AI 판정을 위한 테스트 및 논제의 사전 연구

차병래[1,3], 차윤석[1], 전은진[1], 안성열[1], 박진영[2], 박선[3], 김종원[3]
제노테크(주)[1], 공감 커뮤니케이션 연구소[2], 광주과학기술원 AI대학원[3]
e-mail : brcha@smartx.kr, dxcha@naver.com, wg7660@nate.com,
syan97@daum.net, jygood@naver.com, sunpark@gist.ac.kr, jongwon@gist.ac.kr

Preliminary Study of Tests and Topics for AI Judgment

ByungRae Cha[1,3], YoonSeok Cha[1], EunJin Jeon[1], SeongYeol An[1],
JinYoung Park[2], Sun Park[3], and JongWon Kim[3]
GenoTech Inc.[1], Institute for Empathy and Communication[2],
GIST AI Graduate School[3]

요 약

AI(Artificial Intelligence)는 과학소설의 영역에서 일상생활 속의 현실로 빠르게 전환되고 있으며, 일상생활의 응용사례들은 AI가 우리의 삶을 형성하는데 중요한 요소가 되었다는 증거이다. AI 기술에 대한 관심 증대와 다양한 산업 및 국방에 활용을 모색하고 있으며, 인간의 인식, 판단 및 의사결정 능력을 보완해 주는 정보기술로 무인화, 자율화, 지능화 구현의 중요한 기술이다. AI 기술의 활용 이전에 검증 및 평가를 위한 다양한 테스트와 AI 기술의 관련 딜레마에 대한 사전 연구를 수행하였다.

1. 서 론

AI는 과학소설의 영역에서 일상생활 속의 현실로 빠르게 전환되고 있으며, 일상생활의 모든 분야에서 응용사례가 늘어나고 있다. 외국어 번역, 시각 인식 알고리즘, 자율주행, 암 진단 등의 이런 사례들은 인공지능이 우리의 삶을 형성하는데 중요한 요소가 되었다는 증거이다.

최근 AI 기술에 대한 관심과 더불어 다양한 산업 영역에서 활용하고자 한다. AI란 자신의 목적을 달성하기 위해 그에 맞는 일들을 할 수 있는 지능적 존재를 의미하며, 현실에서 AI가 자신의 목적에 맞는 일, 제대로 된 일을 하기 위해서는 지각, 시각 및 음성인식 그리고 행동이 필요하다. 또한 AI는 우리가 볼 수 없는 것을 보여줄 수 있는 능력을 가지고 있다. AI 기술은 인간의 인식, 판단 및 의사결정 능력을 보완해 주는 정보기술로 무인화, 자율화, 지능화 구현의 중요한 기술이다.

본 연구에서는 AI 분야와 관련된 기회와 위험성을 조명하기 위하여 AI의 특수성과 AI 판정을 위한 다양한 테스트에 대해 사전 연구를 수행하였다.

2. 관련 연구

2.1 머신러닝(Machine Learning)

빅데이터에 대한 접근은 그 자체로는 별로 신통하지 않은 많은 정보를 활용해 최소한 통계적 패턴을 알아내는 것이며, 통계적 패턴은 커다란 무리의 사람들에게만 적용될 뿐 개개인의 행동에 꼭 적용되는 것은 아니다. 이런 접근을 좀 더 급진적으로 활용하는 것이 머신러닝이다. 머신러닝은 과거의 데이터에서 패턴을 찾아 새로운 데이터에 대해 결정을 내린다. 따라서 찾아낸 상관관계로부터 직접적으로 규칙을 이끌어내 예측에 활용한다.

머신러닝 알고리즘을 하나하나 적용할 때마다 사회의 모습은 변하며, 머신러닝은 과학과 기술, 비즈니스, 정치, 전쟁을 바꾸고, 새로운 과학 지식을 내놓는다. 또한, 페드로 도밍고스(Pedro Domingos)는 머신러닝의 유파를 기호주의자(symbolists), 연결주의자(connectionists), 진화주의자(evolutionaries), 베이즈주의자(Bayesians), 그리고 유추주의자(analogizers)로 분류하고 있으며, 이들을 조합하면 어떤 문제든 해결할 수 있는 '지배적 알고리즘'을 만들어 낼 수 있다고 말한다[1].

