2항의 결정은 제9조(2)의 (a)호와 (g)호가 적용되고, 개인정보주체의 권리와 자유 및 정당한 이익을 보호하는 적절한 조치가 갖추어진 경우가 아니라면 제9조(1)의 특정 범주의 개인정보를 근거로 해서는 안 된다.

캘리포니아 소비자 프라이버시법(California Consumer Privacy Act)[19]

- 2018년 6월 캘리포니아주 의회에서 소비자프라이버시법이 통과되어, 2020년 1월 1일부터 발효될 예정임
  - Online 개인정보와 관련하여 해당 법은 특히 '소비자 권리'를 보호하기 위한 강력한 법적 조치임
- CPA는 GDPR만큼 광범위하진 않지만 미국에서는 가장 강력하고 포괄적인 개인정보 보호법임
  - CPA의 주요 타켓은 구글, 페이스북 등 거대기업이며, 향후 다른 주에서도 이와 유사한 조치가 취해질 가능성이 높음
- 주요 내용은 자신의 개인정보에 대한 통제권을 보장하는 데 있음
  - 정보 제공 주체는 개인정보수집에 있어 수집되는 데이터의 종류와 그 변화에 대해 사전에 공지를 받음
  - 또한 개인은 정보를 수집하는 목적과 정보가 공유되는 제3자의 범주에 대해 알 수 있으며, 자신의 정보에 대한 판매를 불이익이 수반됨 없이 거부할 수 있는 권리가 있음
  - 그리고 자신과 관련해 수집한 모든 데이터에 대한 알 권리로서 특정 기업에 1년에 2번씩 무료로 자료를 요청할 수 있으며, 자신의 정보가 입수된 출처(sources)의 범주(categories)를 알 권리가 있음
  - 정보제공자는 자신의 데이터를 삭제할 권리가 있으며, 16세 미만의 어린이에 대한 정보 판매는 반드시 사전 동의(opt-in)를 요하는 것으로 규정하고 있음
  - 이와 더불어 법 집행을 위한 검찰권을 발동할 수 있으며, 개인은 자신의 정보유출에 대해 제소할 수 있는 권리가 있음

19) https://leginfo.legislature.ca.gov; https://www.caprivacy.org

기술기업들의 미국 프라이버시 연방법 (federal Privacy Rules) 제정 요구 (NAVER Privacy & Security, 2018)

- 2018년 9월 26일 미국 상원 의회의 상업 위원회(the Senate Commerce Committee)에서 애플, 아마존, 구글, 트위터, AT&T, Charter 경영진이 출석한 청문회가 개최됨
- 기업 임원진들은 의원들에게 연방차원에서 새로운 단일 프라이버시 보호법이 필요하다는 의견을 제시하였는데, 이는 지난 몇 년 간 미국 기업들의 프라이버시 규제반대 행보와는 상반되는 모습임
- 상업과학교통위원회 소속 상원의원 John Thune은 하이테크 업계의 청원을 '지역별이 아닌 국가 전체의 대응을 바라고' 있는 것으로 해석하는 동시에, "법제화에 임하고는 있지만 연내 입법은 어려울 것"으로 예상함
- 또한 그는 2018년 후반기에 소비자 단체의 의견을 듣기 위한 보다 폭넓은 공청회를 다시 개최할 필요가 있다는 의견을 피력함
- 이에 전자프런티어재단(EFF: Electronic Frontier Foundation) 등 프라이버시 옹호 단체들은 미국 기업들의 연방 프라이버시법 제정 움직임에 대해 회의적인 입장을 내비치며 다음 사항을 우려하거나 지적하기도 함

① 소비자 프라이버시 청문회로 개최된 청문회임에도 불구하고 소비자 대표가 한 명도 참석하지 않았다는 점
② 연방 프라이버시법 제정에 기업의 영향력이 너무 많이 행사될 수 있다는 점
③ 높은 수준으로 제정된 주 단위의 프라이버시 보호법에 우선하여, 보다 완화된 수준으로 제정된 연방 단위의 프라이버시 보호법이 적용되면, 결과적으로 연방 법안이 기업들의 면책 수단이 될 수 있다는 점

4. 정보기술을 통한 인권의 신장

- 인공지능, 빅데이터, 로봇 기술을 이용해 초국가 혹은 국가 규모의 사회적 문제를 완화하고 인권을 신장할 수 있음
  - 자율주행자동차: 장시간 운전에 따른 피로를 줄일 수 있어 교통사고 저감에 기여함
  - 의료 분야에서의 인공지능: 오진으로 인한 사망사고율 감소, 비용절감, 예방적 치료 등에 활용 가능함
  - 로봇 기술의 이용: 돌봄 노동 수행, 인간과 인간 사이에서 (ex. 노인과 노인건강 전문가 사이에서) 매개자 역할을 할 수 있음. 인간이 투입되기 어려운 재해 혹은 산업 현장에서 인력을 대체할 수 있음
  - 블록체인 기술: 중앙 집중적인 권력 구조 없이도 인터넷 거래의 신뢰를 높이고 중개 수수료를 절감할 수 있음
  - 투명한 정부 : 정부 기관들의 재정 운용 및 거래 내역을 실시간으로 추적하고, 이로부터 감사(audit) 활동이 자동화/지능화되도록 하여, 국가 재정이 시민을 위해 보다 충실하게 봉사할 수 있도록 함[20]
- 그러나 이 경우에도 개인정보 보호를 위한 조치, 실업에 대한 대비, 보험에서의 차별 등에 대한 예방조치가 필요함

20) AI For Good Foundation "Corruption: Transparency in Government" https://ai4good.org/corruption-transparency-government/

5. 설명가능성 등 거버넌스를 위한 노력

☑ 신경망 회로 인공지능의 문제: 왜 인공지능이 그런 특정한 결정을 내렸는지를 알고리즘만으로 알 수 없음 (Goodman and Flaxman, 2016)

☑ 설명을 획득할 권리

- (GDPR 제13조) 개인정보가 개인정보주체로부터 수집되는 경우 제공되는 정보: 프로파일링 등, 자동화된 의사결정의 유무[21)]
- (GDPR 제15조[22)]) 개인정보주체의 열람 : 프로파일링 등 자동화된 의사결정의 유무
  * GDPR 제13조, 제15조: 최소한 이 경우, 관련 논리에 관한 유의미한 정보와 그 같은 처리가 개인정보주체에 가지는 중대성 및 예상되는 결과
- GDPR 전문 71항에 명시된 설명을 획득할 권리
  - 정보주체는 자동처리에만 근거하여 정보주체의 개인적인 측면을 평가하는 조치를 포함할 수 있고, 온라인 신용신청에 대한 자동적 거절이나 인적개입 없이 이루어지는 전자채용 관행 등 정보주체에게 법적인 영향이나 이에 상응하는 중대한 영향을 미치는 결정에 적용받지 않을 권리를 갖는다. 어떠한 경우에도, 이러한 처리는 정보주체에게 구체적인 통지전달, 인적개입을 획득할 수 있는 권리, 의사를 표현할 권리, 이러한 평가 이후 도달한 결정에 대한 설명을 획득할 권리 (to obtain an explanation of the decision), 해당 결정에 이의를 제기할 권리 등, 적절한 안전조치를 적용받아야 한다

---

21) 제13조 개인정보가 개인정보주체로부터 수집되는 경우 제공되는 정보

1. 개인정보주체에 관련된 개인정보를 개인정보주체로부터 수집하는 경우, 개인정보처리자는 개인정보를 취득할 당시 개인정보주체에게 다음 각 호의 정보 일체를 제공해야 한다.
2. 제1항의 정보와 함께, 개인정보처리자는 개인정보가 입수될 때 공정하고 투명한 처리를 보장하는 데 필요한, 다음 각 호의 추가 정보를 개인정보주체에 제공해야 한다.

