



인간과 기계

갈등과 공생의 역사



자연 ≈ 거대한 자동시계

“나는 물리적 원인에 대해 많은 연구를 했다. 그 목적은 우주라는 기계가 살아 있는 성스러운 존재가 아니라 시계와 비슷한 것임을 보여주는 것이다.”

요하네스 케플러(1571-1630)

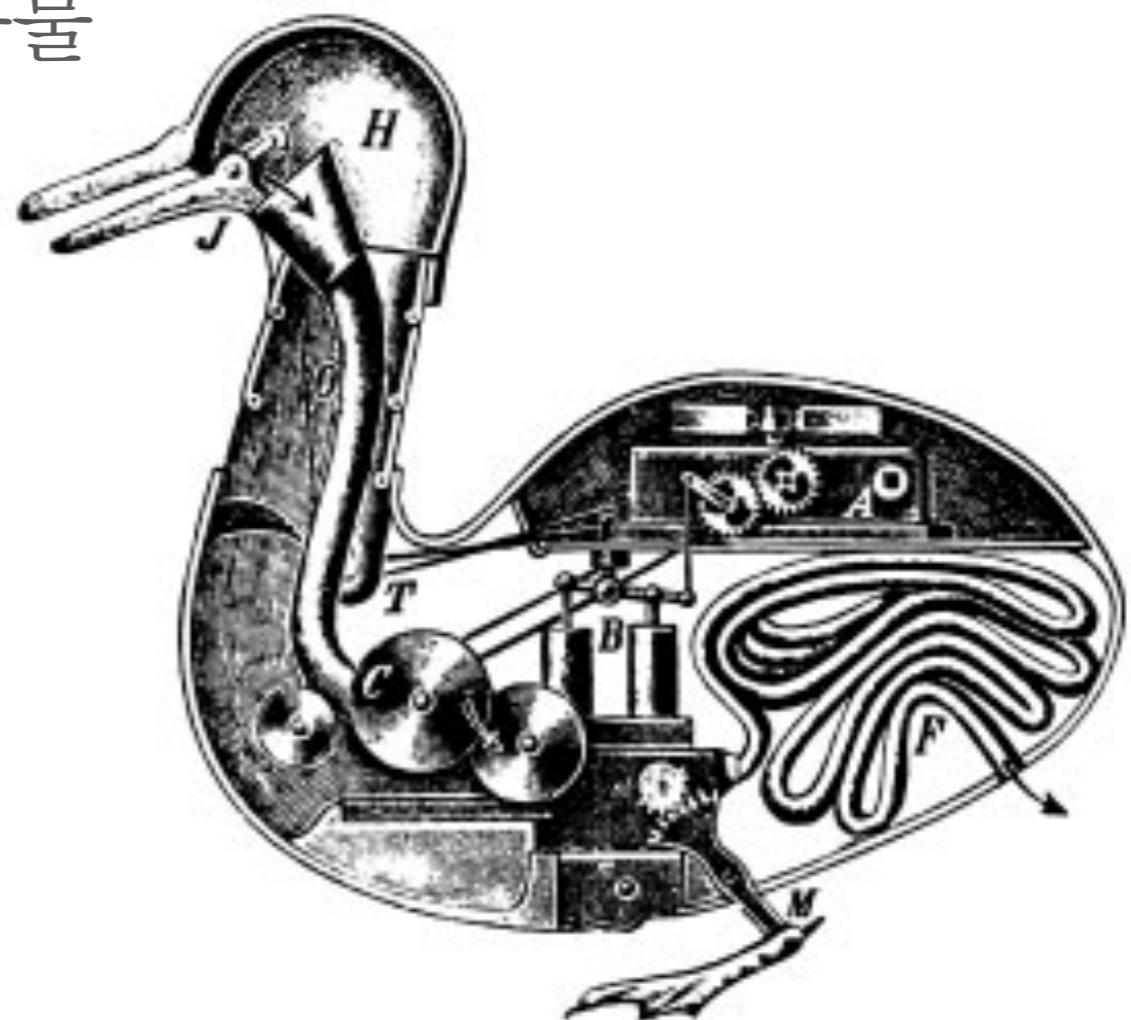
스트拉斯부르크 대성당의 거대한 천문시계
(1574년 완공)

똥싸는 오리

겉보기의 생명력 vs. 사실상의 인공물

복잡하지만 완벽하게 이해 가능

자연 역시 마찬가지로 이해 가능!





