

“인공지능은 인문학이다”:

구성적 정보 철학적 관점에서*

박 충 식**

주제분류 인공지능, 인문학, 정보 철학, 구성주의

주요어 인공지능, 인문학, 정보 철학, 구성주의

요약문

인공지능은 기계가 인간 지능을 가지도록 모색하는 컴퓨터과학의 한 분야이다. 이를 위하여 인간의 지능을 계산적 모델로 이해하기 위한 연구 분야이기도 한다. 그러므로 인공지능은 인간을 이해하기 위한 오랜 학문인 문사철의 인문학과 밀접한 관계가 있다. 더구나 인지과학의 발달로 인하여 인간에 대한 이해는 새로운 국면에 들어섰다고 할 수 있다. 인지과학의 한 분야이기도 한 인공지능은 포괄적으로 인간에 대한 이해를 도모하는 또 다른 측면의 인문학이라고 할 수 있다.

필자는 현재의 인공지능이 아직은 인문학적 인간 이해와는 많은 거리를 가지고 있지만 좀 더 발전된 인공지능을 위해서 인공지능은 인문학에 대한 이해가 필수적이고, 인문학 또한 좀 더 정교한 인간 이해를 위해서는 인공지능에 대한 이해와 인공지능적 상상력이 필요하다고 생각한다. 본 연구는 구성주의에 기반한 정보 철학이 인공지능과 인문학을 포괄적으로 이해하기 위한 방안이 될 수 있다고 논의한다.

* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A6A3A01078538)

** 유원대학교

1. 들어가기¹⁾

통상적으로 인공지능은 기계가 인간 지능을 가지도록 모색하는 컴퓨터 과학의 한 분야로 정의된다. 이를 위하여 인간 지능에 대한 이해를 필요로 하고 이해된 결과를 계산적 모델(computational model)로 만들려고 한다. 그러므로 인공지능은 인간을 이해하기 위한 오랜 학문인 문사철의 인문학과 밀접한 관계가 있다. 더구나 위키피디아에서 “인간의 마음과 동물 및 인공적 지적 시스템(artificial intelligent systems)에서 정보처리가 어떻게 일어나는가를 연구하는 학문”으로 정의하는 인지과학은 심리학, 철학, 신경과학, 언어학, 인류학, 전산학, 학습과학, 교육학, 사회학, 생물학, 로봇틱스 등의 여러 학문과 연관되어 있다. 이러한 인지과학의 발달로 인하여 인간에 대한 이해는 새로운 국면에 들어섰다고 할 수 있다. 그런 점에서 인공지능은 인간에 대한 이해를 도모하는 또 다른 측면의 인문학이라고 할 수 있다.

현재의 인공지능이 아직은 인문학적 인간 이해와는 많은 거리를 가지고 있다. 하지만 좀 더 발전된 인공지능을 위해서 인공지능은 인문학에 대한 이해가 필수적이고, 인문학 또한 좀 더 정교한 인간 이해를 위해서는 인공지능에 대한 이해와 인공지능적 상상력이 필요하다. 본 연구는

1) 도발적인 주제를 긍정적으로 보아 주신 심사위원들께 감사드립니다. 그리고 거친 문장에 대하여 친절하게 대안을 제시해주시고 인문학과 인공지능의 융합 연구 주제에 대하여 하위목차로 구분할 것을 제안해주신 심사 의견에도 감사드리며 대부분 수용하였습니다. 또한 “인공지능은 인문학이다”라는 논문 제목이 논리적으로 바르지 못하므로 인공지능의 인문학적 성격이 드러나는 제목으로 변경할 것을 조언해주신 심사위원의 의견에도 감사드립니다. 다만 이런 비논리적인 제목이 철학 논문의 제목으로써 다소 적절치 않을 수 있지만 주제를 강조하기 위한 수사적인 표현이라고 이해하셨으면 하는 바램입니다. 조언의 타당성을 수용하는 바, 논문의 제목이 수사학적인 표현이라는 점을 강조하기 위하여 따옴표 처리하였습니다. 과학주의적인 인공지능이 인간의 모든 것을 밝혀줄 것이라는 성급한 결론을 경계해주신 논평에 대해서 심분 공감하며 현재의 인공지능이 아니라 인문학적 성찰로 좀 더 나은 미래의 인공지능이 되기 위한 노력의 일환으로 보아주시기를 바라는 마음입니다.

구성주의에 기반한 정보 철학이 인공지능과 인문학을 포괄적으로 이해하기 위한 방편이 될 수 있다고 논의한다. 먼저 우리 사회에 인공지능에 대한 담론을 촉발했던 딥러닝의 원리와 그와 연관된 설명가능 인공지능의 문제를 살펴보고, 설명가능한 인공지능의 핵심이라고 여겨지는 심볼 그라운딩을 알아본다. 다음으로 구성적 정보 철학의 개념과 심볼 그라운드, 그리고 언어와의 관계를 살펴봄으로써 물리주의의 인공지능과 인문학의 연결고리를 생각해본다. 마지막으로 구성적 정보 철학적 관점에서 인공지능과 인문학의 융합적인 연구 주제의 몇 가지를 모색하고자 한다.

2. 인공지능의 화려한 부활?: 딥러닝

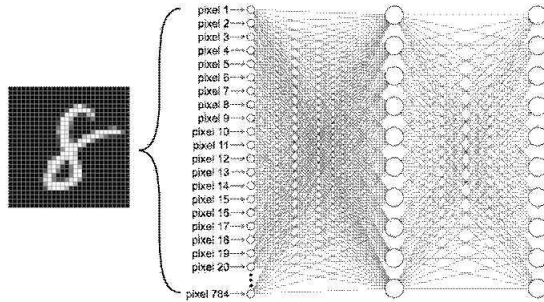
1956년 다트머스 인공지능 하계 연구 학술대회에서 ‘인공지능’이라는 단어가 등장한 이후 이토록 많은 사람들이 이 단어를 알게 된 적은 없었다²⁾. 2016년 온 국민의 전례없는 관심 속에 이세돌과 알파고의 바둑 대결은 이루어졌고 소수 일부를 제외한 많은 인공지능 전문가의 예측과는 반대로 알파고의 승리로 마무리되면서 ‘인공지능’이라는 단어는 유행어가 되어 버렸다. 더불어 ‘구글’과 ‘알파고’라는 상업적 용어뒤에 인공지능 기술용어인 ‘딥러닝’도 알려지고 이것이 알파고 승리의 핵심이라는 것도 알려지게 되었다. 세상 사람들의 들끓는 관심과 언론의 부추김 속에 딥러닝은 인간 기술 발전의 총아로 떠올랐고 드디어 인간과 같은, 나아가 인간을 능가하는 인공지능의 발전 전망과 인공지능으로 인한 실업문제와 교육문제, 나아가 인공지능에 의한 인간 지배라는 SF의 단골 시나리오까지 진지하게 등장하였다.

사실 딥러닝은 길지 않은 인공지능의 역사에서 과거에 있던 인공신경망이라고 불리는 아이디어가 몇 번의 성공과 실패를 거치면서 최근 정보통신기술의 발전에 힘입어 빠르고 많은 컴퓨터들과 많은 양의 데이터를

2) 박충식, 「2장 거꾸로 보는 인공지능의 역사」, 『인공지능의 이론과 실제(포스트휴먼 사이언스 총서 6)』, 서울: 아카넷, 2019, 31쪽

수집하고 저장하고 처리할 수 있게 되고 이를 잘 이용할 수 있도록 하는 인공지능 연구자의 아이디어에 의하여 가능하게 된 것이다. 딥러닝을 실제로 만들기 위해서는 많은 자금과 복잡하고 귀찮은 과정이 필요하지만 그 원리는 비교적 간단하다고 할 수 있다.

딥러닝은 인공지능 분야에서 가장 유명한 교과서인 『인공지능 현대적 접근』에도 아직 등장하지 않은 용어이고, 딥러닝 기술의 모태가 되는 인공신경망 기술에 대한 설명도 많지 않을 정도이므로 원리적으로 복잡한 기술은 아니다. 인공신경망의 원리는 다음과 같이 간략히 살펴볼 수 있다.



<그림 1> 인공신경망 숫자인식³⁾

숫자인식을 위한 인공신경망의 경우(<그림 1>), 가로-세로 28개씩 784개의 점으로 이루어진 0부터 9까지의 숫자 이미지들을 입력으로 하고 해당 숫자가 입력으로 주어졌을 때 출력층의 해당 숫자 출력이 1이 되도록 입력층, 은닉층, 출력층의 노드에 연결된 가중치를 조정하는 것으로 숫자인식이 이루어지게 된다. 입력층의 숫자 이미지와 해당 숫자 출력이 주어졌을 때 출력층의 결과와 기대하는 출력층의 결과의 차이가 없어지도록 수많은 숫자 이미지를 입력으로 제공하고 연결 가중치를 조정하는 반복 작업을 학습이라고 하며, 이러한 학습이 성공적으로 끝나면 임의의 숫자 이미지가 제시되어 해당 숫자의 출력층 결과가 나타나게

3) Looking inside neural nets, https://ml4a.github.io/ml4a/looking_inside_neural_nets/

된다.

