Issue Report

2018. 7. 31. 제2018-002호

범용 인공지능의 개념과 연구 현황

추형석 선임연구원†

SPRi 소프트웨어정책연구소

○ 본 보고서는 「과학기술정보통신부 정보통신진흥기금」을 지원받아 제작한 것으로 과학기술정보통신부의 공식의견과 다를 수 있습니다.

○ 본 보고서의 내용은 연구진의 개인 견해이며, 본 보고서와 관련한 의문사항 또는 수정・보완할 필요가 있는 경우에는 아래 연락처로 연락해 주시기 바랍니다.

- 소프트웨어정책연구소 기술·공학연구실 추형석 선임연구원(hchu@spri.kr)

《 Executive Summary 》

현대 인공지능은 유례가 없을 정도로 빠르게 진화하고 있다. 이러한 속도 때문에, 많은 전문가들은 인공지능이 사람을 지배할 것이라는 예측을 내놓고 있다. 대표적으로, 테슬라 최고경영자인 엘론 머스크는 미래 인공지능을 핵무기에 빗대어 표현하면서 그 위험성을 강조했다. 지난 2018년 3월 작고한 영국의 물리학자 스티븐 호킹 역시 인공지능이 인간을 대체할 수 있다는 것에 두려움을 표명했다. 그렇다면 이처럼 인공지능이 인류를 위협할 수 있다는 사실은 어디에서 기인하는 것일까? 그것은 사람 수준의 인공지능(Human-level intelligence), 즉 범용 인공지능(Artificial General Intelligence)이라고 일컬을 수 있다.

범용 인공지능은 기본적으로 사람 수준의 지능을 구현한 시스템이라고 개념 지을 수 있다. 범용 인공지능은 사람처럼 학습하고, 추론하며, 문제를 인식하고 이것을 해결하기 위해 능동적인 대처를 한다. 그러나 사람 수준의 지능은 전문가마다 보는 시각이 매우 다를 수 있다. 따라서 범용 인공지능의 구체적인 정의는 각 분야마다 다르고, 연구자들은 범용 인공지능을 개발하면서 그 정의를 정립해 나가고 있다.

여기서 중요한 점은 범용 인공지능이 사람 수준을 능가할 수 있는 초지능(Superintelligence)로 바라보는 것이다. 만약 범용 인공지능이 탄생한다면, 이것은 자가 향상(Self-improvement)을 통해 사람보다 훨씬 높은 수준의 지능에 도달한다는 관점이다. 이 관점은 대표적인 미래학자인 레이 커즈와일이 그의 저서에서 밝힌 특이점과도 일맥상통한다. 범용 인공지능을 초지능으로 가는 변곡점으로 여긴다면, 앞서 서술한 세계 석학과 최고경영자의 우려 섞인 예측도 충분히 상상 가능한 미래다.

그러나 현재 범용 인공지능과 관련된 연구 현황을 살펴 볼 때, 아직 범용 인공지능이 탄생하기에는 시기상조로 보인다. 그러나 현대 인공지능의 눈부신 발전은 많은 연구자들에게 범용 인공지능 시대가 도래 할 것이라는 긍정적인 신호를 보내고 있다. 이에 국제사회는 범용 인공지능의 출현에 대비하기 위해 인공지능을 인류에 이롭게 활용한다는 대원칙을 세우고 있다. 우리도 이러한 국제사회의 동정에 적극 참여하여 범용 인공지능 시대를 대비해야 할 것이다.

《 Executive Summary 》

Modern artificial intelligence evolves to unprecedented speed. Because of this speed, many experts are predicting that artificial intelligence will dominate human. Typically, Elon Musk, Tesla CEO, stressed the dangers of future artificial intelligence as nuclear weapons. The British physicist Stephen Hawking, who died in March 2018, also expressed fears that artificial intelligence could replace humans. So where does the fact that artificial intelligence can threaten humankind? It can be called human-level intelligence, that is, Artificial General Intelligence (AGI).

The AGI is basically a system that implements human intelligence. AGI can learn, inference, recognize problems, and take active steps to solve them like human. However, human-level intelligence can be very different for each expert. Therefore, the specific definition of the AGI is different for each field, and the researchers are in the process of establishing the definition while developing the AGI.

What is important here is that the AGI can be regarded as superintelligence that can surpass the human level. If a AGI is born, this is a view that self-improvement will reach a much higher level of intelligence than humans. This view is in line with the singularity laid out by a representative futurist, Ray Kurzweil, in his book. If the AGI is regarded as an point to superintelligence, it is a imaginable future with the worries of the world scholars and CEO as described above.

However, when we look at the present state of research related to the AGI, it seems that it is premature to produce the AGI. However, the remarkable development of modern artificial intelligence has sent a positive signal to many researchers that the era of AGI will come. Therefore, the international society is making a principle of using artificial intelligence beneficially to humanity. We should also actively participate in the trend of the international community and prepare for the age of the AGI.

《 목 차 》

1. 서 론 ····· 1

2. 범용 인공지능의 개념과 접근방법 ····· 4

3. 범용 인공지능의 연구 현황 ····· 8

4. 결 론 ····· 12

《 Contents 》

1. Introduction ····· 1

2. Concepts and approaches of AGI ····· 4

3. AGI R&D Trends ····· 8

4. Conclusions ····· 12

1. 서 론

지난 2016년 3월 구글이 개발한 바둑 인공지능 프로그램 AlphaGo는 세계 최정상 바둑기사인 이세돌 9단에 맞서 4:1로 완승했다. 그 성공의 원동력은 심층학습(Deep learning)이다. 심층학습은 기본적으로 인간의 뇌 구조를 모사한 인공신경망(Artificial Neural Network)을 학습하는 기술이다. 여기서 심층학습은 더 복잡하고 깊은 인공신경망[1]을 학습하는 방법론으로 그 차별점을 찾을 수 있다. 깊은 인공신경망을 성공적으로 학습시키기 위해서는 필연적으로 많은 데이터와 계산 자원이 필요하다. 현대 인공지능의 핵심 기술로써의 심층학습은 그 방법론의 진화[2]뿐만 아니라, 폭발적으로 성장한 연산처리장치의 성능[3]과 빅데이터의 대중화가 서로 맞물려 일어난 결과다.