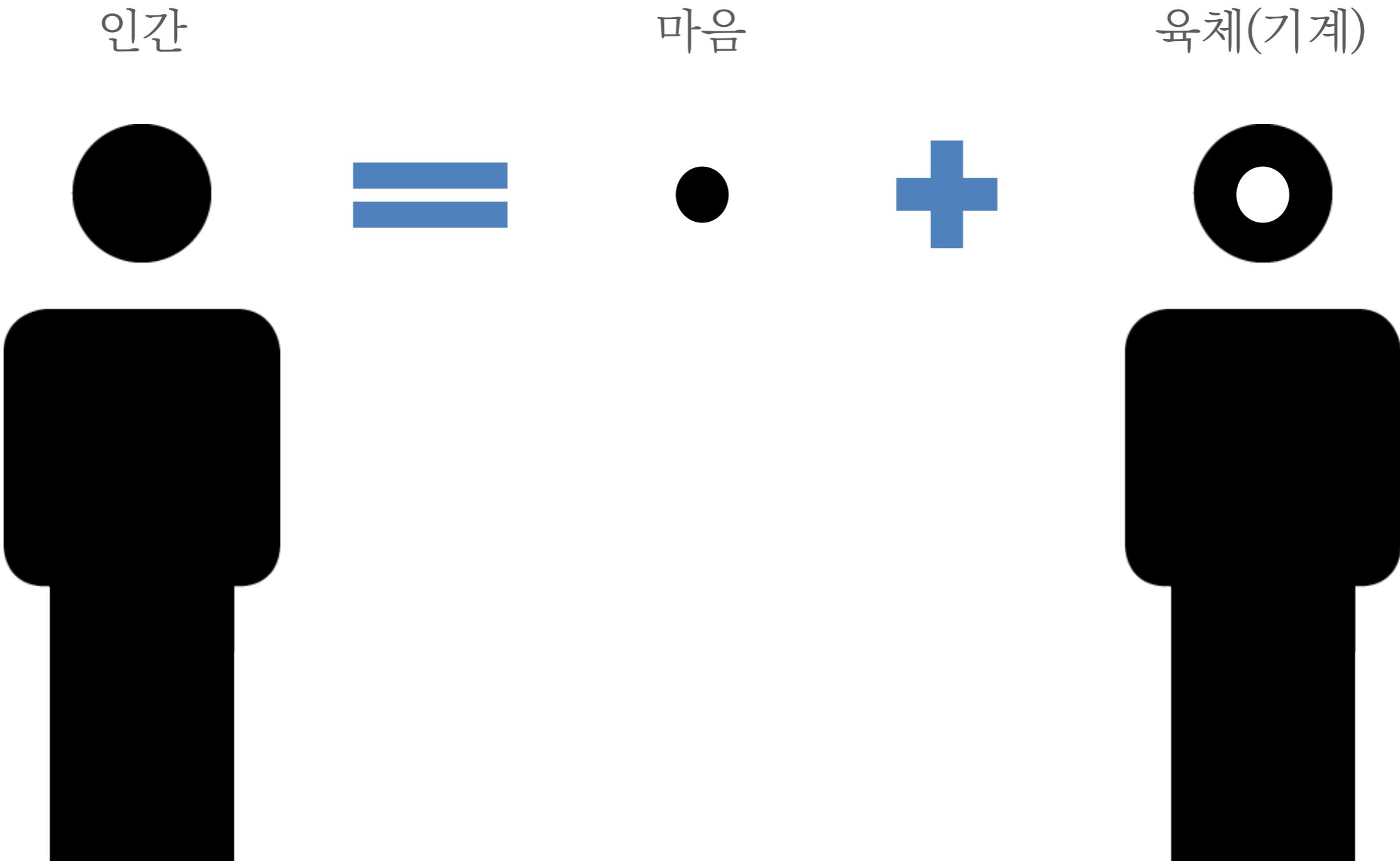
자연을 마치 기계처럼 생각하라!