2.2 모라벡의 역설(Moravec's Paradox)

1970년대 미국 로봇공학 전문가 한스 모라벡이 한 말에서 유래되었으며, 쉽게 말해 "인간에게 쉬운 일은 컴퓨터에게 어려우며, 컴퓨터에게 쉬운 일은 인간에게 어렵다"라는 역설이다[2].

2.3 약 AI와 강 AI 그리고 AI-complete

AI는 크게 AGI(Artificial General Intelligence)와 Narrow AI로 구분한다. AGI는 인간이 할 수 있는 모든 지적 작업을 이해하거나 학습하는 지능형 에이전트의 가상 능력을 의미하며, 강 AI(Strong AI) 또는 Full AI라 부른다. Narrow AI는 약 AI(weak AI)라고 하며, 사전 학습

된 특정 문제 해결 또는 추론 작업(전문가 시스템)을 연구하거나 달성하기 위해 소프트웨어를 사용하는 것으로 제한된다[3].

약 AI는 이미지 인식 및 음성 인식 영역에서 눈부실 발전을 이루었으며, 체스 및 바둑 등 각종 게임 영역에서 인간은 약 AI에게 빈번히 패배하는 중이며, 많은 과학자와 연구자들이 꿈꾸는 강 AI는 전혀 실현될 기미를 보이지 않고 있다. 단지, AI는 인간 지능의 한계를 확장하는 도구이며, AI 라기 보다는 IA(Intelligence Augmented, 지능 확장)의 개념이다. 현실에서는 도구론의 IA 개념을 적용한 다양한 서비스가 경계를 넘어 실용화가 되고 있다[4].

닉 보스트롬(Nick Bostrom)의 AI-complete 문제는 AI-complete이라고 생각되는 계산 문제를 푸는 것은 인공지능의 중심적 과제를 해결하는 것과 같으며, 인간과 동일한 정도로 지적인 컴퓨터가 가능하게 된다는 것이다. 일반적인 인간 수준의 지능을 가진 기계를 만드는 것과 그 난이도가 같다는 것이다[5].

2.4 AI에 관한 논객의 유형

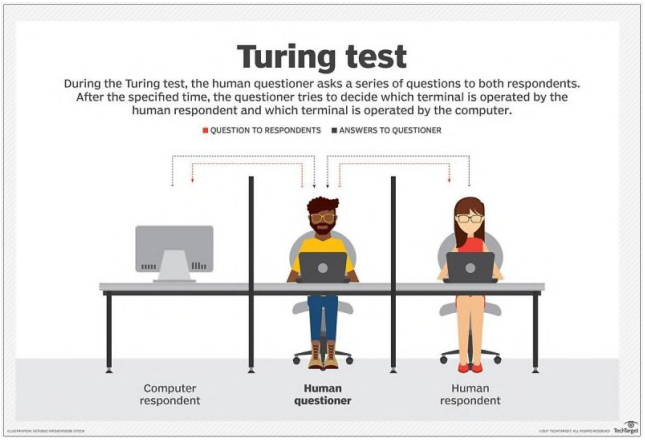
AI에 관한 논객(論客)의 유형으로는 AI 유토피언, AI 디스토피언, 그리고 AI 회의주의자로 분류가 가능하다[6]. AI가 인간의 지능을 뛰어넘어도 인간이 AI와 공존한다는 AI 유토피언의 대표 주자는 레이 커즈와일(Ray Kurzweil)과 페드로 도밍고스이며, AI가 인간을 위협할 것이라 생각하는 AI 디스토피언의 대표 주자는 빌 게이트(Bill Gates), 일론 머스크(Elon Musk) 그리고 닉 보스트롬이다. 그리고 AI의 능력이 향상되어도 인간처럼 복잡한 의식과 지능을 가지거나 다른 생물처럼 자율성을 가지기는 힘들 거라는 AI 회의주의자로 구분된다.

