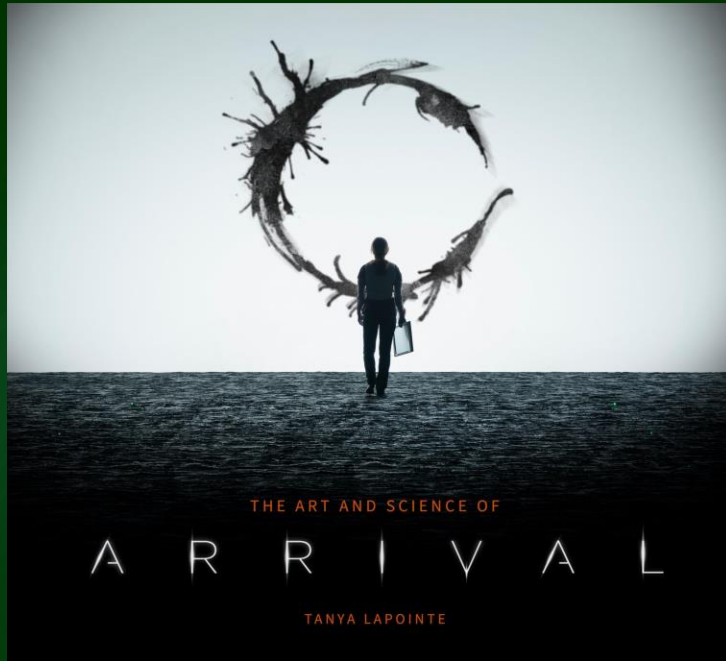


AI 윤리란 무엇인가? '낮선 지능'과 함께 살아가기



이상욱

한양대학교 철학과 & 인공지능학과

유네스코 세계과학기술윤리위원회 의장단

HY 과학기술윤리·법·정책센터장

인공 신경망 기반 AI의 원리

- 인공 신경망(Artificial Neural Network)은 인간 두뇌의 신경망 구조를 모방한 것이기는 하지만 차이점이 공통점보다 많다.
- 핵심은 계산 단위인 노드(~신경세포)는 우리가 이해할 수 있는 의미를 부여할 수 없는 '숫자'이고 학습과 예측 모두 이 숫자들의 계산을 통해 이루어진다는 점이다. ('자각 없는 수행')
- 그러므로 인공 **신경망** 기반 AI는 설계원리적으로 불투명하다. (Cf. 반투명 상자)



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

엄청나게 똑똑한 인공지능

- ▶ [Prompt] What is the reason for one of the meanings of 'bounty' to be subsidies?
- ▶ [Answer] The reason for one of the meanings of 'bounty' to be subsidies is due to the historical evolution of the word's usage. Originally, 'bounty' referred to a reward or payment offered for a specific action, often by a government. This could be for things like killing a harmful animal, capturing a criminal, or even encouraging certain behaviors like enlistment in the military. Over time, this meaning expanded to include financial incentives or subsidies provided by governments to support specific industries or activities.
- ▶ For example, governments might offer bounties (subsidies) to farmers for producing certain crops, to encourage domestic production and support agricultural industries. Similarly, bounties might be given to companies developing renewable energy technologies, to promote innovation and combat climate change.
- ▶ Essentially, the core concept of 'bounty' as a reward or incentive remained, but its application broadened to encompass government-funded support for various economic and social goals.

[Gemini Output, free version, 2024.12.25]

