

세 가지

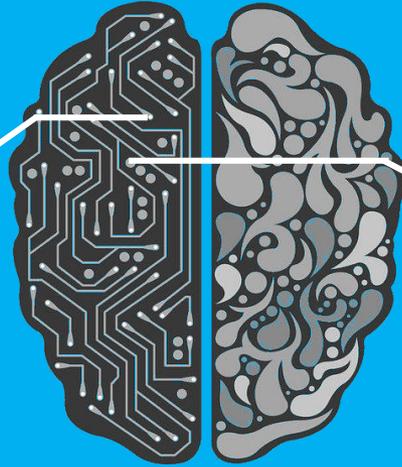
AI 연구소개

2021년 8월 26일 ILP 강연

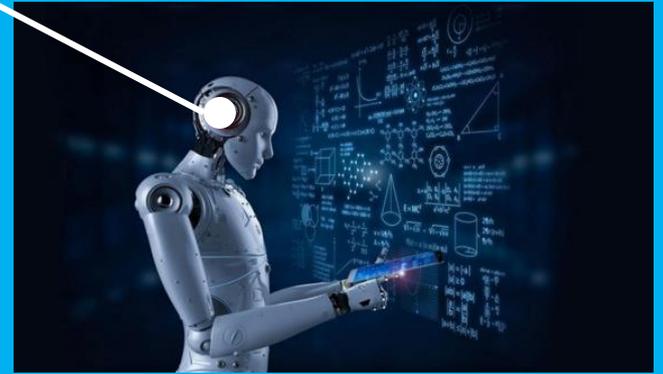
서창호

**KAIST 전기 및 전자공학부
KAIST AI Institute 부원장**

AI



AI 비서



AI 튜터

채용 심사



형량 결정



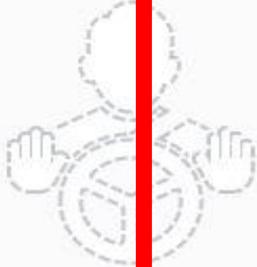
대출 심사



AI 분야 세 가지 게임체인저

1. 자율주행 기술
2. 신뢰할 수 있는 AI
3. 질병 예측 AI

1. 자율 주행

0 단계 비자동화	1 단계 운전자 보조	2 단계 부분 자동화	3 단계 조건부 자동화	4 단계 고도 자동화	5 단계 완전 자동화
					
<ul style="list-style-type: none"> · 운전자는 상황을 파악하고 운전함 	<ul style="list-style-type: none"> · 운전자는 상황을 파악하고 운전함 · 시스템이 운전자의 가/감속 또는 조향을 보조함 · 스마트 크루즈 컨트롤, 차로 유지 보조 등 	<ul style="list-style-type: none"> · 운전자는 상황을 파악하고 운전함 · 시스템이 운전자의 가/감속과 조향을 보조함 · 고속도로 주행 보조, 원격 스마트 주차 보조 등 	<ul style="list-style-type: none"> · 운전자가 시스템의 요청 시 운전함 · 시스템이 상황을 파악하고 운전함 · 교통 혼잡 시 저속주행, 고속도로 주행, 자동 차로 변경 등 	<ul style="list-style-type: none"> · 운전자가 시스템에 개입하지 않음 · 시스템이 정해진 도로와 조건 하에 운전함 	<ul style="list-style-type: none"> · 시스템이 모든 도로와 조건에서 운전함

현 상황

완전자동화를 위한 필수 기술

이유: "테이터" 부족

서울주행: 최근 연구

컴퓨터 시뮬레이터 활용: 사고장면 연출



서창호 교수 연구진, AAAI 2019 (국제인공지능학회)

