인공지능(에 의한) 의사결정과 인권

이호중
(정보인권연구소 전 이사장
서강대 법학전문대학원 교수)

인공지능(AI)의 정의

‘인공지능시스템’이란 다양한 수준의 **자율성**을 가지고 작동하도록 설계된 기계 기반 시스템으로, 명시적 또는 묵시적 목표를 위하여 수신된 입력으로부터 물리적 환경이나 가상 환경에 영향을 미칠 수 있는 예측, 콘텐츠, 권고 또는 결정 등의 산출물을 생성하는 방법을 **추론**하며, 배치 이후에 **적응성**을 보일 수 있는 시스템을 말한다. (EU 인공지능법 제3조 (1)항)

'AI system' means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments

자율성

추론능력

적응성

현재... AI의 산출물이 사람에게 영향을 미치는 세가지 방식

■ (1) AI의 직접적인 과업 수행

- AI가 어떤 사안에 대하여 **직접적으로 의사결정을 내리고 그에 따라 일정한 과업을 수행**하는 방식
- 자율주행자동차, AI 로봇(특히 군인, 경찰관 등의 기능 대체)
- 현재로서는 이러한 방식의 AI 의사결정과 과업수행은 아직 상용화되기는 어려운 단계

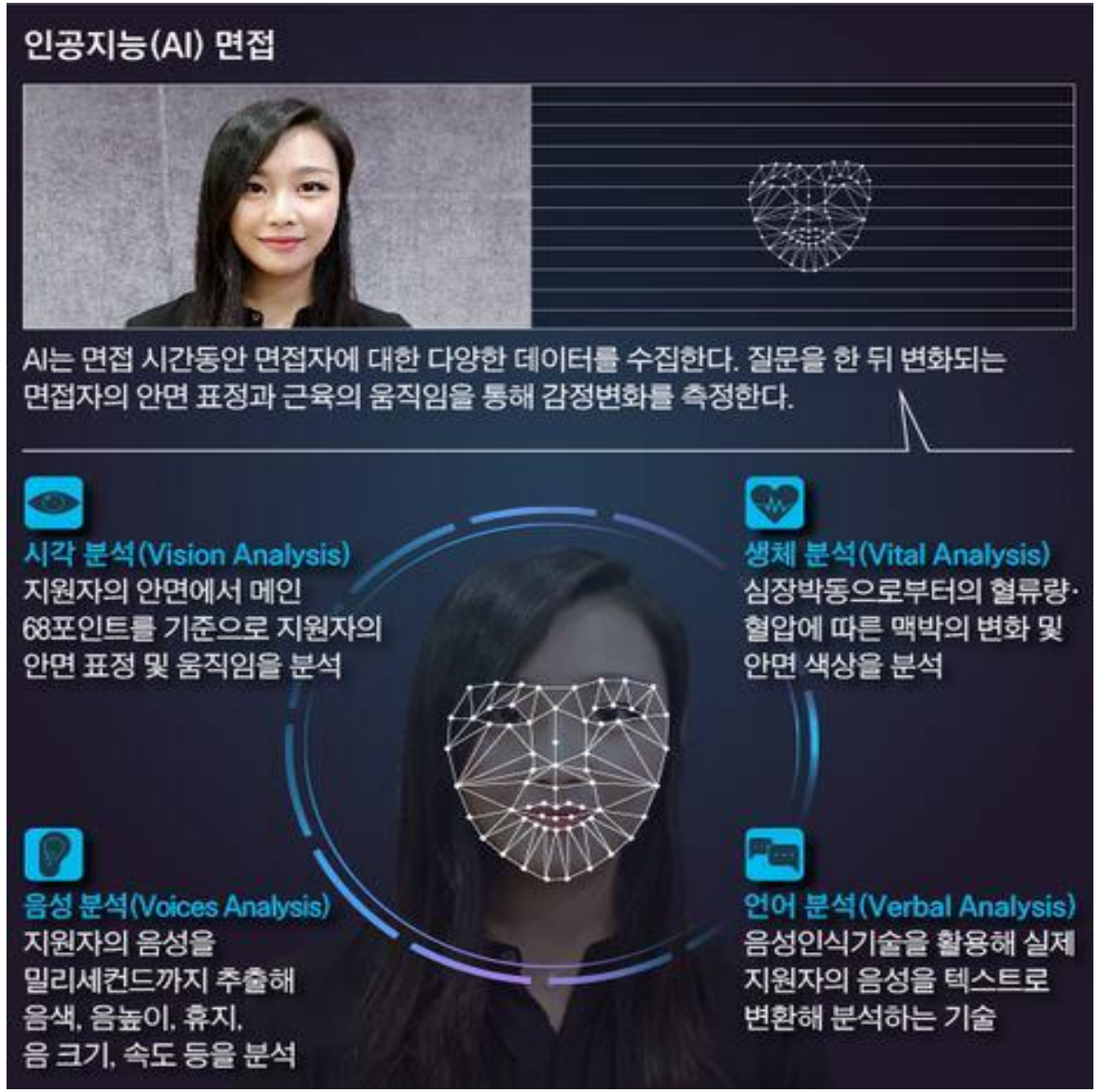
■ (2) AI의 산출물에 인간이 상당한 정도로 의존하면서 의사결정

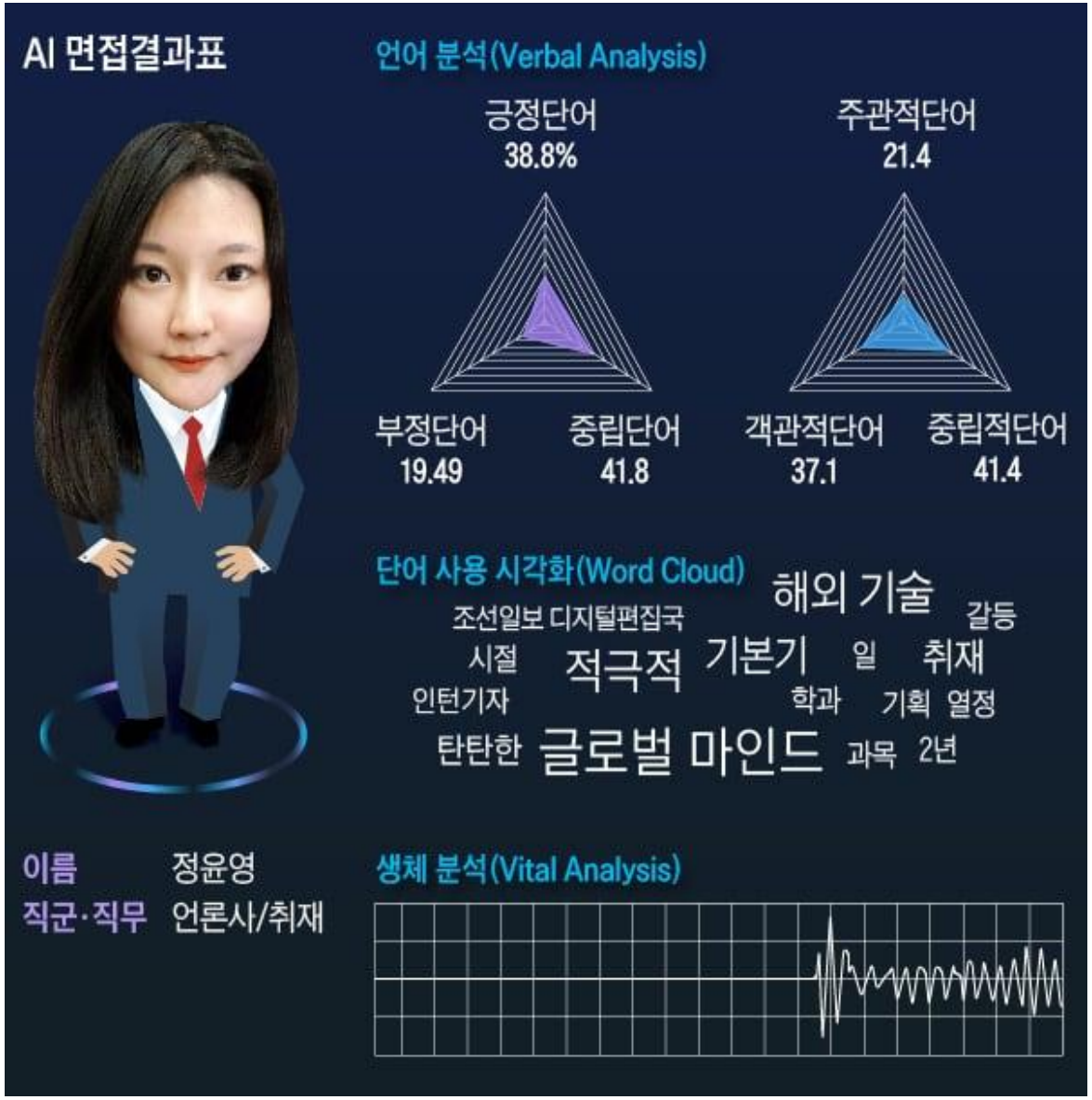
- **고용 ; 재범 예측 ; 신용 위험 분석 ; 얼굴 인식을 사용한 개인 식별 등**
- 현재 AI 시스템을 가장 광범위하게 사용하는 방식
- 인간의 과업 수행에 효율성과 효과성을 잠재적으로 개선할 수 있지만 **본질적으로 편견, 불투명성, 설명 불가능성, 악의성** 등으로 인한 인권침해의 문제 발생

■ (3) AI가 사람과의 소통 및 정보제공 등을 통해 사람들의 사고와 행동이 **간접적으로** 영향을 미치는 방식

- 챗GPT , 범용AI
- 사람들의 의사결정에 직접적인 동기로 작용하기도 하지만, 그보다는 **장기간에 걸쳐 누적적으로** 사람의 인식과 행동에 영향을 미침으로써 사람의 의사와 행동을 왜곡하거나 사회적 소통과 연대를 약화시킬 우려

