언어와 인공지능

그리고 의식

최승락

한림대학교
HALLYM UNIVERSITY
인문학부(철학전공)

목차

1. 소개 및 근황

- 요즘 이러고 있습니다…

2. 인공일반지능과 의식의 문제

- 중국어방 논증과 의식의 문제

3. 인공지능 기술의 발달 득인가 실인가?

- 우리는 잃고 있는 것은 아닌가?

01

소개 및 근황

요즘 이러고 있습니다…

1. 소개 및 근황

이　　름: 최승락
소　　속: 한림대학교 인문학부 철학전공
연구분야: 논리학/논리철학/수리철학
연구방법: 분석철학 일반, 계산가능성이론, 증명론

논리술사
INTP-A

가　족: 아내와 8개월 딸이 있어요.
취　미: 덕업일치(德業一致), 힘들게 운동하기
MBTI: INFJ (학생들은 INTP냐고 물어보지만…)
　　ESTP (아내와 성격이 완전 다르지만 잘 지냅니다…)
종　교: 천주교

0. 소개 및 근황

V. 수업 예제 제작 방법

0. 본 자료는 수업 예제 자료 제작 방법을 투명하게 공개하기 위해 작성되었습니다. 본 수업은 이 수업 외부에서 일어나는 학생들의 ChatGPT 사용과 관련된 어떠한 종류의 의견 및 입장에도 관여하지 않습니다.

1. ChatGPT에 다음과 같은 프롬프트(prompt)를 명령어로 실행.[1)]

As a professional philosopher, you have been asked to write an article explaining [Write a subject in here]. Your article should follow the following guidelines:

(a) You should provide a thought provoking title for the article.

(b) You should consult [Write references in here], referring these references #REF.

(c) If you wish to dispute someone's position, you must first accurately represent their thesis and provide proper citation for their arguments, following Chicago reference citation style.

(d) Your article should cite more than three references, which should be relevant to the topic and support your arguments.

(e) Your article should follow the following outline format and make full paragraphs without mentioning each title of sections and subsections:

For the first section, provide a thought provoking title for an introduction section. Then, you should state your claim #CLAIM to be justified explicitly and why #CLAIM is so interesting.

For the second section, provide a thought provoking title for the second section. Then, you should propose an argument #ARG for #CLAIM. You should state premises #PREM of #CLAIM.

For the third section, provide a thought provoking title for the third section. Then, you should explain the validity of #ARG and how #PREM can support #CLAIM.

For the fourth section, provide a thought provoking title for the fourth section. Then, you should explain why #PREM are true.

For the fifth section, provide a thought provoking title for the fifth section, that is the conclusion of the article, you should propose plausible rebuttals #REB of #CLAIM and your reponses against #REB. Then, you should summarize your article.

At last, you should make a list of #REF following Chicago reference citation style.

3. ChatGPT가 작성한 글을 네이버 번역 프로그램 파파고를 통해 번역함.

4. 그 외에 가독성을 방해하거나 읽기 힘든 부분을 개별적으로 수정함.

5. 2024년 봄학기 앎과인간 수업 논평과제 생성 예시는 다음 링크를 참고하세요.

링크: https://chat.openai.com/share/66f75ea7-1b3f-4a6e-a4a9-b51156ed9053

GPT가 생성한 글을 비판하는 과제를 앎과인간 수업에서 실시

1) 프롬프트란, 컴퓨터에게 원하는 결과 값을 얻기 위해 제시하는 명령어로, 생성형 약한 인공지능 AI 프로그램을 사용한 이후 "원하는 결과를 이끌어내기 위해 만든 (최적화된) 명령어"를 의미하는 것으로 사용되곤 합니다.

VI. 보너스 과제(선택): 10점.

– ChatGPT 3.5와의 대담을 통해 모순을 이끌어 내어 보세요.

1. 보너스 과제 발표를 원하시는 분은 ChoiSR@hallym.ac.kr으로 메일을 주시면 수요일 발표하는 시간을 가지도록 하겠습니다.
2. 보너스 과제 발표 예시.
   (1) (철학적) 대화 형식일 것.
   (2) 특정 개념의 의미를 묻거나 비판하는 형식일 것.
   (3) ChatGPT가 "A"라는 대답을 한 경우, "이전에 너는 A라고 대답했는데 지금은 A가 아니라고 대답하는 구나."라는 반박이 가능한 대화일 것.
3. 보너스과제 예시 참고.

GPT와 논쟁 후 모순 이끌어 내기 앎과인간 수업에서 실시

0. 소개 및 근황

논거의 타당성

① 고품질의 효율적인 글 생성이 가능하다?

현재 상용화 되고 있는 자연어 처리 모델은 비문 등 문법 오류를 걸러내지 못함, 또한 GPT는 결국 수동적인 존재이기에 글 생성에 있어서 인간의 반복적인 작업이 꼭 필요함

