

특별호: 심리학과 4차 산업혁명 2

## 인공지능과 인지과학: 기회와 도전\*

### 정혜선†

인지과학과 인공지능의 연구는 모두 계산과 정보이론의 발전에 힘입어 출현하였고, 서로 밀접하게 상호작용하면서 성장하였다. 인공지능 연구의 초기 단계에 인간인지는 인공지능 연구에 영감의 원천이자 기준으로 작동하였으나 최근 들어 가속화되고 있는 인공지능의 발전은 인간과 인공지능 간의 관계에 대한 재정립을 요구하고 있다. 본 논문에서는 인지과학과 인공지능 연구가 어떻게 함께 발전하였는지를 바탕으로 두 분야 연구가 앞으로 어떠한 식으로 관계를 맺을 수 있을지 살펴보았다. 인공지능의 발달은 인지과학에 도전과 기회를 동시에 제공하고 있는데, 첫째, 인공지능의 발달은 인간 마음의 작동에 대한 이해를 심화할 수 있는 기회를 제공한다. 둘째 인공지능의 발달은 다양한 정보처리 도구의 개발을 촉진하여 이를 통해서 인간이 더 효과적으로 정보를 처리할 수 있도록 도와줄 것이다. 동시에 인공지능은 인간이 속한 정보환경을 필연적으로 변화시킬 것이고, 이는 인간의 인지 능력에도 중요한 변화를 야기할 것으로 보인다. 인공지능의 발달이 가져오는 변화와 영향의 성격이 아직 충분히 드러나지 않았지만 인공지능의 발달이 가져오는 기회를 활용하고 도전에 대처하는데 인지 연구자들의 적극적인 노력과 참여가 필요하다.

주요어 : 인공지능, 인간지능, 알고리즘, 지식, 학습, ITS, 연구지원, 이론개발, 인지도구

\* 본 논문은 한림대학교의 지원을 받아 수행되었습니다(HRF-202011-002).

† 교신저자: 정혜선, 한림대학교 심리학과 교수, 강원도 춘천시 한림대학길 1 사회경영1관

Tel: 033-248-1725, E-mail: heis@hallym.ac.kr

## 배경

4차 산업혁명에 대한 논의가 뜨겁다. 4차 산업혁명의 핵심에는 빅데이터, 딥러닝, 기계 학습을 포괄하는 인공지능이 존재한다. 인공지능과 인간은 다른 하드웨어를 가지고 있지만 정보처리를 한다는 점에서 공통점을 갖는다. 정보이론의 도래와 함께 1950년대부터 인간인지와 인공지능에 대한 연구가 본격적으로 시작되었다. 두 분야의 성립 초기에 인간의 인지는 인공물의 정보처리를 구현하는데 영감의 원천이었고, 인간의 지능을 얼마나 비슷하게 모사할 수 있는가는 인공지능을 평가하는 기준이었다. 이러한 관계는 최근 급격한 변화를 경험하고 있다. 하드웨어의 발달, 빅데이터의 출현 및 프로그래밍의 발달로 인하여 인공지능의 수행이 인간을 넘어서는 것은 물론 그 작동 방식은 더 이상 인간을 모델로 하지 않게 되었다. 이러한 변화에 직면하여 본 논문에서는 인공지능과 인지과학 간의 공생의 역사를 살펴보고 이를 바탕으로 앞으로 어떠한 접점과 관계 맺기가 가능한지를 살펴보았다. 인간과 인공지능 모두 정보처리 시스템이라는 공통점을 지니고 있기 때문에 두 분야는 앞으로 어떤 형태로건 밀접한 관계를 유지할 수 밖에 없다. 인공지능의 발달은 인지과학자들에게 도전과 기회를 동시에 제공할 것으로 보여지는데, 첫째, 인공지능의 발달은 인간의 정보처리에 대한 이해를 심화할 수 있는 기회가 될 것으로 보인다. 둘째, 인공지능의 발달은 인간의 정보처리를 지원하는 다양한 도구/시스템의 개발을 가능하게 하고, 이는 인간의 정보처리 및 기타 과제의 수행을 향상할 수 있는 기회를 제공한다.