(f) 제22조(1) 및 (4)에 규정된 프로파일링 등, 자동화된 의사결정의 유무. 최소한 이 경우, 관련 논리에 관한 유의미한 정보와 그 같은 처리가 개인정보주체에 미치는 중대성 및 예상되는 결과

22) 제15조 개인정보주체의 열람권

1. 개인정보주체는 본인에 관련된 개인정보가 처리되고 있는지 여부에 관련해 개인정보처리자로부터 확답을 얻을 권리를 가지며, 이 경우, 개인정보 및 다음 각 호의 정보에 대한 열람권을 가진다.

(h) 제22조(1) 및 (4)에 규정된 프로파일링 등 자동화된 의사결정의 유무. 최소한 이 경우, 관련 논리에 관한 유의미한 정보와 그 같은 처리가 개인정보주체에 가지는 중대성 및 예상되는 결과

● 설명에 대한 권리는 법적 구속력이 있는가에 대한 견해가 대립하고 있으나 법적 구속력이 있는 권리라는 견해가 다수임(Burt et al., 2018)

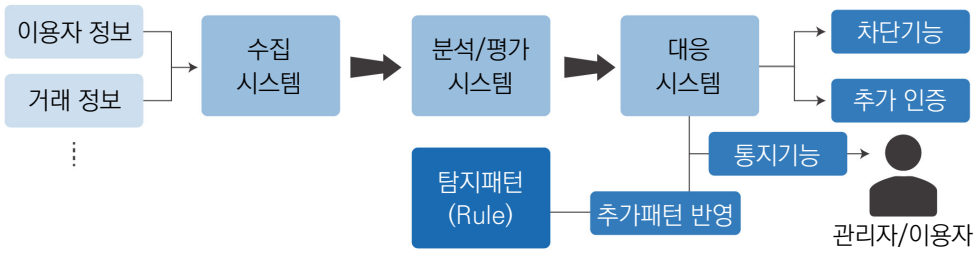
- 찬성: 적절하며 향후 알고리즘 발전에 도움이 될 것이다(Selbst and Powles 2017; Burt et al., 2018)
- 반대: '설명'이 결정의 논리가 아니라, 시스템의 기능에만 국한되어 있으며 실제 인공지능의 neural network을 고려하면 '설명'이라는 것이 불가능하거나 매우 어려움(Wachter et al., 2017; cf. Burrell, 2017)

Ⅳ 국내의 현황과 문제점

1. 국내 알고리즘의 활용 영역

규제행정의 영역

- 인공지능 알고리즘은 규제행정의 영역에서 사람이 판별하기 어려웠던 의심거래, 이상거래, 부당청구 등의 탐지 업무를 수행하고 있음
- 금융정보분석원(FIU: Financial Intelligence Unit)의 의심거래보고
  - STR(Suspicious Transcation Report)필터링 사용
- 금융기관의 이상금융거래탐지시스템(FDS: Fraud Detection System)
  - 전자금융거래 접속정보, 거래내역 등을 분석하여 이상금융거래 탐지 및 차단



- 국민건강보험과 부당청구감지시스템(FDS)
  - FDS는 전자요양기관의 진료비 청구내역, 의료자원 (인력·시설·장비) 신고·변경 정보 및 현지조사 결과, 인지된 부당청구 유형 등을 분석·개발한 여러 항목(Rule)을 통하여 부당청구 개연성이 높은 기관을 감지하는 시스템임
  - 수신자의 부당수급: 요양기관의 부당청구
  - 건강보험심사평가원의 현지조사 : 부당청구감지시스템(인공지능 활용)
- 이러한 자동화시스템은 타당한 결정을 내릴 수도 있지만, 알고리즘의 오류, 편향 가능성이 있기 때문에, 항상 합리적이고 타당한 결과가 도출된다고 보기는 어려운 측면이 있음
  - 자동화시스템을 이용하고자 할 때에는 프라이버시 침해도 있을 수 있고 억울한 사람이 생길 가능성이 있음을 항상 열어두어야 함

- 국회의원 최도자의 2016년 보건복지위원회 전체회의 질의에 따르면, 2015년 FDS에서 11,327개 장기요양기관의 부당위험 기관을 추출하여 이 중 부당위험 점수 상위 150개 기관에 기획현지조사를 실시한 결과 이 중 절반은 부당청구가 없는 것으로 나타났고, 상위 75번째 기관의 부당청구액은 1천원에 불과했으며, 14개 기관은 20만원 이하였음. 결국 수사의뢰가 실시된 기관은 1개에 불과했음[23)]
- FDS의 부당위험 점수가 실제 부당청구 가능성과 동떨어진 이유로는 해당 기관의 서비스 제공이 많다는 이유로 부당위험점수가 높아진 것이 지목되었음. 이에 따라, 최도자 의원은 전면적 시스템 재진단과 감시 모형의 재설계를 요구하였음
- 이후 국민건강보험공단은 FDS를 전면 재설계하겠다고 최도자 의원 측에 보고함 (현지조사 결과를 분석해 부당요인 변수를 정밀화, 전문성 향상을 위한 시스템 고도화 등)[24)]

- 따라서 이러한 알고리즘에 대해 △설계 의도가 무엇이며 △어떤 규칙이 적용되었는지 △해당 의도와 규칙이 적절한지 △다양한 사용 사례에 대해 오작동의 우려는 없는지 등이 평가되고 검증되어 그 내부 시스템의 동작이 이해될 필요가 있음. 그렇지 않고는 이런 시스템이 왜 특정한 결정을 내렸는지 누구도 알 수 없음

자동결정

- 미국과 마찬가지로 국내에서도 대출심사, 신용평가, 인력채용에서 인공지능 알고리즘이 점점 더 많이 사용되고 있고, 이런 경향은 앞으로 계속 증가할 추세임
- 또한, 채용에서는 서류심사, 면접에서 인공지능활용, 취업에서의 맞춤형 매칭 시스템 및 추천 시스템 등에 자동결정 알고리즘을 활용하고자 하는 개인 및 기업이 증가하고 있음
- 대출심사
  - 카카오뱅크와 케이뱅크 등 인터넷전문은행은 개인대출 심사를 자동화했고, 사람이 하는 심사역을 별도로 두고 있지 않음[25)] (이는 유럽에서는 GDPR에 의해 금지되어 있음)
  - 하나은행도 기업대출심사에 인공지능을 도입, 영업점 창구에서도 자동화시스템을 도입했으며, 자동화시스템을 통해 대출 거부된 경우에는 사람이 이를 다시 심사하고 있음

---

23) 최도자 (2016) "멀쩡한 장기요양기관 잡는 건보공단 부당청구감시시스템" (2016.6.23.) 바른미래당 국회의원 최도자 - 의정활동. http://www.choidoja.com/board_act/2454 ;

24) 디지털의사신문 (2016) "건보공단 장기요양기관 현지조사 개선 …'FDS 재설계 필요'" (2016.7.4.) http://www.doctorstimes.com/news/articleView.html?idxno=176050

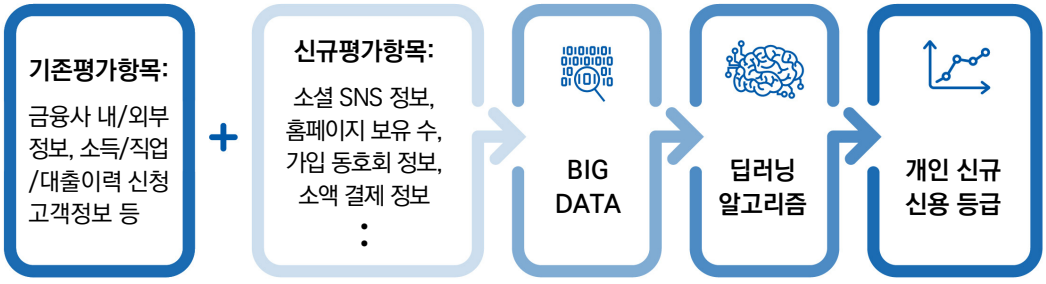
25) KEB하나은행, 기업대출 심사에 국내 최초로 AI 도입, 이학렬 기자, 머니투데이 , 2017. 9. 14. http://news.mt.co.kr/mtview.php?no=2017091217421558098