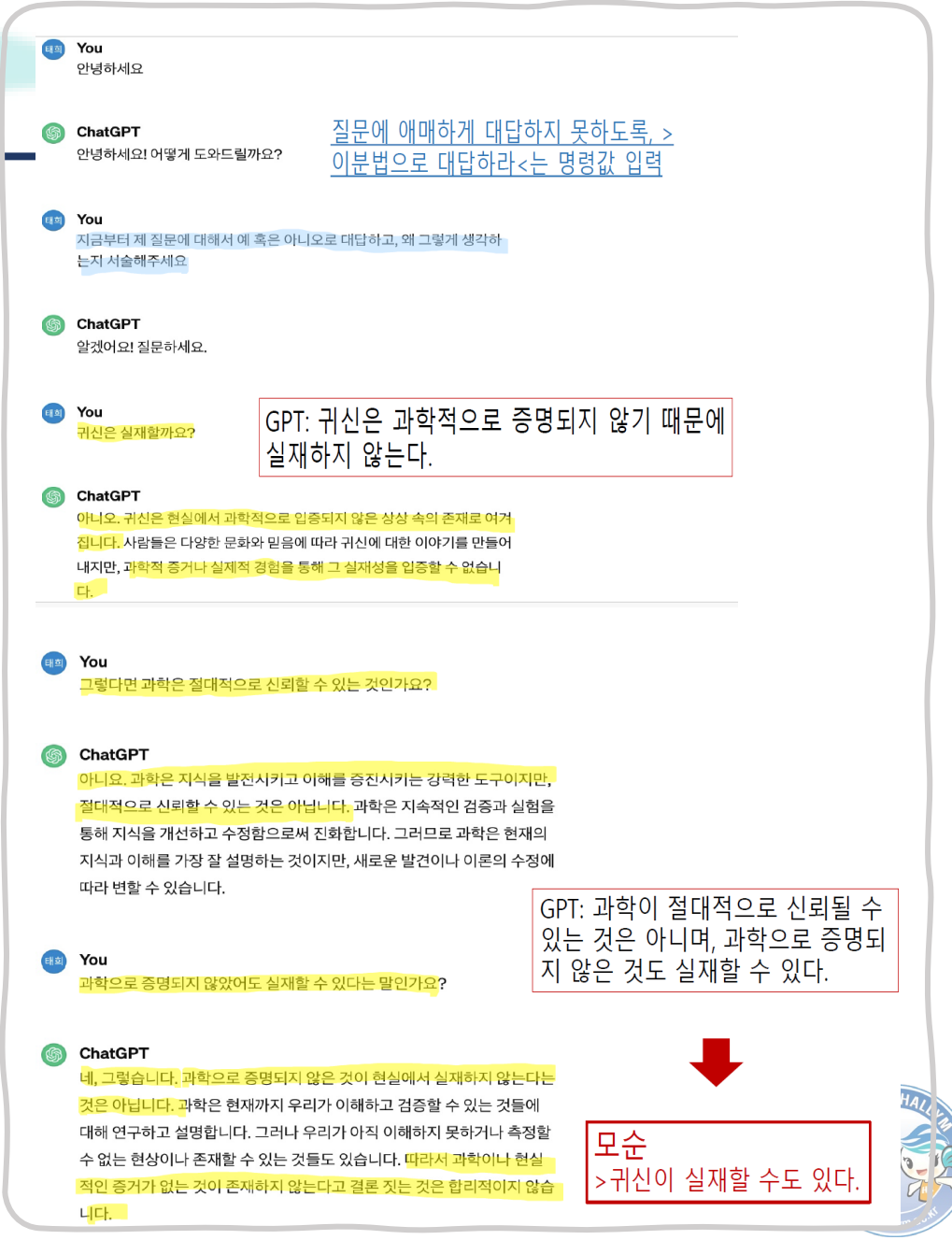
② 선결문제의 오류

논증을 제시할 때 전제를 충분히 논증하지 않은 채 결론을 도출하려고 하고 있음, 따라서 전제가 결론을 정당화하는 데 필요한 충분한 논거 없이 결론을 도출하는 선결문제의 오류를 범하고 있음

학생 결과물. GPT가 생성한 글을 비판

학생 결과물. GPT와 논쟁을 통해 모순 이끌기

학우들이 생각 보다 잘 해서 감동 먹고 있습니다. ☺



0. 소개 및 근황

(1) 작성 및 투고 중인 논문

교비 연구

"Problems of the Proof-Theoretic Criterion for Paradoxicality" – *Synthese*, R&R 심사 중

"Is the liar paradox strictly classical?" – 〈논리연구〉, 심사 중

"Formalizing Negation in Intuitionistic Natural Deduction"(가제) 〈논리연구〉 투고 예정 – 작성 중

"There is no clear distinction between classical, intuitionistic, and minimal inference in proof theory" (가제)

– *Analysis* 투고 예정 – 작성 중

(2) 구상 중인 논문

"Symbolic and Connectionist AI on the Chinese Room Argument"

- 기호주의 AI와 연결주의 AI는 모두 원초 재귀 관계를 기반으로 한 체계라는 측면에서 존 썰의 입장을 평가하는데 큰 차이가 없음을 주장할 예정.

"The Role of Consciousness in Language Understanding"

- '언어의 이해'에 대한 전통적인 정의를 따른다면, 의식은 언어의 이해에 필수 조건이다. 최근 AGI의 성취 문제는 강한AI 성취 여부와 구별될 필요가 있으며 강한 AI의 성취 여부는 의식의 문제가 핵심임을 주장할 예정.

02

인공일반지능과 의식의 문제

중국어방 논증과 의식의 문제

2.1. 이세돌과 알파고

알파고 꺾으며 인류최초의 승리, 이세돌 은퇴

머니투데이 | 김도엽 인턴

VIEW 9,171 | 2019.11.20 11:10

0

1995년 7월 프로 생활 시작한 후 24년 4개월만에 전문 기사직 사퇴

지난 2016년 3월13일 이세돌 9단이 13일 '구글 딥마인드 챌린지 매치' 인공지능(AI) 바둑 프로그램 알파고(AlphaGo)와의 5번기 제4국에서 승리한 뒤 옅은 미소를 띄며 대국장을 나서고 있다./사진=뉴

Kellin Pelrine delivered a decisive defeat to top-ranked AI system for playing the board game Go © Kellin Pelrine

2023년 2월. 최고 수준 바둑 프로그램 카타고에게 15전 14승한 아마추어 바둑기사 켈린 펠라인(Kellin Pelrine)

기사 출처: https://www.ft.com/content/175e5314-a7f7-4741-a786-273219f433a1

2.2. 생성형 인공지능과 인공일반지능

AI 테크

오픈AI, GPT-4 후속 모델 훈련 돌입…“AGI 향한 한 걸음”

이상덕 기자 asiris27@mk.co.kr

입력 : 2024-05-28 19:44:11

가

차세대 인공지능 모델 개발 예고
새로운 안전 및 보안 위원회 구성

AI 타임스

AI&Big Data Show 2024 | 2024 인공지능&빅데이터

19 - 21 June I Coex

2024년 05월 29일 (수)

전체기사 뉴스 브리핑 Feature 인터뷰 Future Energy Art & Living Global

홈 > 산업 > 산업일반

가 가

머스크 "세계 최대 데이터센터 구축으로 AGI 본격 경쟁"

박찬 기자 입력 2024.05.26 10:37 수정 2024.05.27 14:03 댓글 0 좋아요 4

〈2024년 5월 28일 기사〉

8뉴스 SBS 단독보도 경제

[단독] 앤드류 응 "AGI 수십 년 걸릴 것..

