

AI 시대, 인간을 다시 성찰하다



이상욱

한양대학교 철학과 & 인공지능학과

HY 과학기술윤리·법·정책센터장

AI의 '낯선' 특징 3가지

1) '자각 없는 수행'

>> 의식적 경험의 부재

2) '이해할 수 없는 실패'

>> 진정한 이해의 부재

3) '계산과 실재의 간극'

>> '몸'이 없는 사이버 존재

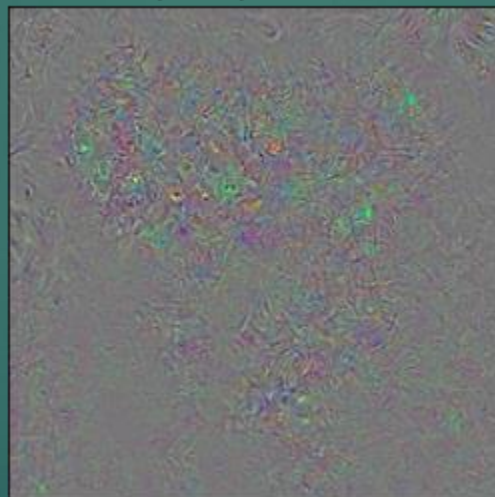


적대적 공격, 상식, 추상화, 유비 추론, 진정한 이해

African bush elephant



amplified perturbation



goldfish



과거의 AI, 미래의 AI



- '낮선' AI의 지능을 의인화하는 경향성
- 잘못된 관행이 훈련 데이터로 사용되어 발생하는 편향성
- AI의 '중간 수준' 직업 대체로 인한 경제적 양극화의 위험
- 인간과 비인간 사이의 엄격한 구분에 근거한 기존 사회제도의 변화 필요성

AI 딥페이크와 자기 데이터 권리

- 이미 현재 AI 기술은 통상적인 음성/이미지 인식을 통한 '개인 확인'이 안전하지 않음을 보여주고 있다.
- 심각한 보안 문제에 대한 대응이 시급하다.
- 보다 근본적인 수준에서 자신의 데이터가 자신도 모르는 사이에 '학습되어' 자신의 아바타가 만들어지는 상황을 막을 수 있는 통제권 권리가 있어야 한다.