- OK저축은행은 인공지능을 활용한 머신러닝 모형을 대출상품 심사에 도입함[26]
  - 보통 금융회사가 10~15개 지표만으로 보수적으로 대출 승인을 내주는 반면, 이 모형은 100여 개 변수를 활용해 변별력을 높임
  - 기존 신용평가시스템(CSS)은 '20대는 연체율이 높다' '고소득자는 신용도가 좋다'는 식의 비교적 단순한 기준으로 대출 여부·한도·금리를 설정하였음
  - 반면 인공지능은 기존의 계좌·카드·연체 정보, 주거지·직장·통신 기록 등 방대한 데이터에 복잡한 패턴을 적용해 한 번에 신용도를 평가하기 때문에 편리하고 합리적이라고 판단함

● 신용평가

- 신한은행, KB캐피탈, AXA손해보험, SBI저축은행 등은 신용평가에서 인공지능을 도입함
  - 보험심사와 대출심사 시스템에 AI를 활용한 '다빈치랩스' (데일리금융그룹이 개발한 AI 데이터 분석솔루션)를 이용함[27]
  - 다빈치랩스는 7가지 이상의 알고리즘을 조합해 기존 리스크 평가기법보다 평균 50% 이상 리스크에 대한 예측 정확도를 높였음
- ASTER: 소셜네트워크 비정형 데이터를 활용한 신용평가 알고리즘으로 행동심리학 등 600개의 국내외 논문을 분석해 비정형데이터에서 항목별 가중치를 주기 위한 이론[28]

그림 Ⅳ-1. 인공지능을 활용한 신용평가 프로세스



※ 온라인 메신저 'QQ(8억명 유저)', 모바일 메신저 'WeChat(4억명 유저)' 등 SNS내역 AI 분석

자료: KT경제경영연구소

26) OK저축은행, 모든 대출상품 심사에 AI 기술 도입, 김경윤 기자, 연합뉴스, 2018. 1. 4. http://www.yonhapnews.co.kr/bulletin/2018/01/04/0200000000AKR20180104059500002.HTML

27) 은행의 미래 ④, 7만개 변수 파악해 AI가 신용평가, 정해용 기자, 조선비즈, 2017. 7. 23. http://biz.chosun.com/site/data/html_dir/2017/03/21/2017032101045.html

28) 페이스북 데이터 기반 AI 신용평가시스템 출격, 이데일리 김현아 기자, 2017. 5. 31. http://www.edaily.co.kr/popup/print_popup.html?newsId=03955686615934496&mediaCodeNo=257

● 한편, 채용에서는 서류심사와 면접에 인공지능이 활용되고 있음

- 서류전형 심사에 AI 활용
  - 롯데그룹 : 백화점, 마트, 정보통신 등 일부 계열사에서 지원자가 서류를 제출하면 AI가 자기소개서 등을 분석해 인재 부합도, 직무 적합도, 표절 여부 등을 평가
  - SK C&C : IBM 왓슨의 한국어 API 기반 AI플랫폼 '에이브릴'을 SK하이닉스 신입사원 서류평가에 시범 활용함. 에이브릴은 인사 담당자 10명이 하루 8시간씩 7일간 살펴야 할 1만 명의 자기소개서를 8시간 정도면 정확히 평가할 수 있음(자기소개서 하나를 3초에 평가)[29]
- AI 면접
  - 마이다스아이티 : AI면접관 '인 에어'는 센서와 질문을 통해 지원자를 평가함. 카메라를 통해 지원자의 얼굴과 실시간 표정을 분석하고 오디오를 통해 목소리의 톤 등을 판단하며 심장 박동과 뇌파를 체크해 지원자의 긴장정도까지 구별 가능함[30]
  - 한국방송통신전파진흥원(KCA): 공공기관 중 처음으로 AI를 활용하여 인·적성 검사, 직무역량 평가 등을 진행하였음
- 취업에서의 맞춤형 매칭 시스템, 추천 시스템
  - 취업포털 사람인 : 사람인에 접속한 이용자 검색, 공고조회 등 행동 패턴과 이력서, 자기소개서 등의 데이터를 분석해 개개인에게 적합한 공고를 선별해 전달함[31]

---

29) SK C&C "신입사원 자기소개서, AI가 3초면 평가", 이정민 기자, 조선비즈, 2018. 1. 25. http://biz.chosun.com/site/data/html_dir/2018/01/25/2018012502143.html

30) 채용시장에 부는 AI 바람…서류전형도 면접도 인공지능으로, 김선경 기자, CBS노컷뉴스, 2018. 5. 19. http://www.nocutnews.co.kr/news/4972145

31) "AI가 지원부터 면접까지 돕는다… 진화한 'HR테크'" 2018. 5. 9. http://www.saramin.co.kr/zf_user/help/live/view?idx=80872&listType=news

2. 차별 행위를 금지하는 법률

☑ 우리나라에는 국가인권위원회법 등 차별을 금지하는 다양한 법률이 존재:
국가인권위원회법, 장애인차별금지법, 남녀고용평등법, 고령자 고용법 등

- 국가인권위원회법: 평등권 침해 차별행위를 정의
  - 차별금지사유[32]를 명시
  - 차별 금지 영역 명시: 고용, 재화, 용역, 교통수단, 상업시설, 토지, 주거 공급, 교육, 직업훈련, 성희롱
- 장애인 차별금지 및 권리구제 등에 관한 법률(약칭: 장애인차별금지법)
- 남녀고용평등과 일, 가정 양립지원에 관한 법률(약칭: 남녀고용평등법)
- 고용상 연령차별금지 및 고령자고용촉진에 관한 법률(약칭: 고령자고용법)

32) 성별, 종교, 장애, 나이, 사회적 신분, 출신 지역(출생지, 등록기준지, 성년이 되기 전의 주된 거주지 등을 말한다), 출신 국가, 출신 민족, 용모 등 신체 조건, 기혼·미혼·별거·이혼·사별·재혼·사실혼 등 혼인 여부, 임신 또는 출산, 가족 형태 또는 가족 상황, 인종, 피부색, 사상 또는 정치적 의견, 형의 효력이 실효된 전과(前科), 성적(性的) 지향, 학력, 병력(病歷) 등을 들고 있음.

3. 알고리즘 차별 · 윤리 관련 국내 윤리 헌장

▣ 국내 공공기관

- 정부에서 알고리즘의 차별 및 윤리에 관련한 기술윤리에 대한 검토는 정부 부처별로 각개약진하고 있는 상황이라 볼 수 있음[33]

표 Ⅳ-1. 국내 알고리즘 차별·윤리 관련 기술윤리 현황

담당 기관 및 부처	알고리즘 차별 및 윤리 관련 기술윤리
과학기술정보통신부	지능정보사회 윤리 가이드라인
방송통신위원회 및 산하 한국정보화진흥원	지능정보사회 윤리
산업자원부	지능형 로봇윤리헌장
행정자치부	지능형 정부 구현 중장기 로드맵 수립사업
한국정보화진흥원	지능형정부 구현을 위한 AI 활용 윤리연구