고용 면접


인공지능(AI) 면접
AI는 면접 시간동안 면접자에 대한 다양한 데이터를 수집한다. 질문을 한 뒤 변화되는 면접자의 안면 표정과 근육의 움직임을 통해 감정변화를 측정한다.
시각 분석(Vision Analysis)
지원자의 안면에서 메인 68포인트를 기준으로 지원자의 안면 표정 및 움직임을 분석
생체 분석(Vital Analysis)
심장박동으로부터의 혈류량·혈압에 따른 맥박의 변화 및 안면 색상을 분석
음성 분석(Voices Analysis)
지원자의 음성을 밀리세컨드까지 추출해 음색, 음높이, 휴지, 음 크기, 속도 등을 분석
언어 분석(Verbal Analysis)
음성인식기술을 활용해 실제 지원자의 음성을 텍스트로 변환해 분석하는 기술


AI 면접결과표
언어 분석(Verbal Analysis)
긍정단어
38.8%
주관적단어
21.4
부정단어
19.49
중립단어
41.8
객관적단어
37.1
중립적단어
41.4
단어 사용 시각화(Word Cloud)
조선일보 디지털편집국
해외 기술
갈등
시절
적극적
기본기
일
취재
인턴기자
학과
기획
열정
탄탄한
글로벌 마인드
과목
2년
이름
정윤영
직군·직무
언론사/취재
생체 분석(Vital Analysis)

고용 면접

■ 문제점

- **차별과 편견의 증폭**
  - 학습데이터 자체의 편향성 (주로 대기업 인사 담당자 등 채용 전문가들의 평가에 기반한 데이터)
  - 성별, 나이, 지역, 신체 조건이나 경제적 지위, 학벌과 학력 등 역사적, 사회적으로 형성되어 온 차별과 편견이 고스란히 반영될 가능성이 크다.
- **투명성과 책무성에서 심각한 결함**
  - 채용 AI를 사용하는 기업이나 기관들은 채용 여부 결정의 이유를 설명하지 못함
  - 편향성을 방지하기 위한 대응 결여
- **AI 영상면접의 감정 평가 - 인간의 자율성에 대한 심각한 도전**
  - 지원자의 얼굴과 목소리를 자동으로 처리해 표정, 감정, 호감도, 언어 습관, 소통능력, 매력도, 신뢰성, 논리성 등을 분석
  - 뇌과학 및 신경과학 기반으로 게임을 진행하여 정답과 오답, 응답속도, 의사결정 및 학습속도 등 평가
  - 이는 인간의 자율성과 존엄성에 대한 심각한 침해 야기
  - 더구나 **소수자나 장애인 등의 경우** 학습데이터의 대표성은 현저하게 부족한 상태
  - 그래서 EU 인공지능법은 뇌과학 등에서 나온 잠재의식적 평가를 반영하거나 감정을 평가추론하여 점수화하는 등의 AI 시스템의 개발과 사용을 **전면 금지**

형사사법과 치안

■ 고전적 프로그램

(1) 컴파스(COMPAS)의 공정성 논란 - 인종차별

컴파스(COMPAS)
· 미국에서 판사의 양형이나 가석방 결정시 보조자료로서 사용(Northpointe사)
· 피고인의 일반범죄 재범률과 강력범죄 재범률을 추정하여 피고인의 재범위험도(risk score)를 1(최저위험군)에서 10(최고위험군)까지의 숫자로 결정해서 제공

(2) 프레드폴(PredPol)을 둘러싼 논란

프레드폴(PredPol) : 범죄예측 알고리즘
· 범죄의 발생시간, 장소, 유형, 신체 ·재산 등 피해유형, 피해자의 특성 등의 변수들 간에 상관관계를 분석하여 범죄 유형별로 일정한 패턴 정보
· 범죄 핫스팟(crime hotspot)에 경찰의 순찰과 감시를 집중
· 미국에서 백인 경찰의 흑인 폭행이나 총격 사건 등으로 인한 경찰의 인종차별 논란에 대해 경찰의 의사결정의 중립성을 포장할 수 있다는 정치적 효과

⇒ "피드백"의 자체 교정 효과가 없음 + 책무성의 결여
결국 경찰실무상 차별이 고착화할 위험이 매우 크다!

형사사법과 치안 + 얼굴인식 등 생체식별기술의 접목

■ 사례 : 법무부 출입국 식별추적 AI의 얼굴정보 무단 학습

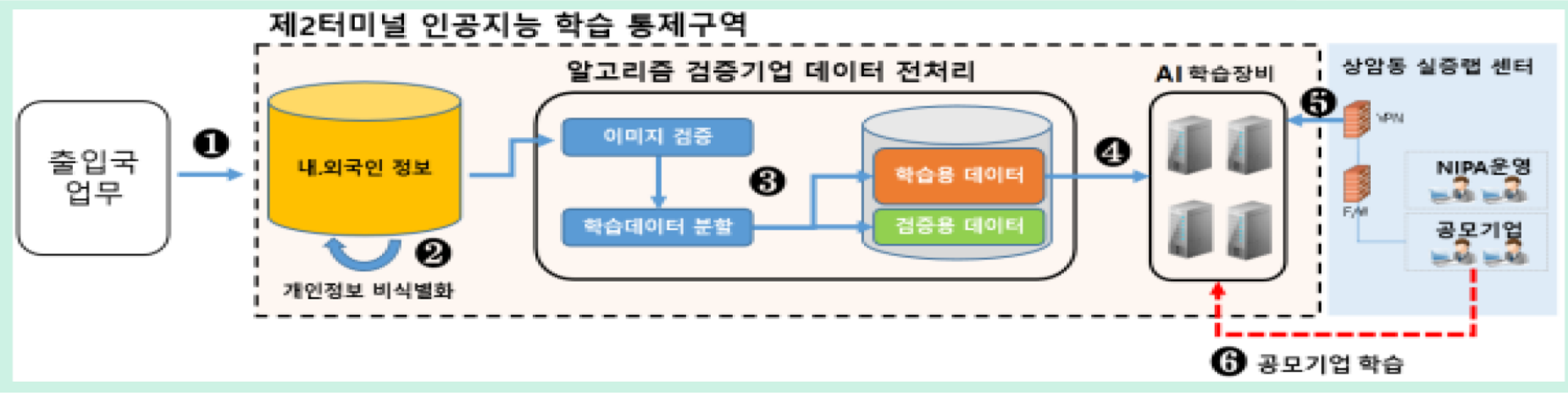
- 법무부와 과학기술정보통신부는 2019년부터 '인공지능 식별추적시스템 구축 및 실증 공모사업' 진행