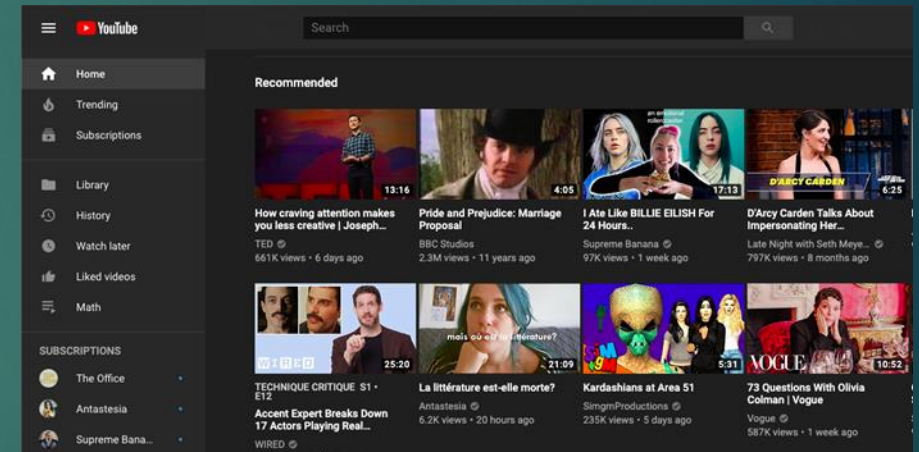
이성적이고 자율적인 근대적 인간관



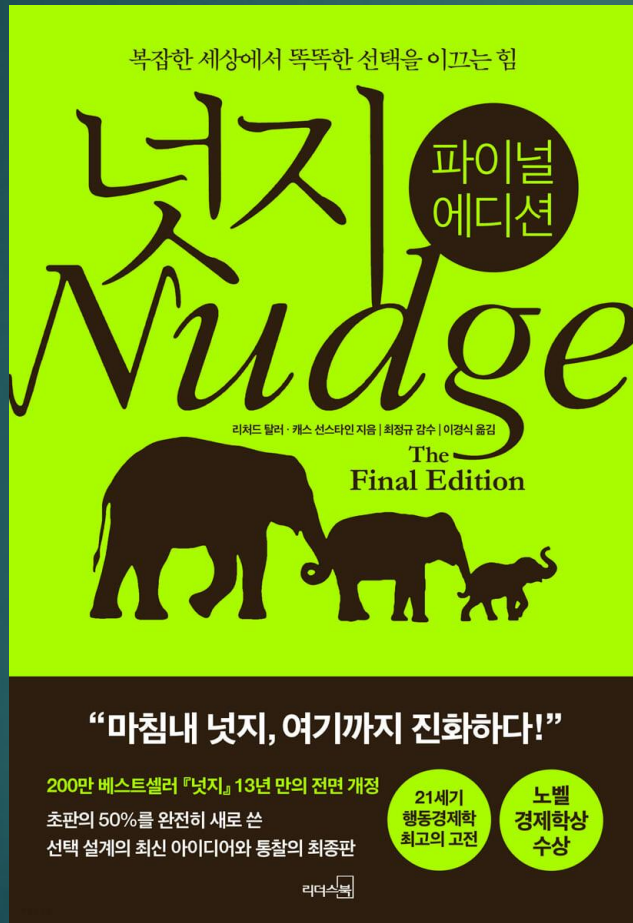
- ▶ 자율성(autonomy): “자기 스스로의 원칙에 따라 어떤 일을 하거나 자기 스스로 자신을 통제하여 절제하는 성질이나 특성” (<표준국어대사전>)
- ▶ 이런 의미에서 ‘자율주행차’는 실은 ‘운전자 없는 차(driverless vehicle)’이라고 부른 것이 맞음

AI 기반 추천시스템의 명암

- ▶ 나보다 더 나를 잘 아는 AI?
- ▶ '메아리방 효과': AI 기반 추천 알고리즘은 개인이 접근가능한 정보나 콘텐츠를 간접적인 방식으로 제한함으로써 개인의 취향이나 정치적 견해를 극단화시키거나 왜곡시킬 수 있다.
- ▶ 자유로운(free) 선택이긴 하지만 자율적(autonomous) 선택이 아닐 수 있다.



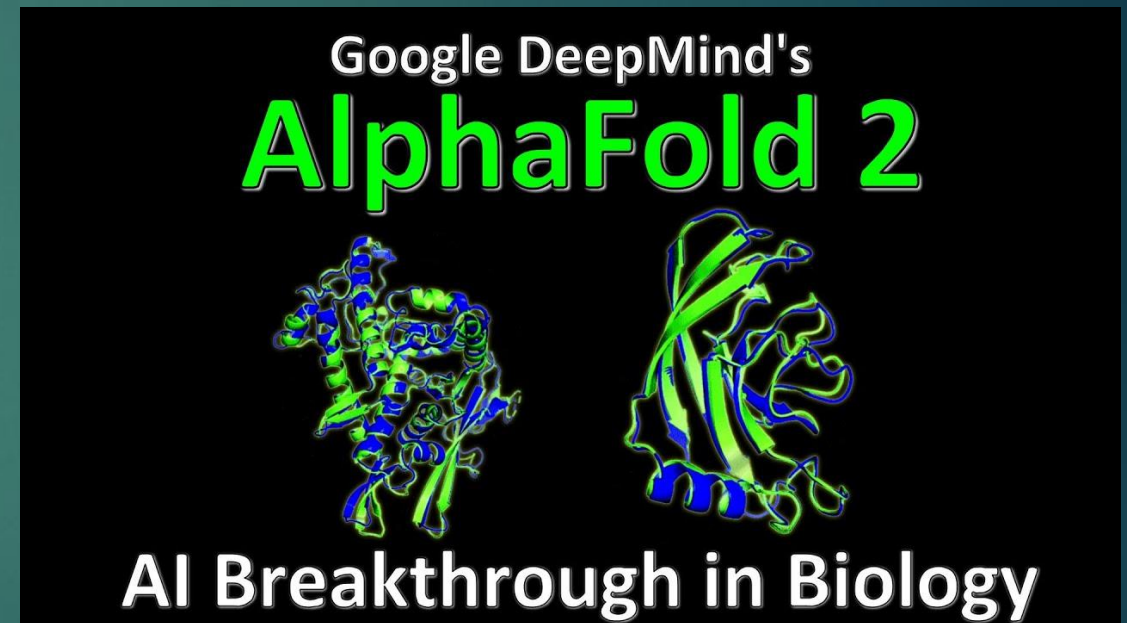
자율적 선택을 강요할 수 있는가?