- **"지능정보사회 윤리 가이드라인" (과학기술정보통신부)**[34]
  - 2018년 4월에는 과학기술정보통신부와 한국정보화진흥원의 지원 아래 정보통신기술 분야 교수·연구원·전문가·업계 관계자 25명으로 구성된 정보문화포럼에서 2년여의 작업 끝에 '지능정보화사회 윤리 가이드라인'을 제작함
  - AI와 로봇, 빅데이터 등 4차 산업혁명 기술 전반에 걸쳐 개발자·공급자의 윤리뿐만 아니라 이용자의 오남용 방지와 권리 보호까지 포괄한 데에 의의가 있음
  - 공공성(Publicness, 차별요소를 배제하고 공공 이익에 부합), 책임성(Accountability, 윤리적 절차의 충실한 이행과 책임원칙), 통제성(Controllability, 기술적 제어와 위험요소 제거), 투명성(Transparency. 정보의 공유와 은닉 기능 금지) 등을 4대 원칙으로 설정하고 개발 기술의 개발·활용 단계별로 적용되는 38개 세부 지침이 작성됨

33) 이원태 (2018) "ICT 사회문제해결 세미나와 로봇윤리포럼이 동시에 개최..." (2018.9.20.) https://www.facebook.com/wontae.lee.9889/posts/1905665249471920

34) 매일경제 (2018) "개발·공급자 윤리부터 이용자 권리까지 포함" (2018.4.9.) http://news.mk.co.kr/newsRead.php?year=2018&no=224223

- **“지능정보사회 윤리” (방송통신위원회 및 산하 한국정보화진흥원)**[35]
  - 지능정보사회 윤리기반 마련’을 위해 윤리 가이드라인 개발, 인터넷윤리문화 조사·연구, 인공지능(AI)의 올바른 활용 교육과정 개발·운영 등을 추진할 예정임
  - 가이드라인에는 인공지능의 안전 및 신뢰성 구현, 개인정보 보호, 오남용 예방, 개발자의 책임성 구현, 인간 가치 존중 등이 담길 수 있게 개발함

표 Ⅳ-2. 지능정보사회 윤리 가이드라인

적용 대상	적용 단계	세부지침
개발자	수요분석	-지능정보기술의 공공성 화곱 노력과 책임의 공유
	제품개발	-윤리적 절차 이행 연구개발 및 기술개발 시 사회적 차별요소 배제 -품질 인증기준 충족, 기술 제어장치 마련 -예외적 상황에 대한 종합적 검토와 위험에 대한 예측 및 공급자와의 공유 -개발자들 간 정보 교류와 기술 갱신 지속 참여
	이용지원	-지속적 품질관리, 위급상황 때 필요데이터 제공
공급자	수요분석	-공익에 부합하는 제품 공급, 이용자 선택권 보장 -상업적 이익과 공공적 기여 간 조화 -지능정보서비스의 자율 의사결정 조건 및 범위 확립 -안전성 검증 및 통제 조치 마련
	공급유통	-책임 공유 및 책임보상 원칙 마련 -이용자 권리보장, 제품 유통과정서 위험 통제
	이용지원	-위험 관련 정보의 공유, 이용자 정보의 부당 이용금지 -위험예방 위한 사회적 공론화 참여
이용자	수요분석	-지능정보기술 이용역량 강화
	공급유통	-소비자 행동원칙 준수 일상화 및 이용자 윤리책임 숙지, 설명 요구할 권리
	이용지원	-이용 제품의 교체·갱신·폐기 시 지침 준수 -소비자 정보의 공유 의무, 공익 위한 제품 개선 참여

출처: 지능정보사회 윤리 가이드라인 (매일경제 2018)

35) 방송통신위원회 (2017) “방통위, 「아름다운 인터넷 세상 2022」 계획 수립” (2017.9.8.) http://www.kcc.go.kr/user.do;jsessionid=KxaYcgAVGIOQJROMwmSAIDal5yUCwHOP8go5x1JZwrZ1dUAU0d6aiT1skKQDkYlN.hmpwas02_servlet_engine1?mode=view&page=A02020600&dc=&boardId=1008&cp=1&boardSeq=46343

- **'지능형 로봇윤리헌장' (산업자원부)**[36]
  - AI 시대에 필요한 윤리기준을 제정하고 구체적인 이행방안을 마련하였음
  - 2018년 12월 완료를 목표로 검토 중인 '지능형 로봇 개발 및 보급 촉진법(약칭: 지능형 로봇법)' 시행령의 개정안에 산업부 장관을 위원장으로 하는 로봇산업정책심의회(심의회) 산하에 설치될 "로봇윤리자문위원회(자문위)"의 구성, 역할, 활동 기간 등 구체적인 운영 방안이 담길 예정임
  - 자문위는 '지능형 로봇윤리헌장' 제정을 위해 필요한 조사·연구와 각계 의견을 수렴하는 역할을 하고 이후 자문위의 권고 내용을 바탕으로 심의회가 이해 당사자들의 의견을 종합해 윤리헌장 제정에 나서게 됨
  - 이는 2018년 5월 지능형 로봇법 개정안이 국회 본회의를 통과한 데 따른 것으로, 개정안은 지능형 로봇의 기능과 지능 발전에 따른 각종 폐해를 방지하고 로봇이 인간의 삶의 질 향상에 이바지할 수 있도록 윤리헌장의 제정 근거를 규정하고 있음
  - 윤리헌장은 지능형 로봇의 개발자·제조자·사용자별 구체적 행동지침을 담을 예정이며, 정부는 단순 '권고' 수준이 아닌 법제화를 통해 윤리헌장에 강제력을 부여하는 방안도 검토하고 있음
  - 한국에서는 과거에도 2007년 정부 주도의 윤리헌장 제정을 시도하여, 초안을 마련했지만 공식 채택은 무산되었음. 당시 관련 전문가들이 합의에 이르지 못했기 때문임
- **"지능형 정부 구현 중장기 로드맵 수립사업" (행정자치부)**[37]
  - 2019년 3월까지 향후 AI 시대에 필요한 윤리기준을 제정하고 구체적인 이행방안을 마련함
  - 지능형 정부의 청사진을 제시해 인공지능(AI) 등 지능정보기술이 전자정부시스템에 전면 적용될 수 있도록 할 예정임
  - 민원처리나 행정업무 처리 등에 있어 핵심 전략과제를 발굴하고 핵심 전략과제별 구체적인 실행계획을 설계함
  - 행정 내부적으로는 공무원의 업무 처리를 도와주는 AI 정책자문관을 도입해 업무 생산성을 높일 예정임
  - 민원처리나 정책자문의 정부서비스에 AI 기술 적용으로 발생하는 윤리적 문제나 법·제도적 부분까지도 검토해 추후 전자정부법 개정 때 반영할 방침임

---

36) 머니투데이 (2018) "정부, '세계 최초' AI 윤리헌장 만든다…자문위 내년 출범" (2018.10.8.)
37) 디지털데일리 (2018) "지능형정부 청사진 나온다 'AI 행정비서·정책자문관 등장'" (2018.9.12.)

- **"지능형정부 구현을 위한 AI 활용 윤리연구" (한국정보화진흥원)**[38]
  - 향후 AI 시대에 필요한 윤리기준을 제정하고 구체적인 이행방안을 마련함
  - AI 기술이 각 산업분야에 어떻게 접목되고 있는지 동향 조사와 분석을 통해 국내 특수성을 반영한 윤리기준을 제시함
  - 해당 가이드라인은 정부 사업 안에서만 통용될 예정으로, 민간 부분에서 참고할 수 있는 가이드라인은 과학기술정보통신부에서 마련하는 것으로 추진함

● 각 부처 간 윤리 원칙이 서로가 상충되거나 역할이나 기능상의 중복은 없는지 상호 비교가 필요함

☑ 국내 민간

● 카카오 알고리즘 윤리헌장 (카카오 2018; 이원태 2018)

- 의의 (이원태 2018)
  - '사전 예방의 원칙(precautionary principle)'에 입각하여 기술 개발의 윤리적 행위 원칙을 설정함으로써 신기술의 잠재적 위험성을 개발자 및 사업자가 스스로 통제하겠다는 사회윤리적 책임성에 부응한 것이라 볼 수 있음
- 지적된 향후 보완점 (이원태 2018)
  - 알고리즘 윤리의 주요 행위자로서 이용자의 역할이 잘 보이지 않음
  - 세부적인 가이드라인으로 구체화된 것은 아니므로 알고리즘 관련 기술 및 서비스의 성격에 따라 좀 더 구체적인 윤리 행위 원칙으로 진전시킬 필요가 있음

38) 디지털타임스 (2018) "인간이 통제하는 AI… NIA `윤리기준` 제정" (2018.9.9.)