## 지능 연구의 전통

지능에 대한 다양한 정의가 존재하지만 통상적으로 “계산이나 문장 작성 따위의 지적 작업에서, 성취 정도에 따라 정해지는 적응 능력, 지능 지수 따위로 수치화할 수 있다”(표준국어대사전) 또는 “새로운 대상이나 상황에 부딪혀 그 의미를 이해하고 합리적인 적응 방법을 알아내는 지적 활동의 능력”(네이버 국어사전)을 지칭한다. 지능은 오랫동안 인간에게 고유한 속성이라고 간주되었으나, 그러한 통념과는 달리 지능은 인간에게만 국한된 능력이 아니다. 침팬지, 개, 쥐, 심지어 문어와 같은 동물도 일정 수준의 지능을 소지하고 있으며, 최근 들어서는 기계의 지능이 눈부신 속도로 발전하고 있다. 20세기에 들어와서 본격적으로 등장한 인공지능(artificial intelligence) 또는 기계지능(machine intelligence)은 통상 기계, 컴퓨터가 소지한 지능을 지칭하며 이해 및 지능적 행동을 할 수 있는 인공물의 생성을 다루는 컴퓨터 과학의 분야를 지칭하기도 한다(Lee & Lee, 2018).

### 개인차로서의 지능과 정보처리

심리학 내에서 지능에 대한 연구는 크게 두 갈래로 나뉘어 진행되었다. 첫째, 임상과 측정을 중심으로 이루어진 지능 연구의 전통이 존재한다. 이들 전통에서는 지능을 개인차 변인으로 간주하고 지능을 정확하게 정의하고 측정하는 데 초점을 맞추었다. 이 과정에서 다양한 지능 이론이 출현하였는데, 지능이 단일 요소가 아닌 여러 하위 요소들로 이루어져 있음이 제안되었고, 일반/범용지능(general intelligence), 특수지능, 결정지능(crystalized

intelligence), 유동지능(fluid intelligence) 등 다양한 유형의 지능이 제안되었다. 초기의 지능 이론들은 정확한 반응을 하는 능력(Thorndike), 추상적인 사고능력(Terman) 등 인간의 인지적인 능력을 지능의 핵심으로 간주하였다. 그러나 이후 이론들은 지능 개념을 정서, 사회관계조절 능력 등을 포함하는 것으로 확장하였고 실제지능(practical intelligence), 정서 지능(emotional intelligence), 다중지능(multiple intelligence) 등이 제안되었다. 이들 지능 연구의 전통은 지능에 대한 다양한 개념화를 제시하는 것은 물론 정교한 측정 이론을 바탕으로 다양한 지능 검사의 개발에 기여하였다.

심리학 내 지능 연구의 또 다른 전통으로 인지심리 또는 인지과학의 연구 전통을 들 수 있다. 정보처리 관점에서 볼 때 인간은 환경에서 정보를 수집하고, 내적으로 처리/조작하여 이를 바탕으로 환경과 상호작용하는 시스템이다. 인간이 환경에서 수집한 정보를 바탕으로 다양한 정보처리가 일어나는데, 감각과 지각, 정보 저장, 추론, 의사결정, 언어이해와 표현, 추리 등은 인간이 수행하는 대표적인 정보처리 활동이다. 정보처리 능력은 인간이 보편적으로 소지하고 있는 능력으로 인간의 지능적 행동을 가능하게 바탕이 된다. 이에 대한 연구는 심리학 내에서는 인지심리학에서 주도적으로 연구되었으나, 인지에 대한 관심은 심리학에만 국한되지 않는다. 컴퓨터 과학, 인류학, 교육학 등의 관심사이기도 한데, 학제간 연구 분야인 인지과학은 다양한 분야에서 수행되는 인지 연구를 연결하고 통합하는 역할을 해왔다.

### 정보처리이론의 대두와 인공지능의 발달

일련의 사건들이 인지과학과 인공지능 분야가 출현하는데 공통의 토대를 제공하였다. 일차적으로 20세기 초에 철학자들에 의해 이루어진 인간의 사고에 대한 형식화, 계산화 작업을 들 수 있다. 러셀(Russell)과 화이트헤드(Whitehead)는 수학의 원리가 논리학의 원리로 환원될 수 있다고 보고 애매한 일상 언어 대신에 기호논리학을 도입하여 수학 체계를 재구성하려고 시도하였다. 이 작업은 이들의 저서 *수학 원리* (Principia Mathematica; 1910-1913)에 집대성되었는데, 이 책은 이후 인공지능의 발달에 중요한 역할을 한다. 복잡해 보이던 수리 논리학의 과정이 형식화됨으로서 기계가 사고를 모사할 수 있는 기반이 되었다.