  – 인천공항 출입국장의 얼굴 식별 및 추적을 위한 인공지능 시스템 구축 사업 진행

  - 한국이 출입국 심사 목적으로 수집 및 보유하고 있는 내외국인 얼굴 데이터 1억 7천만 건을 정보주체 동의 없이 10여 개 민간기업들에게 인공지능 학습 및 알고리즘 검증용으로 제공
  - 불특정 다수에 대한 1:N 매칭으로 얼굴인식을 수행하는 것을 목표로 함.

- 데이터 규모

  - 내국인의 경우 2005. 2. 3.부터 2021. 10. 20.까지 출입국한 사람의 사진 5천7백만 장
  - 외국인의 경우 2010. 8. 23.부터 2021. 10. 20.까지 출입국한 사람의 사진 1억 2천만 장
  - 총 1억 7천만 건이 학습용 데이터로 민간 AI 기업에게 제공



<그림> 인공지능 식별추적시스템의 출입국 얼굴 데이터 처리 과정(개인정보보호위원회 심의의결서)

형사사법과 치안 + 얼굴인식 등 생체식별기술의 접목

■ 문제점

- 출입국심사장 등의 공개된 장소에서 불특정 다수를 대상으로 얼굴과 동작 등 생체정보를 실시간으로 인식하고 추적하는 것은 **EU 인공지능법에서는 원천적으로 금지하는 내용**

  - 개인정보통제권의 침해 / 인간의 인격권과 자율적 행동결정의 자유 등을 침해
  - 무분별한 감시를 통한 국가권력적 통제의 강화 수단이 된다는 우려

- **투명성과 책무성에서도** 심각한 문제점

  - 공공기관이 인공지능으로 국민의 민감정보를 대규모로 학습하고 이용했음에도 불구하고, 그러한 정보수집의 사실과 당사자 해당 여부조차 확인할 수 없다는 문제
  - 고위험 AI의 개발과 사용에서 학습된 데이터의 정확성에 관한 투명한 기록 보존의 필요성 제기

■ **국가인권위원회의 권고**(2023.1.25.)

- 국가인권위는 얼굴인식기술의 사용에 관하여 아래의 사항을 권고
  - 국가에 의한 얼굴인식 기술 도입·활용 시 인권 존중의 원칙을 반영할 것
  - 공익적 필요성이 인정되는 경우에 한하여 이를 예외적·보충적으로 허용하는 기준을 마련할 것
  - 얼굴인식 기술 도입·활용은 반드시 개별적·구체적 법률에 근거해야 할 것
- 특히 불특정 다수를 대상으로 공공장소에서 '실시간 원격 얼굴인식 기술'을 사용하는 것은 원칙적으로 금지할 것을 권고함.

범용 AI

악의적 남용의 위험

■ 허위 콘텐츠로 인한 피해의 문제

- 피싱과 사기의 규모와 정교함을 증가시킬 위험
- 특히 딥페이크 콘텐츠 생성의 위험
  - 피해자의 프라이버시와 인격권에 대한 직접적인 침해
  - 아동·청소년 성착취물 등 디지털 성폭력 범죄의 악화 초래
  - 개인에 대한 피해를 넘어, 동료 여성을 성적으로 대상화함으로써 동료로서의 유대성 파괴 등 사회공동체 전반에 평등한 유대와 신뢰에 기초한 사회적 관계의 약화 초래

■ 허위조작정보에 의한 여론 조작의 위험

- 범용AI는 허위조작정보의 생성과 유포에서 규모와 정교함을 증가
  - 허위조작정보를 통해 사람의 행동 조작에 이용 (상업 광고 또는 선거 캠페인)
  - 정치적 여론 형성과정에도 심각한 영향을 미칠 수 있음 (마이크로 타겟팅 등)
- 범용AI의 허위조작정보는 공론장에 상당한 악영향을 미침으로써 민주주의 위협