카카오 알고리즘 윤리 헌장

1. 카카오 알고리즘의 기본 원칙	카카오는 알고리즘과 관련된 모든 노력을 우리 사회 윤리 안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다.
2. 차별에 대한 경계	알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다.
3. 학습 데이터 운영	알고리즘에 입력되는 학습 데이터를 사회 윤리에 근거하여 수집·분석·활용한다.
4. 알고리즘의 독립성	알고리즘이 누군가에 의해 자의적으로 훼손되거나 영향받는 일이 없도록 엄정하게 관리한다.
5. 알고리즘의 대한 설명	이용자와의 신뢰 관계를 위해 기업 경쟁력을 훼손하지 않는 범위 내에서 알고리즘에 대해 성실하게 설명한다.

2018년 01월 31일
카카오

kakao

4. 알고리즘에 의한 차별의 유형

☑ 알고리즘을 통한 차별에는 직접차별, 간접차별, 기계학습에 의한 차별 등이 있음 (Wagner 2016)

- 직접차별: 채용, 이용 등에서 차별적 요소를 포함
- 간접차별: 직접적 요소를 포함하지는 않았지만 차별금지 사유에 해당하는 집단에 불리한 결과를 초래하는 경우
- 기계학습 알고리즘에 의한 차별: 기계학습을 통해 차별적 결과를 가져오는 것

5. 우리나라에서 알고리즘에 의한 차별의 사례

☑ AI 알고리즘에 의한 서류전형, 면접, 채용

- 점점 더 많은 기업이 채용 과정에서 알고리즘을 사용하고자 하는 추세지만 해당 알고리즘의 직접, 간접 차별 여부를 검증하기 어려움. 기업은 이런 알고리즘이 직접차별, 간접차별의 요소가 없는지 투명하게 공개할 필요가 있음

- 아마존 에딘버러 엔지니어링 팀이 과거 입사 지원자들의 이력서를 바탕으로 머신러닝을 훈련시켜 최적 구직자 추천 모델을 만들어내도록 한 결과, 해당 모델은 남성 편향적인 결과를 보여주었음
  - 이는 10년간의 아마존 지원자 풀에서 남성이 압도적으로 많았기 때문이며, 이와 같은 편향적인 결과로 인해 개발을 중단함[39]

☑ 일자리의 자동화된 추천, 매칭 알고리즘[40]

- 워크넷 일자리포털(한국고용정보원 운영)은 역량에 맞는 일자리와 직무에 적합한 인재를 연결해주는 자동매칭시스템을 도입할 계획[41]

---

39) 머니투데이 (2018) "아마존, 'AI채용시스템' 폐기…알고리즘이 남성 선호" (2018.10.11.)

40) 일자리정보 워크넷이 더 똑똑해진다…빅데이터·AI 도입, 뉴시스, 2018.04.23

41) 머신러닝 기반 일자리매칭 ISP 제안요청서(주관기관 : 한국고용정보원), 2017. 9. (공공기관경영정보공개시스템 http://www.alio.go.kr/informationBidView.do?seq=2320608)

- 워크넷은 가입자 수가 1553만명이며 일평균 접속자 82만명, 구직자 480만명이 이용하고 하루 1만 4천여건의 구인광고가 올라오는 대형 잡 포털(Job Portal)임
- 2019년부터 빅데이터를 적용한 자동 매칭 서비스를 도입하려고 계획하고 있는데, 이 때 매칭 방식에 차별 가능성이 존재함
- 구직자의 성, 학력, 출신, 인종 등과 관련해서 차별이 없는지가 우선적으로 검증되어야 함
- 특히, 이런 알고리즘의 경우에 성공률을 목표로 삼으면 그 목표 자체가 차별적인 결과를 낳을 수 있음

알고리즘 기반 뉴스 추천의 편향

- 인터넷 정보제공자인 플랫폼이 맞춤형 정보를 이용자에게 제공해 이용자는 필터링 된 정보만을 접하게 되는 필터 버블(Filter Bubble)[42] 효과가 확대됨
- 소셜네트워크서비스(SNS)의 확대로 유유상종 현상의 강화에 따른 에코 체임버 효과(Echo Chamber)[43]가 증폭될 가능성이 있음
- 뉴스 소비에 있어서 민주주의의 근간인 다양성이 훼손되며 극단적인 의견들이 득세하는 결과를 낳게 됨
- 극단적인 의견들이 득세하면 다수의 것으로 비치고 다수 의견과 반대되는 의견 피력을 피하는 “침묵의 나선”(Spiral of Silence) 현상이 뒤따르게 되어, 극단적인 의견이 그 중간 스펙트럼의 의견의 입지를 더욱 좁게 만드는 현상으로 이어질 수 있음을 우려해야함 (Matthes, 2018)

규제행정과 알고리즘

- 규제행정 알고리즘의 결과에 대해 이의를 제기하기 힘듦
- 규제 행정 알고리즘에 대한 알 권리를 강하게 보장해야 함

---

42) 필터 버블(filter bubble)은 개인화된 검색의 결과물의 하나로, 사용자의 정보(위치, 과거의 클릭 동작, 검색 이력)에 기반하여 웹사이트 알고리즘이 선별적으로 어느 정보를 사용자가 보고 싶어 하는지를 추측하며 그 결과 사용자들이 자신의 관점에 동의하지 않는 정보로부터 분리될 수 있게 하면서 효율적으로 자신만의 문화적, 이념적 거품에 가둘 수 있게 한다(위키백과, 2018.11.13)

43) 에코 체임버: 반향실에서 소리가 울려 증폭되는 것처럼 정보, 아이디어, 신념 등이 정의된 시스템 내에서 증폭되거나 강화되는 상황을 은유적으로 설명

범죄예측

- 대검찰청 등은 범정부 지능형 범죄예방 협업체계 구현을 위한 시범서비스를 추진하고 있으며[44] 경찰청에서도 비슷한 서비스 추진 중임[45]
- 그러나 이러한 서비스 역시 성, 지역, 국적, 인종 등과 관련하여 차별 가능성 존재함
- 미국에서는 이런 사법, 치안 영역에서의 알고리즘이 큰 사회적 논쟁의 대상이 되었음
- 이 영역 역시 알고리즘에 대한 평가가 어렵다는 문제가 있기 때문임

신용평가와 차별

- 신용평가가 자동화될 때 젊은 계층에 차별적인 결과가 나올 가능성이 큼

---

44) 지능형 범죄예방 협업체계 구현을 위한 시범서비스 구축 사업 제안요청서(대검찰청, 2017. 4. 13.) (공공기관경영정보공개시스템 http://alio.go.kr/mobile/tender_view.do?pageNo=1&idx=2248920&search_opt=&search_text=)

45) 경찰청의 '빅데이터 기반 범죄 분석 프로그램', 보호관찰 재범방지 전략회의의 아동학대사범을 비롯해 성폭력사범 등 고위험 보호관찰 대상자의 관리감독 강화방안과 재범방지 대책에서도 고위험 대상자의 교우관계 등 범죄친화적인 사회관계망을 분석하고, 동시에 축적된 보호관찰대상자 190만명의 정보와 13만 건의 재범사례를 빅데이터 분석을 통해 재범을 선제적으로 막는 시스템을 추진할 방침이라고 함.([종합]법무부, 보호관찰 재범방지 전략회의 개최…"재범 예측 시스템 만들 것" 뉴시스, 2016.01.29. https://news.joins.com/article/19499976)

