

[Proposal Form for Venture Research Program for KAIST Ig-Nobel Prize]

<제안서는 국문 또는 영문으로 작성 (총 5페이지 이내)>

인공지능의 개념적 기초에 대한 새로운 철학적 접근 방법 제시

1. Information on Research Team

Name	Student ID	Department	Career	Role
임규성	20118087	물리학과	석박사통합	

2. Project Summary

Objective	<p>인공지능의 개념적 기초에 대해 기존과는 다른 관점의 철학적 접근 방법을 제시한다.</p> <ul style="list-style-type: none"> - 인공지능의 발전에 따른 용어의 재정의 - 인식적 · 경험적 측면에서 인공지능으로의 접근 - 인간과 인공지능의 차이에 기초한 새로운 개념 정립
Description	<p>2016년 3월 알파고와 이세돌의 바둑 대결이 벌어지기 전 많은 사람들이 이세돌의 승리를 예측했다. 체스는 1997년에 컴퓨터에 정복당했지만(IBM이 개발한 딥블루(Deep Blue)가 체스 세계 챔피언 가리 카스파로프에 승리), 바둑은 룰이 오히려 간단하기 때문에 고려해야 할 경우의 수가 체스보다 월등히 많다는 것이 승부 예측의 이유 중 하나였다. 결과는 알파고의 승리였다. 이를 통해 인공지능의 능력을 과소평가했다는 지적이 나왔고 알파고의 진정한 능력은 예측하기 힘들다는 얘기도 나왔다. 이 대결을 통해 인간은 인공지능이 과거에 비해 얼마나 비약적으로 발전했는지 직접 목격할 수 있었다.</p> <p>1956년 다트머스 학회에서 인공지능(Artificial Intelligence)라는 용어를 처음 사용한 John McCarthy는 인공지능에 대해 "It is related to the similar task of using computers to understand human</p>

	<p>intelligence” 라고 언급했다[1]. Merriam-Webster 사전에 따르면 인공지능은 1. a branch of computer science dealing with the simulation of intelligent behavior in computers, 2. the capability of a machine to imitate intelligent human behavior 와 같이 정의된다. 인공지능의 정의는 분야에 따라 조금씩 다르지만 일반적으로 인간의 지능 또는 합리성을 모방한다는 것을 골자로 하고 있다. 그러나 알파고의 사례에서 볼 수 있듯 인공지능은 특정 분야에 한해서는 인간보다 더 좋은 성과를 낼 수 있게 되었고 그 능력도 정확히 파악하기가 어렵다. 물론 지능을 개체의 통합적인 능력으로 파악하고는 있으나 체스, 바둑 뿐 아니라 다른 여러 분야에서 인간보다 좋은 성과를 보일 인공지능의 점진적 개발을 예상한다면[2] 인공지능에 대한 현재의 정의는 부족한 면이 있으며 기초적 단위에서 용어의 재속고가 필요함을 알 수 있다. 또한 하사비스 구글 답마인드 CEO는 ‘인공지능을 통해 인간들이 보지 못하고 있는 문제의 해결 포인트를 발견할 수 있다는 잠재력을 발견했다고 밝혔다[3]. 인공지능이 인간과는 다른 방식으로 문제에 접근한다는 점으로부터 다른 인식 체계를 가지고 있을 수 있음을 도출할 수 있으며, 이러한 경우 ‘인식’을 어떻게 정의해야 할 것인가 또한 의문으로 남게 된다. 본 연구에서는 이런 의문들에 대한 답을 제시하려는 것이다.</p> <p>제 46회 다보스 포럼에서 핵심 의제로 ‘제4차 산업혁명의 이해’를 꼽았는데, 제4차 산업혁명의 대표적 기술분야 중 하나인 인공지능에 대한 철학적 배경이 선행되어야 앞으로 다가올 제4차 산업혁명의 개념적 기반을 견고히 하는 데 도움을 줄 수 있다. 본 연구에서는 알파고가 방대한 양의 데이터를 통해 머신러닝으로 학습한 점에서 출발해, 경험적(‘경험’이 인공지능을 어떤 식으로 규정짓는지 등과 관련)·인식적 측면에서 인공지능에 대해 철학적으로 접근하고자 한다.</p>
Novelty / unusualness	<p>1. 관련 선행 연구</p> <p>1950년 앨런 튜링이 제안한 튜링 테스트(Turing Test)는 지능적 개체인 인간과 ‘구분할 수 없음’이라는 개념을 이용하여 설계되었다. 텔레프린터를 통한 대화에서 기계가 사람인지 기계인지 구별할 수 없을 정도로 구분이 되지 않는다면 그 기계가 생각할 수 있는 것으로 간주할 근거가 된다는 것이다[4]. 하지만 이 테스트는 사람처럼 대화를 할 수 있다는 점에만 초점을 맞추어 특정 영역에서 컴퓨터의 수행 능력은 반영하지 않는다는 면이 있다.</p> <p>역사적으로 AI의 정의에는 크게 4가지 접근 방식이 있다[5]. 사고와 행동, 그리고 인간의 능력과 합리성에 초점을 맞춘 것인데, 인간 및 지능의 이상적 개념인 합리성으로 접근하기 때문에 딥러닝을 통한</p>

경험적 측면, 그리고 인식 개념의 변화 측면에 기반을 둔 정의와는 차이가 있다고 할 수 있다..