홍영재 기자 작성 2024.05.24 21:05 수정 2024.05.24 22:12 조회 1,192

〈2024년 5월 24일 기사〉

최근의 화두는 “인공일반지능”을 성취하느냐?

2.2. 생성형 인공지능과 인공일반지능

존 썰
(John Searle)

인간의 인지 능력을 컴퓨터로 시뮬레이션하려는 최근의 노력에 어떤 심리적, 철학적 의미를 부여해야 할까? 이 질문에 답할 때, 나는 “강한” 인공지능과 “약한” 또는 “신중한” AI(인공 지능)를 구분하는 것이 유용하다고 생각한다. 약한 AI에 따르면, 마음 연구에서 컴퓨터의 주요 가치는 그것이 우리에게 매우 강력한 도구를 제공한다는 것이다. 예를 들어, 컴퓨터는 보다 엄격하고 정밀한 방식으로 [마음에 관한] 가설을 수립하고 테스트할 수 있게 해준다. 그러나 강한 AI에 따르면, 컴퓨터는 마음 연구에 사용되는 단순한 도구가 아니다. 오히려, 적절하게 프로그램된 컴퓨터는 말 그대로 마음이다. 올바르게 프로그램된 그 컴퓨터는 말 그대로 [인간과 같은 수준으로] 이해하고 다른 인지 상태를 지닌다. 이러한 점에서 강한 AI인 프로그램은 심리적 설명을 테스트할 수 있게 해주는 단순한 도구가 아니라 그 자체가 설명의 대상이다.

Minds, brains, and programs, *The Behavioral and Brain Sciences*, 1980, p. 417.

존 썰은 유명한 중국어방 논변을 제시하는 논문에서 최초로 ‘강한 인공지능’ 개념을 소개한다. 그리고 용어의 의미가 다소 차이가 있으나 최근 ‘인공일반지능’(artificial general intelligence, AGI)이란 명칭이 ‘강한 인공지능’ 대용으로 사용되고 있다.

2.2. 생성형 인공지능과 인공일반지능

1. 전통적 컴퓨터 프로그램: 명확한 규칙과 절차를 코드로 작성하여 작업 수행.

예. 라면 끓이기 프로그램: 컴퓨터에게 다음 과정을 코드로 작성해 프로그램을 만든다.

1단계. 물 500ml를 냄비에 붓고 끓인다.
2단계. 물이 끓으면 라면과 스프를 넣는다.
3단계. 4분 동안 끓인다.
4단계. 라면이 다 끓으면 불을 끄고, 그릇에 담는다.

주동률 교수님 비법 레시피 (1인분)

1단계. 물 500ml과 진라면 순한맛 스프 그리고 무와 마늘 한 개 반을 넣고 끓인다.
2단계. 물이 끓으면 면을 넣고 4분간 끓인다.
3단계. 면을 넣고 2분 30초가 지났을 때 계란과 파를 넣는다.

실패가 없는 맛 :b

Midjourney가 그린 규칙을 따라 라면을 끓이는 존 썰(John Searle)

장점 1. 목표를 위한 효율적 프로그래밍이 가능함.
(예. 효율적으로 라면 끓이기 학습이 가능)
장점 2. 프로그램의 안정성이 보장됨.
(예. 라면을 끓여 달라고 하면, 맛이 예측 가능함)
단점 1. 다른 목표를 위해서는 새로운 프로그래밍이 필요함.
(예. 가르쳐준 라면 밖에 못 끓임. 맨날 같은 거 먹어야 함.)
단점 2. 다른 종류의 프로그래밍을 하려면 그 목표를 위한 명확한 규칙을 프로그래머가 직접 파악해야 함.
(예. 다른 종류의 라면을 끓이도록 가르치려면 명확한 규칙을 알고 있어야 함.)

2.2. 생성형 인공지능과 인공일반지능

2. 기계학습(machine learning) 기반 프로그램:

방대한 자료를 보여주고 특정 목적을 수행하기 위한 패턴을 찾는 작업을 수행.

예. 자기주도 학습(self-supervised learning)으로 만든 라면 끓이기 프로그램

사례 1. 물의 양: 500ml, 재료: 라면, 스프, 끓이는 시간: 4분
사례 2. 물의 양: 450ml, 재료: 라면, 스프, 채소, 끓이는 시간: 5분
사례 3. 물의 양: 550ml, 재료: 라면, 매운 스프, 고추, 끓이는 시간: 3분
...
사례 n. 물의 양: 500ml, 재료: 라면, 스프, 계란, 끓이는 시간: 5분
사례 n+1. 물의 양: 400ml, 재료: 라면, 스프, 치즈, 끓이는 시간: 4분 + 30초

라면이 무엇인지, 잘 끓인 라면이 무엇인지 등을 전혀 설명하지 않고 많은 예시를 통해 라면 끓이는 법에 대한 패턴을 익히도록 반복함.

Midjourney가 그린 라면 끓이기를 무수히 반복하는 존 썰

2.2. 생성형 인공지능과 인공일반지능

2. 기계학습(machine learning) 기반 프로그램:

방대한 자료를 보여주고 특정 목적을 수행하기 위한 패턴을 찾는 작업을 수행.