“시계, 그리고 이런 종류의 다른 기계들은 비록 사람이 만든 것이기는 해도 그렇다고 그것들이 각자의 방식으로 스스로 움직이는 힘이 결여된 것은 아니다.”

어째서 인간의 호흡과 소화, 운동, 감각에 대해서는 시계, 분수대, 제분기의 운동을 설명하는 것과 같은 방식으로 설명하지 않는가?

르네 데카르트(1596-1650)

데카르트의 마음(MIND)-육체(BODY) 이원론





그러나

.....
만약 인간을 감쪽같이 닮은 자동 기
계가 있다면 기계와 인간을 어떻게
구별할 수 있을까?

〈블레이드 러너〉의 리플리칸트



산업혁명 : 노동의 기계화

- 기계가 세상을 파괴해서 “생명이 없는 태엽 장치만 남아서, 분해 조립이 가능한 부품들의 집합적이고 기계적인 삶”만을 남길 것이다. - 프리드리히 살러
- “인간은 손뿐 아니라 머리와 가슴까지 기계적이 되었다” - 칼라일
- <프랑켄슈타인> - 매리 셀리
- 노동이 기계에 종속됨으로써, 노동 과정에서 인간이 소외되고 있다 - 칼 마르크스

새뮤얼 버틀러



기계에 의한 인간 지배 가능성

-
- ▶ 거듭된 진화를 통해 의식을 가진 기계가 자신과 비슷한 새끼를 만들게 될 것이다.
 - ▶ “매일매일 우리는 점점 더 그들에 의존하고 있다. 더 많은 사람들이 그들의 노예가 되고 있고, 더 많은 사람들이 기계적 생명을 발전시키는 데 더 많은 에너지를 쓸고 있다. 이 결과는 단순히 시간의 문제일 것이지만, 기계들이 이 세계와 그 구성원들 모두에 대해서 확실한 우위를 점하는 시간이 도래할 것이다.” (1865)

KEANU REEVES LAURENCE FISHBURNE



<매트릭스> 3부작

- 새로운 인간-기계 관계의 모색
- 기계에 저항하는 공동체 시온
역시 거대한 엔진, 동력 기계,
APU와 같은 기계들에 의존.
- 영화의 결말은 인간 해방이 아
닌 기계와의 타협을 통한 불안
전한 공생 관계의 지속
- 인간의 특징으로 부각된 면모
- 나약한 육체/나약한 마음
(이성보다 감정이 중요)
- 인간 선택의 예측 불가능성



기계에 대한 두려움의 증대

- ▶ 기계의 복잡성과 불투명성 증대
- ▶ 기계에 대한 통제력 상실의 느낌
- ▶ 기계가 인간을 쓸모 없는 존재로 만든다는 두려움
- ▶ 기계의 발전과 함께 없어지는 직업보다 새로 생기는 직업이 더 많지만, 직업의 생성과 소멸은 불균등.
- ▶ 숙련에 근거한 공장 노동자는 점차 줄어드는 대신, 고액 지식 노동자 증가, 비정규직 서비스 노동자는 더 많이 증가.