딥러닝은 이런 인공신경망에 엄청난 양의 입출력 학습 자료를 제공하고 빠르게 처리하면서도 은닉층을 계층적으로 처리할 수 있도록 여러 층으로 늘리고, 은닉층의 개수를 자동으로 조정하는 등의 부가적인 은닉층을 아주 여러 층으로, 즉 “깊게(deep)”함으로써 딥러닝이라는 이름이 붙게 된 것이다. 하지만 이러한 방법이 항상 성공하는 것이 아니기 때문에 여러 번 조금씩 학습 알고리즘의 실행을 조정할 필요가 있다. 수식을 쓰지 않고 설명한다면 이것이 인공 신경망 알고리즘에 대하여 알아야 하는 전부이다. 하지만 누구라도 인공 신경망 알고리즘이 이루어지는 간단한 원리만 이해한다면 이러한 방법으로 딥러닝이 인간과 같은 지능을 가지게 될 것 같지는 않다는 것을 쉽게 알 수 있다. 인간의 두뇌가 신경세포들로 이루어져 있기는 하지만 이러한 신경세포들은 “지능”이라는 현상을 위하여 아주 복잡한 기능을 하는 많은 덩어리(모듈)들로 이루어져 있다. 딥러닝은 그러한 단하나의 아주 단순한 덩어리조차 충분히 구현하지 못하고 있는 것이다.

하지만 딥러닝을 사용하는 알파고가 이세돌과의 바둑에서 승리하였으나 왜 그 자리에 바둑돌을 놓았는지 설명해줄 수는 없다. 그 이유는 바둑 착점의 모든 내용들이 딥러닝의 숫자로 표시되어 있기 때문이다. 현재 가장 각광받는 딥러닝을 사용하는 모든 인공지능 시스템은 모두 같은 사정이다.

인간은 모든 내려진 결정을 이해하기 원하지만 딥러닝은 작동원리상 그렇게 할 수 없다. 바둑의 착점이나 사진속의 대상이 무엇인지는 그냥 그렇다고 하고 넘어갈 수도 있지만 자신의 해고가, 수백만불의 투자가, 핵무기의 발사가 아무 설명없이 결정되고 실행된다면 사람들은 그냥 그렇다고 하고 넘어갈 수 있을까? 이러한 문제의식은 인공신경망 연구 초기부터 제기되었던 문제이지만 이제 인공신경망에서 발전한 딥러닝 기술이 놀라운 성과를 보이게 되면서 이 문제가 다시 제기되었고 의사결정과정을 드러나게 할 수 있는 설명가능한 인공지능(XAI: eXplainable

AI)에 대한 관심이 활발히 제기 되고 있는 실정이다.

인간의 모든 의사결정도 신경세포수준에서 설명할 수 없으며 의사결정 이후 설명을 만들어내는 것이라는 인지과학적 연구도 있으려니와 자신의 의사결정 과정과 설명이 일치하지 않을 수 있다. 설명 또한 학습해야 할 또 다른 능력이라고 할 수 있다. 딥러닝이 자신의 의사결정을 인간에게 설명하려면 먼저 자신의 의사결정 과정을 스스로 관찰할 수 있어야 하고 관찰의 대상들과 대상들의 관계를 개념화할 수 있어야 한다. 또한 그 개념화가 인간의 개념화와 어느 정도 유사해야만 하고 의사소통할 수 있는 수단을 가져야만 한다. 그것이 바로 인간에게는 언어이다. 인간의 언어는 대상과 대상들의 관계를 표현할 수 있도록 음성적이든 문자적이든 심볼을 사용한다. 인공지능에서 심볼이 대상에 대하여 의미를 가지게 되는 과정을 심볼 그라운드(symbol grounding)이라고 한다⁴⁾. 결국 딥러닝이 인간에게 자신의 의사결정과정을 설명하려면 인간의 언어를 가져야 하고 그 관건은 심볼 그라운드인 것이다.

3. 심볼 그라운드

물리학이나 생물학이 설명하는 인간과 그것으로 만족하지 못하는 인문학은 어떻게 인간을 이해하고 싶어하는 것일까? 인간을 포함한 세상은 물리적 토대에서 존재할 수밖에 없다고 생각하는 필자와 같은 물리주의자들은 소위 자연과학적으로 설명할 수 있다는 신념을 가지고 있다. 하지만 풍요롭다 못해 너무나 복잡한 인간의 정신과 사회를 생각할 때 인간을 한갓 물질이라거나 한갓 생물이라는 이해를 받아드릴 수 없다. 그렇다고 인간의 정신과 사회가 현재 과학적 발전이 보여주듯이 물리적 토대에 생물학적 진화에 의하여 이루어졌다는 사실을 부인하기도 어렵다.

물리학자 카를로 로벨리는 의미나 지향성과 같은 개념들을 ‘의미있는

4) 박충식, 「구성적 인공지능」, 『인지과학』 15(4), 2004, 65쪽, Harnad, S., “The Symbol Grounding Problem”, *Physica D* 42, 1990, 338쪽

정보’라는, 순전히 물리적인 정의로부터 이끌어 내기를 원한다. 단순화하여 설명한다면 로벨리는 ‘의미 = 정보 + 진화’라는 전략아래 두 개체간의 연결 중단이 열적 평형을 낮출 확률이 커진다면 서로에게 ‘의미있는 정보’라고 정의한다. 로벨리는 자연을 기본적인 법칙을 따르는 기본적인 구성요소들의 앙상블에 의하여 형성되는 것이고 이를 ‘창발적’이라고 부르며 ‘의미있는 세계’에 대한 우리의 서로 다른 수준의 이해에 연결을 제공할 수 있다고 제안한다. 이러한 창발, 즉 물리적 창발과 생물적 창발 위에 물리적 존재만도 아니고 생물적 존재만도 아닌 인간은 단연코 ‘언어적 존재’라는 것이다.

인간이 언어적 존재가 되는데 핵심적 개념이 심볼 그라운드이며 딥러닝이 인간에게 자신의 의사결정 과정을 설명하도록 만드는 핵심적인 개념이 심볼 그라운드이라는 것이다. 물론 심볼 그라운드가 가능한 인공지능은 설명이상의 것을 할 수 있게 될 것이고 인간에 대한 인문학적 이해도 진일보할 것이다.

심볼 그라운드는 관찰자가 대상을 지칭하여 의미를 가지게 되는 것을 의미하기 때문에 인간이 아닌 생물적 체계들도 정보처리 수준에 따라 심볼 그라운딩을 할 수 있다. 이러한 생물적 체계들 중에 특정 생물체들은 공동사회를 이루면서 의사소통을 할 수 있기도 하다. 전자를 1차 심볼 그라운딩(first symbol grounding)이라고 하고, 후자를 2차 심볼 그라운딩(second symbol grounding)이라고 할 수 있을 것이다. 이러한 2차 심볼 그라운딩이 심볼로 외재화될 수 있으면 3차 심볼 그라운딩(third symbol grounding)이라고 구별할 수 있다. 이러한 3차 심볼 그라운딩이 바로 인간의 언어이다.

생명체의 모든 앎은 원초적으로 감각기관을 통하여 이루어지고, 환경의 모든 대상에 대하여 감각기관이 만들어낸 차이에 의해서만 감지할 수 있다. 모든 개체는 허용된 심볼 그라운딩만을 할 수 있기 때문에 스스로에게 단혀 있다고 할 수 있다. 모든 생물들이 이러한 정보를 처리하지만 인간만이 현재와 같은 언어적인 심볼 그라운딩을 할 수 있게 된 것은

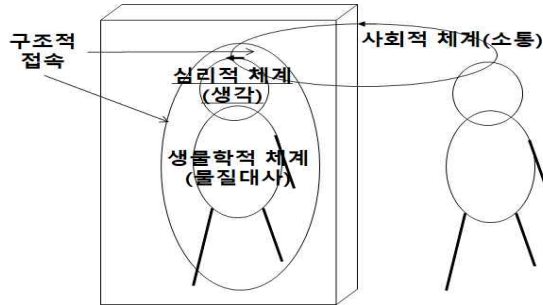
“우연과 필연”으로 설명되는 오직 “진화”에 의한 것이다. 다른 말로 하면 구성된 것이라고 할 수 있으며, 이러한 구성은 특정 생물 개체의 당대에 만 이루어지는 “학습”뿐만 아니라 오랜 세월을 거친 “진화”에 의하여 유전적으로 각인된다.

심볼 그라운드할 수 있는 생물들은 같은 심볼 그라운딩을 할 수 있는 다른 생물들과 제한된 교류를 통하여 공적인 심볼 체계, 즉 문화를 형성하게 된다. 그 문화는 다시 심볼화하는 체계에 영향을 미치는데, 심지어 해당 생물이 욕망할 수 있는 것과 욕망할 수 없는 것에 영향을 미치고, 이러한 반복은 계속된다.