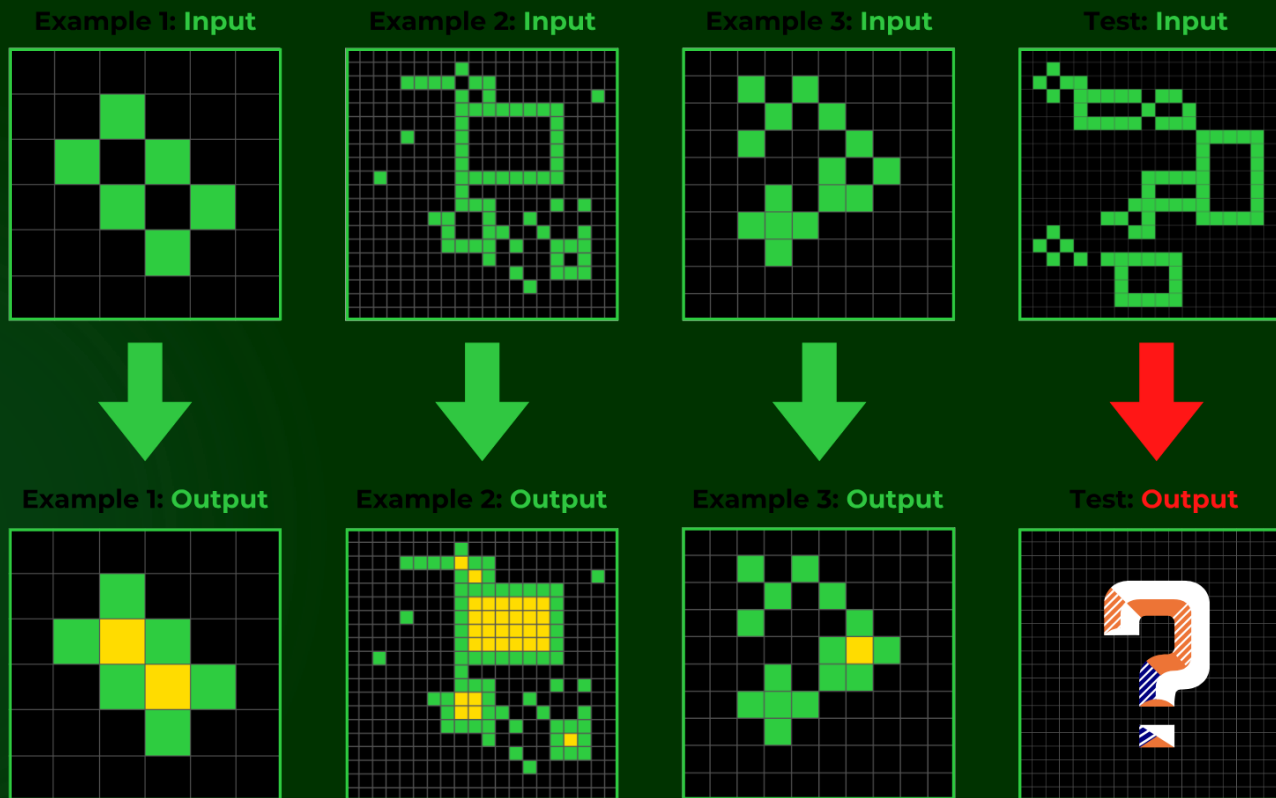
당황스럽게(?) 멍청한 인공지능

- ▶ [Prompt] Show me an image of Paris at night, which does not include Eiffel tower.
- ▶ [Answer] You got it! Here's an image of Paris at night, showcasing its beauty without the Eiffel Tower:
- ▶ As you can see, Paris still exudes a captivating charm even without its iconic landmark. The city's streets, buildings, and bridges are adorned with twinkling lights, creating a magical atmosphere.

[Gemini Output, free version, 2024.12.25]



인간과는 다른 방식으로 지적인 AI (ARC Prize)