인공지능

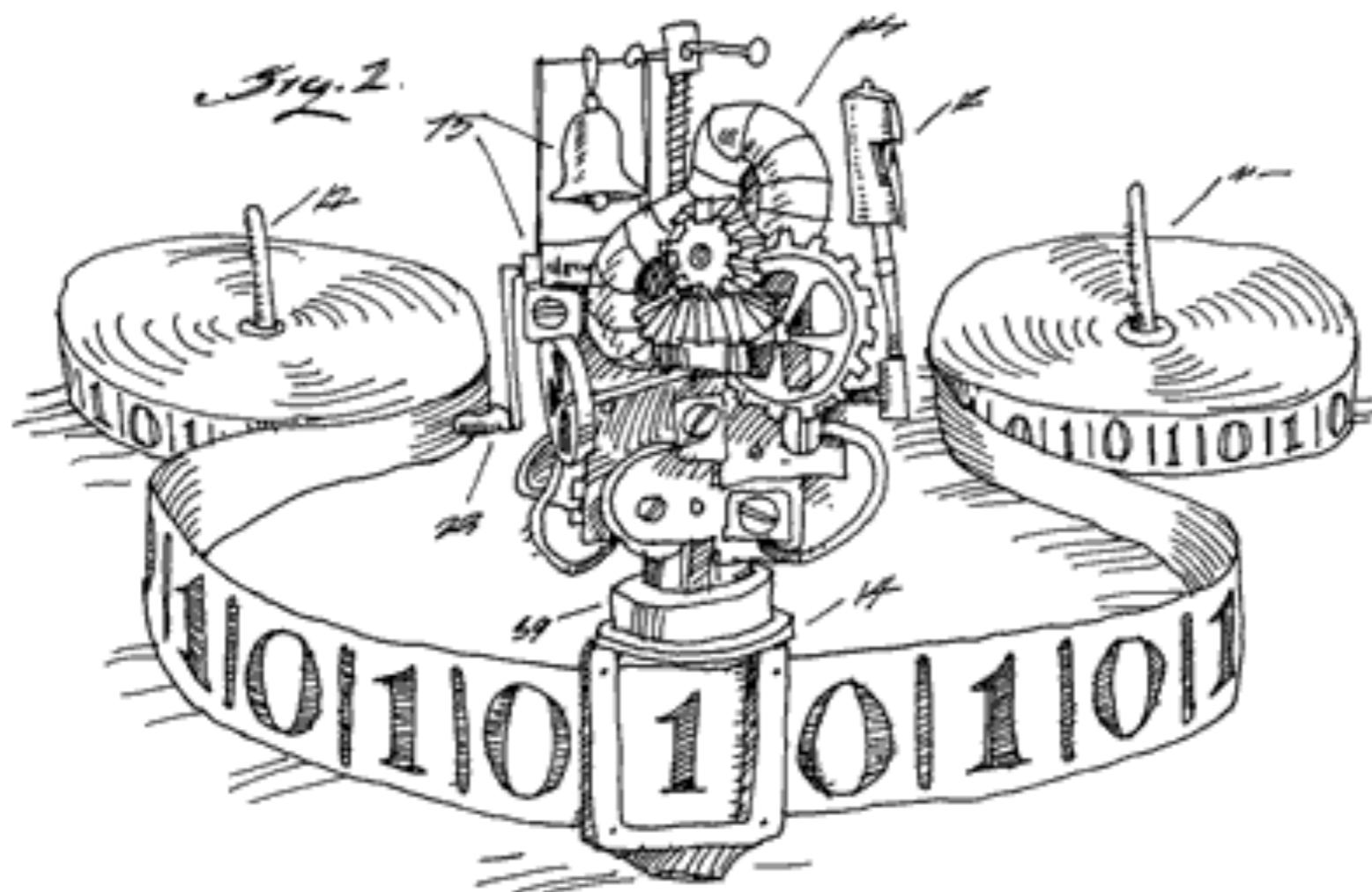
- ▶ I. J. 굿의 ‘초지능 기계’ (1965)
- ▶ 컴퓨터의 발전으로 언젠가 인간의 지능을 초월하는 인공지능을 가진 기계 만들어질 것.
- ▶ 최초의 초지능 기계가 만들어지면, 그 기계는 점점 더 뛰어난 기계를 제작함에 따라 순식간에 인류는 인공지능 기계보다 한참 열등한 종이 될 것.
- ▶ 두 가지 질문
 - ▶ 의식을 가진 기계는 가능한가?
 - ▶ 초지능 기계를 만들 수 있다면, 인간은 무엇에 대비해야 하는가?