예. 자기주도 학습(self-supervised learning)으로 만든 라면 끓이기 프로그램

사례 1. 물의 양: 500ml, 재료: 라면, 스프, 끓이는 시간: 4분
사례 2. 물의 양: 450ml, 재료: 라면, 스프, 채소, 끓이는 시간: 5분
사례 3. 물의 양: 550ml, 재료: 라면, 매운 스프, 고추, 끓이는 시간: 3분
...
사례 n. 물의 양: 500ml, 재료: 라면, 스프, 계란, 끓이는 시간: 5분
사례 n+1. 물의 양: 400ml, 재료: 라면, 스프, 치즈, 끓이는 시간: 4분 + 30초

라면이 무엇인지, 잘 끓인 라면이 무엇인지 등을 전혀 설명하지 않고 많은 예시를 통해 라면 끓이는 법에 대한 패턴을 익히도록 반복함.

장점 1. 여러 가지 목적을 위한 프로그램을 얻을 수도 있음.
(예. 한 가지 이외에 다양한 종류의 라면을 끓일 수 있음. 심지어 맛있기도 함.)
장점 2. 프로그램이 예상하지 못 했던 기능을 발휘하기도 함.
(예. 갑자기 라면 이외의 다른 음식도 요리해 줌.)

단점 1. 단순한 프로그램을 만드는데도 지나치게 많은 데이터나 전기 등이 필요함.
(예. 라면 끓이는 데이터가 많지 않으면 가르칠 수 없음. 효율적이지 않음.)
단점 2. 원하는 결과를 얻지 못하는 경우도 많음.
(예. 라면을 끓여 달라고 했는데 아주 이상한 음식이 나올 수 있음.)

이상한데 맛있다…

2.2. 생성형 인공지능과 인공일반지능

1. 전통적 컴퓨터 프로그램

장점 1. 효율적으로 라면 끓이기 학습이 가능
장점 2. 라면을 끓여 달라고 하면, 맛이 예측가능함

단점 1. 가르쳐준 라면 밖에 못 끓임. 맨날 같은 거 먹어야 함.
단점 2. 다른 종류의 라면을 끓이도록 가르치려면 명확한 규칙을 알고 있어야 함.

실패 없는 맛 :b

2. 기계학습기반 프로그램

장점 1. 한 가지 이외에 다양한 종류의 라면을 끓일 수 있음. 심지어 맛있기도 함.
장점 2. 갑자기 라면 이외의 다른 음식도 요리해 줌.

단점 1. 라면 끓이는 데이터가 많지 않으면 가르칠 수 없음. 효율적이지 않음.
단점 2. 라면을 끓여 달라고 했는데 아주 이상한 음식이 나올 수 있음.

이상한데 맛있음 :0

2.2. 생성형 인공지능과 인공일반지능

존 썰
(John Searle)

2022년 경부터 주어진 범주 내에서의 예측이나 결정을 내리는 프로그램이 아닌 새로운 결과를 제시하는 “생성형 인공지능 프로그램”이 대세를 이뤄가고 있다. ‘인공지능 프로그램’이라고 이 발표에서 사용하는 것은 기계학습 기반 ‘생성형 인공지능’이라 불리는 프로그램을 지칭할 것이다. 예. ChatGPT, Midjourney 등.

약한/강한 AI에 대한 통상적인 이해

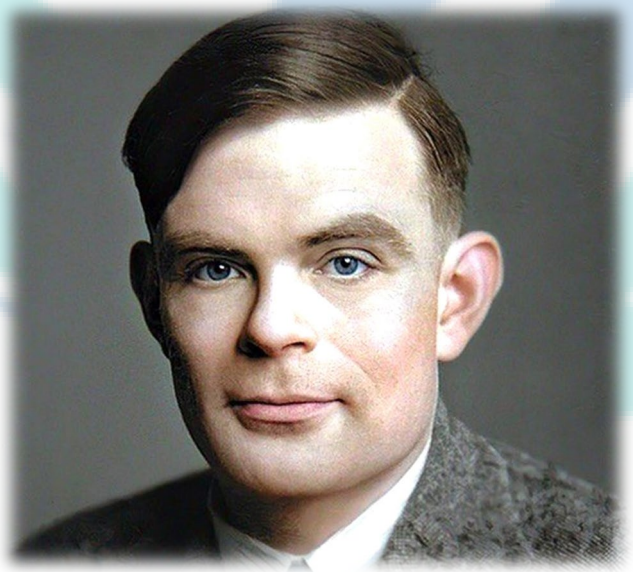
(1) 약한 인공지능(Weak/Narrow AI)

- 특정 문제 해결이나 특수한 목적을 위해 만들어진 기계학습 기반 프로그램.

(2) 강한 인공지능(Strong AI)

- 인간과 같은 보편적인 인지 능력을 갖춘 AI 프로그램을 일컫는다.
- 존 썰은 강한 인공지능을 의식과 자각(self-awareness) 능력도 지니고 있는 것으로 고려하며 이는 아직은 실현되지 않은 이론적 개념이다.

2.2. 생성형 인공지능과 인공일반지능



앨런 튜링
(Alan Turing)

'범용 컴퓨터'(General purpose computer)에 대한 이론적 기반을 최초로 제시한 논리학자이다. 2차 세계대전 당시 독일의 에니그마 암호(the Enigma code)를 해독하는 업적을 세웠다.

Turing(1936), "On Computable Numbers, with an Application to the Entscheidungsproblem".