V 결론

- 빅데이터, 인공지능 시대에 새로운 유형의 인권문제가 대두됨. 자동 프로파일링, 얼굴인식 기술의 발전 등을 통한 개인정보의 수집, 알고리즘을 통한 차별의 문제가 이에 해당됨

- 개인정보의 수집과 자동화 측면에서는 인권보호의 차원에서 유럽연합의 GDPR이나 캘리포니아의 소비자 보호 프라이버시 법처럼 기업, 병원, 정부기관의 개인정보 수집을 규제하는 더 강력한 법안이 인준되고 있는 상황임

- 그러나 국내에서는 이런 외국의 법안에 대한 논의만 진행되고 있을 뿐 오히려 정부, 기업, 정보기술 전문가 일부는 현재의 개인정보보호법이 강력해 빅데이터 기업 활동을 방해한다고 주장하며, 대표적인 규제완화 대상으로 간주하는 등 개인정보 보호가 인권 차원의 문제라는 문제의식이 결여되어 있음
  - 개인정보 활용의 원칙은 공공적인 연구를 목적으로 하는 경우에는 상대적으로 자유롭게 사용할 수 있게 허용하지만, 기업이 사적 이윤을 목적으로 사용하는 개인정보는 엄격하게 규제하고 제한해야 함
  - 개인정보에 대한 규제를 풀어야 한다는 주장은 개인정보 보호를 강화하는 세계적인 흐름과 역행하고 있음
  - 빅데이터 기술의 발달로 고도화되는 정보화시대에 오히려 개인정보보호 규제를 완화하는 것은 궁극적으로 한국 기업의 국제 경쟁력을 떨어뜨릴 수 있음
  - 이를 고려하여 개인정보보호법의 개정 등의 문제는 충분한 사회적 공론화를 통해 합의점을 찾아야 함

- 알고리즘에 의한 차별에 대해서는 지난 2-3년 동안 미국과 EU의 여러 나라에서 언론을 통한 문제제기가 있었고 사회적으로 큰 이슈가 되었음
  - 미국은 이러한 문제를 전문적으로 다루는 대안언론, 알고리즘의 투명성을 강조하는 대항 전문가 및 시민사회의 참여를 통한 규제라는 거버넌스의 구조가 만들어지는 과정에 있음

- 알고리즘에 의한 차별은 보통 사람이 이해하기 어렵고 알고리즘에 접근하기도 어려움
  - 특히 신경망 네트워크 인공지능의 경우에는 이를 만든 사람도 왜 그런 결정이 내려졌는지를 정확히 알 수 없다는 문제가 있음
- 시민사회의 비판에 직면해서 정부와 기업은 알고리즘의 투명성, 책임성, 설명가능성 등 문제를 해결하거나 완화하기 위한 다양한 개념, 해법을 제시하고 있음. 이런 해법이 문제를 궁극적으로 해결하지는 못할지라도 알고리즘에 대한 신뢰를 회복하기 위한 노력이라는 점에서 긍정적으로 평가할 수 있음
- 반면 국내의 경우 4차 산업혁명의 일환으로 인공지능과 빅데이터에 대한 정부의 지원은 급격하게 늘어나고 있으나, 이에 수반되는 사회적 문제와 차별가능성에 대한 검토와 법적 규제에 대한 논의는 부족한 실정임
  - 사법, 치안, 구직과 채용에서의 인공지능 알고리즘의 사용도 급격하게 확산될 것이 전망되나 인공지능의 응용이 새로운 산업의 창출이라는 점에서 장려되고만 있음
  - 이에 수반되는 부작용, 특히 알고리즘에 의한 차별 가능성 등에 대해서 조사하고 검토해 보는 활동이 필수적임
- 이를 위해 시민사회, 규제기관, 전문가 집단이 포함된 자발적인 거버넌스의 구축이 바람직함
- 그러나 인공지능 알고리즘 자체가 전문가가 아니면 이해하기 힘들고, 정부 부처 주도의 각종 알고리즘 윤리 가이드라인이 권고 이상의 실질적인 효력을 담보할 수 있을지 보장하기 어려우며 기업 등은 알고리즘에 접근하기 힘든 특성이 있음
- 따라서 정부는 개인정보보호위원회 등에 알고리즘에 대한 감시의 권한 등을 부여하고 시민사회 측의 관점과 우려를 반영할 수 있도록 참여 범위를 확장시켜 해당 논의를 진행하여야 함[46] (Andrews et al. 2017)

---

46) 외국의 경우로는 영국 하원이 Algorithm in Decision Making 주제에 대해 도입한 청문회(https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2015/inquiry9/)와 보고서(https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf)를 참조

참고문헌

Acemoglu, D. & P. Restrepo (2017). "Robots and Jobs: Evidence from US Labor Markets." NBER Working Paper No. 23285. March.

Acemoglu, D. & P. Restrepo (2018). "Artificial Intelligence, Automation and Work." NBER Working Paper No. 24196.

ACLU (2018) "Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots" by Jacob Snow (18.7.26.) https://www.aclu.org/blog/ privacy-technology surveillance-technologies/amazons-face-recognition-falsely-matched-28?page=3

Andrews, L. et al. (2017). *Algorithmic Regulation*. Center for Analysis of Risk and Regulation.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016,). "Machine bias." ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Athalye, Anish, Logan Engstrom, Andrew Ilyas, Kevin Kwok (2018) "Synthesizing Robust Adversarial Examples" in *Proceedings of the 35 th International Conference on Machine Learning (ICML)*. https://arxiv.org/abs/1707.07397

Barocas, S., and A. D. Selbst. 2016. "Big Data's Disparate Impact." California Law Review 104: 671-732.

Bojarski, Mariusz Bojarski, Philip Yeres, Anna Choromanska, Krzysztof Choromanski, Bernhard Firner, Lawrence Jackel, Urs Muller (2017) "Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car" https://arxiv.org/abs/1704.07911

Burrell, J. (2016). "How the machine 'thinks': Understanding opacity in machine learning algorithms." *Big Data & Society,* 3(1).

Burt, A. et al. (2018). Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models. Future of Privacy Forum.

Caplan, Robyn, Joan Donovan, Lauren Hanson, and Jeanna Matthews (2018). "Algorithmic Accountability: A Primer." Available at: https://datasociety.net/output/algorithmic-accountability-a-primer/

Chouldechova, Alexandra. (2017) "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments" *Big Data* 5.

DARPA (2016) "Explainable Artificial Intelligence (XAI)" https://www.darpa.mil/program/explainable-artificial-intelligence

De Stefano, V. (2016) 'The rise of the just-in-time workforce: on-demand work, crowdwork and labour protection in the "gig-economy"', ILO Conditions of Work and Employment Series, Working Paper No. 71, Geneva, ILO.

DiMaggio, P. and E. Hargittai (2001). "From the 'Digital Divide' to 'Digital Inequality':. Studying Internet Use as Penetration Increases." Working Paper Series, 15. Center for Arts and Cultural Policy Studies. Princeton University.

European Parliament (2017). "Civil Law Rules on Robotics" http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P8-TA-2017-0051&language=EN&ring=A8-2017-0005

Gillespie, T. L. (2014). "The Relevance of Algorithms."in Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot (ed.), *Media Technologies: Essays on Communication, Materiality, and Society.* MIT Press.