튜링 테스트 (1950)



<계산 기계와 지능> (1950)

- ▶ 테스트에서 기계와 인간을 구별할 수 없다면,
그 기계는 생각을 한다고 할 수 있다.
- ▶ 그 기계의 후보는 추상적으로는 튜링 기계,
구체적으로는 디지털 컴퓨터(프로그램에 따른 기호 처리기)

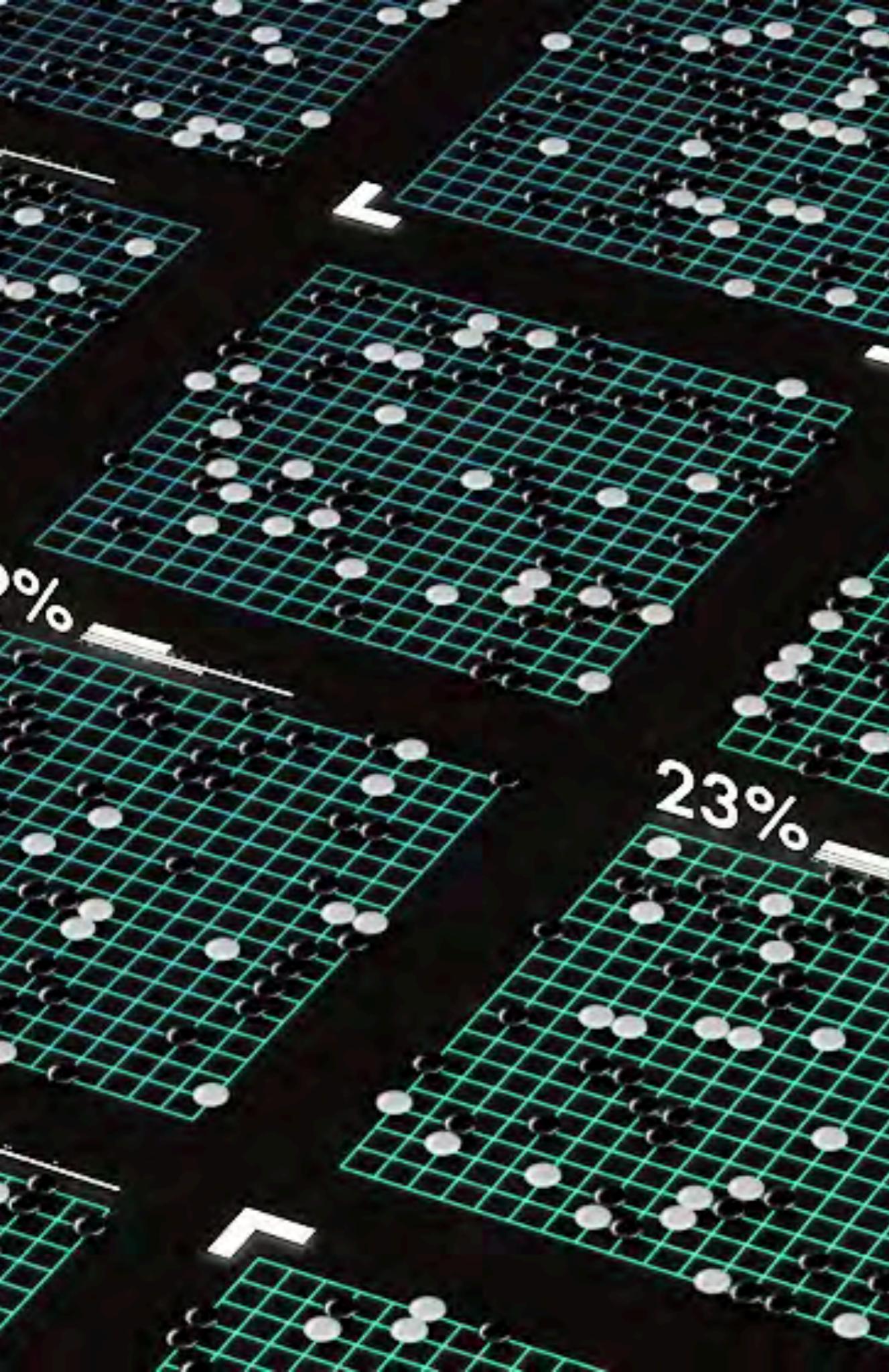


가능한 반박과 튜링의 대응 1

- ▶ 수학적 반박 : 일관된 논리 체계에 기반한 디지털 컴퓨터도 답할 수 없는 질문이 반드시 존재하기 마련이다. (괴델의 증명에 의존)
- ▶ 튜링의 반론 : 똑같은 한계가 사람에게도 있을 수 있다.
- ▶ 의식에 근거한 반박 : 아무리 훌륭한 기계도 기쁨이나 슬픔과 같은 감정을 느낄 수 없을 것이다.
- ▶ 튜링의 반론 : 다른 ‘사람’이 그런 의식적 특징들(이해, 기쁨, 슬픔, 흥분, 분노, 우울 등)을 가진다는 것은 어떻게 믿느냐? 만약 기계가 그런 것을 가진 듯한 답변을 훌륭하게 해낸다면?

가능한 반박과 튜링의 대응 2

- ▶ 기계의 여러 결함에 근거한 반박 : 컴퓨터는 인간 사고가 가진 여러 특징들(주체성, 도덕성, 실수, 경험으로부터의 학습, 창조성 등)을 가지지 못한다.
- ▶ 튜링의 반론 : 과거의 기계로부터 잘못 귀납한 결과이다.
- ▶ 행위의 비형식성에 근거한 반론 : 인간의 경우 모든 상황에 대한 행위 규칙이 정해져 있지 않다. 따라서 인간은 규칙에 따라 작동할 뿐 인 기계일 수 없다.
- ▶ 튜링의 반론 : 의식적 ‘행위 규칙’과 자연 법칙으로서의 ‘행동 법칙’을 구분할 때, ‘행동 법칙’이 존재하지 않는다는 것은 확신할 수 없다.



인공지능 연구의 성과들

- ▶ 그레이 월터의 ‘거북’ (1959)
- ▶ 튜링 테스트 통과를 위한 다양한 채팅봇들 (2009)
- ▶ 체스 컴퓨터 딥 블루
- ▶ 패턴 인식, 기계 학습
- ▶ 알파고, 자율 주행 자동차 ...