심볼 그라운딩은 개체 안에서 일어나는 현상이지만 모여 있는 개체 사이에 일어나는 사회적 현상이기도 하다. 니클라스 루만은 이러한 사회를 이해하기 위해서 ‘사회는 인간들로 이루어진다’는 상식적인 통념을 거부하고 사회는 인간들에 의하여 행해지는 소통에 의하여 자기생산적으로 이루어진다는 테제를 전면에 내세운다.

마투라나와 바펠라에 의하여 연원한 생물과 인지의 자기생산체계는 루만에 이르러 자기생산적 사회체계이론으로 발전한다. 자기생산체계는 환경과의 차이로부터 생기는 것이고 사물들이 아니라 작동들로 이루어진다. 이러한 체계들은 생물적 체계(living systems), 심리적 체계(psychic systems), 사회적 체계(social systems)로 구별할 수 있고, 사회적 체계에는 사회(societies), 조직(organizations), 상호작용(interactions)이 있다. 그리고 사회적 체계는 근대에 오면서 경제 체계, 정치 체계, 법 체계, 학문 체계, 예술 체계, 등으로 분화되었다고 본다. 인간에게는 생물적 체계와 심리적 체계가 존재하고 두 체계가 서로 영향을 주고받기는 하지만(구조적 접속) 두 체계는 엄연히 다른 체계이다. 그리고 심리적 체계와 사회적 체계, 또한 두 체계가 서로 영향을 주고받기는 하지만(구조적 접속) 심리적 체계가 사회적 체계의 구성요소는 아니다. 심리적 체계는 생각과 생각에 의하여 구성되고 사회적 체계는 심리적 체계가 아니고 심리적 체계 사이에 일어나는 소통과 소통에 의하여 구성된다. 하지만 심리적 체계와

사회적 체계 공히 “의미”라는 매체에 의하여 작동된다.



<그림2> 3개의 자기생산체계 모델⁵⁾

3차 심볼 그라운딩은 언어에 의한 사회적 현상이다. 언어는 루만의 사회체계이론에서 사회적 체계를 구성하는 소통의 매체로서의 역할을 수행한다. 언어의 심볼 그라운딩은 생각에 의하여 작동하는 심리적 체계와 심리적 체계에 의하여 발생하는 소통에 의하여 작동하는 사회적 체계 사이에 서로 영향을 주는 ‘의미’라는 구조적 접속에 의하여 발생하는 것이라고 할 수 있을 것이다.

4. 구성적 정보 철학

구성적 정보 철학은 마투라나와 바렐라, 그리고 폰 피르스터의 급진적 구성주의와 루만의 작동적 구성주의에 기반한 정보 철학이다. 정보 철학은 이 용어를 처음 사용한 플로리디(Floridi)는 “정보의 개념적 성질과 원칙, 그리고 그의 다이내믹스, 활용과 과학을 포함하는 중요한 연구, 그리고 정보-이론적 그리고 계산적 방법론들을 철학적 문제들에 적용하고 정교화하는 새로운 철학분야로 정의한다”. 플로리디는 정보를 물리적 신호

5) Brier, Søren, “Levels of Cybersemiotics : Possible Ontologies of Signification”. *Cognitive Semiotics* 4, 2009, 46쪽.

의 패턴(참이든 거짓이든), 환경적 정보와 같은 실재로서의 정보(as reality), 진리적으로 자격이 있는 의미적 정보의 실재에 대한 정보(about reality), 그리고 유전정보, 알고리즘, 명령들, 요리법과 같은 명령어들과 같은 실재를 위한 정보(for reality)로 구별하였다⁶⁾. 이러한 정보에 대한 구별은 사용으로서의 정보의 다양한 개념을 보여줄 수는 있지만 정보에 대한 명확한 정의를 내려주지 못하고 그로 인하여 정보의 개념을 혼란스럽게 하는 면이 있다.

구성주의에 기반한 구성적 정보 철학은 ‘관찰자의 내부에 구성된 정보’만을 ‘정보’로 정의한다. 이러한 관점은 퍼스의 기호학에서 기호 삼각형을 염두에 둔 것이다. 기호 삼각형은 기호가 지칭하는 대상체, 기호 자체인 표현체, 관찰자가 이해한 해석체로 이루어진다. 퍼스는 기호 자체보다는 기호가 의미를 가지게 되는 과정, 즉 기호작용(semiosis)을 중요하게 생각한다. 그러므로 관찰자가 이해한 해석체는 또 다른 대상체가 될 수 있고 이를 대상으로 다시 표현체(기호)와 해석체의 또 다른 기호 삼각형이 이루어진다.

이런 관점에서 플로리디의 정보 구별을 생각해보면 실재로서의 정보는 기호의 표현체를 이루는 매체, 실재에 대한 정보는 해석체, 그리고 유전정보, 알고리즘, 명령들, 요리법과 같은 명령어들과 같은 실재를 위한 정보는 매체로 구성된 표현체로 간주할 수 있다.

구성적 정보 철학에서 관찰자에게 구성될 수 있는 정보는 “차이를 만드는 차이”로써 관찰자가 구별할 수 있는, 허락된 감각적 기능에 의해서만, 그리고 관찰자의 욕망(가치)에 관여되는 한에서만, 또한 관찰자에게 형성된 기존의 정보와의 관계 하에서 포착할 수 있는 범위 내에서만 가능하다. 이것을 마투라나와 바렐라는 자기생산체계(autoietis system)의 ‘인지적 폐쇄(cognitive closure)’라고 하였다. 필자는 퍼스의 기호작용과 마투라나의 바렐라의 인지적 자기생산체계가 인공지능에 가장 중요한 문

6) 박충식, 「빅 히스토리와 인공지능: 정보적 관점에서」, 『인간·환경·미래』 2, 2019, 39쪽.

제라고 생각하는 심볼 그라운딩의 중요한 실마리를 제공하며 3차 심볼 그라운딩은 바로 언어이고 인공지능과 인문학을 연결하는 핵심개념이라고 생각한다.

표현체로 정보의 외재화가 가능한 심볼 그라운딩이 인간의 언어이기 때문에 인간의 언어는 사회적 현상이기도 하지만 이 언어의 심볼 그라운딩은 여기에 그치지 않는다. 관찰자는 외부 환경을 관찰하기도 하지만 관찰자는 역시 자신이 의식하는 생각들을 관찰하고 그 차이를 구별하고 지칭할 수 있기 때문에 퍼스의 기호 작용에서 언급했듯이 해석체가 다시 대상체가 될 수 있다. 이것을 인문학에서는 ‘성찰적’ 또는 ‘반성적’이라고 한다. 철학에서 ‘성찰’은 ‘반성’과 같이 쓰이는 것으로 보이는데 칸트적 의미에서 반성을 다음과 같이 풀이한다. “일반적으로 이성이 “반성(성찰)한다는 것은, 주어진 표상들을 다른 표상들과 또는 자기의 인식능력과, 그에 의해 가능한 개념과 관련해서, 비교하고 대조하는 것이다”⁷⁾.

현재 딥러닝은 성찰적이지 않다. 알파고의 경우, 이세돌과의 그 이후 세계최고수 바둑기사들과의 대국에서 놀라운 승리를 보여주었지만 알파고의 한수, 한수가 어떤 이유에서 그렇게 되었는지에 대해서는 알파고 스스로도 알파고를 만든 공학자들도 아무 설명을 해주지 않는다. 설명을 해주지 않는 것이 아니고 해줄 수 가 없는 것이다. 인공지능경망의 학습 알고리즘은 특정한 구조를 가진 변수들을 일정한 규칙에 따라 변경하는 프로그램으로 이루어져 있기 때문에 학습 프로그램을 작성한 공학자들도 학습의 방법만을 만든 것이지 학습의 내용을 알지는 못한다. 알파고도 자신의 프로그램을 수행하기는 하지만 자신의 학습의 내용을 이해하고 그 학습의 내용을 설명할 수 없는 것이다.

자신의 결과를 타인에게 설명하는 것은 자신이 무엇을 어떻게 생각해서 그런 결과를 내놓게 되었는지 기술하는 것으로 ‘성찰’이나 ‘반성’의 의미를 볼 때 전형적인 ‘성찰적’ 능력이다. ‘설명’이라는 개념도 인문학적으로 볼 때 매우 다양한 의미와 복잡한 현상이지만 이 ‘설명’도 ‘성찰’과

7) 백중현, 『이성의 역사』 서울: 아카넷, 2017, 525쪽

관련된 한 가지 사례일 뿐이다. ‘성찰’은 진정으로 지능적인 또는 인간을 인간답게 하는 면모라고 할 수 있다.

구성적 정보 철학의 입장에서 볼 때 성찰적 과정은 해석체인 정보를 처리하는 과정이며 이미 알고 있는 정보들에 의해서만 이루어질 수밖에 없다. 이미 알고 있는 정보들은 환경에 적응하고 살아가기 위한 세상에 대한 모델로서 세상에 대해 부분적일 수밖에 없고 알고 있는 정보들로 이루어진 세상 모델조차 일관적이지 않고 모순된 상태로 존재하게 된다. 이에 대한 성찰적 작업은 자신이 가지고 있는 세상 모델을 정교하게 하는 일이 된다. 이러한 성찰적 과정이 루만이 말하는 관찰에 대한 관찰, 즉 2차 관찰이고 계속적인 관찰에 대한 관찰을 통하여 자신을 포함한 좀 더 나은 세상 모델을 가질 수 있게 되지만 결코 관찰자의 자기지시적 맹점으로 인하여 결코 완전한 관찰은 원리적으로 가능하지 않다.