Goodfellow, I. Yoshua Bengio, and Aaron Courville (2016) *Deep Learning,* MIT Press. http://www.deeplearningbook.org

Goodman, Bryce, and Seth Flaxman. (2016) "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" arXiv:1606.08813

Grgić-Hlača, N., Redmiles, E.M., Gummadi, K.P., and A. Weller. (2018) "Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction" arXiv:1802.09548

Hargittai, E. and Y. P. Hsieh (2013). "Digital Inequality," The Oxford Handbook of Internet Studies

ILO (2016) Non-standard employment around the world: Understanding challenges, shaping prospects, November. Available at: http://www2.ilo.org/global/ublications/ang—en/index.htm

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). "Accountable algorithms." *University of Pennsylvania Law Review*, 165, 633-705.

Lüge, Christoph (2017). "The German Ethics Code for Automated and Connected Driving." in *Philosophy & Technology* 30 (4):547-558. September.

Marcus, Gary (2018) "Deep Learning: A Critical Appraisal" https://arxiv.org/abs/1801.00631

Matthes, J., Knoll, J., & von Sikorski, C. (2018). "The 'spiral of silence' revisited: a meta-analysis on the relationship between perceptions of opinion support and political opinion expression." *Communication Research,* 45(1), 3-33.

McKinsey & Company. (2016) Automotive revolution – perspective towards 2030: How the convergence of disruptive technology-driven trends could transform the auto industry.

Mcsweeny, T. and B. O'dea (2017) "The Implications of Algorithmic Pricing for Coordinated Effects Analysis and Price Discrimination Markets in Antitrust Enforcement" Antitrust 32.

MIT CSAIL (2017) "Fooling neural networks w/3D-printed objects" (2017.11.2.) https://www.csail.mit.edu/news/fooling-neural-networks-w3d-printed-objects

Moses, Lyria Bennett, and Janet Chan. "Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability," *Policing and Society* DOI: 10.1080/10439363.2016.1253695

NAVER Privacy&Security (2018), "기업들의 미국 프라이버시 연방법(Federal Privacy Rules) 제정 요구" in NAVER 개인정보 보호 (2018.10.11.) https://blog.naver.com/n_privacy/221375275420

Northpointe Inc. Research Department (2016) "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity | Performance of the COMPAS Risk Scales in Broward County" (July 8, 2016) http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf

NVIDIA Korea (2016) "2012년 이미지넷에서 알파고까지… 딥 러닝의 모든 것" (2016.3.21.) http://blogs.nvidia.co.kr/2016/03/21/all_of_deeplearning/

Perry, Walter L., Brian McInnis, Carter C. Price, Susan C. Smith, John S. Hollywood, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation. Available at: https://www.rand.org/pubs/research_reports/RR233.html

Selbst, A.D. and J. Powles. (2017). "Meaningful information and the right to explanation" *International Data Privacy Law*, Volume 7.

Elsayed-Ali, Sherif (2017). "Artificial intelligence and the future of human rights" in Amnesty Insights Investigating today's human rights issues (Oct 19, 2017) https://medium.com/amnesty-insights/artificial-intelligence-and-the-future-of-human-rights-b58996964df5

Shetty, Salil (2017). "Artificial Intelligence for good" in Amnesty International (9 June 2017) https://www.amnesty.org/en/latest/news/2017/06/artificial- ntelligence for-good/

Silberman, S., and Irani, L. (2016) 'Operating an employer reputation system: lessons from Turkopticon, 2008-2015', *Comparative Labor Law & Policy Journal,* 37 (3) Spring.

The White House (2016) "Artificial Intelligence, Automation, and the Economy." Available at: https://obamawhitehouse.archives.gov/blog/2016/12/20/rtificial-intelligence-automation-and-economy

Wachter, S., Mittelstadt, B., and L. Floridi (2017), *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*. Oxford Internet Institute, University of Oxford.

Wagner, B. (2016). "Algorithmic regulation and the global default: Shifting norms in Internet technology" *Etikk i praksis* 1. DOI: http://dx.doi.org/10.5324/eip.v10i1.1961

디지털의사신문보스트롬, 닉 (2017)『슈퍼 인텔리전스 - 경로, 위험, 전략 (*Superintelligence: Paths, Dangers, Strategies* (2014)』조성진 역. 까치.

브린욜프슨, 에릭· 앤드루 맥아피 (2014),『제2의 기계시대: 인간과 기계의 공생이 시작된다(*The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, 2014) 이한음 역. 청림출판.

오닐, 캐시 (2017),『대량살상수학무기 (Weapons of Math Destruction (2016)』김정혜 역. 흐름출판.

이원태 (2018) "'카카오 알고리즘 윤리 헌장'의 의미 : 개발자 자율규제 의지 천명… '이용자 윤리' 보완 필요" 『신문과방송』통권 568호 (2018년 4월호) pp.52-55.

카카오 (2018) “카카오 알고리즘 윤리 헌장” in 카카오 AI, https://www.kakaocorp.com/kakao/ai/algorithm

커즈와일, 레이(2007).『특이점이 온다 - 기술이 인간을 초월하는 순간(*The Singularity is Near: When Humans Transcend Biology* (2005)』. 장시혁 외 역. 김영사.

테그마크, 맥스, [2015] 2017,「라이프 3.0」, 백우진 역, 동아시아.

포드, 마틴 (2016)『로봇의 부상: 인공지능의 진화와 미래의 실직 위협(*Rise of the Robots: Technology and the Threat of a Jobless Future*, 2015)』이창희 역. 세종서적.

한국고용정보원 (2017)『기술혁신을 반영한 중장기 인력수요 전망 2016-2030』.

Turkoptikon (n.d.), https://turkopticon.ucsd.edu/

https://futureoflife.org/bai-2017/

https://futureoflife.org/ai-principles-korean (2018.8.8. 검색)

https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375 (2018.7.20. 검색)

https://www.caprivacy.org/post/ab-375-signed-californians-for-consumer-privacy

프로젝트 소개

1. 연구 개요

■ 연구과제명

- 3개 한림원 연구·정책협의회* 운영을 통한 과학기술 공동정책연구·자문
  (2018년 과학기술종합조정지원사업)

  * 한국과학기술한림원, 한국공학한림원, 대한민국의학한림원은 2017년 4월, 국가 과학기술 발전을 위한 중장기 정책에 대한 3개 한림원의 공동의견을 제시함으로써 일관되고 효율적인 정책자문을 수행하기 위해 '3개 한림원 연구· 정책협의회'를 출범함. 분기별 회의를 개최해 공동과제 발굴 및 협력방안 논의를 진행 중. 미국한림원연합회(National Academies of Science, Engineering, and Medicine)와 같이 국가 과학기술 중장기 정책 수립과 관련한 역할을 하는 것이 목표.

■ 연구목표 및 주제

- 3개 한림원 회원, Y-KAST 회원, 각 분야 전문가 등 과학기술 전 분야 석학으로 구성된 Pool을 활용해 과학기술 중·장기 정책 아젠다 발굴 및 정책연구·자문
- 2018년 연구주제: ①**과학기술과 인권**, ②미래사회를 열어갈 12가지 과학기술

2. '과학기술과 인권' 프로젝트

■ 배경 및 목적

- 연구개발의 결과가 국가사회의 발전은 물론 개인의 삶에 미치는 영향력이 커짐에 따라 과학기술사회가 사회적 책임감을 가져야 한다는 분위기가 조성되고 있음
- 한국과학기술한림원은 국내 과학기술계의 인권의식을 높이고, 과학기술인을 포함한 모든 국민들의 인권 신장을 위한 중·장기 정책 아젠다를 발굴하고자 함

"기술발전이 초래할 불평등의 심화와 소외의 확대는 미래사회의 주요 문제점이다. 인공지능과 자동화가 창출할 혜택이 소수의 개인이나 기업에 의해 장악될 수 있으며, 의학 발달의 혜택 역시 부유한 소수에게 집중되는 현상이 초래될 수 있다. 과학기술의 발전은 멈춰지지 않을 것인데, 이것이 계속해서 삶의 질을 높이는 토대가 되려면 인간존중의 이념과 사회구조의 창출이 수반돼야 한다."