심층학습이 본격적으로 학계나 산업계에 널리 퍼지게 된 계기는 2012년에 개최된 이미지 인식 경진대회[4]의 결과로부터 출발한다. 2011년 경진대회의 우승자는 약 25.8%의 객체 인식 오차를 기록한 반면 2012년에는 심층학습 기법 중 하나인 합성곱신경망(Convolutional Neural Network)을 활용하여 16.4%의 오차율을 기록했다. 이것은 컴퓨터 비전 분야의 획기적인 연구 성과로 주목받음과 동시에 다른 학계에서의 관심까지 이끌어냈다. AlphaGo에서 활용된 인공신경망 역시 이미지 인식에 활용된 합성곱신경망이다.

심층학습의 성공사례는 비단 바둑이나 이미지 인식에 그치지 않는다. 게임 인공지능, 음성 인식, 기계 번역 등 심층학습은 다양한 분야에서 가장 좋은 성능을 기록하고 있다. 그러나 심층학습은 특정 분야에 최적화된 기술이다. 이미지 인식을 잘 하는 인공신경망이 음성 인식을 잘 할 수 없기 때문이다. 그럼에도 불구하고 심층학습의 높은 성능 덕에 심층학습에 대한 학계나 산업계의 관심은 나날이 높아지고 있는 실정이다.

1) 인공신경망은 노드(퍼셉트론)과 링크(퍼셉트론 간의 연결)로 구성되며, 다수의 노드를 하나의 층(layer)으로 지칭한다. 인공신경망은 많은 수의 층(layer)를 쌓을수록 깊이가 깊어지고, 일반적으로 심층신경망일수록 높은 예측성능을 기대할 수 있다.
2) 심층학습의 선구자인 토론토 대학의 제프리 힌튼 교수가 개선한 비지도학습, 전자공학에서 사용되는 활성함수(Rectified Linear Unit)의 차용 등 방법론 자체도 진화
3) ‘매 18개월 마다 연산처리장치의 성능은 2배 향상된다’는 무어의 법칙(Moore’s Law)아래 연산처리장치는 지수적으로 그 성능이 향상됐으며, 대규모 병렬처리에 적합한 그래픽연산처리장치(Graphical Processing Units)의 대중화로 심층학습에 필요한 시간을 대폭 감소시킴
4) Imagenet Large Scale Visual Recognition Challenge(ILSVRC)는 120만 장의 이미지에서 1000가지의 객체를 분류하고 위치를 찾는 경진대회

많은 전문가들은 심층학습의 발전 속도가 유례없이 빠르다고 강조한다. 이세돌 9단과 대결했던 AlphaGo는 경기 당시 이상한 실수를 하는 등 완벽하지 않았다. 또한 많은 전문가들은 심층학습의 낮은 설명 가능성으로 인해 AlphaGo가 더 높은 수준으로 진화하기에는 수년의 세월이 필요할 것이라고 예측했다. 그러나 AlphaGo 개발진인 구글 딥마인드는 불과 1년 만에 인류의 바둑지식을 초월한 AlphaGo Zero를 공개했다. AlphaGo Zero가 더욱 충격적이었던 사실은 AlphaGo Zero가 어떠한 바둑 기사의 기보도 학습에 활용하지 않았다는 점이다.

이러한 심층학습의 놀라운 성과는 반대로 다양한 우려를 낳고 있다. 테슬라 모터스의 최고경영자인 엘론 머스크나 천재 물리학자 스티븐 호킹은 미래의 인공지능이 인류를 멸망시킬 것이라고 강조했다. 또한 사회적으로는 인공지능이 직업을 대체하여 인류의 일자리를 위협할 것이라는 걱정도 팽배해 있다. 이러한 부정적인 관점에 대한 원인은 현대 인공지능의 빠른 발전 속도로 인해 정말 사람 수준(Human-level)의 인공지능이 수 십 년 내에 만들어 질 것이라는 점이다. 사람 수준의 인공지능, 즉 범용 인공지능(Artificial General Intelligence)의 개발은 정말로 가능한 것일까? 현재 범용 인공지능과 관련된 연구는 무엇이 있을까? 이 보고서는 이 물음에 대한 답을 찾기 위해 범용 인공지능의 개념과 연구현황을 살펴보고자 한다.

본격적으로 범용 인공지능을 다루기에 앞서 심층학습에 대해 조금 더 살펴보자. 먼저 심층학습의 장점은 복잡한 데이터에서 패턴을 인식하고 예측하는 성능이 우수하다는 것이다. 예를 들어, 기계 번역은 end-to-end라는 심층학습 방법을 활용한다. 한국어를 영어로 번역하는 과정에서 한국어 문장과 그 번역에 해당하는 영어 문장을 대응시켜 방대한 데이터 셋을 만들고, 이를 학습시키는 과정이 end-to-end 방법이다. 이것은 전통적인 자연언어처리 방법과는 다르게 문장의 구조나 특정 문맥에서의 단어의 의미를 전혀 파악하지 않아도 된다. end-to-end 심층학습은 과거 전통적인 자연언어처리 기술을 활용한 번역보다 월등히 성능이 뛰어나기 때문에 현재 기계 번역 분야에서 많이 활용되고 있다.

심층학습은 이러한 장점과 대비하여 그 한계도 분명하다. 먼저 심층신경망[5]의

5) 심층신경망(Deep Neural Network)은 개념적으로 깊은 인공신경망(Artificial Neural Network)이다. 또는, 신경망 층(layer)을 2~3개 적층한 구조를 얕은 신경망(Shallow Network)에 대비되는 개념으로도 이해할 수 있다. 일반적으로 심층학습의 인공신경망은 심층신경망을 의미한다.

성능은 특정 분야에 국한돼있다. 앞서 소개한 기계 번역의 end-to-end 심층신경망은 이미지나 음성을 인식할 수 없다. 인간은 뇌라는 하나의 신경망으로 사물을 인식하고 미래를 추론하며 바둑을 둘 수 있다. 심층학습은 이러한 관점에서 매우 제한적인 범용성을 가지고 있다. 흥미로운 것은 기술적인 진전이 조금씩 진행되고 있다는 것이다. AlphaGo Zero는 자체대국을 통한 자가 학습 과정을 일반화하여 AlphaZero라는 방법을 제안했다. AlphaZero는 바둑 이외에 체스, 쇼기 등의 보드게임에서도 최정상 실력을 입증하여 그 가능성을 시사했다. 그러나 AlphaZero는 여전히 보드게임이라는 틀에서 벗어나지 못한 것이 현실이다.