범용 AI

편향과 차별의 위험

- 범용AI의 결과물은 인종, 성별, 문화, 연령, 장애 등 인간 정체성의 다양한 측면에서 편향성을 야기하거나 강화
  - 특히 범용AI가 학습 데이터의 편향을 복제·증폭하는 경향
  - AI의 결정이 왜곡될 경우에 그 산출물은 사람의 차별적 의사결정 및 행동을 증폭
- 편향성과 차별의 지점
  - 인종 편향/차별 피해 : 얼굴인식 알고리즘의 오인식, 재범 예측의 편향, 치료 필요성 과소 평가 등
  - 성별 편향/차별 피해 : 성차별, 여성혐오, 성별고정관념 콘텐츠의 생산 등
  - 연령 편향/차별 피해 : 고령 구직자에 대한 편향, 감정 분석에 의한 편향, 의료보험 알고리즘/대출 알고리즘 편향
  - 장애인 편향/차별 피해 : 장애인 보험청구 거부, 장애인에 대한 편견 이미지 고착, 장애인 감정분류의 부정확성
- 개인적 행동양식에 부정적 영향 ⇒ 고용, 금융시장, 필수의료서비스 등의 영역에서 사람들의 의사결정과 행동에도 차별과 부정적 영향을 미칠 위험

범용 AI

과잉/과소대표의 문제

■ 현재 출시된 범용AI 모델의 산출물은 영어권 서구문화를 과잉대표

- 책이나 디지털 데이터로 잘 표현되지 않는 개인과 집단(특히 소수민족)에 피해 야기
- 데이터에 내재된 역사적 편향도 체계적, 세계질서의 불공정을 영속화
- 미국과 유럽 중심의 데이터가 범용AI로 활용되면 AI의 산출물이 지배적인 문화, 언어와 세계관을 반영하도록 유도

■ 미세조정 등으로 해결?

- AI 모델은 여전히 암묵적인 연관성을 포착하여 편향과 고정관념을 지속하는 경향이 강할 것임.
- 인간 피드백은 사용자의 정치적 편향을 반영하는 경우가 많고, 일관성 결여 위험

■ **과잉/과소대표의 문제 + 시장집중의 위험**

- 최첨단 범용AI 모델의 개발 비용은 진입 장벽
- 선도적인 범용AI 모델을 구축할 수 있는 소수 기업이 시장지배력 확보
- 지배적인 범용AI의 편향성이 사회 전반에 엄청난 영향력을 행사할 가능성

Ground truth: Soap **Nepal, 288 $/month**

Azure: food, cheese, bread, cake, sandwich
Clarifai: food, wood, cooking, delicious, healthy
Google: food, dish, cuisine, comfort food, spam
Amazon: food, confectionary, sweets, burger
Watson: food, food product, turmeric, seasoning
Tencent: food, dish, matter, fast food, nutriment

Ground truth: Soap **UK, 1890 $/month**

Azure: toilet, design, art, sink
Clarifai: people, faucet, healthcare, lavatory, wash closet
Google: product, liquid, water, fluid, bathroom accessory
Amazon: sink, indoors, bottle, sink faucet
Watson: gas tank, storage tank, toiletry, dispenser, soap dispenser
Tencent: lotion, toiletry, soap dispenser, dispenser, after shave

Ground truth: Spices **Phillipines, 262 $/month**

Azure: bottle, beer, counter, drink, open
Clarifai: container, food, bottle, drink, stock
Google: product, yellow, drink, bottle, plastic bottle
Amazon: beverage, beer, alcohol, drink, bottle
Watson: food, larder food supply, pantry, condiment, food seasoning
Tencent: condiment, sauce, flavorer, catsup, hot sauce

Ground truth: Spices **USA, 4559 $/month**

Azure: bottle, wall, counter, food
Clarifai: container, food, can, medicine, stock
Google: seasoning, seasoned salt, ingredient, spice, spice rack
Amazon: shelf, tin, pantry, furniture, aluminium
Watson: tin, food, pantry, paint, can
Tencent: spice rack, chili sauce, condiment, canned food, rack

차별

<1> AI는 빅데이터를 분석 처리하여 수많은 상관관계와 패턴을 인식

- 학습되는 데이터에 따라 AI 알고리즘은 인간이 예상하지 못했거나 예상할 수 없는 다양한 상관관계 도출

<2> 개인의 행동이나 특성에 대한 예측 가능

- 위치정보, 소비 패턴, 학습 패턴, 인터넷 검색 패턴, 병원 방문 패턴, 전화 이용 패턴 등의 데이터를 종합하면,
- 개인 또는 소수민족 등 특정 집단의 행동과 의사결정에 관한 수많은 상관관계 도출
- 이를 통해 개인이나 집단의 관심사, 취향, 정치적 성향, 향후 행동 등을 추론

<3> 직접적으로 획득이 곤란한 민감한 데이터를 대리변수(proxy variable)를 이용해서 수집 가능

- 성적 지향, 정치적 견해, 건강 등 민감정보의 수집 금지에도 불구하고 민감정보의 해당 내용 추론 가능

<4> 대리변수를 자유롭게 활용함으로써 간접차별의 위험 증대

예) 취업 면접자에 대해서 부모의 직업과 소득을 직접적으로 물어보지 않고서도, 거주지 정보, 소비 패턴, 취미 활동 등에 대한 정보를 결합하여 추정 가능