계산과 실재의 간극

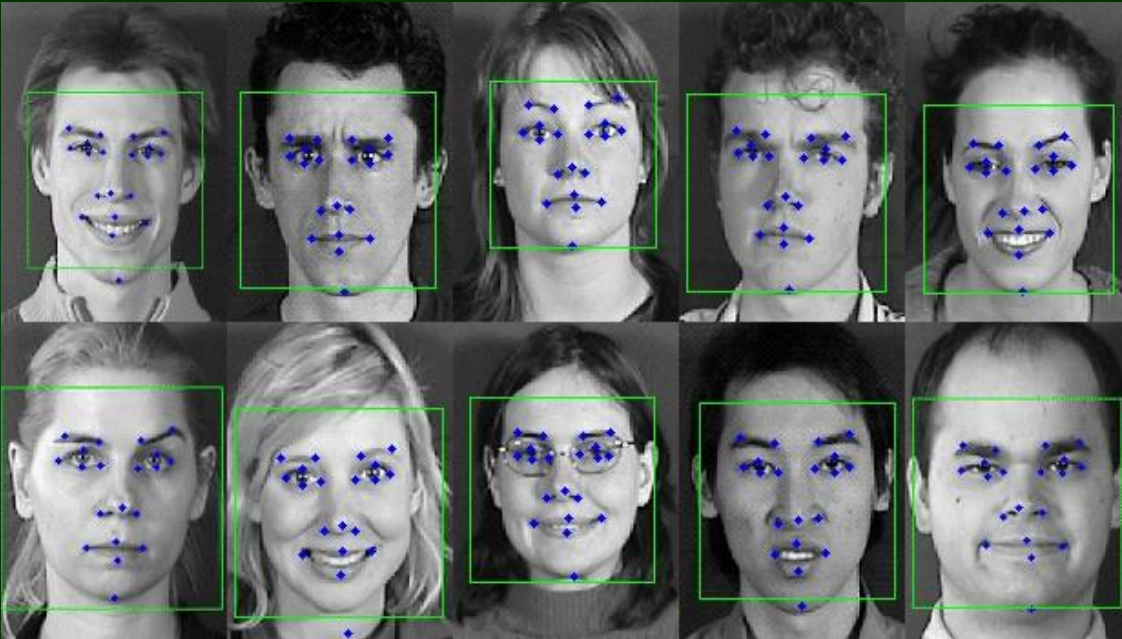


윤리(倫理) vs. Ethics

- 윤리: “사람으로서 마땅히 행하거나 지켜야 할 도리” (표준국어대사전)
<< 우리말에서 윤리란 개인이 누구에게나 명백하게 도리에 어긋나는 행동을 하는 것과 긴밀하게 관련되는 개념
- Ethics: “A set of moral principles, especially ones relating to or affirming a specified group, field, or form of conduct” (OED)
<< 어원을 따져 보면 ethics는 고대 그리스어에서 ‘인격^{character}’을 뜻하는 단어 ethos, 그리고 라틴어에서 ‘관습^{customs}’을 뜻하는 단어 mores와 깊은 관련됨. Mores는 moral(도덕적)의 어원이기도 함.
>> 우리말의 ‘사회적’과 거의 같은 범위를 다루되 본질적으로 규범적 판단을 요구한다는 차별점이 있음.

참고자료: 고등과학원 웹진 <HORIZON> <https://horizon.kias.re.kr/17815/>

인간만이 공감할 수 있을까?



- ▶ '공감'으로 무엇을 의미하는지에 따라 답이 다르다.

Cf.) '감정 없는 감정 로봇/인공지능'의 등장 (Affective Computing)

- ▶ MIT 컴퓨터 공학자 Weizenbaum의 'ELIZA' 프로젝트(1965): 사람들은 '진짜' 사람보다 사람이 아닌 것이 분명한 인공지능/로봇과 내밀한 이야기를 나누는 것을 더 편해할 수 있다.
- ▶ ChatGPT에게 '위로'를 얻는 사람들...
- ▶ 현명하게 공감하는 능력을 키우는 것이 중요하다!

윤리란 무엇인가?

- ▶ 우리말의 '윤리(倫理)'와 영어의 ethics의 의미 차이에 주목할 필요가 있다.
- ▶ 서양의 ethics는 고대 그리스의 (제한된) 직접민주주의 전통에서 도시국가의 운영을 위해 필요한 시민의 핵심역량에서 유래
- ▶ 예를 들어, 같은 그리스 도시국가 중에서도 아테네와 스파르타의 에토스는 상당히 달랐다!