심층학습의 또 다른 한계는 설명가능성에 있다. 심층학습의 결과물은 매우 뛰어난 성능을 보여주나, 이 결과가 왜 산출되는가에 대해 명확히 설명 불가능하다는 것이다. 이것은 본질적으로 인공신경망의 구조에서 그 이유를 찾을 수 있다. 인공신경망은 다수의 인공신경세포를 연결해서 만든 구조다. 하지만 어떻게 연결해야 최적의 성능을 보장할 수 있는가에 대한 이론적인 근거가 없다. 이 사실로부터 심층학습 방법론은 경험적인 결과에 의존한다고 추론할 수 있다. 가능한 한 많은 경우의 수를 수행하여 최적의 결과를 도출하는 심층학습은 접근 방법 자체에서의 태생적인 한계가 존재한다는 것이다.

이러한 심층학습의 단점에도 불구하고, 심층학습에 대한 관심은 인공지능뿐만 아니라 과학기술 전반에 걸쳐 매우 높아지는 추세다. 그 첫 번째 이유는 바로 성능이다. 설명가능성이 높은 방정식 기반의 고전적인 모델링보다 데이터를 학습해 추론하는 심층학습의 성능이 월등하기 때문이다. 하지만 심층학습은 만능이 아니다. 심층학습은 일반적으로 학습을 위해 방대한 데이터를 필요로 한다. 또한 경험적으로 많은 시도를 해야 하기 때문에 필연적으로 많은 계산을 요구한다. 이렇게 심층학습을 객관적으로 보자면 심층학습이 발전하여 범용 인공지능으로 진화한다는 것은 시기상조일 수 있다. 그렇다면 심층학습과 범용 인공지능의 관계는 어떻게 될까? 심층학습이 범용 인공지능으로 가는 길에 가장 핵심적인 기술일까? 아니면 요소 기술 중 하나일까?

이 보고서는 범용 인공지능의 개념과 연구 현황에 대해서 논의할 것이다. 먼저 범용 인공지능의 개념과 접근방법을 소개하고, 분류별로 활발히 수행되고 있는 연구 내용을 살펴보겠다. 마지막으로 범용 인공지능의 미래에 대해 결론지을 것이다.

2. 범용 인공지능의 개념과 접근방법

범용 인공지능은 1956년 다트머스 회의에서 인공지능의 개념이 처음 정립됐을 때부터 출발한다. 여기서 논의된 궁극적인 인공지능의 형태가 바로 범용 인공지능이기 때문이다. 그러나 범용 인공지능으로 가기 위한 기술적 한계[6]는 극명했다. 이로 인해 인공지능은 지난 60여 년 간 부침의 역사를 겪었고, 결국 인공지능은 컴퓨터 과학의 한 분야로 인간의 지능적 행동의 일부를 모사하기 위한 접근을 취했다. 최근 범용 인공지능이라는 키워드가 다시 부상한 이유도 심층학습의 눈부신 성장 때문이라고도 볼 수 있다.

하지만 범용 인공지능은 아직 개발되지 않은 기술이기 때문에, 그 정의 역시 구체적이지 않은 것이 사실이다. 많은 학자들은 범용 인공지능을 인간 수준의 지능(Human-level intelligence)로 일컫는다. 이것은 인간처럼 정보를 학습하고, 의미를 부여하며, 다양한 문제를 인지하여 해결책을 도출하는 과정을 구현함에 있다. 물론 범용 인공지능을 단순히 사람 수준의 지능이라고 바라보는 것은 어폐가 있다. 범용 인공지능이 자의식(self-consciousness)이나 감정(emotion)이 필요한가? 혹은 범용 인공지능을 위해 물리적인 몸체(physical body)가 필요한가?에 대한 대답은 분야별 전문가에 따라 상당히 다르기 때문이다.

따라서 범용 인공지능에 대한 정의는 아직도 논의가 진행 중이다. 그 정의가 모호하기 때문에, 오히려 연구자들이 범용 인공지능에 대한 실체를 만들어 나가는 중이라고 볼 수 있다. 앞서 언급했듯이 범용 인공지능을 인간 수준의 지능을 구현한 기계장치로 바라본다면, 범용 인공지능은 다양한 목표를 달성하기 위해 능동적으로 해결책을 찾는 시스템이라고 볼 수 있다. 범용 인공지능 분야에서 활발하게 활동 중인 벤 괴르첼(Ben Goertzel) 박사는 이것을 핵심 범용 인공지능 가설(The Core AGI Hypothesis)이라고 제시했다. 이 가설이 맞다면 범용 인공지능은 좁은 인공지능(narrow AI)[7]의 반대개념으로 구분될 수 있다는 관점이다.

6) 인간의 뇌가 어떠한 방식으로 동작하는가에 대한 생리학적인 규명의 부재, 의식(consciousness)과 감정(emotion)에 대한 구현 방안, 인간이 학습하고 추론하는 방식에 대한 메카니즘 등

7) 좁은 인공지능(narrow AI)은 세계적인 미래학자인 레이 커즈와일(Ray Kurzweil)의 저서 『The Singularity Is Near』에서 소개된 개념으로 특정한 상황에서 특정한 지능적 행동을 하는 시스템으로 정의된다. 좁은 인공지능은 상황이나 행동을 약간 변화시킨다면, 일정 수준의 프로그래밍과 재구성이 필요하다. 심층학습은 좁은 인공지능의 대표적인 예로 볼 수 있다.

The Core AGI Hypothesis

- 충분히 넓고(인간 수준) 강한 범용성을 갖는 인공지능의 창조나 연구. 질적인 측면에서 상당히 좁고 약한 범용성과 반대되는 개념
- 원문 : the creation and study of synthetic intelligences with **sufficiently broad** (e.g. human-level) **scope and strong generalization** capability; is at bottom qualitatively **different from** the creation and study of synthetic intelligences with **significantly narrower scope and weaker generalization** capability

자료 : Artificial General Intelligence: Concept, State of the Art, and Future Prospects, Ben Goertzel (2017)

이러한 개념상의 범용 인공지능을 고려해보면, 심층학습은 아직 그 수준에 도달하지 못했다. 그러나 심층학습으로 촉발된 인공지능의 성능 발전 속도를 보면 범용 인공지능의 출현은 그리 먼 미래가 아닐 수도 있다. 실제로 비영리단체인 Future Life Institute(FLI)는 인공지능 전문가 1,634명을 대상으로 범용 인공지능의 출현 시기를 설문조사 했다. 그 결과 45년 뒤에 50%의 확률로 인공지능이 인간을 초월할 것이라고 전망했고, 인간의 모든 직업을 대체 하는 데는 120년이 걸릴 것으로 예측했다.[8)]