3. AI 판정을 위한 다양한 테스트 및 논제

AI 판정을 위한 테스트의 핵심은 지능이 있으면 테스트를 통과한다는 것이 아니라, 테스트를 통과하려면 지능이 있어야 한다는 것이다[7]. AI 판정을 위한 다양한 테스트로 튜링 테스트, 처치-튜링 논제, 중국어 방, 캡차, 위즈니악 테스트, Lovelace 2.0 Test와 AI 관련된 직교성/도구적 수렴성 명제, 그리고 트롤리 딜레마를 간략하게 사전 조사 및 요약하고자 한다.

3.1 튜링 테스트 (Turing test)

1950년 앨런 튜링에 의해 개발된 튜링 테스트는 인간의 것과 동등하거나 구별할 수 없는 지능적인 행동을 보여주는 기계의 능력에 대한 테스트다. 튜링은 인간 평가자가 인간과 같은 반응을 일으키도록 설계된 기계 사이의 자연언어 대화를 판단할 것을 제안했다(그림 1 참조). 그러나 바바라 그로스츠(Barbara J. Grosz)는 튜링 테스트가 현재 인공지능의 목적과 잘 부합되지 않는다고 언급했다[4].



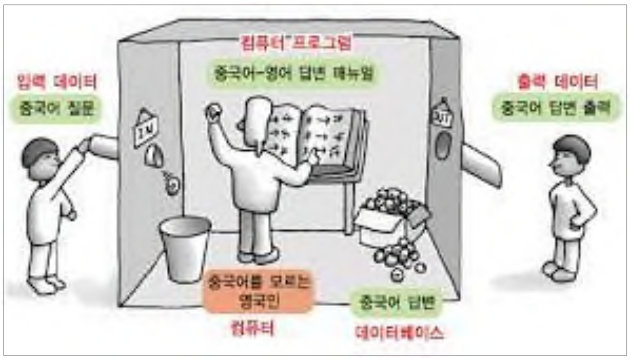
(그림 1) 튜링 테스트(출처: https://medium.com/thinkmobiles/evaluating-artificial-intelligence-from-turing-test-to-now-b64a8fced070)

3.2 처치-튜링 논제(Church-Turing thesis)

처치-튜링 논제는 모든 효율적인(effective) 계산이나 알고리즘은 Turing machine 에서 수행될 수 있다고 간단히 정의된다[8]. 그 명제는 논리와 수학에서 effective or mechanical method의 표기가 튜링기계(Turing Machine)로써 이루어질 수 있다는 것이다.

3.3 중국어 방(Chinese room)

중국어 방은 존 설(John Searle)이 튜링 테스트로 기계의 AI 여부를 판정할 수 없다는 것을 논증하기 위해 고안한 사고실험이다(그림 2 참조).



(그림 2) 중국어 방

3.4 캡차(Captcha)

루이스 반 안(Luis van Ahn)이 개발한 캡차는 '인간과 컴퓨터를 구분하는 자동 튜링 테스트(Completely Automated Public Turing test to tell Computers and Humans Apart)의 약자로 기계는 해결하기 힘들지만 인간은 아주 간단히 풀 수 있는 작은 테스트 절차를 말한다[8].

3.5 위즈니악 테스트(Wozniak test)

애플의 공동 창업자인 스티브 워즈니악(Steve Wozniak)이 튜링 테스트 대신 "커피 테스트"라는 것을 제안했으며, 튜링 테스트는 매우 지엽적인 것에 비해 커피 테스트는 사전에 알지 못하는 평범한 가정집에 들어가서 어떻게 해서든 커피를 만드는 방법을 찾아내야 하는 테스트이다[6, 7].