- ▶ 미성년자가 아닌 성인의 자유로운 선택은 무조건 존중되어야 할까?
- ▶ 자유주의적 간섭주의의 '넛지' 전략
- ▶ AI의 선호기반 추천 시스템에서 '넛지' 전략을 사용하기 쉽지 않다. (무엇이 객관적으로 바람직한 선택지인지 자체가 논쟁적이다.)
- ▶ 그럼에도 불구하고 민주주의적 가치를 위협하는 AI 악용 사례에 대한 대응책은 필요

인공 지능의 강점 혹은 특별함

- ▶ 인공지능은 기본적으로 '알고리즘'이어서, 미리 정해진 규칙에 따라 데이터를 처리하여 결과를 산출한다.
- ▶ 하지만 최근 초거대 신경망 인공지능의 경우 처리 단위가 의미론적 기호가 아니고 그 복잡도가 너무 커서 최종 결과를 미리 예측하기 어렵다.
- ▶ 일반적으로 인공지능은 '목표'가 수학적으로 잘 정의될 수 있고, 정형화되고 균질적인 학습 데이터가 많을 때 잘 작동한다.



인간 지능의 강점 혹은 특별함



- ▶ 인간 지능은 동시에 처리할 수 있는 인지능력의 '폭'이 넓다는 강점을 갖는다. (그에 비해 속도나 '깊이'는 인공지능에 비해 불리하다.)
- ▶ 예를 들어 공항 관제사가 하는 일 각각(레이더 상에서 항공기 위치 확인, 활주로 상황 점검, 항공기 조종사와의 교신을 통한 상황 파악 등등)을 수행하는 인공지능은 개발가능하지만 이런 일을 모두 할 수 있는 능력은 당분간은 인간만이 가지고 있다.

서로 다른 두 지능의 협업

- ▶ 아마존의 물류 창고가 변화하는 양상이 시사하는 바가 크다.
- ▶ 원칙적으로 거대언어모형인 ChatGPT 같은 것이 일반인공지능이 될 가능성은 존재한다.
- ▶ 하지만 당분간은 인간과 인공지능의 협업이 대세가 될 것이다.
- ▶ 단, 단순하고 반복적인 일은 인공지능이 하고 창의적인 일은 인간이 한다는 생각은 더 이상 타당하지 않다!