만약 범용 인공지능이 실제로 출현하면 인류는 인공지능의 지배아래 놓이게 될까? 세계적인 미래학자인 레이 커즈와일은 그의 저서에서 밝혔듯이 범용 인공지능이 출현하는 시점을 특이점(Singularity)으로 정의했다. 특이점 이후의 인공지능은 자가 발전(self-improvement)하여 인간을 뛰어넘는 초인공지능(superintelligence)이 되는 것이다. MIT의 물리학과 교수인 맥스 테그마크 역시 범용 인공지능의 출현은 곧 초인공지능을 의미한다고 밝혔다.[9)] 이 관점은 인간과 범용 인공지능은 단순한 주종관계가 아니라, 서로 공존하는 관계로 정립돼야 함을 내포한다. 이에 인공지능 연구자들은 2017년 아실로마 학회에서 인공지능을 이롭게 활용하기 위한 23개 원칙 <표 1> 을 정립했다.[10)]

8) When Will AI Exceed Human Performance? Evidence from AI Experts (2018)
9) How to get empowered not overpowered by AI, Max Tegmark, TED talk (2018)
10) 아실로마 AI원칙, Future of Life Institute (2017), https://futureoflife.org/ai-principles-korean/

<표 1> 아실로마 23대 AI 원칙

용 어	내 용
연구목표	인공지능 연구의 목표는 방향성이 없는 지능을 개발하는 것이 아니라 인간에게 유용하고 이로운 혜택을 주는 지능을 개발하는 것이다.
연구비 지원	인공지능에 대한 투자에는 컴퓨터 과학, 경제, 법, 윤리 및 사회 연구 등의 어려운 질문을 포함해 유익한 이용을 보장하기 위한 연구비 지원이 수반되어야 한다.
과학정책 연결	인공지능 연구자와 정책 입안자 간에 건설적이고 건전한 교류가 있어야 한다.
연구문화	인공지능 연구자와 개발자 간에 협력, 신뢰, 투명성의 문화가 조성되어야 한다.
경쟁 피하기	인공지능 시스템 개발팀들은 안전기준에 대비해 부실한 개발을 피하고자 적극적으로 협력해야 한다.
안전	인공지능 시스템은 작동 수명 전반에 걸쳐 안전하고 또 안전해야 하며, 적용 가능하고 실현 가능할 경우 그 안전을 검증할 수 있어야 한다.
장애 투명성	인공지능 시스템이 손상을 일으킬 경우 그 이유를 확인할 수 있어야 한다.
사법적 투명성	사법제도 결정에 있어 자율시스템이 사용된다면, 권위 있는 인권기구가 감사 할 경우 만족스러운 설명을 제공할 수 있어야 한다.
책임	고급 인공지능 시스템의 디자이너와 설계자는 인공지능의 사용, 오용 및 행동의 도덕적 영향에 관한 이해관계자이며, 이에 따라 그 영향을 형성하는 책임과 기회를 가진다.
가치관 정렬	고도로 자율적인 인공지능 시스템은 작동하는 동안 그의 목표와 행동이 인간의 가치와 일치하도록 설계되어야 한다.
인간의 가치	인공지능 시스템은 인간의 존엄성, 권리, 자유 및 문화적 다양성의 이상에 적합하도록 설계되어 운용되어야 한다.
개인정보 보호	인공지능 시스템의 데이터를 분석 및 활용능력의 전제하에, 사람들은 그 자신들이 생산한 데이터를 액세스, 관리 및 통제할 수 있는 권리를 가져야 한다.
자유와 개인정보	개인정보에 관한 인공지능의 쓰임이 사람들의 실제 또는 인지된 자유를 부당하게 축소해서는 안 된다.
공동이익	인공지능 기술은 최대한 많은 사람에게 혜택을 주고 힘을 실어주어야 한다.
공동번영	AI에 의해 이루어진 경제적 번영은 인류의 모든 혜택을 위해 널리 공유되어야 한다.
인간의 통제력	인간이 선택한 목표를 달성하기 위해 인간은 의사결정을 인공지능 시스템에 위임하는 방법 및 여부를 선택해야 한다.
비파괴	고도화된 인공지능 시스템의 통제로 주어진 능력은 건강한 사회가 지향하는 사회적 및 시정 과정을 뒤엎는 것이 아니라 그 과정을 존중하고 개선해야 한다.
인공지능 무기 경쟁	치명적인 인공지능 무기의 군비 경쟁은 피해야 한다.
인공지능 능력에 관한 주의	합의가 없으므로 향후 인공지능 능력의 상한치에 관한 굳은 전제는 피해야 한다.
중요성	고급 AI는 지구 생명의 역사에 심각한 변화를 가져올 수 있으므로, 그에 상응 한 관심과 자원을 계획하고 관리해야 한다.
위험	인공지능 시스템이 초래하는 위험, 특히 치명적인 또는 실존적 위험에는, 예상된 영향에 맞는 계획 및 완화 노력이 뒷받침되어야 한다.
재귀적 자기 개선	인공지능 시스템이 재귀적 자기 복제나 자기 개선을 통하여 빠른 수적 또는 품질 증가를 초래한다면, 설계된 시스템은 엄격한 안전 및 통제 조치를 받아야 한다.
공동의 선	초지능은 널리 공유되는 윤리적 이상을 위해, 그리고 몇몇 국가나 조직이 아닌 모든 인류의 이익을 위해 개발되어야 한다.

자료 : 아실로마 AI원칙, Future of Life Institute (2017),
https://futureoflife.org/ai-principles-korean/

범용 인공지능의 접근 방법은 크게 두 가지가 있다.[11] 첫 번째 접근 방법은 기호적 범용 인공지능(Symbolic AGI)이다. 기호적 범용 인공지능의 근간인 기호적 인공지능(Symbolic AI)은 과거 전문가 시스템에서 현대의 IBM 왓슨에 이르기까지 지속적으로 활용돼왔다. 사람은 물체를 인식할 때 기호를 부여한다. 그리고 그 기호 간의 관계를 정립함으로써 논리적인 구조를 만들고 비슷한 상황을 추론할 때 활용한다. 이렇게 기호적 인공지능은 지능이 기호를 다루는 행위로 정의하는 분야다. 기호적 범용 인공지능의 장점은 접근방법 자체가 일반화에 가장 적합한 도구라는 것이다. 따라서 심층학습과는 다르게 다양한 지능적 활동을 할 수 있다. IBM 왓슨이 질문응답 시스템(Q&A 시스템)으로 출발했지만, 현재 응용분야는 의료, 금융 등 다양하게 그 적용 영역을 넓히고 있기 때문이다. 그러나 기호는 간단한 지각이나 동기로 인해 변화하거나 진화할 수 있기 때문에, 단순히 기호와 규칙만으로 범용 지능을 구현하기에는 한계가 존재한다.