유사한 시기에 기계가 인간과 유사한 방식으로 정보처리를 할 수 있다는 생각이 대두되게 되었다. 산업 혁명 이후 다양한 기계가 사용되었지만, 대부분 물리적인 작업을 지원하는 용도였다. 기계가 인간처럼 사고하고 정보처리를 할 수 있다는 생각은 알란 튜링(Alan Turing)에 의해서 본격화되었다. 알란 튜링은 1950년 출간한 논문 “계산하는 기계와 지능 (Computing machines and intelligence)”에서 기계가 0과 1의 값을 변환시켜가면서 수학적 사고를 할 수 있음을 보여주었다. 처치(Church)와 함께 제안한 처치-튜링 논제(Church-Turing thesis)는 기계에게 충분한 시간과 메모리가 주어지면 어떤 알고리즘도 계산할 수 있다는 것이었다. 기계가 정말로 사고가 가능한지를 알아볼 수 있는 방안으로 튜링 검사(Turing test)가 제안되는데, 평가자가 인간과 기계의 답변을 구분할 수 없다면 기계가 사고할 수 있는 것으로 판별할 수 있다는 것이다.

당시 시작되던 뇌에 대한 연구 또한 기계와 인간 간에 존재하는 처리 유사성을 드러내는데 기여하였다. 뇌의 신경원이 흥분과 휴지의 두 상태를 오가며 작동한다는 것을 바탕으로 맥컬록-피츠(McCulloch-Pitts) 모델이 제안되었는데, 이들은 신경원이 흥분-휴지의 두 상태를 오가며 작동하는 것은 일종의 참-거짓의 명제 논리 상태에 상응하는 것이라고 보았다. 기계와 인간의 하드웨어는 매우 다르지만 공통적으로 흥분-휴지 또는 0-1의 상태를 오가면서 계산을 수행한다는 것이고, 이후 인공신경망(Artificial Neural Network ANN) 모형의 토대가 되었다(Lee, 2018; Russell & Novig, 2016).

이러한 배경 하에 정보이론(information theory)과 계산이론(theory of computation)이 발달하게 되었다. 클로드 섀넌(Claude Shannon)이 1948년에 출간한 논문 “통신의 수학적 이론”은 정보이론의 발달에 지대한 역할을 하였다. 섀넌은 정보(information)를 엔트로피(entropy), 즉 불확실성을 줄일 수 있는 것으로 정의하고 정보의 불확실성이 커질수록 정보의 양이 많아진다고 제안하였다. 정보는 더 이상 추상적인 개념이 아니라 측정 가능한 대상이 되었고, 정보의 측정 단위로 비트(bit)가 제안되었다. 정보이론의 발달은 자연스럽게 정보처리와 계산의 주체인 인간과 인공지능에 대한 관심으로 이어졌다. 인지과학과 인공지능의 연구는 거의 같은 시기에 시작되었는데, 인지과학의 경우 1956년 미국 MIT에서 개최되었던 정보이론(information theory)에 관한 심포지움이 인지혁명(cognitive revolution)의 시발점으로 간주된다. 이 행사에서 노암 촘스키(Noam Chomsky), 조지 밀러(George Miller) 등의 학자들이 모여 행동주의를 배격함과 동시에 인간의 정보처리 특징과 과정에 대한 논의를 시작한 것이 인지 연구가

본격화되는 계기가 되었다. 같은 해 몇 달의 시차를 두고 미국의 다트머스대학에서는 인공지능 연구의 시작을 알리는 워크숍이 진행되었다. 당시 산발적으로 인공지능 연구를 수행하고 있던 마빈 민스키(Marvin Minsky), 클로드 섀넌(Claude Shannon) 같은 연구자들이 모여 두 달간 연구 성과를 공유하는 자리를 가졌고, 이때 인공지능이라는 용어가 처음으로 사용되었다.