내적(성찰적)이든 외적(사회적 또는 환경적)이든 관찰자의 모든 관찰은 생존이든 호기심이든 어떤 종류의 욕망(가치)을 충족하기 위한 행위이고 이를 위해서 세상과 자신에 대하여 언어와 유사하게 구조화된 모델을 구성한다. 이 모델이 바로 정보이다. 그러므로 정보는 직간접적으로 자신이 추구하는 가치와 연결되어 있고 미래의 행동을 위한 앎으로 존재한다. 이렇게 적응해나가는 능력을 지능이라고 한다면 지능은 세상에 대한 모델을 기반으로 하는 예측인 것이다. 이 예측은 환경에 대한 불확실성과 미래에 일어날 수 있는 사건의 불확실성을 감소시키기 위한 것이다.

샤논의 확률론적 정보이론은 의도적으로 정보의 양만을 다루기로 한 이론으로서 정보는 사건의 불확실성을 줄이는 것이므로 정보량은 해당 사건의 발생 확률의 역수로 정의하였다. 하지만 이러한 정의에 따른 정보량 측정은 어떤 상황에서 어떤 사건들이 어떤 확률로 일어나는지를 알 수 있어야 가능하다. 이에 대한 설명 없이는 정보의 의미는 고사하고 정보의 양조차 측정할 수 없게 된다. 하지만 콜모고로프(Kolmogorov)와 카이틴(Chaitin), 각각에 의하여 만들어진 알고리즘 정보이론(algorithmic information theory)⁸⁾은 언어(기호)와 정보의 양에 대한 놀라운 통찰을 준

다. 그들의 이론은 “문자열(문자의 나열, string)의 정보의 양은 그 문자열 자체 안에 포함된 최대 압축 가능한 표현의 길이와 같다”는 것이다. 이 말을 다시 풀어보면 정보가 문자열로 표현된다고 할 때 해당 문자열을 발생시킬 수 있는 가장 짧은 다른 문자열이 정보의 양이라는 의미이다. 이때 다른 문자열이 바로 특정 언어에 의한 표현이라고 할 수 있다. 이 특정 언어는 대상들과 대상들의 관계를 표현하고 정보를 표현하는 문자열을 발생하는 절차를 표현하게 된다. 가령 해당 문자열이 “abcabcabcabcabcabcabc”이라면 또 다른 문자열은 “abc를 7번 반복하라”가 될 수 있다. 이 다른 문자열은 한국어의 문자열이 되는 것이다. 그러므로 해당 문자열은 21개이지만 정보의 양은 빈칸을 포함하여 문자의 개수인 12라고 할 수 있다. 당연히 영어이거나 컴퓨터 프로그래밍 언어인 C언어라면 정보의 양은 달라질 것이다.

사실 표현체인 문자열을 생성하는 또 다른 문자열인 언어는 해당 언어를 사용하는 관찰자의 정보인 세상에 대한 모델이다. 하지만 이러한 정보는 사회적 소통을 통하여 이루어진 정보 구성의 차이로 인하여 매우 상이할 수 있고 소통할 수 있는 범위와 관찰자의 이미 구성된 정보들에 의하여 허용되는 범위 내에서만 같은 또는 유사한 의미를 가질 수 있다. 그 이유는 당연히 관찰자들은 각기 고유한 환경 속에서 이루어지는 심볼 그라운드 때문이다.

관찰자 내부의 정보인 해석체는 항상 대상체를 가지게 되는 이러한 관계를 ‘지향성(intention)’이라고 한다. 지향성 관련 연구에서 해석체는 ‘심적 표상’, 대상체는 ‘지향 대상’이라고 한다. 지향 대상의 의미를 설명하는 이론의 하나인 개념역할 의미론(Conceptual Role Semantics)은 지향 대상의 의미는 해당 심적 표상과 이미 가지고 있는 다른 심적 표상들 사이의 어떤 추론적 관계에 의하여 획득된다는 것이다. 개념역할 의미론은 ‘언어의 의미는 사용에 의하여 이해된다’는 비트겐슈타인의 이론에 기반

8) Chaitin, G. J., “On the Length of Programs for Computing Finite Binary Sequences”, *J. Association for Computing Machinery*, 13(4), 1966, 547쪽.

한 추론주의 의미론과 연결된다. 추론은 다양한 가능성을 가지고 있는 역동적인 과정이기 때문에 정보의 의미가 항상 단일하고 명확할 수 없다. 관찰자의 세계 모델도 세상에 대하여 부분적인 정합성을 가진 모순되고 잠정적인 정보들로 이루어져 있기 때문에 새로이 편입되는 정보는 역동적인 지각과 인식을 주장하는 메를로-퐁티의 기대지평이나 마이클 폴라니의 암묵적 지식 개념처럼 복잡다단하게 이루어진다. 좋은 정보의 좋은 추론은 관찰자의 세계 모델을 확장시키거나 정합적으로 만들 수 있다.

관찰자 정보의 이러한 속성으로 인하여 루만의 사회적 체계의 소통은 명확하게 이루어지기 보다는 우연적으로 이루어질 수밖에 없다. 소통의 송신자는 정보를 임의적으로 선택하고, 선택된 정보를 임의적인 언어의 구성으로 통보를 선택한다. 소통의 수신자는 다양한 의미로 해석될 수 있는 의미의 하나를 선택하는 것으로 소통이 일어나게 된다.

5. 인공지능과 인문학, 그리고 정보 철학

플로리디는 이 시대가 정보적 전회(information turn)를 맞이하고 있다고 생각한다. 코페르니쿠스의 전회는 지동설을 통하여 우주의 중심이 인간이 아님을 알게 하였고, 다윈의 전회는 진화를 통하여 인간이 우연히 발전하게 된 생물에 지나지 않음을 알게 하였고, 프로이트의 전회는 무의식을 통하여 우리가 의식하는 것이 나의 전부가 아님을 알게 하였다. 정보적 전회는 컴퓨터와 인공지능을 통하여 인간이 복잡한 기계일 수도 있다는 것을 알게 하였다. 필자는 인간이 단지 물리적으로 복잡한 기계에 지나지 않는다고 하여 인간의 가치와 이상, 그리고 존엄성이 폄하된다고 생각하지 않는다. 오히려 자연 속에 물리적인 몸을 가진 인간을 사실대로 잘 이해함으로써 인간으로서의 편견과 독단을 이기고 좀 더 나은 삶을 영위할 수 있는 토대가 될 것으로 믿는다⁹⁾. 이전에 있었던 3번

9) 박충식, 「1장 생명으로서의 인공지능: 정보 철학적 관점에서」, 『인공지능의 존재론 (포스트휴먼 시대의 인공지능 철학 1)』, 경기: 한울아카데미, 2018, 38쪽

의 전회를 통하여 인간이 어떠한 존재인지 좀 더 잘 알게 된 것처럼 정보적 전회도 그러할 것이기 때문이다.

멀지 않은 과거이지만 그 언제인가부터 ‘정보’라는 용어는 일상에서부터 여러 전문분야까지 다양하게 사용되어 왔지만 인공지능 대두 이후에는 그 의미가 보다 능동적인 의미를 가지게 되었다. 더구나 융합과 통섭이 유행어처럼 부르짖어지고 있는 상황에서 정보가 여러 분야를 아우를 수 있는 키워드로 보이기도 한다¹⁰⁾.

인문학계에서도 이러한 조류에 힘입어 디지털 인문학(Digital humanities)¹¹⁾이라는 이름으로 다양한 시도들이 이루어지고 있다. 디지털 인문학은 통상적으로 인문학 자원을 디지털화하여 체계적으로 정리하고 디지털화된 자료들을 분석한다. 하지만 이러한 디지털 인문학 연구는 다소 수동적인 작업 방식이라고 여겨진다. 오랜 전통의 인문학적 이론과 직관들이 반영되는 모델을 기반으로 인공지능의 기술을 좀 더 능동적으로 활용할 필요가 있다.

인공지능이 특정 기능의 성능을 향상시키기 위하여 아주 기술적인 문제에만 매달려 있으면서 오랜 전통의 인문학적 이론들을 기술과는 너무나 동떨어진 얘기로 치부하고 인문학은 인공지능에 의해 만들어진 성과를 인정하면서도 그 일부 성과에 대하여 과도한 인문학적 의미를 부여하는 용감함에 경계를 가지지 않을 수 없을 것이다. 인공지능이 충분히 성숙된 기술도 아니고 인문학이 추상적이기 때문에 인문학에서의 인공지능 기술에 대한 이해와 인공지능에서의 인문학적 이해를 위해 서로 노력하는 정도로는 간극을 좁히기는 쉽지 않다. 근본적으로 물질과 정신의 대립이라는 오래된 테제가 문제일 수도 있지만 근본적인 의미에서 일부라도 공유할 수 있는 개념과 기반이 필요한데 그 개념은 정보이고 그 기

10) 박충식, 「빅 히스토리와 인공지능: 정보적 관점에서」, 『인간·환경·미래』 2, 2019, 35쪽.