범용 인공지능의 두 번째 접근방법은 창발적 범용 인공지능(Emergentist AGI)이다. 창발적 범용 인공지능 접근방법은 인간의 뇌의 작동 방식을 다양한 방식으로 모사하여 구현하는 인공지능을 의미한다. 특히, 뇌의 구조와 기능을 역공학(reverse engineering)으로 해석하여 기계로 구현하는 방법이 일반적이다. 심층학습 역시 큰 의미에서 보자면 창발적 접근방법이다. 심층학습에서 활용되는 인공신경망은 본질적으로 뇌 신경망을 모사하여 만들었기 때문이다. 흥미로운 것은 인공신경망의 구조가 조금 더 실제 뇌를 반영하도록 진화하고 있다는 것이다. 그러나 뇌의 생리학적 규명이 아직 완벽하지 않기 때문에, 더불어 창발적 범용 인공지능의 물리적 한계가 존재할 수 있다고 볼 수 있다.

위에 소개한 두 가지 접근방법은 물과 기름처럼 정확히 나뉘는 것은 아니다. 실제로 두 가지를 혼합한(hybrid) 접근방법도 존재한다. 또한 위의 접근방법 이외에 다른 방식으로 범용 인공지능을 해석하는 관점도 존재한다. 이 보고서에서는 위에서 소개된 두 가지 접근 방법을 준용하여, 이어지는 3장에서는 접근 방법별로 대표적인 범용 인공지능 연구 현황을 소개하겠다.

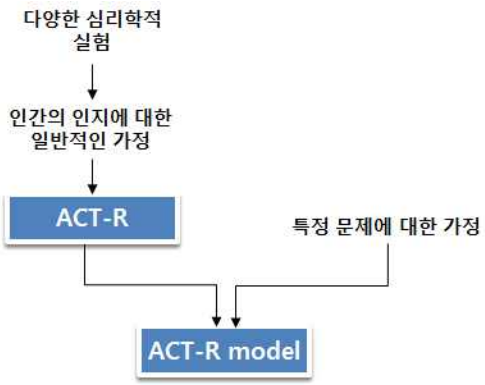
11) 『Artificial General Intelligence: Concept, State of the Art, and Future Prospects』에서 소개된 내용을 정리

3. 범용 인공지능의 연구 현황

(1) ACT-R (Adaptive Control of Thought - Rational)

ACT-R은 대표적인 기호적 범용 인공지능 접근방법의 인지 아키텍처(Cognitive Architecture)[12]로 1973년부터 시작된 유서 깊은 인공지능 연구다. ACT-R은 인간의 인지(cognition)가 어떻게 작동하는지에 대한 이론을 컴퓨터를 활용해 구현하여 다양한 임무를 해결하는데 접목됐다. ACT-R의 이론은 심리학 실험에서 파생된 수많은 사실과 인간의 인지의 과정에 대한 가정(assumption)을 통해 정립된다.

ACT-R은 일종의 SW 프레임워크로 정립된 이론을 통해 다양한 문제[13]를 해결하는 환경을 제공한다. 연구자들은 특정 임무를 해결하기 위해 ACT-R을 활용하여 모델(컴퓨터 프로그램)을 만들 수 있고, 여기에 연구자들은 문제에 대한 가정을 추가한다. 이 가정은 실제 사람이 동일한 임무를 수행한 것과 비교하여 검증하는 단계를 거친다. 비교의 측정 기준은 임무를 해결하는 데 소요된 시간과 정확도, 최근에는 사람의 기능적 자기 공명 영상(FMRI)도 활용한다.



[그림 1] ACT-R의 개괄적인 흐름

자료 : ACT-R homepage에서 재구성, http://act-r.psy.cmu.edu/about/

ACT-R을 활용한 연구결과는 약 700건 이상에 이르며, 대표적인 응용 분야는 인간의 기억, 자연언어처리, 인지 뇌과학, 교육 등 다양하게 활용된다. ACT-R은 개선을 거듭하여 2017년 7월 기준 7.5 버전이 공개돼 있다.[14]

12) 인지 아키텍처는 인간의 사고를 구조화하는 이론을 기계장치에 구현하는 분야
13) 하노이 탑 문제, 문서나 단어의 기억, 언어 이해, 의사 소통, 비행기 조종 시뮬레이션 등
14) ACT-R software, http://act-r.psy.cmu.edu/software/

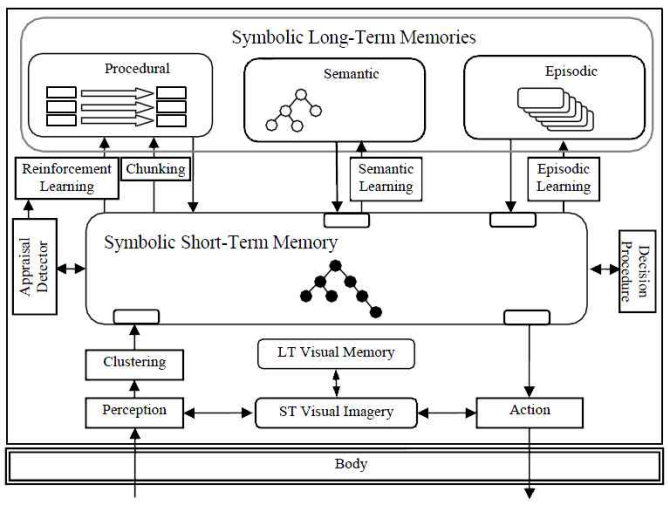
(2) Soar

Soar는 ACT-R과 마찬가지로 대표적인 기호적 범용 인공지능 연구이며 인지 아키텍처이다. Soar는 1983년 처음 연구 프로젝트가 시작됐고 ACT-R과 함께 전통적인 인공지능 연구를 대표한다. Soar는 범용 지능을 갖는 에이전트를 만들기 위해 그 구성 요소기술(building blocks)을 개발하는 것을 목표로 한다. 이 에이전트는 인간의 거의 모든 지적활동을 대체할 수 있는 것으로 의사 결정, 문제 해결, 계획 수립, 자연언어처리 등 광범위한 일을 할 수 있는 범용 인공지능이다. Soar는 ACT-R과 마찬가지로 인간의 인지(cognition)에 대한 이론의 정립과 이를 컴퓨터 공학적으로 구현함에 있다. Soar는 범용 인공지능의 기능과 효율성에 집중을 한 반면, ACT-R은 인간의 인지에 대한 세부적인 모델링에 중점을 뒀다는 것으로 그 차별점을 찾을 수 있다.