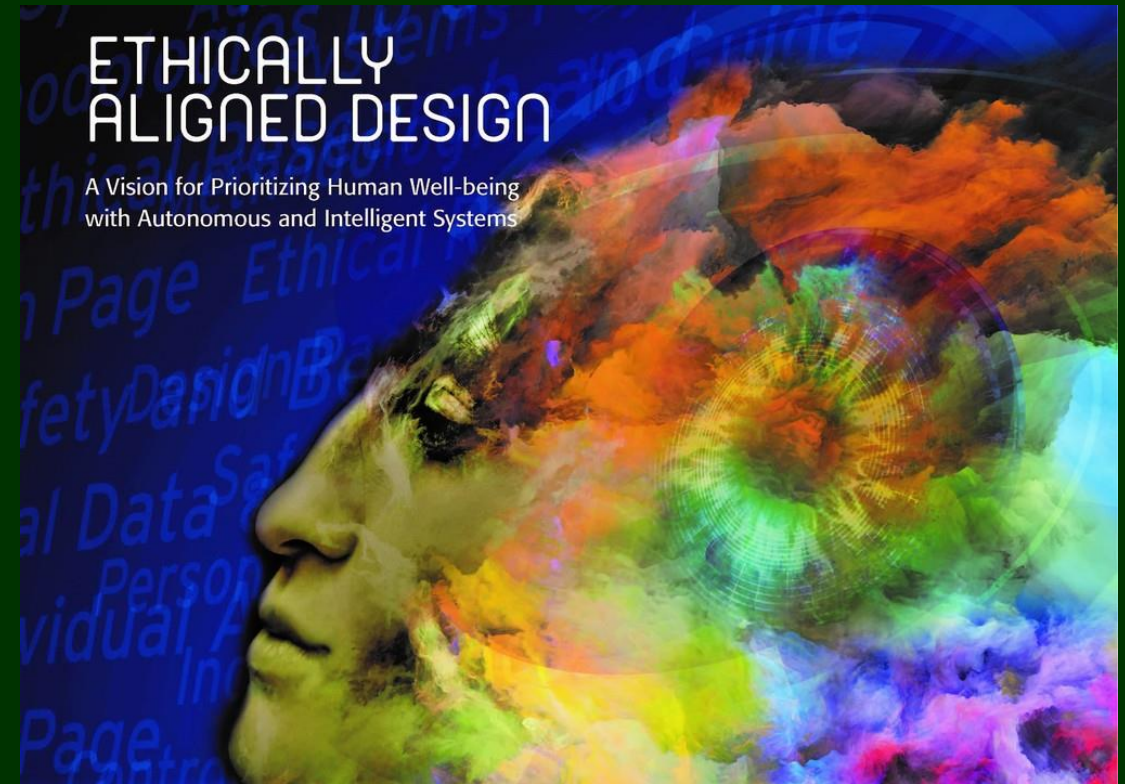
AI 윤리의 생생한 사례



- ▶ 이처럼 AI 윤리란 바람직한 방향으로 인공지능을 개발하고 활용하고 위한 여러 규범적 원칙과 과정에 대해 논의하고 사회적 합의를 통해 이를 실천해 나가는 적극적 의미로 이해되어야 한다.
- ▶ OpenAI CEO 샘 알트만의 해임과 복직 사건은 실리콘 밸리 내에서 '안전한' 인공지능 개발을 두고 서로 다른 윤리관이 충돌하고 있음을 잘 보여준다.
- ▶ 실리콘 밸리에서도 서로 다른 '에토스를 가진 연구자들이 경쟁을 벌이고 있음을 잘 보여준다.

전문직 윤리와 윤리적 설계(Ethics By Design)

- ▶ 전문직 윤리(professional ethics): 특정 전문가 '집단'에게 제한적으로 요구되는 책무와 내부 규율, 일반적으로 사회적 영향력과 지위가 높은 집단에게 책무성으로서 요구됨. (예: 의료 윤리, 법률가 윤리)
- ▶ 역사적으로 전문가 집단 윤리는 집단의 결속력과 사회적 지위 향상을 위해 도입됨. (예: 공학 윤리)
- ▶ IEEE의 '윤리적으로 정렬될 설계' 노력 >> 인공지능 윤리 설계 표준 정립으로 이어지고 있음. (Cf. ISO)



강의 내용 정리

- ▶ 인공지능은 분명 ‘똑똑하지만’ 인간에게는 매우 낮은 방식으로 지적인 결과물을 산출한다. 인공지능과 상호작용할 때 이 점을 항상 유의해야 한다.
- ▶ 온라인 프로그램인 인공지능과 물리적 공간에서 작동하는 로봇은 상당히 다른 기술이라는 점도 기억해 둘 필요가 있다.
- ▶ 인공지능의 ‘탁월한’ 공감 능력은 위험할 수 있다! (인간다운 ‘현명한’ 공감 능력을 길러야 한다.)
- ▶ 인공지능 윤리란 인류에게 막대한 영향을 끼칠 잠재력을 갖춘 새로운 기술인 인공지능을 어떻게 개발하고 활용하는 것이 바람직한지에 대해 사회적 논의 과정을 거쳐 규범을 만들어 나가는 과정 전체에 관련된다.
- ▶ 인공지능 윤리는 이런 의미에서 어떤 것을 가로막는 것이 아니라 바람직한 방향으로 인공지능이 개발될 수 있도록 돕는 역할을 한다!