자동화의 오랜 역사와 그 '불편함'



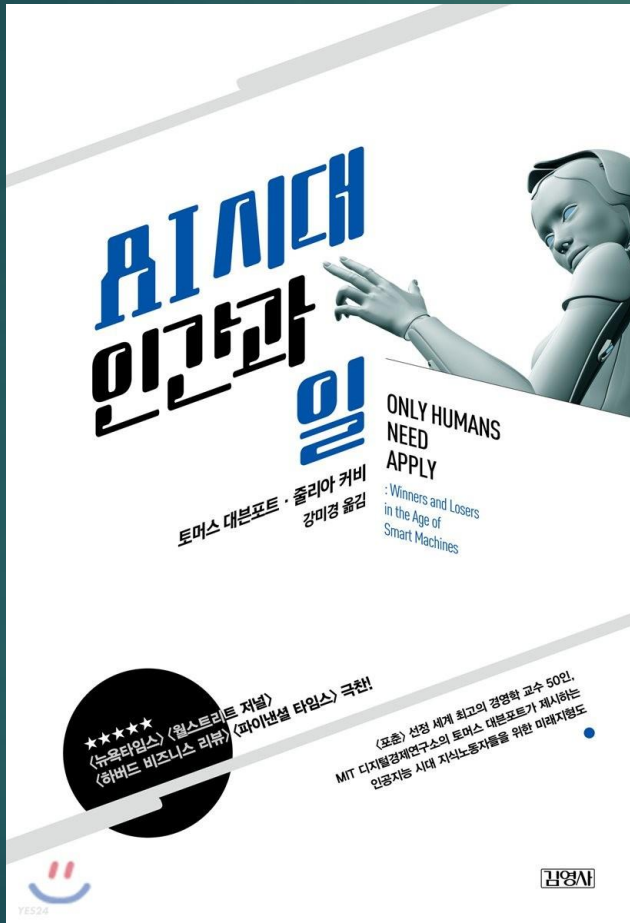
- 인공지능은 자동화(automation) 기술 진화의 오랜 역사의 연장성 상에 있다고 볼 수 있다.
- 자동화는 비숙련 육체 노동 대체에서 숙련 육체 노동 대체로 진행
- 인공지능은 인간의 정신 노동을 대체할 수 있는 잠재력을 가짐
- 자동화로 얻는 이득은 '비용 절감'이지만 실제로는 '비용 전가'인 경우가 더 많다!
- 결국 '자동화'는 기술적 필연이기 보다는 사회적, 제도적 결정이다.

자동화된 결정의 편리함과 위험성

- ▶ 인공지능 기반 추천 시스템의 편리함 (그런 자료 있는 줄도 몰랐던 자료들을 찾아주는 유튜브!)
- ▶ 2016년 12월 베를린에서 트럭의 자동 브레이크 장치(AEB)가 테러로 인한 대형 참사를 막았음.
- ▶ 2018년 10월 인도네시아 & 2019년 3월 에티오피아, 보잉 737 맥스 기종 추락 사고
- ▶ 각도 센서 오작동 & MCAS의 자동화된 결정 >> 기장이 이를 수정하려고 노력했으나 자동장치의 결정을 번복하는데 어려움 겪음.



자동화 대신 증강(augmentation)



- 증강: (1) 인간과 기계의 상호보완적 관계, (2) 더 많은 가치 창출 및 개인의 비약적 성장
- 물류 관리에 바코드를 도입한 것은 '증강'의 좋은 사례
- 계산원이 보다 효율적으로 일할 수 있도록 도움을 준 기술
- 도구를 사용하는 인간에 대한 스티브 잡스의 통찰: 자전거를 탄 인간이 콘도르를 앞지른다. 컴퓨터는 인간의 두뇌를 위한 자전거와 같다.

인류의 우주적 의미

- ▶ 다수의 (아마도 매우 낯선) 지적인 존재로 가득 차 있을 우주, 그리고 그 우주에서 활동할 인류의 미래를 고려하면...
- ▶ 어쩌면 '낯선 지능'을 갖춘 AI의 등장은 우리에게 그 미래를 대비할 수 있게 해주는 유익한 경험을 제공해 주는 것일 수 있다.
- ▶ 이 도전을 '성공적으로' 대응했을 때 인류는 새로운 단계로 도약할 수 있을지도...



감사합니다!