Soar는 인간의 인지에 대한 가설 몇 가지를 제시한다. 첫 번째 가설인 문제 공간 가설(Problem Space Hypothesis)은 아무리 복잡한 임무도 세부적으로 분할될 수 있으며, 이것은 간단한 의사결정의 연속(sequence)로 볼 수 있다는 것이다. 두 번째 가설은 간단한 의사결정이 순차적이 아닌 병렬적으로 이루어진다는 점이다. 또한 Soar의 기본 구조는 모듈화 되어있고, 하나의 모듈이 언어나 계획 등 기능적인 지능 활동이라기보다는 장단기 기억, 시각적인 기억 등 기억에 관련된 모듈로 구성돼있다.

[그림 2]는 Soar의 메모리 모듈을 나타낸다. 기간별, 종류별로 기억을 세분화하여 다양한 지능적 행동을 수행하는 것이 가장 큰 특징이다.

Soar의 대표적인 응용 분야는 추론을 요구하는 퍼즐과 게임, 파일럿 시뮬레이션, 자연언어 이해, 로보틱스 등이 있다. Soar 역시 ACT-R과 마찬가지로 인공지능 프로그램을 공개SW 형태로 제공한다.



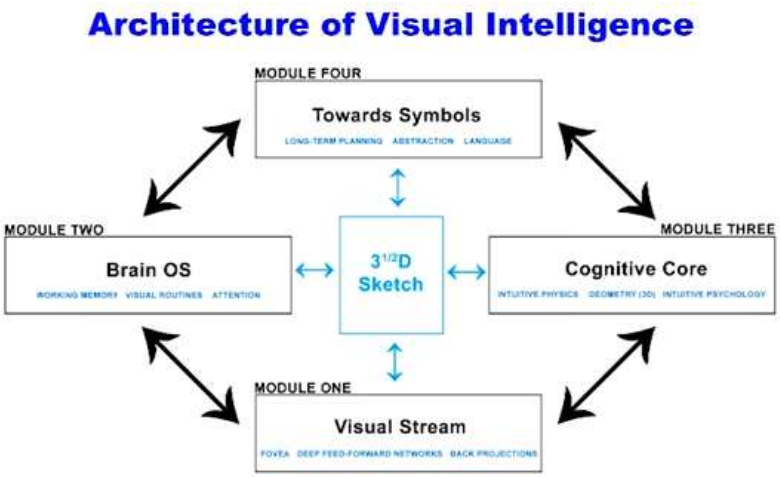
[그림 2] Soar의 메모리 모듈 구조

자료 : Extending the Soar Cognitive Architecture, John E. Laird (2008)

(3) CBMM (Center for Brains Minds+Machines, MIT)[15]

CBMM은 MIT의 연구소로 인간의 뇌구조를 역공학으로 구현하는 연구를 수행한다. CBMM은 궁극적으로 범용 인공지능을 개발하는 것을 목표로 하고 있고, 현재는 시각 지능(Visual Intelligence)에 집중하고 있다. 현재 심층학습 기술은 이미지에서 사물을 인식하는데 매우 뛰어난 성능을 보유하고 있다. 그러나 사람이 가지고 있는 시각적인 지능은 복합적인 의사결정의 결과다. 예를 들어 500명이 앉아있는 강당의 첫 번째 줄에 있는 사람에게 "당신 뒤에 몇 명이나 있나요?" 혹은 "당신에서 가장 가까운 벽은 얼마나 떨어져 있나요?" 에 대한 질문에 대해 뒤나 벽을 보지 않고도 대답할 수 있을 것이다. 이처럼 사람의 시각 지능은 비단 시각적인 정보뿐만 아니라 경험과 직관에 의해 구성된다는 점이다. 따라서 현대의 심층학습은 지능이라기보다 고성능의 패턴 인식 기술이라고 볼 수 있다.

사람처럼 행동할 수 있는 시각 지능을 구현하기 위해 CBMM은 다양한 분야의 전문가와 협업을 하고 있다. 특히 유아의 행동 발달과정을 분석하는 심리학자와 협업하여 시각지능을 구현하기 위한 가설을 세우고 실험을 통해 증명해 나가고 있다. 그 결과 유아의 행동을 역공학적으로 분석하여 아래 [그림 3]과 같은 가설을 세웠다. 이것을 구현하는 방법론으로는 확률적 프로그램(Probabilistic program)으로 소개했다. 확률적 프로그램은 기호적 지능 표현과 확률적 추론, 그리고 심층학습 구조를 융합하여 만든 개념이다.



[그림 3] CBMM의 시각지능 구조 가설

자료 : MIT AGI: Buliding machines that see, learn, and think like people (2018)

15) MIT의 AGI 강의 "MIT AGI: Buliding machines that see, learn, and think like people (Josh Tenenbaum)" 에서 요약

(4) Meta-learning and Self-play[16)]

메타 학습 방법은 기본적으로 심층학습에서 파생되어 나온 분야다. 따라서 큰 의미에서 메타 학습은 창발적 범용 인공지능의 접근 방법으로 해석할 수 있다. 메타 학습의 출발은 강화학습으로부터 시작한다. 강화학습은 AlphaGo의 자체 대국(Self-play)에서도 활용 됐듯이, 행위에 대한 보상으로 전략을 가다듬는 역할을 한다. 강화학습의 가장 큰 관건은 보상에 대한 정의다. 바둑의 경우 승패가 명확하기 때문에 그 보상에 대한 설계도 쉽다. 그러나 사람이 걷는 행위나 물건을 집는 행위는 단순히 맞다 틀리다로 구분 짓기 어렵다. 또한 강화학습은 승리하기 위한 전략에 치중하여 성공을 통해서는 학습할 수 있으나 실패를 통해 학습하기는 쉽지 않다. 그 이유는 성공과 실패의 경우의 수가 같지 않은 경우가 대부분이기 때문이다.