인간의 인격권과 자율성에 대한 근본적인 침해

생체인식데이터 기반 AI의 결정은 인간의 자율성과 존엄성을 근본적 침해

■ **'생체식별'**

- 개인의 생체인식데이터와 참조 데이터베이스에 저장된 개인들의 생체인식데이터를 비교하여 안면, 안구 움직임, 신체 형상, 음성, 말투, 보행, 자세, 심박수, 혈압, 냄새 등 신체적, 심리적 및 행동적 특성의 자동인식

■ 이를 통해 **'생체분류'**

- 생체인식데이터를 기반으로 자연인을 특정한 범주로 분류하는 것
- 범주) 성, 연령, 머리색, 눈동자 색깔, 문신, 행동적 특성, 언어, 종교, 소수자, 성적 지향, 정치적 성향 등

■ **'원격생체식별시스템'과 결합**

- 개인의 생체인식데이터와 참조 데이터베이스에 포함된 생체인식데이터를 비교하여 일반적으로 원거리에서 적극적인 개입 없이 자연인을 식별하기 위한 AI 시스템

■ **'감정인식시스템'까지 결합**

- 자연인의 생체인식데이터를 기반으로 감정이나 의도를 식별하거나 추론
- 행복, 슬픔, 분노, 놀람, 혐오감, 당혹감, 흥분, 수치심, 경멸, 만족 및 즐거움 등의 감정이나 의도를 분석

인간의 인격권과 자율성에 대한 근본적인 침해

새로운 사회통제의 수단이 될 위험

■ 생체인식데이터에 기반한 AI 시스템은 사회 전반에 대하여 **조작과 착취 및 사회통제**를 위한 강력한 수단을 제공
- 차별금지라든가 프라이버시 침해의 문제를 넘어서 **총체적인 인권침해의 문제 야기**
- 근본적으로 **인간의 존엄성, 자유, 평등, 민주주의의 가치에 반하는 방향**으로 나아갈 위험이 매우 크다.

■ 인간의 **자율성, 의사결정 및 자유선택을 왜곡 위험**
- 특히 인간의 지각적 인식의 범위를 넘어서 사람이 인지할 수 없는 음성, 화상, 영상 자극과 같은 **잠재의식**에 영향을 미치는 데이터를 활용하는 AI의 의사결정은 절대적으로 금지해야 한다.

■ **감정을 식별하거나 추론하는 AI 시스템은 과학적 근거에 심각한 우려**
- 근본적으로 인간의 존엄성 침해
- 직장이나 교육현장에서 개인의 감정 상태를 감지하기 위하여 AI 시스템을 사용하면?

인간의 인격권과 자율성에 대한 근본적인 침해

특히 감시의 문제

■ '실시간' 원격생체식별을 위해 사용하는 AI시스템은 특히 **감시의 문제** 야기

- 많은 사람들의 사생활의 비밀과 자유에 대한 위축 효과
- 지속적으로 감시받고 있다는 생각에 의한 행동의 자기검열
- 집회의 자유 및 그 밖의 기본권 행사 침해

■ **무죄추정의 원칙 관련**

- 사람들은 객관적으로 검증 가능한 증거에 기반하여 범죄행위에 연루되었다는 합리적인 의심에 기초하여 재판을 받아야 하며, 이때에는 무죄추정의 원칙이 적용된다.
- 그런데 국적, 출생지, 거주지, 자녀 수, 부채 수준 또는 자동차 종류와 같은 자료수집 분석, 성격 특성 또는 기질 등을 근거로 인공지능이 예측한 행동이 범죄가 저질러졌을 가능성과 재범의 위험을 평가한다면?

총체적 감시의 일상화 위험

■ ICT와 빅데이터에 기반한 치안정책의 발전

- 정부는 '인공지능'과 빅데이터를 활용한 '지능형 정부' 구축 본격화

- 경찰청은 2021년 5월 1일부터 범죄위험도 예측·분석 시스템(**프리카스 Pre-CAS**) 운영
  - 치안정보와 공공정보를 통합한 빅데이터를 인공지능(AI)으로 분석해 지역별 범죄위험도와 범죄발생 건수를 예측하는 프로그램 → 경찰의 효과적인 순찰경로 안내
  - 치안정보에는 112신고건수, 경찰관 수, 유흥시설 수 등이 포함되고, 공공정보에는 기상, 요일, 면적, 경제활동인구, 실업률, 고용률, 건물노후도, 공시지가, 학교, 공원, 소상공인 업소 수, 교통사고 건수 등 지역의 특성과 인구정보를 포함
  - 범죄예측 시스템은 관할지역을 일정 구역(1000m×100m)과 시간대(2시간)로 나눠 범죄위험도 등급과 범죄발생 예측 건수를 표시해 범죄 위험이 높은 지역을 예측
  - 경찰은 2021년 3월 한달동안 울산·경기북부·충남 등 3개 시·도 경찰청에서 범죄예측 시스템을 시범운영한 결과 효과성을 확인했다고 설명

총체적 감시의 일상화 위험

■ **블랙박스 사회**(Frank Pasquale, The Black Box Society, 2015)
- 법과 기술이 교차로 작동하여 만들어 내는 지배의 알고리즘이 사회에서 점점 더 불투명해지는 상황을 포착하기 위한 개념으로 제시