인공지능 연구 초기에 가장 주목을 받은 것은 뉴웰과 사이몬이 개발한 Logic Theorist(LT)라는 추론프로그램이었다. 최초의 인공지능 프로그램으로 불리우는 이 프로그램은 러셀(Russell)과 화이트헤드(Whitehead)의 *수학 원리*(Principia Mathematica; 1910-1913)의 2장에 나오는 정리들을 대부분 증명할 수 있었다. LT 및 그와 유사한 시스템들의 성공은 인공지능에 대해서 매우 낙관적인 전망을 하게 되는 근거가 되었다. 당시 사이먼은 10년 이내에 디지털 컴퓨터가 세계 체스챔피언이 될 것은 물론 20년 이내에 사람이 할 수 있는 모든 일을 기계가 할 수 있을 것으로 예측하였다. 이러한 장미빛 예측은 곧 자취를 감추게 되었는데, 인공지능은 논리적 문제를 푸는 데에는 능숙했으나 인간에게는 훨씬 더 간단한 시지각(vision)의 문제조차 해결하지 못하였다. 기타 일상의 다양한 문제에 대해서 효과적인 해결책을 제시하는데 실패하면서 인공지능 연구는 침체기를 경험하게 된다.

한동안 침체기에 들어갔던 인공지능 연구는 1980년대 두 번째 융성기를 맞게 되는데 그 중심에는 전문가 시스템(expert system)이 존재한다. 전문가 시스템은 의사 등의 영역 전문가(domain experts)가 가지고 있는 지식을 구현하여 이들이 수행하는 과제를 대신하는 인공

지능 시스템이다. 많이 알려진 MYCIN의 경우 전염성 혈액 질환을 진단하는 시스템이었다. 환자의 질문에 대한 일련의 질문에 답변을 제공하면 이를 바탕으로 시스템이 진단과 함께 치료에 필요한 항생제를 처방하였다. 당시 초보 의사와 비교해서 유사한 수행을 하는 것으로 보고되면서 주목을 받았다(Buchanan & Shortliffe, 1984). 그러나 시간이 지나면서 전문가 시스템의 한계가 드러나기 시작하였다. 전문가 시스템은 전문가가 소지한 명시적 지식은 잘 구현하였지만 암묵적 지식은 잘 구현하지 못하였다. 명시적 지식과는 달리 암묵적인 지식은 별도의 방법론을 사용하여 전문가로부터 추출하지 않으면 안 되는데, 그 자체가 상당한 시간과 비용을 요구하는 작업이었다(Chi, 2006). 또한 인간 전문가의 경우 새로 발견되는 지식과 제도의 변화를 반영하여 자신의 지식 기반을 스스로 업데이트하는 반면 전문가 시스템은 스스로 학습하는 능력이 부재하여 엔지니어가 끊임없이 변화된 내용을 입력해 주지 않으면 안 된다는 문제가 존재하였다. 전문가 시스템의 이러한 문제점이 노출되면서 인공지능 연구는 다시 암흑기를 맞게 된다.

인공지능 연구는 2010년 이후 현재 3차 융성기를 맞이하고 있다. 이 시기의 인공지능은 딥러닝(Deep Learning)으로 대표되는 기계학습 기술의 발전으로 특징지어진다. 딥러닝은 GPU(Graphical Processing Units) 프로세서의 개발, 인터넷에 의해 가능하게 된 방대한 양의 빅데이터를 기반으로 인공신경망을 활용하는 기계학습 기법을 지칭한다. 딥러닝 기법에 힘입어 난제로 여겨지던 시지각(vision) 분야에서 비약적인 발전이 일어났고, 기계의 이미지 인식 능력은 일부 과제에서는 인간을 넘어서는 수준에 도달하고 있다. 또한 딥블루(Deep Blue),

알파고(Alpha Go), 왓슨(Watson) 등은 인공지능이 여러 영역에서 인간의 최고 전문가의 수행을 뛰어넘는 수준까지 도달했음을 보여주었다. 이러한 성과에 힘입어 인공지능은 일반 대중을 비롯하여 정책입안자들에게도 많은 주목을 받게 되었고, 4차산업혁명의 핵심 기술로 간주되면서 빠른 속도로 성장하고 있다. 최근 이루어지고 있는 인공지능 연구개발은 다른 어떤 시기보다 기업들에 의해서 주도되고 있는데, 현실에서 사용 가능한 형태의 인공지능을 개발하고 이를 상용화하는 데 노력이 집중되고 있다.