강화학습은 컴퓨터 시뮬레이션을 통해 학습하고, 이것을 물리적인 로봇으로 이식하는 것이 가능하다. 하지만 실제 물리적인 로봇에서는 중력이나 관성 등 다양한 현실적인 모수들이 존재하기 때문에, 실제로 적용하려면 다시 학습해야 하는 어려움이 있다. 이것은 대처하기 위해 다양한 시뮬레이션 환경을 무작위로 부여하여 학습하는 방법론이 소개됐다. 한 가지 흥미로운 연구는 보상을 사람이 직접 개입하여 주는 형태로도 성공적인 학습이 가능하다는 것을 시사했다.

신경망 구조 탐색(Neural Architecture Search)은 최적의 심층신경망의 구조를 도출하기 위해 심층학습을 활용한 기술이다. 쉽게 말하자면 학습하는 방법을 학습하는 기술이다. 신경망 구조 탐색은 [그림 4]와 같이 순환신경망을 통해 파생되는 신경망을 학습하여 최적의 인공신경망 구조를 탐색한다. 인공신경망을 일종의 다변 문자열(variable-length string)로 표현하여 [그림 4]의 controller는 일정 확률 p로 인공신경망 문자열(child network)을 생성한다. 이 문자열로 구성된 인공신경망은 학습 데이터로 학습되고, 그 정확도를 되먹임(Feedback)하여 최적의 신경망을 찾는다. 신경망 구조 탐색은 심층신경망 구조 자체가 미지수이기 때문에 매우 높은 자유도(degree of freedom)를 갖고 있다. 이로 인해 상당한 수준의 계산 자원을 필요로 한다. 최근 연구결과에서는 고성능 그래픽카드 한 장으로도 높은 성능을 이끌어 냈다.[17)]

16) MIT의 AGI 강의 “OpenAI Meta-Learning and Self-Play (Ilya Sutskever)” 에서 요약
17) Pham, Hieu, et al. “Efficient Neural Architecture Search via Parameter Sharing.“ arXiv preprint arXiv:1802.03268 (2018).

[그림 4] 신경망 구조 탐색의 개념

자료 : Neural architecture search with reinforcement learning에서 재구성

강화학습을 비롯한 심층학습이 범용 인공지능으로 가기 위해서는 해결해 할 숙제는 크게 두 가지가 있다. 먼저 학습과 시험용 데이터의 본질적인 차이다. 우리가 지금까지 측정한 심층학습의 성능은 학습 및 시험용 데이터가 거의 동일한 구조를 가지고 있다는 것을 가정하기 때문이다. 위의 로봇 시뮬레이션 예시에서도 알 수 있듯이, 실제 시험용 데이터는 학습용 데이터와 상당히 다를 것이다. 두 번째 숙제는 현재 심층학습 기술은 지속적인 학습이 거의 불가능하다는 점이다. 이미 학습된 인공신경망에 데이터를 추가적으로 학습하기 위해서는 상당히 많은 데이터가 필요하기 때문에, 이를 극복하기 위한 노력이 필요하다는 것이다.

4. 결 론

지금까지 범용 인공지능의 개념과 접근방법, 연구 현황을 살펴봤다. 현재 범용 인공지능 연구와 관련된 문헌이나 활동을 종합해보면 아직 범용 인공지능의 출현은 먼 미래로 보인다. 다만, 심층학습을 필두로 한 인공지능의 성능은 눈부시게 발전하고 있다는 것은 사실이다.

그러나 전통적인 기호적 인공지능의 연구 사례를 살펴보면 심층학습과는 전혀 다른 접근 방법을 취하고 있다. 그렇기 때문에 과연 심층학습의 발전이 범용 인공지능과 상관관계가 있을까?라는 의문을 품게 된다. 전문가들은 기호적 인공지능과 심층학습의 유기적인 결합을 강조한다. 각 분야에 특화된 임무가 서로 다를 수 있기 때문이다. 한편으로는 뇌를 역공학으로 분석하여 구현하는 창발적 범용 인공지능 연구가 진행되고 있다. 이 접근 방법은 유아의 행동 발달 과정이나 뇌의 네오코텍스(neo-cortex), 혹은 뇌 전체를 슈퍼컴퓨터로 구현하는 것을 목표로 한다.

많은 전문가들이 강조했듯이 범용 인공지능이 출현한다면 그것은 인간의 수준을 쉽게 뛰어넘는 초인공지능이 될 것이다. 이 사실은 사람과 인공지능이 더 이상 주종관계가 아니고 공존하는 관계여야 함을 시사한다. 이에 많은 인공지능 연구자들은 미래 인공지능의 윤리와 정책에 대해서도 심도 있는 논의를 진행하고 있다. 또한, 인공지능은 반드시 사람에게 이롭게 활용돼야 한다는 대 원칙아래, 전 세계 인공지능 연구자는 인공지능을 이롭게 활용하기 위한 23가지 원칙을 정립했다. 이러한 움직임은 그만큼 범용 인공지능에 대한 파급력이 크다는 사실을 반증한다.

우리도 인공지능을 이롭게 활용해야 한다는 대 전제에 동참하여 앞으로의 인공지능 R&D 방향을 설정해야 할 것이다. 특히 인공지능에 대한 윤리는 본격적인 연구가 필요하다고 본다. 또한 현재 인공지능은 그 활용 측면에서 산업계의 확산이 매우 빠르게 일어나고 있다. 인공지능을 통한 산업 진흥도 중요한 분야지만, 더 막대한 파급력을 갖고 있는 범용 인공지능에 대한 투자도 아끼지 말아야 한다고 본다.

[별 첨] 범용 지능에 대한 다양한 해석

앞서 범용 인공지능은 사람 수준의 지능이라고 소개했다. 그렇다면 사람 수준의 지능이라는 실체는 무엇일까? 사람의 일반적인 지능적 행동을 지칭하는 단어로는 범용 지능(General Intelligence)이 있다. 이 범용 지능을 구체화하기 위해 다양한 분야의 연구자들은 자신들의 해석을 밝혔다.[18)]

먼저 실용적 관점(Pragmatic Approach)이 있다. 이 관점의 대표적인 예는 1950년 앨런 튜링이 제안한 '튜링 테스트'이다. 튜링 테스트는 사람과 인공지능의 대화에서 누가 사람인지 인식할 수 있는지의 여부를 판단하는 것이다. 이것을 확대해서 보면 사람의 지능과 인공지능의 기능적 측면을 비교하여 범용 인공지능을 판단하는 접근이다. 다시 말하자면, 사람의 지능적 행동에 방해될 만큼 인공지능 기술이 발전했다면 이것이 곧 범용 지능이라는 관점이다.