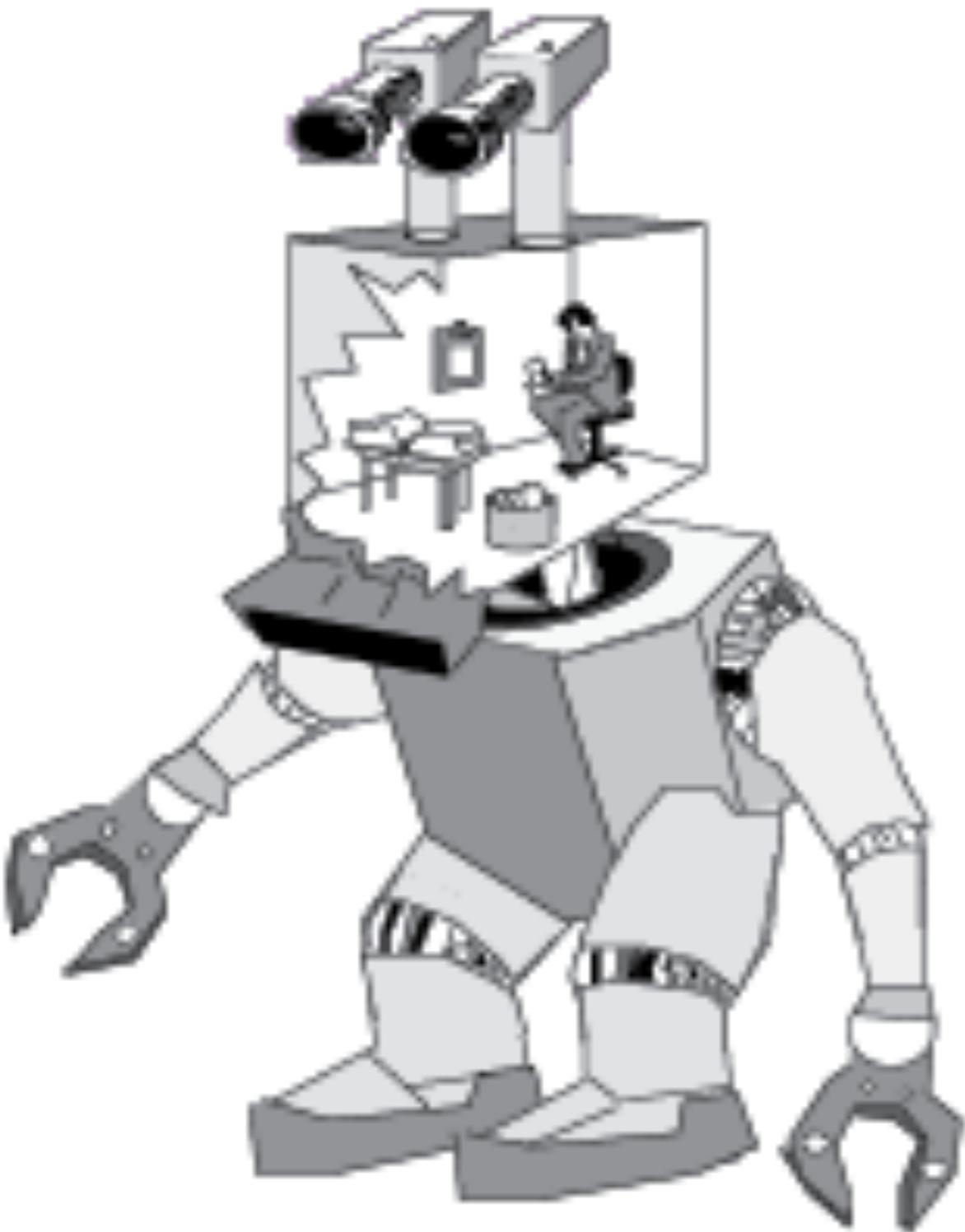
즐거움과 고통의 문제

- 배터리가 떨어져 충전기로 돌아가는 ‘거북’ : “배가 고파서 충전의 욕구가 생긴다”?
- P(leasure)-P(ain) 계수 기법
- 그러나 기계가 p-p 계수의 변화에 따라 기쁨이나 아픔을 ‘느낀다’고 할 수 있는가?
- Yes : 강한 인공지능
- No : 약한 인공지능

존 설의 중국어 방 사고실험



존 설의 주장



- ▶ 나의 입력과 출력은 중국인과 똑같음에도, 컴퓨터 프로그램처럼 일을 수행한 나는 스토리를 전혀 이해하지 못한다!
- ▶ 마찬가지로, 프로그램에 따라 기호를 조작할 뿐인 컴퓨터가 인간과 똑같은 입력을 받아 출력을 준다 하더라도, 인간처럼 이해를 가질 수 없다! (강한 인공지능 비판)
- ▶ 진정한 마음을 위해서는 기호 조작을 넘어서는 의미와 이해 필요