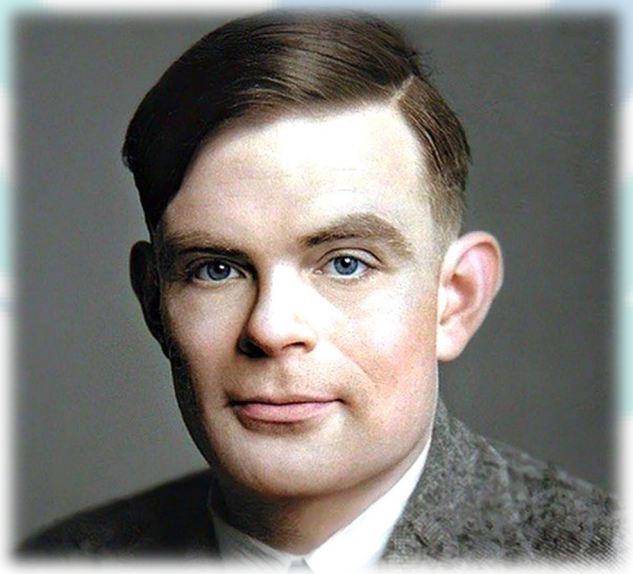
또한 튜링은 20세기 말에 인간 지능과 동일한 수준의 컴퓨터가 개발될 수 있을 것이라고 예견하기도 했는데 이때 컴퓨터의 수준을 판단할 수 있는 시험으로 제시한 것이 '튜링 테스트' 혹은 '모방 게임'(imitation game)이다.

Turing(1950), "Computing Machinery and Intelligence"

튜링 테스트: 평가자가 인간과 기계와 대화를 나눌 때, 평가자가 누가 인간이고 누가 기계인지를 파악하지 못 한다면 그 기계는 인간 수준의 지성을 지니는 것이라고 판단하는 시험.

썰은 중국어방 논변을 통해, 기계가 튜링 테스트를 통과하더라도 인간 수준의 지성을 지닐 수 없다고 주장한다.

2.2. 생성형 인공지능과 인공일반지능



앨런 튜링(Alan Turing)

‘범용 컴퓨터‘(General purpose computer)에 대한 이론적 기반을 최초로 제시한 논리학자이다. 2차 세계대전 당시 독일의 에니그마 암호(the Enigma code)를 해독하는 업적을 세웠다.

Turing(1936), “On Computable Numbers, with an Application to the Entscheidungsproblem”.

범용 컴퓨터는 특정 목적을 위해 사용하는 컴퓨터가 아닌 다양한 목적을 위해 사용되는 컴퓨터로 이후, 인공지능의 아버지라 불리는 존 맥카시(John McCarthy)와 마빈 민스키(Marvin Minsky)가 ‘인공지능’(artificial intelligence) 컴퓨터 개념을 생각하는 데에도 큰 영향을 미쳤다. ‘인공지능’이라는 개념은 맥카시와 민스키 등이 다트머스 학회를 열며 최초로 제안한 개념으로 알려져 있다.

McCarthy and Minsky et all (1955), “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”

존 맥카시(John McCarthy)

마빈 민스키(Marvin Minsky)

2.3. 중국어방 논변과 언어의 이해

레이먼드 커즈와일
(Raymond Kurzweil)

1990년부터 미래학자이자 컴퓨터과학 분야의 전문가인 레이 커즈와일(Raymond Kurzweil)은 컴퓨터과학 기술이 기하급수적으로 발전할 것임을 주장하며, 2005년에는 그의 책 〈특이점이 온다〉(Singularity is near)을 통해, 유명한 특이점 이론(singularity theory)을 주장한다.

2029년은 인공지능이 유효한 튜링 테스트를 통과하여 인간 수준의 지능을 달성하는 시점으로 제가 일관되게 예측한 날짜입니다. 저는 2045년을 '특이점'으로 설정했는데, 이는 우리가 만든 지능과 합쳐져 유효 지능이 10억 배로 늘어나는 시점입니다.

Futurism과의 인터뷰 중
(https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045)

커즈와일은 14세 때, MIT의 민스키에게 편지를 보내 대학에 초청 받은 적이 있었고 그 이후에도 민스키를 멘토로 생각하며 살아왔다. 2012년 커즈와일이 "기계학습과 언어" 프로젝트를 구글에서 전임직으로 진행하면서 '강한 AI'와 '약한 AI' 등의 철학 개념을 자주 사용했었다.

마빈 민스키(Marvin Minsky)

'강한 인공지능'과 '인공일반지능' 개념의 사용

'인공일반지능' 개념은 초기 컴퓨터 과학자와 철학자들의 전통에서 사용했던 용어는 아니었다.

2.3. 중국어방 논변과 언어의 이해

벤 괴르첼
(Ben Goertzel)

‘강한 인공지능’과 ‘인공일반지능’ 개념의 사용

컴퓨터 과학자 벤 괴르첼(Ben Goertzel)과 카시오 페나친(Cassio Pennachin)은 인간 수준의 보편적 AI에 대한 컴퓨터 기술의 접근 방식에 대한 책을 2005년도에 편찬하는데 이때 인간의 IQ와 같은 보편적 지능을 의미하는 ‘인공일반지능’(artificial general intelligence, AGI)라는 개념을 사용한다. 괴르첼은 이 개념은 컴퓨터 과학 및 공학적 입장에 집중하며 존 썰의 철학적 입장에 의존하지 않는다는 점에서 장점이 있다고 여긴다.