11) 김바로, 「해외 디지털인문학 동향」, 『인문콘텐츠』 33, 2014, 229-254쪽, 김영룡, 「디지털 시대의 창의력과 인문학」, 『카프카연구』 33, 2015, 243-265쪽 김현, 「디지털 인문학」, 『인문콘텐츠』 29, 2013, 9-26쪽

반은 구성적 정보 철학이 될 수 있을 것으로 생각한다.

이러한 관점에서 시도해볼 수 있는 융합적 연구의 실마리를 얘기해볼 수 있다. 해당 연구들에 대한 자세한 논의는 상당한 연구가 필요하기 때문에 간략히 그 주제들을 살펴보고자 한다.

1) 성찰과 설명가능한 인공지능

먼저 ‘성찰’에 대해서 논의해볼 수 있다. 현재 인공지능에서는 설명가능한 인공지능(XAI: eXplainable AI)에 대한 연구가 활발히 진행되고 있지만 설명가능한 인공지능은 단순한 인공지능 기술의 아이디어로 해결될 수 있는 문제가 아니다. 설명은 ‘성찰적’ 또는 반성적 ‘기능’이라고 할 수 있다. 이 문제는 오랫동안 인문학이 천착해온 문제이기도 하고 설명의 문제가 성찰이나 반성과 연계되면 인문학적으로도 깊은 논의가 필요하다. 설명가능한 인공지능과 성찰 기능을 가진 인공지능은 우선적으로 인공지능과 인문학의 학제적 연구가 필요한 문제이다. 설명가능한 인공지능 연구자들이 궁금해 하는 ‘사람들은 무엇을 설명이라고 여기는가?’에 대한 아이디어는 사회과학 분야에서 나오기도 하였다¹²⁾. 근본적으로 설명가능한 인공지능 문제는 수많은 단순한 데이터와 그 데이터의 조합으로부터 어떻게 개념이 형성될 수 있는 지에 대한 이론이 필요하고 이 이론은 심볼 그라운드와 밀접히 관련된다. 과거로부터 인공지능이나 컴퓨터 과학에서 이러한 문제를 직접적으로 다루고 있다는 사실은 잘 알려져 있지 않다. 성찰을 쉽게 말하면 대상에 대한 자신의 생각을 대상으로 하는 생각이다. 이러한 자신에 대한 생각으로서의 성찰은 인공지능 초기부터 논의되어 왔으며 놀랍게도 컴퓨터 과학에서도 이미 보편적으로 사용되고 있는 기술이다. 컴퓨터 과학에서 리플렉션(reflection)을 넘어서 실행중인 프로그램 자신의 구조와 행동을 조사하고(examine), 인트로스펙트하고(introspect), 변경(modify)할 수 있는 능력을 의미하며, 누구나 이름은 들

12) Miller, Tim, “Explanation in Artificial Intelligence: Insights from the Social Sciences”, arXiv, 2018, 1706.07269, 2018.

어 본적이 있는 자바(java), 파이썬(python), 고(Go)와 같은 컴퓨터 프로그램 언어들이 모두 이러한 기능을 가지고 있다. 또한 ‘메타프로그래밍(metaprogramming)’은 프로그램을 자료로 다룰 수 있는 능력을 가진 컴퓨터 프로그램에서 사용되는 프로그램 기술이며 이것은 하나의 프로그램이 다른 프로그램을 읽고, 생성하고, 분석하고, 변형하며 심지어 수행 중인 프로그램도 변경할 수 있다는 것을 의미한다.

인공지능이란 말을 처음 만든 존 맥카시는 사람들이 자신의 정신적 과정을 관찰하는데 제한된 능력을 가지고 있지만 인공지능은 보다 나은 성찰적 능력이 필요하다고 하였다. 그리고 마빈 민스키는 인간의 이해는 필수적으로 세계의 모델을 실행하는 과정이라고 하였다. 그의 논의에서 흥미로운 것은 세계의 모델이 주변 세계뿐만 아니라 행위자 자신도 포함한다는 것이다. 이러한 아이디어는 그의 유명한 책 『마음의 사회(Society of Mind)』의 말미에 기술되었다. 인공지능에서 성찰에 관련한 연구는 인공신경망과는 다른 소위 ‘기호적 인공지능(Symbolic Artificial Intelligence)’이라는 분야에서 이루어졌다. 현재 가장 유명한 인공지능 교과서 중에 하나인 『인공지능: 현대적 접근』은 1000쪽이 넘는 분량에도 불구하고 인공신경망에 대한 내용은 일부에 지나지 않고 대부분 철학, 특히 논리학에서 다루어지는 내용들이 인공지능적인 관점으로 기술된 기호적 인공지능에 관한 내용을 담고 있다. 이러한 사실이 인공신경망이 중요하지 않다는 것을 의미하는 것은 아니고 설명하기에 비교적 단순한 아이디어라는 것이다. 과거에도 그러했지만 이제 인공신경망에 기반한 딥러닝이 좋은 결과를 보이고 있는 상황에서 설명가능한 인공지능을 위하여 기호적 인공지능과의 결합된 연구가 긴요하고 기호적 인공지능은 인문학과의 깊은 논의가 매우 절실하다.

인간의 인지적 과정을 컴퓨터로 만드는 인지 시뮬레이션(cognitive simulation)연구 또한 인공지능과 인문학의 학제 간 연구가 필요한 분야이다. 퇴계의 사단칠정이론을 유학의 도덕적 심성모델로 상징하고 이루어진 연구는 인지 시뮬레이션 연구가 그러한 사례라고 할 수 있다¹³⁾. 현재

에도 인공지능 도덕 행위자(AMA: Artificial Intelligence Agent)에 대한 연구¹⁴⁾는 활발하게 이루어지고 있으며, 인공지능 로봇들이 개발되고 있는 실정에서 이 분야에 대하여 인공지능 연구자와 인문학 연구자의 실제적인 연구가 필요하다고 할 수 있다.

2) 사회 속의 인간과 행위자 기반 시뮬레이션

인간에 대한 이해가 인간이 속해있는 사회를 도외시키고는 이루어질 수 없다. 인간의 사고와 행동은 사회 속에서 영향받고 이러한 인간들은 사회에 다시 영향을 미치게 된다. 구성적 정보 철학의 심볼 그라운드에서 살펴본 바와 같이 인간의 구성된 모든 정보는 사회적인 맥락하에서 이루어진다. 사회과학 분야에서도 컴퓨터 발전에 힘입어 빅데이터나 사회 네트워크 분석(Social Network Anaysis)을 전문적으로 하는 사회과학에 대한 연구기관들이 등장하였다. 하지만 디지털 인문학과 비슷하게 데이터를 위주로 하는 수동적 연구 방법들이 사용된다. 해외에서는 계산 사회과학(computational social science)이라는 좀 더 넓은 개념으로 데이터 분석과 더불어 사회를 구성하는 행위자들을 모델링하여 사회 현상을 능동적으로 분석하는, 행위자 기반 시뮬레이션(Agent-based Simulation)을 활용하는 연구도 많이 행해지고 있다¹⁵⁾. 이러한 사회과학 연구 방법에서 아쉬운 점은 사회를 구성하는 행위자들을 너무 단순하게 모델링하고 사

-
- 13) 유권중·박충식, 「性理學的 心性모델 시뮬레이션을 이용한 유교 禮敎育 방법의 효용성 분석」, 『東洋哲學』 16: 2002, 269-298쪽, 유권중·박충식·장숙필, 「인지과학 시뮬레이션을 위한 유교 예敎育 심성모델의 구성(1)」, 『민족문화연구』 37집. 고려대 민족문화연구원, 2002., 유권중·박충식, 외, 「인지과학적 시뮬레이션을 통한 조선 성리학의 예敎育심성모델 개발(2)」, 『東洋哲學』 39, 2014, 297-341쪽.
 - 14) 최현철, 변순용, 신현주, 「인공적 도덕행위자(AMA) 개발을 위한 윤리적 원칙 개발」, 『윤리연구』 111권, 2016, 31-53쪽, 박충식, 「윤리적인 인공지능 로봇: 구성적 정보 철학 관점에서」, 『과학철학』 21(3), 2018, 39-65쪽.
 - 15) 최계영·박충식, 「BRMS를 이용한 루만의 사회체계이론 구현방안」, 『2013년 한국지능정보시스템학회 춘계 학술대회 논문집』, 2013, 255-260쪽, 최계영·박충식, 「행위자 기반 미시-거시 연계 경제 시뮬레이션과 루만의 사회체계이론 구현」, 『2014년 한국지능정보시스템학회 추계 학술대회 논문집』, 2014, 229-232쪽

회적인 속성만을 가진 사회적 인간만을 연구한다는 것이다. 사회 속의 인간이 사회적 특성만으로 정의되는 것은 아니기 때문에 행위자 기반 시뮬레이션의 행위자들에 대하여 인문학적인 고려가 가능할 수 있도록 행위자를 인지 시뮬레이션 가능한 모델로 결합할 필요가 있다.