Systems that think like humans.	Systems that think rationally.
Systems that act like humans	Systems that act rationally

그림 1. AI 에 대한 4가지 접근 방식

2. 연구의 독창성

알파고 뿐만 아니라 IBM의 슈퍼컴퓨터 '왓슨'은 현재 세계적 암 전문 병원인 MD앤더슨센터에서 암 진단을 위한 과정 중에 있는데 이러한 인공지능들은 방대한 양의 데이터에 그 학습 기반을 두고 있다. 데이터의 양이 기하급수적으로 증가할 뿐만 아니라 데이터를 처리할 인프라도 발전하며 인공지능의 학습은 인간의 학습과는 질적으로 다른 면을 보여주고 있다. 이런 측면에서 인식의 수준 자체도 인공지능과 인간은 다르다는 것을 알 수 있으며 현재 인공지능을 정의 내리는 많은 경우에는 이러한 점들을 고려하지 않고 있다. 즉, 받아들이는 데이터의 양적 차이로 인해 생기는 인식적 배경을 어떻게 반영하고 인지할 것인가 등이 새롭게 고려해야 할 사항이 될 것이다. 본 연구에서는 특히 인공지능과 인간이 나타내는 차이에 초점을 두고, 인공지능에 있어서의 지능과 인식 등의 재정의로부터 접근하여 철학적 개념들을 제안하고자 한다.

Scholarly profundity

장하석 교수는 저서 '온도계의 철학'에서 "철학과 역사는 현재의 전문가적 과학에서 배제된 세계에 관한 물음을 찾아내거나 거기에 기여하는 데 함께 일을 한다. 철학은 조직된 회의와 비판이라는 유용한 습관의 제공자로서, 역사학은 잊힌 물음과 답의 제공자로서 기여한다. 과학사와 과학철학은 과학 지식을 확장하고 풍부화하는 과정에서 땀 수 없는 동반자이다. 나는 둘이 함께 이루는 학제를 일러 상보적 과학이라고 부르려 제안한다. 그것은 전문가적 과학에 중요한 보완으로서 존재해야 하기 때문이다."라고 했다[6]. 이런 측면에서 과학적 인식의 기본문제에 관하여 고찰하는 철학적 질문은 과학이 나아감에 있어 항상 제기되어야 하며 답을 고민해봐야 할 문제임을 알 수 있다. 과학이론에 대한 입장을 제시하는 '칸트의 비판철학', '논리실증주의', '비판적 합리주의' 등이 그러한 예라고 할 수 있다. 그러므로 인공지능과 관련하여 그 기초적인 개념들에 대해 다시 한번 질문하고 현재의 관점에서 답을 생각해 보는 것은 인공지능의 정립에 있어 보

완적 역할을 할 수 있을 것이다.

두 번째는 응용적 측면이다. 인공지능이 제4차 산업혁명의 주요 기술로 활용될 것으로 명백히 예상할 수 있음에도, 인공지능의 명확한 법적 개념이나 권리·의무 관계 등은 정립되지 않은 실정이다. 워싱턴대 라이언 칼로 교수는 로봇에 대한 법적 규제가 필요하다고 주장했으며[7] 스티븐 호킹 박사를 비롯한 많은 명사들이 인공지능의 위험성을 경계하고 있다. 뿐만 아니라 인공지능이 초래할 일자리의 엄청난 변동에 있어서도 우려만 하고 있을 뿐 실질적인 대책을 제시하지는 못하는 실정이다. 본 연구에서 제시하는 인공지능에 대한 새로운 접근 방법은 철학적 질문에 대한 경험적·인식적 측면의 답을 제시할 뿐 아니라 이러한 인식을 토대로 인공지능과 관련한 법적·윤리적·사회적 접근에 대한 하나의 대안을 제시할 수 있을 것으로 기대된다. 이는 인간의 존재성에 대해서도 다시 생각해 볼 수 있는 기회를 제공할 것이다.

제1목표는 과학철학을 통해 과학적 질문에 어떻게 답변을 내릴 수 있는지 살펴보는 일이 될 것이다. 이에 대한 명확한 인지가 앞으로의 연구에 있어 기초적인 역할을 할 것으로 생각된다. 인공지능에 대한 종래의 정의 및 접근이 어떠한지를 연구하는 것은 두 번째 목표로 놓을 수 있다. 이를 통해 인공지능이 지향하는 바를 인간적 측면에서 파악하게 되고 인공지능에 대한 이해도를 높일 수 있다. 제3의 목표는 알파고에 적용된 학습 방법이 경험적·인식적 측면에서 기존의 인공지능에 대한 개념과 어떤 차이가 있는지에 대한 분석이다. 철학적 분석이 주 연구과제이기 때문에, 연구의 전체 단계에서 경험적 측면에서 과학을 바라보는 경험주의와 인식론에 대한 이해가 병행되어야 할 것이다. 위의 일련의 단계를 통해 인공지능에 대한 새로운 접근 방법을 제시할 수 있을 것으로 생각된다. 아래는 각각의 단계에 대한 대략적인 예상 연구기간이다.

Research pla

목표	5월	6월	7월	8월	9월	10월	11월	12월
과학 철학	←→							
A.I. 이해			←→					
접근법 제시						←→		
철학적 이해		←→						

참고 문헌

- [1] John McCarthy, What is Artificial Intelligence?, 2007, Stanford University.
- [2] 유발 하라리, 사피엔스 관련 기자간담회, 2016년 4월 26일.
- [3] '알파고 아빠' 하사비스가 밝힌 '세기의 대결' 시사점은, 머니투데이, 2016년 3월 18일.
- [4] A.M. Turing, Computing Machinery and Intelligence, Mind 49: 433-460, 1950.
- [5] Stuart J. Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, Prentice Hall.
- [6] 장하석, 온도계의 철학, 동아시아.
- [7] 인간을 닮아가는 그들... 새로운 법 필요하다, 교수신문, 2015년 9월 1일.