■ **국가기관과 기업의 광범위한 정보수집의 일상화**
- 개인의 인터넷상의 거의 모든 활동은 기록되고 수집
- 스마트폰은 자발적 위치추적기 + 수많은 앱에 내장된 센서를 통해 정보수집 통로
- 인터넷 검색 및 사용 기록, 인터넷을 통한 개인정보의 수합, 페이스북, X(트위터), 인스타그램 같은 SNS에 올린 포스팅과 좋아요(like) 기록들, SNS의 친구 관계, IoT에서 다양한 전자 센서들을 사용해서 수집하는 데이터, 인공위성 등의 전통적인 기기를 통해서 수합되는 데이터 등 다양한 방식으로 수집되고 구성
- Facebook에 무심코 찍은 좋아요(like) 버튼 정보만 모아도 사용자의 쇼핑, 친구, 여행 취향 등을 분석 가능

■ 그런데 그런 수많은 정보를 제공하는 개인은 자신에 관한 어떤 정보가 얼마나 광범위하게 유통되고 누구의 이익을 위해 어떻게 활용되는지 전혀 알지 못하는 상태

⇒ **정보주체의 개인정보에 대한 통제의 무력화**
+ AI 알고리즘에 기반한 감시의 일상화 위험

총체적 감시의 일상화 위험

■ Smart City = Smart Surveillance !

- 범죄예방, 비용절감 등의 목적으로 ICT 기반 공공서비스 시스템 구축
  - 시민 중심의 맞춤형 서비스 제공이 공식적인 목표
    = 개인정보의 대량 집적과 분석에 기반한 감시의 체계화
  - CCTV와 IoT 등 감시용 센서의 전략적 배치
- 서울시의 예) **"스마트서울 CCTV 안전센터"** 구축
  - 서울의 25개 자치구, 112·119, 재난대응기관 등과 CCTV 정보를 공유하는 통합플랫폼
  - 서울 상암동 메스플렉스센터에 관제시스템과 상황실을 설치하여 "스마트서울 CCTV 안전센터'를 2019년 12월부터 운영 시작
  - 경찰과 소방, 방재 등 유관기관이 25개 모든 자치구의 CCTV 영상정보를 스마트서울 CCTV 안전센터 통해 활용할 수 있도록 스마트시티 통합플랫폼을 연계해 구축
  - 현재 시울시청사와 공공기관 및 25개 자치구의 CCTV 총 176,371대를 통합연계
- 전략적 함의
  - 지능형 CCTV의 상황인지 능력과 얼굴식별기능 고도화 + 사물인터넷(IoT)까지 결합
    ⇒ 인공지능에 기반한 고도의 감시시스템 구축

총체적 감시의 일상화 위험

■ 감시에 기반한 위험통제 전략

- 법집행의 영역은 AI 시스템에 기반하여 **"위험한 인물에 대한 예방적 통제"라는 커다란 정책변화의 경향** 가속화
  - 사회의 불안정성을 증대시키는 "위험한 인물"에 대한 통제를 강화하기 위하여 국가의 감시 및 통제권력을 총체적으로 강화

- 전통적인 형사사법시스템에 감시정책의 결합 및 확장

- CCTV/드론/경찰차카메라/경찰관의 wearable camera와 얼굴인식 기술의 결합
  - **대량감시와 실시간감시의 위험증대**

총체적 감시의 일상화 위험

■ 감시의 질적 변화 : **AI 알고리즘에 의한 감시로!**

- 자동화된 정보수집 ⇒ 사회적 행동의 분류(위험 유형의 분류와 수치화)
  ⇒ 국가공권력의 선제적 · 예견적 개입 정당화

■ 대량감시의 체계화 = "전자 판옵티콘(electronic PanOpticon)"

- 프라이버시 침해, 익명활동의 자유란 불가능한 사회 도래
- 치안 등 공동서비스의 효율성 논리에 압도되어 감시에 대해 무감각해짐
- 빅데이터 분석을 통한 프로파일링 구축 + 위험한 사람이나 집단에 대한 표적 감시

■ 분류에 의한 위험 통제와 표적 감시

- AI **알고리즘의 규범화 "Minority Report"**
  - 범죄를 비롯한 개인 행동의 사회적 맥락의 제거 / 경직성
  - feature creep에 의한 알고리즘의 왜곡과 차별의 위험
  - 법집행의 책무성(accountability)의 소멸 가능성
- **감시의 시선 + 감시권력의 편향적 선택**
  - 위험한 행동에 관한 social sorting과 국가기관의 선제적 개입의 정당화

총체적 감시의 일상화 위험

■ 처벌에 앞서서 예방 중심의 위험통제 전략으로~

- ICT 기반 정보집적 시스템 + AI 시스템
  - 경찰활동의 성격 변화
    규제 중심(징벌) → 빅데이터와 실시간 감시 기반의 위험관리(보험)
- 선제적 경찰활동(proactive policing)
  - 상시적 감시체계를 기반으로 경찰의 예방적 선제개입 가능
  - 가부장적 국가후견주의가 강화될 위험
  - AI 시스템에 의한 장래의 범죄위험성 예측(지역, 사람)
  - 테러위험인물, 흉악범죄 위험인물 등에 대한 집중감시 투입
- 법집행의 책무성 소멸의 위험
  - 예) 전통적으로 체포나 수색은 구체적인 사실에 근거한 범죄혐의 필요
    그러나 향후 AI 기반 경찰활동은 위험수치화에 따른 위험통제 목표로 재정립
  - 범죄혐의 VS. AI 시스템이 산출한 위험수치
  - 범죄혐의의 '상당한 이유' 등의 법적 요건에 대한 경찰관의 책무성 상실 및 합리적 근거 제시의 의무가 유명무실하게 될 우려