계산 사회과학에서의 이러한 방법은 역사학에도 그대로 사용되고 있는데 이러한 연구방법을 계산 역사학(computational history)이라고 한다. 디지털 인문학에서 역사학은 주로 사료들을 체계적으로 정리하고 시각화하고 텍스트를 분석하는 방식으로 이루어진다. 계산 역사학은 전술한 계산 사회과학처럼 기존의 사료를 기반으로 행위자 기반 시뮬레이션을 통하여 다양한 역사 해석과 스토리텔링으로 역사에 대한 이해를 높일 수 있다¹⁶⁾.

이러한 행위자 기반 시뮬레이션을 활용하는 역사 시뮬레이션은 연구자들뿐만 아니라 역사를 배우는 학생들에게도 새로운 방식의 역사 공부를 적용할 수 있으며 컴퓨터 시뮬레이션을 활용함으로써 홈페이지와 로봇을 움직이는 단순한 코딩 교육을 넘어서 인문학적 코딩교육을 시도해볼 수 있다.

3) 감정 연구와 인지 시뮬레이션

인공지능과 연계된 중요한 또 다른 연구 대상으로 감정을 들 수 있다. 감정은 가장 인공지능적이지 않은 것으로 여겨지고 감정에 대한 인공지능 연구는 가장 반인문적이라고 생각될 수 있다. 인공지능이 감정을 가진 것처럼 보이거나 인간의 감정을 아는 것처럼 보이는 것은 인공지능이 실제로 감정을 가진다는 것과 완전히 다른 문제이다. 미래의 지능적인 로봇들은 굳이 감정을 가지도록 만들어지지 않는 것이라고 예견되기도 한다. 인간의 일을 대신하는데 이성적인 기능만으로 충분하다고 생각하는 것이다.

16) 박충식, 「빅 히스토리와 인공지능: 정보적 관점에서」, 『인간·환경·미래』 2, 2019, 31-59쪽.

감정은 정신에서 이성과 상대적인 것으로 또는 열등한 것으로 치부되면서 합리적 판단의 장애물처럼 여겨지는 면도 있다. 하지만 최근 인지과학에서 감정이 의사결정에서 매우 중요한 역할을 하는 것으로 밝혀지면서 그 중요성이 부각되고 있다. 감정은 마주 대하는 사태에 대한 신체적인 평가 시스템이기 때문에 감정중추에 문제가 생기면 의사결정 장애를 일으키게 된다. 감정은 현재 직면하는 사태에 대하여 과거로부터의 축적된 경험에서 생겨나는 의식할 수 있는 신체적 평가 반응이다. 감정은 기본적으로 욕구에 비추어진 평가이고 파생적인 욕구를 발생시키며 욕구는 행동을 유발한다. 감정이 매우 친숙한 것도 있지만 생경한 것도 있어서 자신의 감정을 잘 헤아리지 못할 수도 있다. 의식은 이러한 감정을 이성의 대상으로 삼아 분석할 수 있고 분석의 결과에 대하여 다른 감정을 유발할 수도 있다. 일어난 행동을 나중에 의식이 합리화한다는 것은 분리 뇌 실험으로 널리 알려진 사실이다. 마주하는 사건의 평가는 사건의 순수한 논리적인 구조의 모든 요소가 평가를 필요로하기 때문에 감정은 의사결정에서 중요한 내용물인 것이다. 그러므로 흄은 ‘이성은 감정의 노예’라고 하였다. 감정 없는 이성은 내용 없는 논리적 구조만을 다루는 셈이 된다¹⁷⁾.

사람들이 생각하는 것처럼 인공지능이 인간의 일을 처리함에 감정없이도 잘 처리할 수 있는 일도 있겠지만 감정없이 인간처럼 일을 할 수는 없을 것이다. 우리 인간은 같은 사태를 경험함으로써 다른 사람의 감정을 이해한다. 항상 맞는 것은 아니지만, 심리학에서 ‘마음의 이론(Theory of Mind)’이라고 부르는 것이다. 인간은 심지어 거울 뉴런을 통하여 선천적으로 다른 사람을 공감할 수 있는 생물학적 능력을 타고 났다고 한다. 무감정(Apathy)이나 감정표현불능(Alexithymia)은 정상적인 사회생활을 불가능하게 한다. 감정이입(Empathy)이 불가능한 사람이 사이코패스이다.

17) 박충식의 인공지능으로 보는 세상: “인공감정” <http://www.econovill.com/news/articleView.html?idxno=337463>, 2018.

가장 심각하게 논의되고 있는 인공지능의 윤리 문제도 인공지능이 인간의 도덕적 감정을 가지고 있지 못하기 때문이다. 거장 그림과 비슷한 그림이나 거장 음악과 비슷한 음악을 만들어내는 인공지능이 예술적 감정을 가지고 있을 리 없다. 뜨거운 열정에 휩싸인 또는 쓰라린 짝사랑의 감정이 없는 인공지능이 쓰는 소설이나 시도 마찬가지로이다. 색과 모양, 질감, 음율, 단어와 문자의 뉘앙스들이 복잡한 느낌으로 복잡한 감정을 불러일으키는 것은 복잡한 정보이기 때문이다. 그러한 복잡함의 총체적 평가가 감정이고 감정은 기억과 기대의 복잡한 정보처리일 것이다(토노니(Tononi)의 ‘의식의 정보통합이론’). 왜냐하면 감정은 의식되는 경험이기 때문이다¹⁸⁾.

논의한 바와 같이 인공지능에서의 감정 연구는 단순히 감정을 가진 것처럼 보이도록 하는 것으로는 부족하다. 구성적 정보 철학에서 심볼 그라운드(ground)는 관찰자의 정보가 ‘의미’를 가지게 되는 것인데 하지만 그 의미는 단순히 ‘지시하는 것이 참’이라는 논리적인 것만을 의미하지는 않는다. 관찰자에게 의미는 관찰자의 욕구(가치)에 의하여 만들어지는 것이므로 대상에 맞닥뜨린 관찰자에게 즉각적이고도 본능적인 의미는 감정으로 평가된다. 감정은 매우 복잡한 인문학적 주제이지만 구성적 정보 철학의 관점에서 분석 가능한 현상이다. 이러한 감정에 대한 연구는 인지 시뮬레이션에도 중요하게 반영된다.

4) 구성주의적 언어 연구와 인문코딩 교육

또 다른 연구 주제로서 구성적 정보 철학을 기반으로 한 언어 연구를 생각해 볼 수 있다. 인문학과 인공지능의 공통 키워드가 심볼 그라운드이고 언어는 심볼 그라운드 그 자체이기 때문이다. 인간의 대화나 문학 작품, 등의 텍스트 언어들은 인간의 정보인 해석체의 표현체이다. 이러한 표현체를 통한 발화자의 해석체를 추적하는 작업은 다른 관찰자가 의미를 재구성하여 자신의 해석체로 만드는 과정이다. 구성적 정보 철학은

18) Emotion, <https://en.wikipedia.org/wiki/Emotion>

이러한 의미들을 추론주의 의미론으로 설명할 수 있다. 심볼 그라운드, 언어습득¹⁹⁾, 그리고 시와 같은 문학 언어 연구를 위하여 점증하는 인지 과학을 기반으로 한 인지언어학의 연구 성과를 수용하고, 빅데이터 처리에까지 다가간 대용량 코퍼스(말뭉치) 기반의 언어연구가 용이하면서도 가장 급진적인 언어이론이라고 할 수 있는 비트겐슈타인의 언어사용이론을 시험할 수 있는, 보다 유연한 문법이론이 구문문법(Construction Grammar)이다. 구문문법은 전통적인 언어학과는 다른 관점의 언어학으로써 생성문법과는 대척점에 존재하는 언어이론이다. 구문문법은 동사 위주의 기존 문법에서 구문의 중요성에 초점을 가진다. 구문문법은 통사부와 어휘부의 명확한 구분을 지양하고 양자를 통사-어휘 연속체로 본다. 구문문법에서 문법적 구문은 내부 구조의 복잡성과는 상관없이 형식과 의미의 짝으로 간주된다. 그러므로 단어 역시 구문의 일종이며 구문문법은 형식과 의미의 짝이라면 모두 구문이라고 주장하기 때문에, 구, 관용어, 단어, 형태소 등이 모두 구문으로 간주된다. 구문문법에서 구문은 형식(발화)과 의미, 두 개의 극(pole)으로 소쉬르의 시니피앙과 시니피에에 해당된다고 할 수 있다²⁰⁾. 이러한 구문의 적용은 구성적 정보 철학의 관점에서는 전체 정보에 대한 추론을 통하여 이루어지는 추론주의 의미론으로 이루어진다.

현재 자연언어 대화 서비스를 제공하는 많은 챗봇들이 만들어 지고 있고, 이러한 챗봇을 만들기 위한 많은 무료 소프트웨어들도 존재한다. 인문학적 코딩 교육의 일환으로 챗봇 만들기를 활용할 수 있다. 이러한 교육은 인공지능과 인문학을 같이 이해할 수 있는 좋은 사례가 될 수 있다²¹⁾.