참고. 괴르첼의 블로그: https://web.archive.org/web/20181228083048/http://goertzel.org/who-coined-the-term-agi/

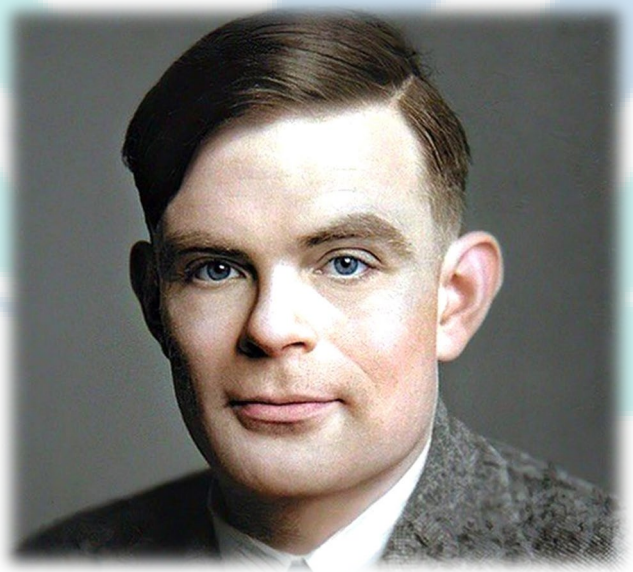
카시오 페나친
(Cassio Pennachin)

‘강한 인공지능’과 ‘인공일반지능’은 그 기원에서 각자가 목표로 하는 바가 다를 수 있다. ‘강한 인공지능’은 썰의 철학적 입장을 다룰 때, 더욱 적합하며 2024년 현재의 인공지능 기술 발전을 언급함에 있어서는 ‘인공일반지능’을 사용하는 것이 더 옳을 수 있다. 하지만 두 용어는 현재 사회적/학술적으로 혼용되어 사용되고 있기에 이 글에서는 같은 뜻으로 사용하기로 하며 이와 관련된 논의는 추후를 기약하기로 한다.

존 썰
(John Searle)

레이먼드 커즈와일
(Raymond Kurzweil)

2.3. 중국어방 논변과 언어의 이해



앨런 튜링
(Alan Turing)

'범용 컴퓨터'(General purpose computer)에 대한 이론적 기반을 최초로 제시한 논리학자이다. 2차 세계대전 당시 독일의 에니그마 암호(the Enigma code)를 해독하는 업적을 세웠다.

Turing(1936), "On Computable Numbers, with an Application to the Entscheidungsproblem".

또한 튜링은 20세기 말에 인간 지능과 동일한 수준의 컴퓨터가 개발될 수 있을 것이라고 예견하기도 했는데 이때 컴퓨터의 수준을 판단할 수 있는 시험으로 제시한 것이 '튜링 테스트' 혹은 '모방 게임'(imitation game)이다.

Turing(1950), "Computing Machinery and Intelligence"

튜링 테스트: 평가자가 인간과 기계와 대화를 나눌 때, 평가자가 누가 인간이고 누가 기계인지를 파악하지 못 한다면 그 기계는 인간 수준의 지성을 지니는 것이라고 판단하는 시험.

썰은 중국어방 논변을 통해, 기계가 튜링 테스트를 통과하더라도 인간 수준의 지성을 지닐 수 없다고 주장한다.

2.3. 중국어방 논변과 언어의 이해

튜링 테스트(모방 게임)과 중국어방 논증

튜링 테스트: 평가자가 인간과 기계와 대화를 나눌 때, 평가자가 누가 인간이고 누가 기계인지를 파악하지 못 한다면 그 기계는 인간 수준의 지성을 지니는 것이라고 판단하는 시험.

중국어방 논증: 튜링 테스트는 강한 AI를 평가하는데 적합하지 않고 강한 AI도 가능하지 않다.

(1) 중국어를 전혀 모르는 썰이 중국어 대답을 하는 규칙이 적힌 책을 들고 좁은 방안에 갇혀 있다.

(2) 방 밖의 중국인이 질문 메모를 주면 썰은 책 속 기계적 규칙을 찾아 질문에 대답하고 이 메모를 방 밖 중국인에게 전달한다.

(3) 방 밖의 평가자는 중국어를 전혀 모르는 썰에게 메모를 받았지만 그가 중국어를 제대로 이해하고 있다고 여긴다.

이해했네~

2.3. 중국어방 논변과 언어의 이해

중국어방 논증에 대한 비판: 당신이 중국어를 이해 못하더라도 중국어 방 전체라는 체계는 이해라는 작용을 하는 것이 아니냐?

반박: 썰은 중국어를 이해하지 못 하더라도 썰과 중국어 규칙 책을 합친 것은 중국어를 이해하는 어떤 작용을 하고 있는 것이 아니냐? 썰과 중국어 규칙 책을 하나의 체계로 본다면 바로 그 체계는 중국어에 관한 이해를 구현하는 것이라고 볼 수 있는 것 아니냐? 참고. System Reply

썰의 대답: 중국어 대답을 하는 기계적 방식을 내 머리 속에 내재화 한다고 하더라도 그러한 기계적 조작 방식을 따르는 것과 용어의 의미를 아는 것은 서로 다르다. 컴퓨터는 결국 기호체계의 기계적 방식을 구현하는 구문론적(syntactic) 체계일 뿐이고 이는 의미론(semantics)을 구현할 수 없다.

참고. 구문론(syntax)과 의미론(semantics)의 물음

구문론적 물음: 기호를 어떻게 체계적으로 나열할 것인가? 예. 한글 글자 배열

의미론적 물음: 나열된 기호에 의미를 어떻게 부여할 것인가? 예. 정의, 형식 의미론(formal semantics)

2.3. 중국어방 논변과 언어의 이해

중국어방 논증에 대한 비판: 당신이 중국어를 이해 못하더라도 중국어 방 전체라는 체계는 이해라는 작용을 하는 것이 아니냐?

반박: 썰이 휴대전화를 들고 구글이나 딥플 번역기를 사용하는 것은 어떤가? 여기서 더 나아가 ChatGPT를 발전시켜 외부 자극을 보고 듣고 느끼고 냄새 맡는 등의 로봇을 만든다면 그 로봇은 강한 AI(혹은 AGI)라고 할 수 있지 않은 가?