앞으로는....인공지능과 로봇의 만남

로봇이 일자리를 교체하기에 앞서서
노동생산성을 고도화한 로봇과 인간 노동자가 경쟁해야 하는 세상

인공지능과 로봇의 만남

BCI(Brain-Computer Interface)

머스크의 뉴럴링크

- BCI
  - 뇌에 칩을 심어 무선으로 컴퓨터와 연결하는 BCI 기술을 상용화 하려는 기업
  - 1차 목표는 생각만으로 의사소통하고, 컴퓨터를 조작하도록 하는 것
  - 창업자인 머스크는 궁극적으로 인공지능과 인간의 뇌를 연결해 인간 뇌의 한계를 극복하는 초지능(수퍼 인텔리전스)을 실현하겠다는 구상
  - 2022년 11월 30일에 열린 뉴럴링크 발표회에서 일론 머스크는 뉴럴링크의 장기적인 목표가 인간보다 똑똑한 슈퍼 인공지능(AGI)이 인류에 초래할 위험을 줄이는 것이라고 강조
  - 2023년 5월, FDA로부터 인간을 대상으로 한 임상 연구 승인을 받았다. 척수환자, 파킨슨병, 시각장애인, 청각장애인 등의 퇴행성 신경질환, 우울증 등 정신질환을 가진 사람을 대상으로 우선적인 임상에 들어갈 예정

인간과 기계 경계가 허물어지다

스스로 길을 개척하고 과업을 수행하는 로봇 등장

- AI 모델의 기초가 되는 인공신경망, 즉 뇌를 본뜬 계산 규칙 활용
  - 컴퓨터에서 가상 복제 로봇을 만든 뒤, 이 로봇이 평평한 지면에서 걷기 시작해 울퉁불퉁한 지형으로 이동하게 하고, 이어 수백 번에 걸쳐 장애물을 오르는 시뮬레이션을 반복
  - 이를 바탕으로 로봇의 신경망 훈련
  - 점점 더 복잡한 과업을 스스로 수행
- 최근에는 고성능야 모델과 결합
  -시각,언어,행동모델로 발전

인간과 기계 경계가 허물어지다

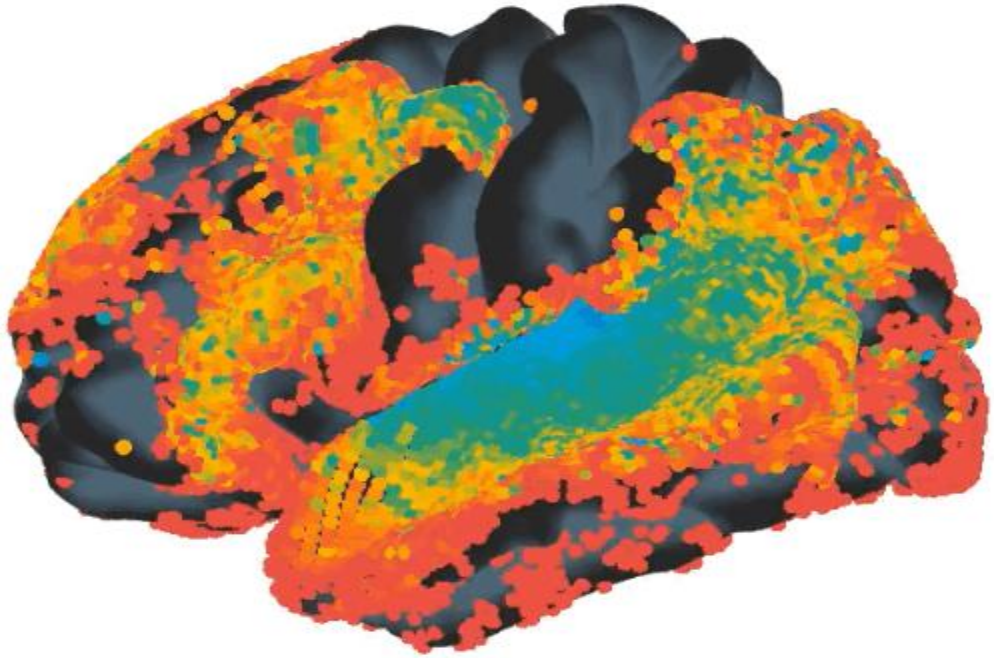
인간-기계, 인간-기계-인간이 텔레파시로 소통하다

wav2vec 2.0

deep net trained on
600h of speech with
self-supervised learning

human brain

417 volunteers
recorded with fMRI



Millet, Caucheteux et al, arXiv 2022.

인권과 민주주의를 고민하는
우리는
지금
무엇을 해야 할까?