사고 위험도 예측 성능

Demonstration Video

Each number represents the predicted probability of collision of the vehicle

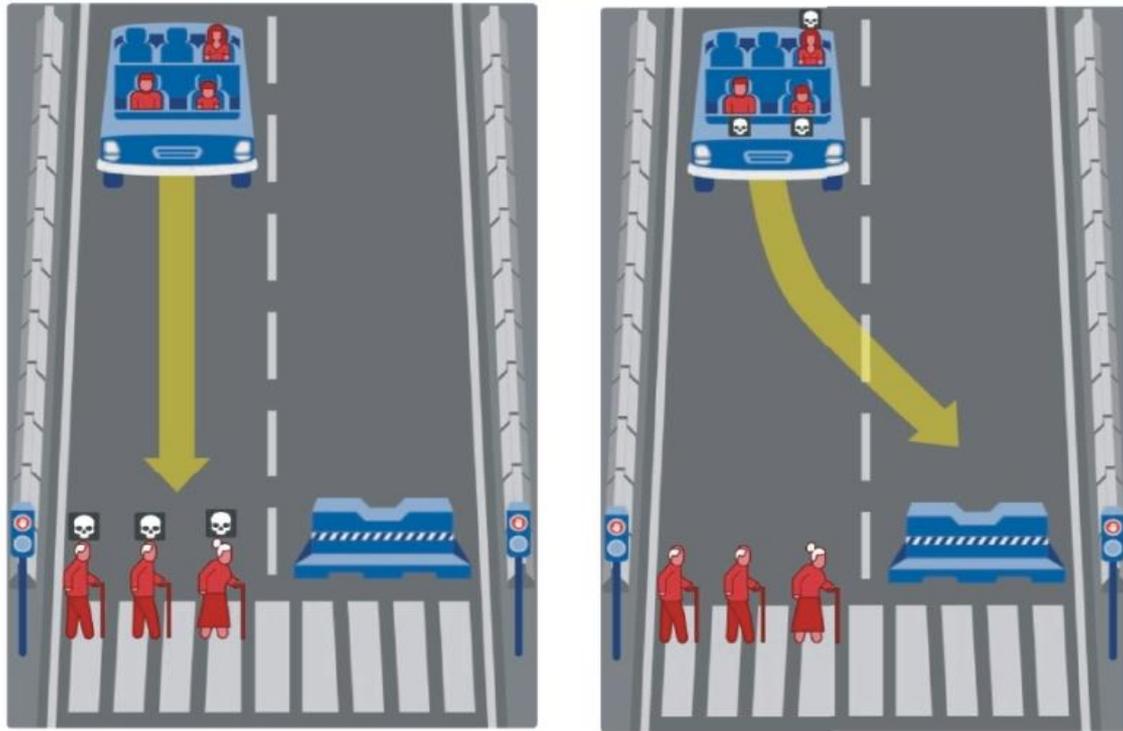
 : Ground truth dangerous vehicle

Predicted result of our algorithm



도전적 과제

트롤리 전차의 예와 같은 딜레마 상황



정당한 주행이란?

2. 신뢰할 수 있는 AI

필요성:

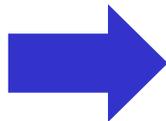
1. 인권이나 공정성 관련된 영역까지 AI가 인간을 대체하고 있음
2. 기계가 지배하는 세상을 막기 위해 AI는 의도된대로 설계되어야 함

현 상황

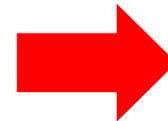
많이 부족한 상황

공정성 논란 사건:

Amazon scraps secret AI recruiting tool that showed bias against women



남성



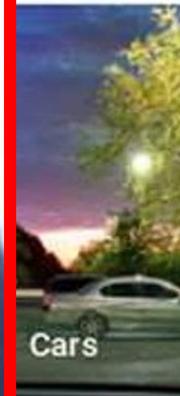
여성

출처: Reuters

현 상황

공정성 논란 사건:

Google



Google Mistakenly Tags Black People as 'Gorillas,'
Showing Limits of Algorithms

출처: The Wall Street Journal

공정성 논란의 원인

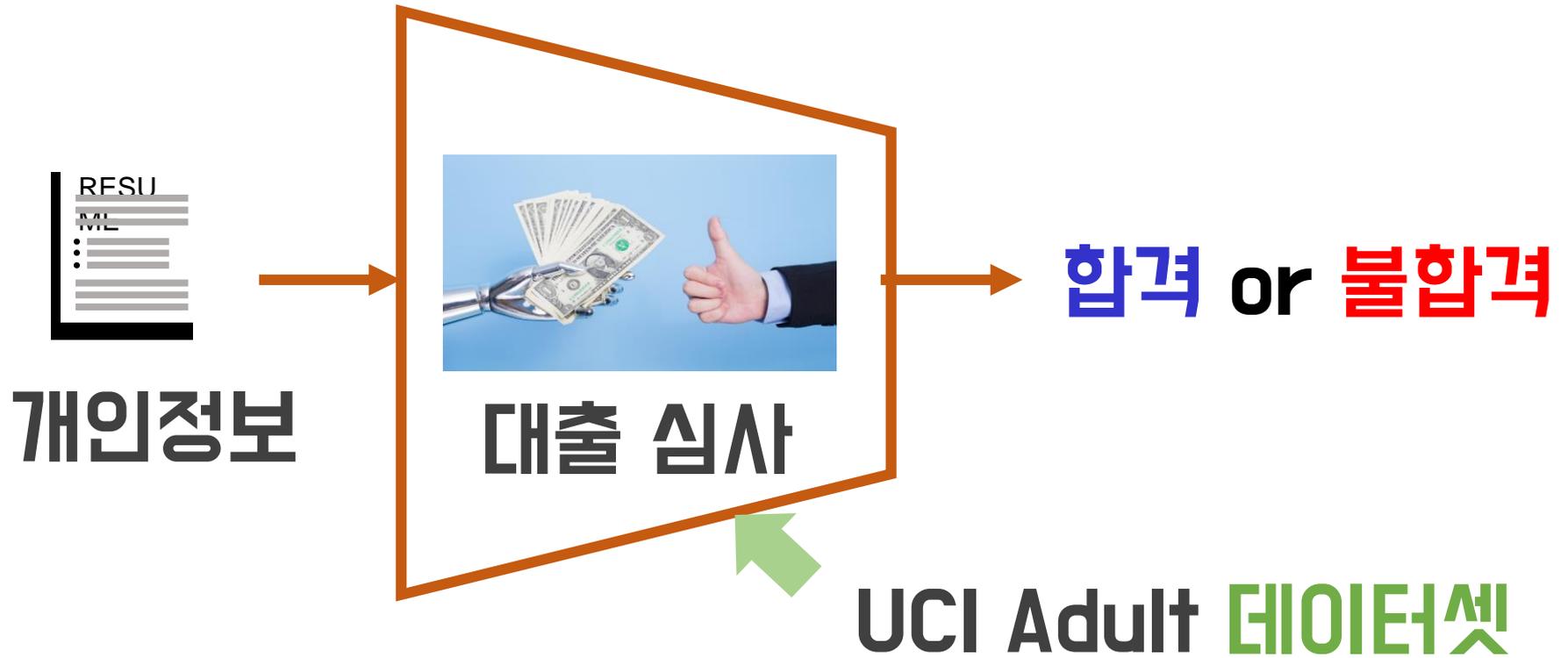
데이터 편향성!

신뢰할 수 있는 AI: 최근 연구

편향된 데이터로도 공정성을 보장하는 AI 기술 개발

서창호 교수 연구진, NeurIPS 2020 (국제신경망학회)

대출 심사 사례



VS



성능

서창호 교수 연구진, NeurIPS 2020 (국제신경망학회)