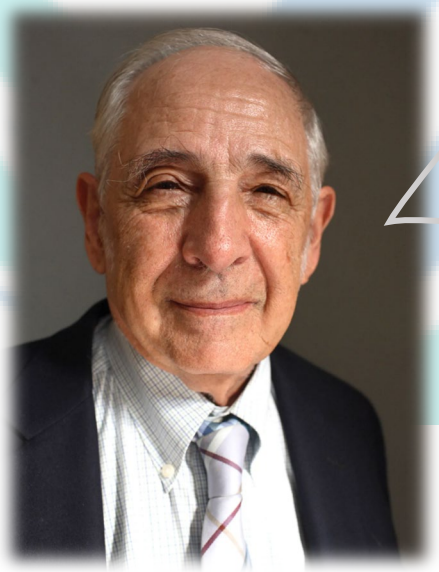
참고. Robot Reply

썰의 대답: 그 로봇이 보고, 듣고, 느낀다고 여기는 정보 역시 0과 1로 구성된 기계적 과정에 의해 처리된다. 이는 로봇의 체계가 구문론적 체계일 뿐임을 말하므로 그 로봇이 언어를 이해한다고 볼 수 없다. 구문론은 의미론을 함축하지 않기 때문이다.

형식적인 기호 조작은 그 자체로는 어떠한 지향성(intentionality)도 지니지 않는다. 그것들은 의미가 없을 뿐이다. 또한 기호는 아무것도 상징하지 않기 때문에 그것들은 심지어 *기호* 조작도 아니다. 언어학적 용어로 표현하자면, 구문론만 있을 뿐 의미론이 없다. 컴퓨터가 지니고 있는 것처럼 보이는 지향성은 전적으로 그 컴퓨터를 프로그래밍하는 사람과 사용하는 사람, 입력을 보내는 사람과 출력을 해석하는 사람의 마음 속에만 존재한다.

(Minds, brains, and programs, *The Behavioral and Brain Sciences*, 1980, p. 422)

2.3. 중국어방 논변과 언어의 이해



존 썰
(John Searle)

형식적인 기호 조작은 그 자체로는 어떠한 지향성(intentionality)도 지니지 않는다. 그것들은 의미가 없을 뿐이다. 또한 기호는 아무것도 상징하지 않기 때문에 그것들은 심지어 *기호* 조작도 아니다. 언어학적 용어로 표현하자면, 구문론만 있을 뿐 의미론이 없다. 컴퓨터가 지니고 있는 것처럼 보이는 지향성은 전적으로 그 컴퓨터를 프로그래밍하는 사람과 사용하는 사람, 입력을 보내는 사람과 출력을 해석하는 사람의 마음 속에만 존재한다.

(Minds, brains, and programs, *The Behavioral and Brain Sciences*, 1980, p. 422)

지향성은 정의상 세계의 대상과 사태로 향하는 특정 심적 상태이다. 그러므로 믿음, 욕구 그리고 지향성은 내재적 상태이다.

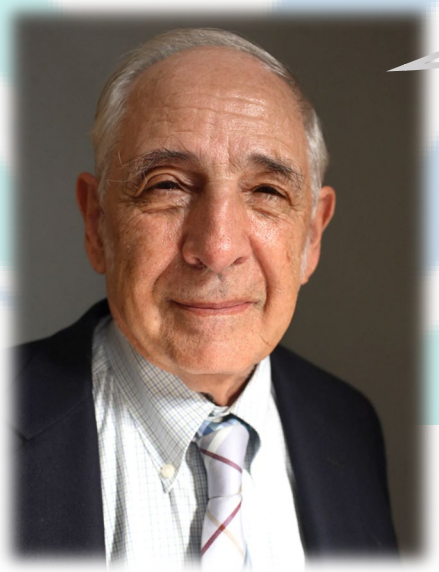
(Minds, brains, and programs, *The Behavioral and Brain Sciences*, 1980, footnote 3.)

※ 참고. 어떤 대상이 의식이 있다는 것은 그 대상이 보고 듣고 생각하는 것과 같은 주관적 경험을 한다는 것이다.

나는 수년 전 컴퓨터 프로그램의 구현이 그 자체로는 의식이나 지향성에 충분하지 않다는 것을 이른바 중국 방 논증을 통해 설명한 바 있다.(Searle 1980). 계산은 순전히 형식적 혹은 구문적으로 정의되는 반면, 마음은 실제 정신적 또는 의미론적 내용을 가지고 있기 때문에 구문론적 연산만으로는 구문론에서 의미론으로 넘어갈 수 없다. 조금 더 기술적으로 설명해서, 나는 '동일하게 구현된 프로그램'이라는 말을 특정 물리적 실현과 독립적으로 규정된 동치 집합으로 정의한다. 하지만 이렇게 규정된 것은 인지 과정을 유발하는 뇌의 생물학적 능력을 배제할 수 밖에 없다. 예를 들어, 중국어 화자의 행동을 시뮬레이션한 컴퓨터 프로그램의 단계를 거치는 것만으로는 중국어에 대한 이해를 습득할 수 없다.

Searle, J. "Why Dualism (and Materialism) Fail to Account for Consciousness", p. 17.

2.3. 중국어방 논변과 언어의 이해



존 썰
(John Searle)

의식 작용 혹은 지향성 없이는 언어를 이해할 수 없다.

최근의 사회적 관심. ChatGPT는 언어를 이해하지 못 하는 것인가?

구상 중인 논문

물음 1. 썰의 중국어방 논변은 기계학습 기반 프로그램에도 적용되는가?

"Symbolic and Connectionist AI on the Chinese Room Argument"

- 기호주의 AI와 연결주의 AI는 모두 원초 재귀 관계를 기반으로 한 체계라는 측면에서 존 썰의 입장을 평가하는데 큰 차이가 없음을 주장할 예정.

물음 2. 언어의 이해에 의식 작용은 필수적인가?

"The Role of Consciousness in Language Understanding"

- '언어의 이해'에 대한 전통적인 정의를 따른다면, 의식은 언어의 이해에 필수 조건이다. 최근 AGI의 성취 문제는 강한AI 성취 여부와 구별될 필요가 있으며 강한 AI의 성취 여부는 의식의 문제가 핵심임을 주장할 예정.