19) 김종도, 「인지문법의 국어에의 적용」, 『한국어학한국어학』 제5권, 1997, 1-44, Anderson, Michael, “Embodied Cognition: A Field Guide”, Artificial Intelligence, 2003, 139-159쪽

20) 장재인·이호준·박충식·이옥희, 「인지문법을 활용한 연애시물레이션게임 담화분석」, 『한국지능정보시스템학회 2013 추계학술대회논문집』, 2013, 226-231쪽, 고재현·전인후·박충식·이옥희, 「인지문법을 활용한 연애시물레이션게임 담화분석」, 『한국지능정보시스템학회 2014 추계학술대회논문집』, 2014, 237-240쪽

5) 인공지능 거버넌스

또 다른 주제로써 4차 산업혁명으로 떠오른 인공지능에 대한 사회적 문제를 고려해볼 수 있다. 인공지능 자체를 윤리적으로 만들어야 하는 문제와는 별개로 인공지능에 대한 많은 우려들이 존재한다. 윤리적인 문제없이 잘 동작하는 인공지능이 대체하는 노동으로 생기는 실업 문제, 이러한 사회 환경 변화에 대처하기 위한 교육 문제, 등이 있을 수 있다. 이러한 문제들만큼이나 인공지능에 의한 산업 발전과 노동으로부터의 자유와 해방을 기대하는 사회적 바람도 존재한다. 인공지능과의 공존 또는 인공지능 거버넌스로 언급되는 이러한 주제들은 인공지능 자체의 문제라기보다는 인공지능을 둘러싼 인간 이해 당사자들의 문제라고 할 수 있다. 구성적 정보 철학의 관점에서 보면 무엇이 옳은 것인지에 대한 논의와는 무관하게 모든 인간은 고유한 욕망(가치)을 추구하는 가운데 고유한 심볼 그라운딩을 통하여 생각하고 행동할 수밖에 없다. 서로의 욕망들이 복잡하게 얽혀 있는 공동체에서 무엇인가를 해야 한다면 소통을 통한 타협이 중요할 수밖에 없다. 하지만 이러한 소통이 쉽지 않은 것은 고유한 심볼 그라운딩을 유지하고 있기 때문일 것이다. 이러한 상황에서 효과적인 소통을 위하여 ‘소셜 머신(social machine)’이라는 정보통신 기술을 사용할 수 있다. 소셜 머신은 ‘인간과 기술이 상호작용하여 어느 한쪽이 없다면 불가능한 결과나 행동을 만들어 내는 환경’으로 정의된다²¹⁾.

우리나라의 ‘청와대 국민청원’도 정부주도이긴 하지만 일종의 소셜 머신이라고 할 수 있다. 민간 주도적으로 이루어지고 있는 소위 ‘디지털 데모크라시(Digital Democracy)’ 또는 ‘시빅테크(Civic Tech)’도 이러한 유형이라고 할 수 있다. 소셜 머신의 이론과 실천을 주제로 하는 영국의

21) 문경훈, 박충식, 「챗봇 제작을 통한 인공지능교육 체험」, 『한국지능정보시스템학회 2018 추계학술대회논문집』, 2018, 123-124, 박충식의 인공지능으로 보는 세상: “인공지능 코딩교육”, <http://www.econovill.com/news/articleView.html?idxno=358640>, 2019, 박충식, 「심리상담 챗봇 만들기?」, 『2018 한국심리학회 학술대회 자료집』, 123-124쪽.

22) Social Machine, https://en.wikipedia.org/wiki/Social_machine

SOCIAM 프로젝트(sociam.org)는 옥스퍼드 대학, 그리고 사우스햄튼 대학, 에딘버러 대학이 공동으로 진행하고 있다²³⁾. 이곳의 소셜 머신 연구는 네트워크상에서 상호작용하는 사람들의 콘텐츠를 단순히 분석하는 것을 넘어서 이러한 상호작용들이 어떻게 이루어지는지 분석하고 이러한 상호작용들이 윤리적이고 안전하게 이루어질 수 방법을 모색한다.

소셜 머신이 정치적 기계로서 작동하게 되겠지만 진리 주장의 격투장은 아니다. 펠렐만의 신수사학이 논증적 이성을 통해 사회 속에서 진리가 아닌 의견 가운데 가장 큰 공약수, 즉 ‘언어 공동체가 판단, 또는 명제에 부여하는 합의’를 통해서 다양성을 나타내고자 하는 것으로 보는 것이 전제에 대한 논의를 새롭게 한다는 의미에서 더욱 타당해 보인다²⁴⁾. 구성적 정보 철학은 이러한 소셜 머신에서 좀 더 공통적인 심볼 그라운드 환경이 될 수 있도록 하는 이론적 토대가 될 수 있도록 노력할 수 있다. 덧붙여 이러한 소셜 머신에서는 계산 사회과학의 행위자 기반 시뮬레이션을 통하여 다양한 가정 하에 이루어지는 주장들을 같이 검토하고 거짓 정보들을 탐지할 수 있는 기반으로 사용할 수도 있을 것이다²⁵⁾.

구성적 정보 철학에 기반하여 시도해볼 수 있는 몇 가지 주제들을 살펴보았다. 주제들이 따로 논의되기는 했지만 서로 관련된 부분이 많고 같이 연구되어야 할 주제들이기도 하다.

23) SOCIAM, <https://sociam.org/>

24) 양현정, 「신수사학과 데카르트 철학-카임 펠렐만의 관점에서 바라본 데카르트의 철학」, 『수사학』 12: 2010, 125-145쪽.

25) 박충식, 「기계들과의 공존: 소셜머신」, 신상규 (편), 『4차 산업혁명시대 인문학에 묻다』, 2018년 길 위의 인문학, 인문포럼 성과물 서울: 문화체육관광부, 한국도서관협회, 이화인문과학원, 2018, 96-115쪽, 박충식, 「3장 공존의 존재, 인공지능: 소셜 머신 2.0」 한국포스트휴먼연구소, 한국포스트휴먼학회(편), 『포스트휴먼 사회와 새로운 규범 (포스트휴먼 사이언스 총서 5)』, 서울: 아카넷, 2019, 93-113쪽.

7. 나가기

인공지능과 인문학의 관계는 자동차와 인문학의 관계와는 같지 않다. 자동차에 대한 인문학적 성찰을 위하여 자동차에 대한 이해가 필요하지만 인공지능에 대한 인문학적 성찰을 위해서는 인공지능에 대한 이해뿐만 아니라 인간에 대한 이해가 필요하다. 인간과 같이 지능적인 기계를 만드는 것이 가까운 미래 안에 이루어지지는 않겠지만 이러한 지능적인 기계는 자유의지, 감각질, 직관, 도덕, 예술, 등과 같은 인문학적 문제를 다시 촉발시키고 있다. 인간이 만든 비행기의 비행 원리가 새의 비행 원리를 따르고 있지 않은 것처럼 인공지능의 원리가 인간 지능의 원리를 따를 필요는 전혀 없지만 우리가 아는, 충분히 지능적인 유일한 존재가 인간이고 인공지능이 인간처럼 지능적인 일을 해내기를 원하기 때문에 인간에 대한 탐구, 즉 인문학은 인공지능을 이루기 위한 필수적인 과정이라고 할 수 있다. 이러한 인문학은 물리적 존재로서의 인간, 생물적 존재로서의 인간, 사회적 존재로서의 인간에 대한 성찰 또한 포함한다.

적당한 노력이 따른다면 인문학자들도 인공지능이 작동하는 방식을 이해할 수 있다. 그러한 이해는 복잡한 수식 안에 있는 것이 아니다. 아무리 복잡해 보이는 수식도 주어진 입력 자료 이상으로 행동하거나 학습할 수 없다. 인문학적인 개념 분석이 복잡한 수식보다 더욱 중요하다. 자주 거론되는 자율 자동차, 의료 로봇이나 전쟁 로봇들의 도덕적 판단과 행위도 마찬가지로 접근이 필요하다. 인공지능과 인문학의 이러한 융합적 접근은 인공지능을 위해서도 인문학을 위해서도 필요한 일이다²⁶⁾.

본 연구는 구성주의에 기반한 정보 철학이 인공지능과 인문학을 포괄적으로 이해하기 위한 방편이 될 수 있다고 논의하였다. 우리 사회에 인

26) 박충식, 「인공지능 인문학을 위한 인공지능 기술 이해」, 『인공지능 인문학 연구』 제1권, 2018, 173-182쪽, 박충식, “인공지능은 인문학이다?”, 수유너머104 강연, 2019년 트랜스인문학연구소 [제25회 화요토론회] 인공지능은 인문학이다(박충식), 2019년 1월 15일 화요일, <http://www.nomadist.org/s104/TuesdayDiscuss/103698>.