공정성 고려 x

85%

정확도

0.37

공정성

제안 알고리즘

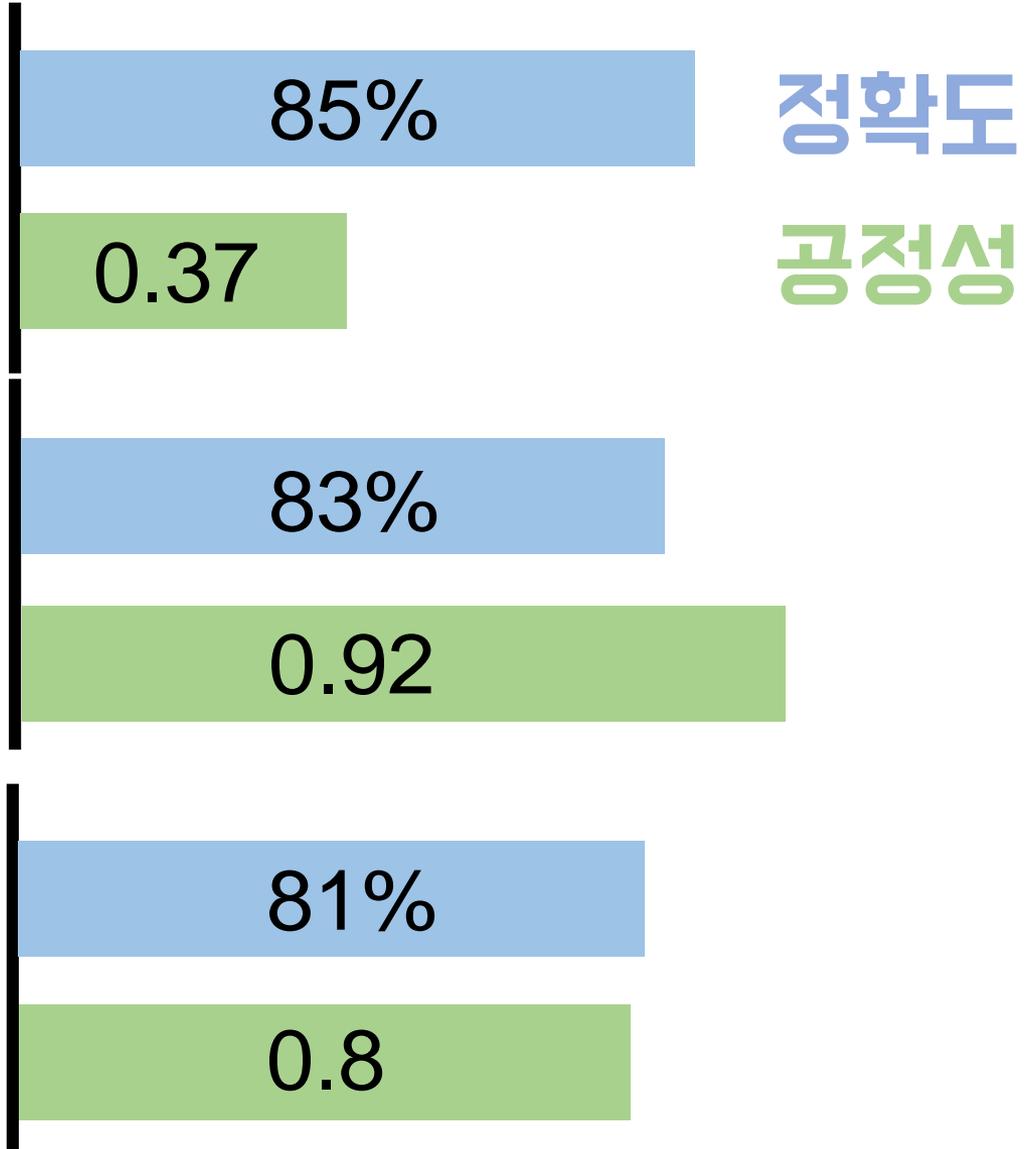
83%

0.92

최신 알고리즘
(Google)