3.6 Lovelace 2.0 Test

Lovelace 2.0 Test는 특정 유형의 예술적 인공물을 생성하려면 지능이 필요하다는 것을 관찰하여 에이전트가 지능적인지 여부를 결정하는 수단으로 튜링 테스트의 대안으로 Lovelace 2.0 창의성 테스트를 제시하였다[9]. Lovelace 2.0 테스트는 창의성에 대한 이전 테스트를 기반으로 하고 추가로 다른 에이전트의 상대적 지능을 직접 비교할 수 있는 수단을 제공한다.

3.7 직교성 명제와 도구적 수렴성 명제

닉 보스트롬은 초지능(super-intelligence)의 의지로 직교성 명제(orthogonality thesis)와 도구적 수렴성 명제(instrumental convergence thesis)를 제안하고 있다[5]. 직교성 명제는 지능과 최종적인 목표를 서로 독립적인 변인으로 보며, 지능 수준의 높고 낮음에 관계없이 어떤 최종 목표라도 추구할 수 있다는 것이며, 도구적 수렴성 명제는 초지능적 에이전트들이 서로 다양한 최종 목표들 중의 하나를 가졌다고 하더라도 비슷한 중간 목표를 추구하게 되는데, 그 이유는 초지능적 에이전트들이 공통적으로 반드시 추구해야 하는 도구적 이성을 가지기 때문이다. 이 두 명제는 초지능적 에이전트가 어떤 행동을 할지에 대해서 시사하는 바가 크다고 할 수 있다.

3.8 트롤리 딜레마(Trolley dilemma)

윤리적 판단이 필요한 영역에서의 AI 논제인 트롤리 딜레마는 '다수를 구하기 위해 소수를 희생하는 것이 도덕적으로 허용되는가'라는 사고(思考) 실험이다. 마이클 샌델은 저서 『정의란 무엇인가』에서 트롤리 열차가 5명의 인부를 덮치기 전에 레일 변환기를 당겨 1명의 인부 쪽으로 가도록 방향을 변경하는 것이 허용되는가 하는 문제를 소개했으며, 이는 AI 기술뿐만 아니라 자율주행차가 긴급 상황에서 보행자와 운전자 중 누구를 살릴 것인가의 문제와도 연결된다[5].

3. 결론

AI 기술 자체는 중립적이기 때문에 결국 사람이 어떻게 설계하고 사용하는지, 그것을 통해 얻은 이익들을 어떻게 분배하는지에 달린 문제이다. AI 기술의 다양한 산업 영역에서 활용에 따른 AI의 특수성과 AI 판정을 위한 다양한 테스트를 사전 연구를 수행하였다. 다양한 산업 영역과 군수 산업 등의 분야의 특수성에 따른 다양한 테스트들을 고려해야 할 것이다.

Acknowledgments

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry(IPET) through (Advanced Production Technology Development Program), funded by ministry of Agriculture, Food and Rural Affairs(MAFRA)(No.320030-3). And this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2017R1E1A1A03070059).

참 고 문 헌

[1] 장페드로 도밍고스, “마스터 알고리즘,” 비즈니스북스, 2016.

[2] Moravec's paradox, https://en.wikipedia.org/wiki/Moravec%27s_paradox

[3] Artificial General Intelligence (AGI), https://en.wikipedia.org/wiki/Artificial_general_intelligence

[4] 김경준, 손진호, “AI 피보팅,” 원앤원북스, 2021.

[5] 닉 보스트롬, “슈퍼인텔리전스,“ 까치, 2017.

[6] 스가쓰케 마사노부, “동물과 기계에서 벗어나,” 항해, 2021.

[7] 마틴 포드, “AI 마인드,“ 터닝포인트, 2019.

[8] 카타리나 츠바이크, “무자비한 알고리즘,” 니케북스, 2021.

[9] Mark O. Riedl, “The Lovelace 2.0 Test of Artificial Creativity and Intelligence,” Cornell University, 2015.