두 번째는 심리학적 관점이다. 범용 지능의 심리학적 관점은 실용적 관점보다는 더 원론적인 접근을 취한다. 심리학 관점에서는 범용 지능을 크게 두 가지 원칙으로 해석한다. 먼저 개인이 형성한 도메인 지식(Knowledge domains)은 서로 상관성이 있기 때문에, 해당 도메인별 지식의 수준(level)이 크게 다르지 않다. 이것을 개인 내 다양성(intra-individual variability)이라고 한다. 두 번째 원칙은 개인 간의 도메인 지식은 상당한 수준으로 상이할 수 있다는 것으로, 개인 간 다양성(inter-individual variability)라고 지칭한다. 심리학자 가드너(Gardner)는 이것을 바탕으로 인간의 지능에 대해 8가지[19)]로 구분 지었다. 심리학적 관점에서 인간 수준의 범용 지능 논의하기 개최된 2009년 AGI Roadmap workshop에서는 14가지[20)]의 분류체계를 제시했다.

다음은 인지 아키텍처(cognitive architecture) 관점의 범용 지능이다. 인지 아키텍처는 SW의 속성이 강하다. 따라서 인지 아키텍처 연구자들은 범용 지능을

18) 『Artificial General Intelligence: Concept, State of the Art, and Future Prospects』에서 소개된 내용을 정리

19) ① 언어적(linguistic), ② 논리적·수학적(logical-mathematical), ③ 음악적(musical), ④ 운동감각적(bodily kinesthetic), ⑤ 공간적(spatial), ⑥ 대인관계(interpersonal), ⑦ 자기 내부적(intrapersonal), ⑧ 자연주의적(nautralist) 지능, 다중 지능이론, Gardner(1999)

20) 인지(perception), 구동(actuation), 기억(memory), 학습(learning), 추론(reasoning), 계획(planning), 주목(attention), 동기(motivation), 감정(emotion), 모델링(modeling self and other), 사회적 상호작용(social interaction), 소통(communication), 정량적(quantitative), 창조(building/creation)

구현하기 위해 SW적인 관점에서의 요구사항을 논의한다. 예를 들면, 다양한 작업을 수행함에 있어 SW 구조는 고정되어야 한다는 점이다. 새로운 작업을 수행하기 위한 논리나 알고리즘이 SW상에 구현돼 있어 그 구조는 동일해야함을 의미하는 것이다.

네 번째는 수학적 방법이다. 수학적 방법의 주된 가정은 범용 지능을 구현하기 위해서 무한대의 계산이 필요하다는 점이다. 이것은 제한된 계산 자원으로는 범용 지능을 달성할 수 없으나, 특정 시스템은 다른 시스템보다 더 지능적일 가능성이 있다고 해석할 수 있다. 이 가정은 사람 역시 완벽한 범용 지능이라고 볼 수 없는 모순에 직면한다. 그러나 사람은 다른 시스템(곤충, 파충류 등)보다 더 지능적이라고 볼 수 있기 때문에 수학적 접근으로 지능을 해석하는 것이 일견 타당한 면이 있다.

마지막은 주변 환경과 밀접하게 관련된 적응주의적(Adaptationist) 관점에서 범용 지능을 접근하는 방법이다. 인공지능 연구자인 Pei Wang은 범용 지능을 '제한된 자원을 활용해 환경에 적응하는 것'으로 개념화했다. 지금까지 범용 지능을 특징화한 다섯 가지 접근에 대해서 살펴봤다. 이것을 정리하면 다음 <표 2>와 같다.

<표 2> 범용 지능을 특징화하기 위한 다양한 접근방법

접근 방법	내 용
실용적 관점	- 사람의 지능과 인공지능의 기능적 측면을 비교하는 접근 - 대표적인 예 : 튜링 테스트
심리학적 관점	- 지능의 해석 : 개인 내 다양성, 개인 간 다양성 - 인간의 지능 분류 : Gardner의 8가지 분류
인지 아키텍처 관점	- SW의 관점에서 범용 지능의 해석 - 다양한 작업을 수행함에 있어 SW구조는 고정돼야 함
수학적 관점	- 범용 지능을 구현하기 위해서는 무한대의 계산 자원이 필요 - 제한된 계산 자원으로는 상대적으로 우수한 범용 지능을 구현
적응주의적 관점	- 제한된 자원을 활용해 환경에 적응한다는 접근

자료 : Artificial General Intelligence: Concept, State of the Art, and Future Prospects, Ben Goertzel (2017)

[참고문헌]

1. 국외문헌

Goertzel, Ben. “Artificial general intelligence: concept, state of the art, and future prospects.“ Journal of Artificial General Intelligence 5.1 (2014): 1-48.

Grace, Katja, et al. “When will AI exceed human performance? Evidence from AI experts.“ arXiv preprint arXiv:1705.08807 (2017).

Kurzweil, Ray. The singularity is near. Gerald Duckworth & Co, 2010.

Laird, John E. “Extending the Soar cognitive architecture.“ Frontiers in Artificial Intelligence and Applications 171 (2008): 224.

Pham, Hieu, et al. “Efficient Neural Architecture Search via Parameter Sharing.“ arXiv preprint arXiv:1802.03268 (2018).

Zoph, Barret, and Quoc V. Le. “Neural architecture search with reinforcement learning.“ arXiv preprint arXiv:1611.01578 (2016).

2. 국내문헌

AlphaGo의 인공지능 알고리즘 분석, 이슈리포트 2016-002, 소프트웨어정책연구소 (2016)

AlphaGo Zero의 인공지능 알고리즘, 이슈리포트 2017-009, 소프트웨어정책연구소 (2017)

3. 기 타

ACT-R software, http://act-r.psy.cmu.edu/software/

How to get empowered not overpowered by AI, Max Tegmark, TED talk (2018)

MIT AGI Lectures (2018), https://agi.mit.edu/

아실로마 AI원칙, Future of Life Institute (2017), https://futureoflife.org/ai-principles-korean/

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.