81%

0.8

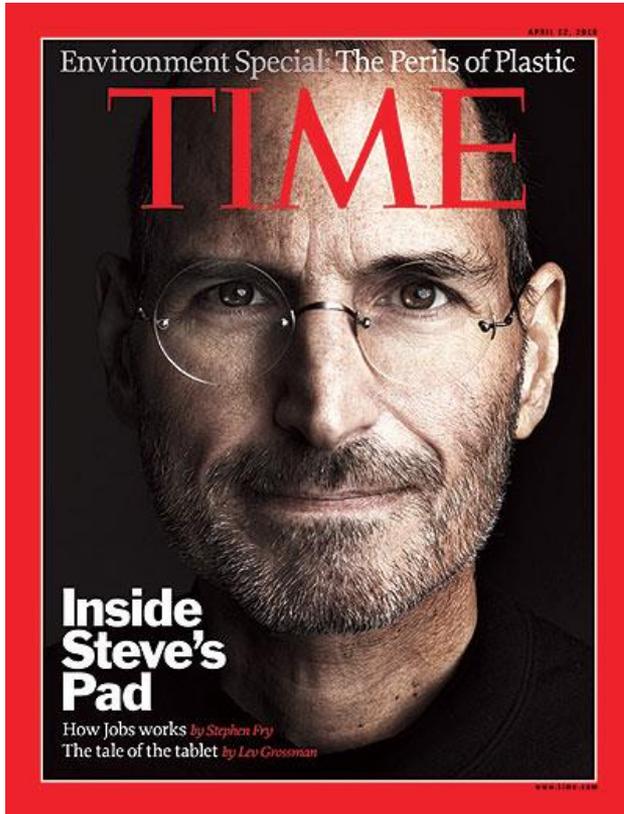


향후 계획

1. 공정성/윤리에 대한 올바른 개념 확립
2. 이를 AI 알고리즘으로 구현

3. 질병 예측 AI

“예측”의 중요성



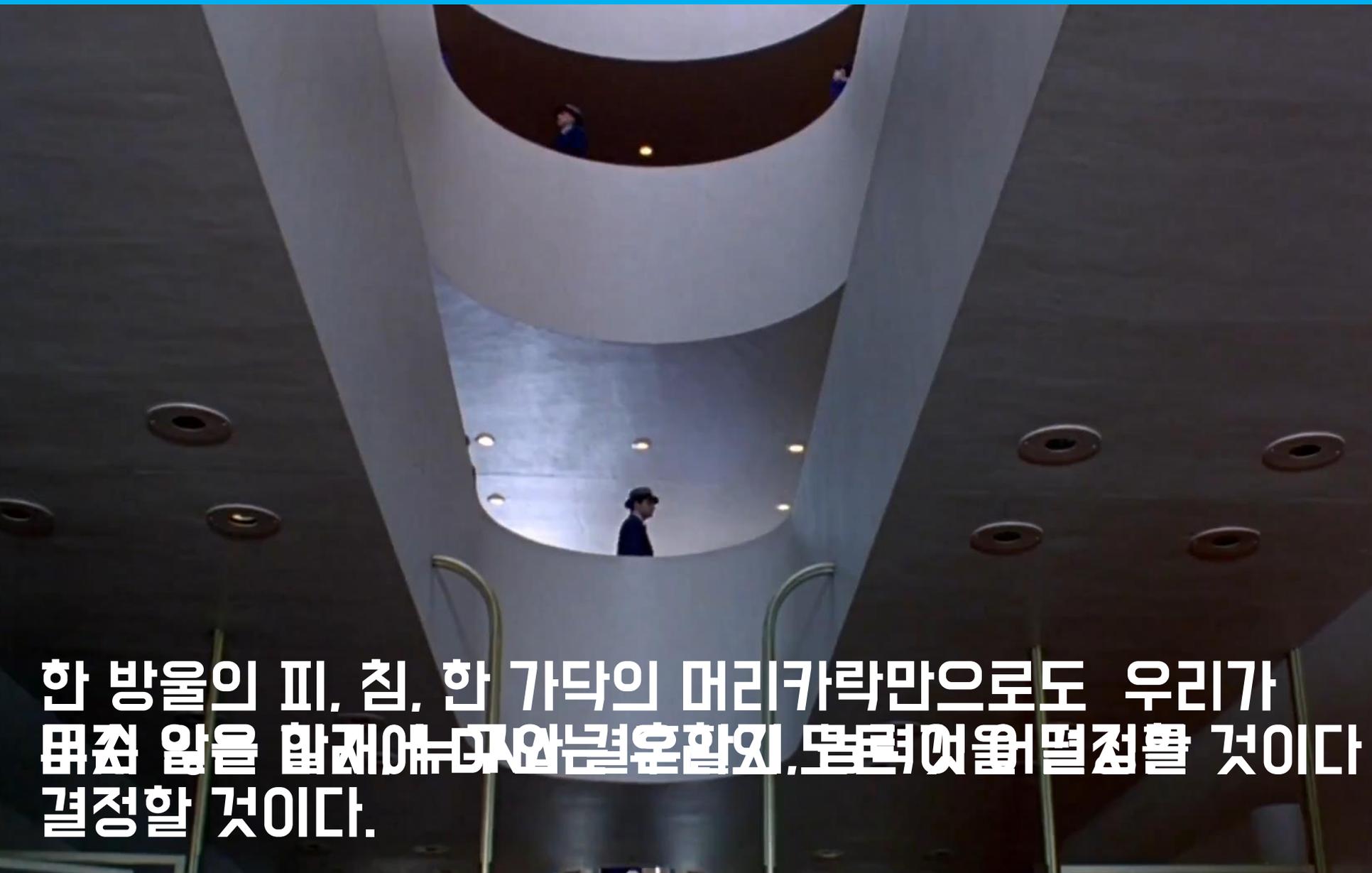
진단 → 개선



예측 → 예방

핵심 데이터?

출처: 영화 "가타카" 예고편



한 방울의 피, 침, 한 가닥의 머리카락만으로도 우리가
떠준 앓을 할래야 아니냐는 걸 혼합지, 모든 걸 어울러 펼쳐볼 것이다
결정할 것이다.

융합 기술



유전자



암 진단



암 발병 확률

현 상황

어려움이 있음:

- 1. 유전자 시퀀싱 기술의 한계**
- 2. 개인정보문제로 인한 유전자 정보 수집의 어려움**

질병 예측 AI: 최근 연구

유전자 시퀀싱 기술 보유

서창호 교수 연구진, ICML 2016 (국제머신러닝학회)

기존 시퀀싱 기술 대비 비용 절감

향후 계획

개인정보 문제 해결을 위한 병원/정부와의
협력

마무리

1. 자율주행 기술

2. 신뢰할 수 있는 AI

3. 질병 예측 AI