03

득(得)인가? 실(失)인가?

우리는 잃고 있는 것은 아닌가?

3. 우리는 잃고 있는 것은 아닌가?

Midjourney생성: 합리성을 파는 샘 알트만

한 달에 20달러면 “야 너두 합리적인 인간이 될 수 있어!”

3. 우리는 잃고 있는 것은 아닌가?

2023년 4월. 영어 교정 서비스를 받아봤으나 논문 게재 반려의 핵심 문제로 지적 받음.

이후에 수정 후 재투고 논문을 요청했을 때도, 에디티지 교정자는 동일한 실수를 반복했고 형식적인 정의를 매우 애매한 용어로 수정해 두어 제가 직접 그런 부분을 찾아서 수정해야하는 문제가 있었습니다. 그리고 얼마전 논문을 투고했던 JPL 측으로 부터 다시 수정 후 재투고 결정을 받았고 Reviewer는 영어 교정 문제를 심각하게 제기하며 다음과 같은 조언을 제게 주었습니다.

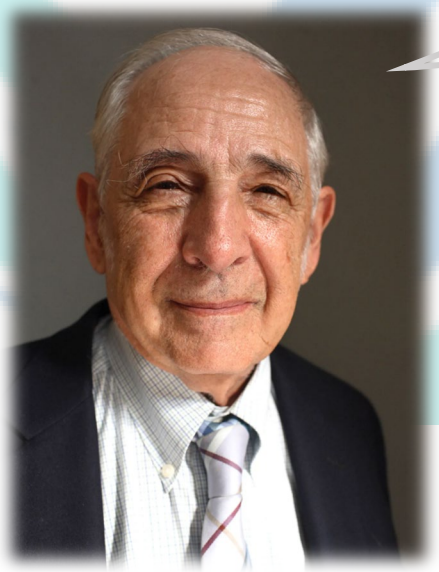
Finally, unfortunately, this draft is much less clear than the first, with certain places where the authors meaning is obscure. I will express the possibility that having the paper proofread by someone who does not understand the material might be the cause of this change. As a native English speaker who has had paper proofread, I can report the need to be very vigilant about the proof reader "correcting" sentences into nonsense.

이후 친구의 권유로 GPT를 알게 되어 사용하게 됨.

사용했던 GPT 프롬프트

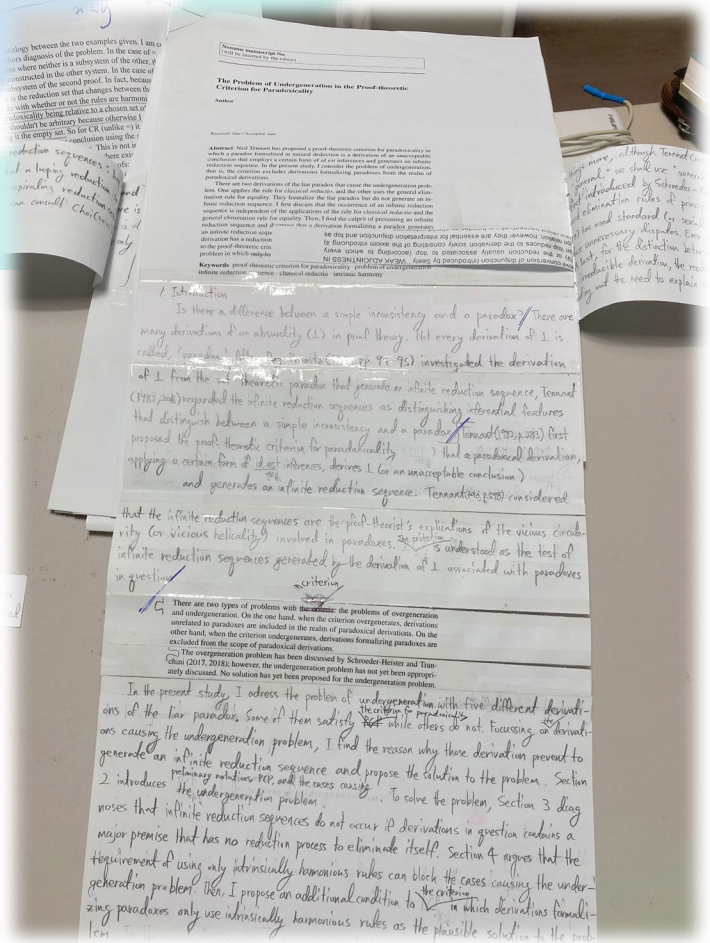
Please create three paraphrased versions of the text provided below. These versions should maintain the style of a professional philosopher, with flawless grammar in American English. Ensure compliance with these specific requirements:
(1) Retain all LaTeX codes present in the text.
(2) Preserve the following technical terms without alteration:
(3) Avoid the use of first-person pronouns, such as 'we' and 'I'.
(4) Keep expressions within quotations unchanged. If any grammatical errors are detected within these quotations, bring them to my attention for revision.
Here is the paragraph for paraphrasing:

3. 우리는 잃고 있는 것은 아닌가?



존 썰
(John Searle)

의식 작용 혹은 **지향성** 없이는 언어를 이해할 수 없다.



손글로 논문을 쓰다가 GPT를 돌려 글을 쓰는 지금…

내 글은 나의 욕구, 직관, 느낌을 잘 반영하고 있는가?

내 글에는 나의 의식 작용, 지향성이 잘 담겨 있는가?

나는 글을 쓰고 있는가? 찍어내고 있는가?

3. 우리는 잃고 있는 것은 아닌가?

GPT를 사용해 글을 쓰는 나는 글을 "쓰는" 것인가? "찍어내는" 것인가?

처음 영어를… 글을… 배웠을 때의 그건 아니었어요…

우리는 "편리"라는 이유로 우리 인간의 본질적인 부분을 잃어가는 것은 아닐까?

감사합니다.