- 클라우스 슈밥(Klaus Schwab) 세계경제포럼 (World Economic Forum) 회장

"과학기술과 세상의 연결을 강화해 나가야 한다. 여성, 고경력 과학기술인 등 잠재인력이 활동할 수 있는 기회를 확대해 나가야 하며, 국민관심이나 사회적 파장이 큰 이슈에 대해 과학기술계가 전문성을 바탕으로 자발적으로 참여하는 '사이언스 오블리주(Science Oblige) 운동을 과학기술계와 함께 연구해 펼쳐 나갈 계획이다."

- 과학기술인재 정책 추진방향(2017~2022) 중

3. '과학기술과 인권' 프로젝트 참여위원 및 주요 주제

■ 참여위원 명단

정책연구회 (8인)

위원장: **유욱준** 한국과학기술한림원 총괄부원장
이무하 회원담당부원장(서울대학교 명예교수)

분과별 집필위원회 대표(6인 가나다 순): 박병주, 윤정로, 이주영, 이중원, 최무영, 홍성욱

집필위원회 1분과 과학기술과 인권 (7인)

이주영 서울대학교 인권센터 위원(위원장)
이중원 서울시립대학교 교수(위원장)
최무영 서울대학교 교수(위원장)
박상욱 서울대학교 교수
송세련 경희대학교 교수
송위진 STEPI 선임연구위원
이현숙 서울대학교 교수

집필위원회 2분과 정보기술과 인권 (6인)

윤정로 KAIST 교수(위원장)
홍성욱 서울대학교 교수(위원장)
이병호 서울대학교 교수
이은우 정보인권연구소 이사(변호사)
이진우 KAIST 교수
이호영 정보통신정책연구원 연구위원

집필위원회 3분과 의생명과학과 인권 (7인)

박병주 서울대학교 의과대학 교수(위원장)
윤정로 KAIST 교수(위원장)
김정훈 서울대학교 의과대학 교수
박상민 서울대학교 의과대학 교수
박형욱 단국대학교 의과대학 교수(변호사)
정규원 한양대학교 법학전문대학원 교수(의사)
하대청 GIST 기초교육학부 교수

집필위원회 4분과 과학기술인의 인권 (6인)

홍성욱 서울대학교 교수(위원장)
김소영 KAIST 교수
윤태웅 고려대학교 교수
이대희 한국생명공학연구원 선임연구원
이인우 한국과학창의재단 연구위원
정승은 가톨릭대학교 의과대학 교수

간사 (4인)

김하정 서울대학교 대학원생
이지혜 서울대학교 학부생
조동현 서울대학교 의과대학 연구교수
조장현 서울대학교 대학원생

원고검수 (2분과)

오요한 렌셀러폴리테크닉대학교(RPI) 박사과정생

주요 주제

'과학기술과 인권' 프로젝트 Key Issue

[Part I 총론]
과학기술과 인권

목 표: 과학기술이 인권을 신장할 수 있도록 하는 제도와 정책 발굴

- 현대 과학기술과 인권에 대한 전체적인 조감
- 인권신장을 위한 21세기 과학기술의 역할과 과학기술인들의 사회적 책임
- ICESCR(사회 · 경제 · 문화적 권리에 관한 국제규약) 15조와 관련, '이로운 과학'을 국민들이 평등하게 누리고, '과학의 잠재적 해악'을 방지할 수 있는 구체적인 실행 방안과 제도 제시

Key Issue

과학기술에 대한 인식변화, 인권규범과 과학기술, 과학기술의 책임, 과학기술과 인권의 관계

[Part II 실천 각론①]
정보기술과 인권

목표: 정보기술의 발전이 사회에 미치는 긍정·부정적 영향을 분석하고, 과학기술사회에서 자율적인 시스템을 구축할 수 있는 방향 제시

- 현대 정보기술 개발이 인권이 미칠 수 있는 잠재적 해악을 예측하고 이를 방지할 수 있는 연구시스템 제시
- 정보기술이 사회적 계층 간 불평등을 좁힐 수 있는 방안 마련

Key Issue

새로운 정보기술의 발전, 정보기술이 낳는 인권의 문제, 알고리즘의 활용과 차별

[Part II 실천 각론②]
의생명과학과 인권

목표: 의생명과학의 발전이 야기할 수 있는 잠재적 위험과 인류 복지에 기여하는 부분을 동시에 분석하고, 인권과 연계된 연구 방향 제시

- 의약품 개발 및 생명과학, 배아실험, 유전공학 연구 등에 있어 연구윤리와 가이드라인에서 나아가, 실제 인권과 연결하여 과학연구 수행을 위한 지침과 실천방안 마련

Key Issue

의생명과학과 인권의 역사, 유전체의학의 기술적 · 제도적 현황, 유전체의학의 불확실성, 이해상충, '아래로부터의' 우생학, 건강권 보장

[Part II 실천 각론③]
과학기술인의 인권

목표: 청년 과학기술인 들이 창의적이고 안전한 환경에서 연구할 수 있도록 하는 제도 제시

- 실험실 환경, 연구문화, 제도 등으로 침해 받는 인권 문제 논의
- 과학기술인, 특히 주니어 과학자들이 창의적이고 안전한 환경에서 연구할 수 있도록 하는 제도 도출

Key Issue

청년 과학인의 경제적 처우 및 인권침해, 청년 과학인의 처우개선과 인권보호

한림원의 인권수호를 위한 노력

“

국내 과학기술계 주도 ‘인권선언문’ 마련

과학기술인들의 사회적 책무와 과학기술계 내부 인권문제에 대한 자발적 성찰

‘국제한림원·학회인권네트워크’개최 등 국제사회 인권수호 앞장

국민 모두의 인권 신장을 위한 과학거버넌스 구축에 기여

”

◆ 2013년	과학인권위원회 발족
◆ 2014년~2018년	국제과학인권회의 참여
◆ 2018년 4월	‘과학기술과 인권’ 주제 한림원의 창(매거진) 발간
◆ 2018년 4월	‘과학과 인권’ 주제 제124회 한림원탁토론회 개최
◆ 2018년 5월	‘과학기술과 인권’ 주제 정책연구사업 착수 (2018년 과학기술 종합조정사업)
◆ 2018년 5~10월	‘과학기술과 인권’ 주제 정책연구사업 운영위원회 운영
◆ 2018년 10월	‘과학기술, 인권에 독인가 약인가’ 오픈포럼 개최
◆ 2018년 10월	제13회 세계과학인권회의(IHRN Biennial Meeting 2018)
◆ 2018년 11월	‘한림원의 목소리 제75호 과학기술과 인권의 조화로운 발전’ 발간
◆ 2018년 11월	‘정보기술과 인권’ 주제 이슈페이퍼 발간
◆ 2018년 11월	‘의생명과학과 인권’ 주제 이슈페이퍼 발간
◆ 2018년 11월	‘청년 과학기술인의 인권’ 주제 이슈페이퍼 발간
◆ 2018년 11월	과학기술자 인권선언문 최종 성명 발표
◆ 2018년 12월	‘과학기술과 인권’ 주제 이슈페이퍼 발간

이슈페이퍼 2018-01호 석학정책제안서

새로운 정보기술과 인권

자동화된 프로파일링 등을 통한 개인정보 침해와
인공지능 알고리즘을 이용한 차별을 중심으로

발행일 2018년 12월

발행처 한국과학기술한림원, 대한민국의학한림원

발행인 이명철, 정남식

전화 • 031) 726-7900

Fax • 031) 726-2909

홈페이지 • http://www.kast.or.kr

E-Mail • kast@kast.or.kr

기획·편집 한국과학기술한림원 정책연구팀

디자인·제작 (주)아미고디자인

ISBN 979-11-86795-33-0

본 사업은 과학기술정보통신부의 과학기술종합조정지원사업의 지원을 받아 진행되었습니다.