공지능에 대한 담론을 촉발했던 딥러닝이 설명가능한 인공지능이 되기 위해서는 자신의 구조적인 한계를 넘어 설 수 있는 다른 방안이 필요하다. 그 다른 방안은 기호가 의미를 가지게 되는 심볼 그라운드이다. 인공지능과 인문학이 공유할 수 있는 개념으로서 정보와 그 정보에 대한 전반적인 그림으로서 구성적 정보 철학이 유의미하다고 할 수 있다. 당연히 그 핵심은 심볼 그라운드이고 심볼 그라운드는 인문학의 가장 중요한 주제인 언어 현상의 핵심이기도 하다. 이러한 구성적 정보 철학의 관점에서 인공지능과 인문학의 몇 가지 융합적 연구 주제들을 모색해보았다.

참고문헌

- 고재현·전인후·박충식·이옥희, 「인지문법을 활용한 연애시물레이션게임 담화 분석」, 『한국지능정보시스템학회 2014 추계학술대회논문집』, 2014, 237-240.
- 김바로, 「해외 디지털인문학 동향」, 『인문콘텐츠』 33, 2014, 229-254.
- 김영룡, 「디지털 시대의 창의력과 인문학」, 『카프카연구』 33, 2015, 243-265.
- 김중도, 「인지문법의 국어에의 적용」, 『한국어학』 제5권, 1997, 1-44.
- 김현, 「디지털 인문학」, 『인문콘텐츠』 29, 2013, 9-26.
- 당반치엔, 트란트랑틴, 김종욱, 「인지 에이전트 아키텍처인 Soar의 소셜 로봇 적용에 관한 연구」, 『한국지능정보시스템학회 논문집』 26(1), 2016, 145-146.
- 문경훈, 박충식, 「챗봇 제작을 통한 인공지능교육 체험」, 『한국지능정보시스템학회 2018 추계학술대회논문집』, 2018, 123-124.
- 박충식, 인공지능은 인문학이다?, 수유너머104 강연, 2019년 트랜스인문학연구소 [제25회 화요토론회] 인공지능은 인문학이다(박충식), 2019년 1월 15일 화요일, <http://www.nomadist.org/s104/TuesdayDiscuss/103698>
- 박충식, 「1장 생명으로서의 인공지능: 정보 철학적 관점에서」, 이종원(편), 『인공지능의 존재론(포스트휴먼 시대의 인공지능 철학 1)』, 경기: 한울아카데미, 2018, 19-40.
- 박충식, 「2장 거꾸로 보는 인공지능의 역사」, 한국포스트휴먼연구소, 한국포스트휴먼학회(편), 『인공지능의 이론과 실제(포스트휴먼 사이언스 총서 6)』, 서울: 아카넷, 2019, 29-63.
- 박충식, 「3장 공존의 존재, 인공지능: 소셜 머신 2.0」, 한국포스트휴먼연구소, 한국포스트휴먼학회(편), 『포스트휴먼 사회와 새로운 규범(포스트휴먼 사이언스 총서 5)』, 서울: 아카넷, 2019, 93-113.
- 박충식, 「3장 성찰적 인공지능」, 한국포스트휴먼연구소, 한국포스트휴먼학회(편), 인공지능과 새로운 규범(포스트휴먼 사이언스 총서 4), 경기: 아카넷, 2018, 135-156.

- 박충식, 「4장 생명과 몸과 마음으로서의 인공지능」. 한국포스트휴먼연구소, 한국포스트휴먼학회(편), 『제4차 산업혁명과 새로운 사회윤리(포스트 휴먼 사이언스 총서 3)』, 경기: 아카넷, 2017, 131-156.
- 박충식, 「구성적 인공지능」, 『인지과학』 15(4), 2004, 61-66.
- 박충식, 「기계들과의 공존: 소셜머신」. 신상규 (편), 『4차 산업혁명시대 인문학에 묻다』, 2018년 길 위의 인문학, 인문포럼 성과물 서울: 문화체육관광부, 한국도서관협회, 이화인문과학원, 2018, 96-115.
- 박충식, 「빅 히스토리와 인공지능: 정보적 관점에서」, 『인간·환경·미래』 2, 2019, 31-59.
- 박충식, 「심리상담 챗봇 만들기?」, 『2018 한국심리학회 학술대회 자료집』, 2018, 123-124.
- 박충식, 「윤리적인 인공지능 로봇: 구성적 정보 철학 관점에서」, 『과학철학』 21(3), 2018, 39-65.
- 박충식, 「인공지능 인문학을 위한 인공지능 기술 이해」. 『인공지능 인문학 연구』 제1권, 2018, 173-182.
- 박충식·정광진, 「포스트휴먼 시대의 이해: 루만의 사회체계이론적 관점에서」, 『2017년 제1회 한국사회체계이론학회 정기학술대회 논문집』, 2017.
- 백중현, 『이성의 역사』 서울: 아카넷, 2017.
- 양현정, 「신수사학과 데카르트 철학-카임 펠렐만의 관점에서 바라본 데카르트의 철학」, 『수사학』 12: 2010, 125-145.
- 유권중·박충식, 외, 「인지과학적 시뮬레이션을 통한 조선 성리학의 예교육 심성모델 개발(2)」, 『東洋哲學』 39, 2014, 297-341.
- 유권중·박충식, 외, 『유교적 마음모델과 예교육』, 서울: 한국학술정보, 2009.
- 유권중·박충식, 「性理學的 心性모델 시뮬레이션을 이용한 유교 禮교육 방법의 효용성 분석」, 『東洋哲學』 16: 2002, 269-298.
- 유권중·박충식·장숙필, 「인지과학 시뮬레이션을 위한 유교 예교육 심성 모델의 구성(1)」, 『민족문화연구』 37집. 고려대 민족문화연구원, 2002.
- 장세인·이호준·박충식·이옥희, 「인지문법을 활용한 연애시뮬레이션게임 담화 분석」, 『한국지능정보시스템학회 2013 추계학술대회논문집』, 2013,

226-231.

- 정상윤·윤석영·박충식, 「챗봇 제작을 통한 SW교육방안」, 『한국지능정보시스템학회 2017 추계학술대회논문집』. 2017, 143-145.
- 최제영·박충식, 「BRMS를 이용한 루만의 사회체계이론 구현방안」. 『2013년 한국지능정보시스템학회 춘계 학술대회 논문집』. 2013, 255-260.
- 최제영·박충식, 「행위자 기반 미시-거시 연계 경제 시뮬레이션과 루만의 사회체계이론 구현」. 『2014년 한국지능정보시스템학회 추계 학술대회 논문집』, 2014, 229-232.
- 최현철, 변순용, 신현주, 「인공적 도덕행위자(AMA) 개발을 위한 윤리적 원칙 개발」, 『윤리연구』 111권, 2016, 31-53.
- Anderson, Michael, “Embodied Cognition: A Field Guide”, *Artificial Intelligence*, 2003, 139-159.
- Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press. 2016.
- Brier, Søren, “Levels of Cybersemiotics: Possible Ontologies of Signification”. *Cognitive Semiotics* 4, 2009, 28-62.
- Brier, Søren, “Luhmann Semioticized”, *Journal of Sociocybernetics* 3(2), 2003, 13-22.
- Chaitin, G. J., “On the Length of Programs for Computing Finite Binary Sequences”, *J. Association for Computing Machinery*, 13(4), 1966, 547-569.
- Harnad, S., “The Symbol Grounding Problem”, *Physica D* 42, 1990, 335-346.
- Hendler, James, Mulvehill, Alice, *Social Machines: The Coming Collision of Artificial Intelligence, Social Networking, and Humanity*, APress, 2016.
- Miller, Tim, “Explanation in Artificial Intelligence: Insights from the Social Sciences”, arXiv, 2018, 1706.07269
- O'Hara, Kieron, “Social Machine Politics Are Here to Stay”, *IEEE Internet*

Computing 17(2): 2013, 87-90.

Rovelli, Carlo, “Meaning = Information + Evolution”, eprint arXiv:1611.02420.2016.

Sowa, John F., “Signs and reality”, *Applied Ontology* 10(3-4), 2015, 273-284.

Spier, F., *Big history and the future of humanity*, Chichester: Wiley/Blackwell, 2010.

**AI is the Humanities:
From the perspective of Constructive Philosophy of Information**

Choong-Shik Park (U1 Univ.)

Artificial intelligence is a branch of computer science that seeks to make machines have human intelligence. To this end, it is also a field of research to understand human intelligence as a computational model. Therefore, artificial intelligence is closely related to the humanities of Moon Sa-chul(Literature-History-Philosophy), a long-standing study to understand humans. Moreover, with the development of cognitive science, understanding of human beings has entered a new phase. Artificial intelligence, which is also an area of cognitive science, is another aspect of humanities that comprehensively promotes understanding of human beings.

I think that although current artificial intelligence still has a lot of distance from human understanding, for more advanced artificial intelligence, an understanding of humanities is essential, and for more sophisticated human understanding of humanities also requires an understanding of artificial intelligence and an artificial intelligence imagination. This study discusses that information philosophy based on constructivism can be a way to comprehensively understand artificial intelligence and humanities.

Key words: artificial intelligence, humanities, philosophy of information, constructivism

철학탐구 제56집

박충식 e-mail: leciel@ul.ac.kr

투 고 일	2019년 10월 14일
심 사 일	2019년 11월 14일
게재확정	2019년 11월 14일