인공지능의 성공에도 불구하고 인공지능이 보이는 행동이 과연 진정으로 지능을 의미하는 것인지에 대한 논쟁이 끊이지 않는다. AI 효과(AI effect)는 지능의 지표로 간주되는 행동(예, 체스, 인간과의 대화)을 기계가 막상 구현하고 나면 사람들은 더 이상 그 행동이 지능적이지 않다고, 기계가 단순한 계산을 수행한 것에 불과하다고 주장하는 현상을 지칭한다(Haenlein & Kaplan, 2019). 이러한 현상은 인간이 여전히 특별한 존재이고자 하는 동기에서 기계의 수행을 평가절하하기 때문에 발생할 수 있으나 동시에 인간이 ‘지능’이라고 하는 개념에 여전히 어떤 신비함을 부여하기 때문일 수도 있다. 복잡한 과제 수행도 일련의 계산과정으로 분해되고 나면 신비함이 사라지게 되고, 기계의 수행은 일련의 절차와 계산의 적용일뿐 더 이상 지능적으로 느껴지지 않을 수 있다. 최근 부각되는 약한 인공지능(weak AI)과 강한 인공지능(strong AI)의 구분은 인공지능이 정말 인간처럼 사고할 수 있는가, 지능을 가지고 있는가에 대한 논쟁의 또 다른 모습이다(Searle, 1980). 약한 인공지능은 다양한 계산을 수행할 수 있으나, 스스로 사고하고

문제를 해결하는 능력이 없는 인공지능을 지칭한다. 알파고나 왓슨 등 현재 많이 알려진 시스템들은 이 범주에 속한다. 반면 강한 인공지능은 단순히 계산을 잘 하는 것뿐만 아니라 자신의 인지 상태(cognitive states)를 이해하고 자각하는 의식(consciousness)을 가진 인공지능을 지칭하고 인간의 지능에 좀 더 근접한 것으로 간주된다.

범용 인공지능(Artificial General Intelligence; AGI)과 특수 인공지능(Artificial Narrow Intelligence; ANI)을 구분하기도 하는데, 이는 인공지능의 수행이 특정 문제 영역에 국한된 것인지 아니면 일반적인지를 기반으로 이루어진다. 인간의 지능은 범용이다. 즉, 인간의 경우 바둑을 잘 두는 것뿐만 아니라 책도 읽고, 운전도 하고, 계산을 하는 등 여러 영역에서 지능적 수행을 보일 수 있다. 현재까지 개발된 인공지능은 대부분 바둑, 게임 등의 특정 영역에서의 문제해결과 수행에 전문화되어 있다. 특수인공지능 시스템은 해당 영역에서의 문제해결에는 뛰어난 능력을 발휘하지만, 그 밖의 영역에서는 전혀 지능적인 수행을 보이지 못한다.

강한 인공지능, 범용 인공지능이 과연 가능한지, 언제 가능한지 과연 바람직한지의 논쟁은 현재 진행 중이다(Fjelland, 2020; Hildt, 2019). 하지만 현재 구현된 약한 인공지능, 특수 인공지능만으로도 이미 인간 삶의 영역에 방대한 영향을 미칠 것으로 기대된다. 인공지능이 인간에게 미치는 영향을 검토하고 인간 심리 및 인간인지에 갖는 함의를 고민하고 준비하지 않으면 안 된다.

## 인간지능과 인공지능의 공통점: 인공지능의 역사가 주는 통찰

뇌를 기반으로 작동하는 인간의 지능과 차가운 무기물로 이루어진 인공지능은 일견 매우 상이한 존재이나 하드웨어 상의 차이에도 불구하고 인간과 인공지능은 정보처리 시스템으로서의 공통점을 가지고 있다. 인간의 정보처리는 컴퓨터의 작동에 종종 비유된다. 외부의 정보를 수집하는 인간의 감각 기관은 정보를 컴퓨터의 입력체계인 키보드에, 작업기억은 RAM, 장기기억은 하드디스크로 비유되곤 한다. 인간지능과 인공지능의 공통점과 차이점은 비교의 수준에 따라 달라진다. 데이비드 마(David Marr)는 정보처리 시스템을 3개 수준으로 구분하였는데, 계산론적(computational) 수준, 알고리즘/표상(algorithmic/representational) 수준, 그리고 물리적/구현적(physical/implementation) 수준이 그것이다. 계산론적 수준은 정보처리 체계가 하는 일, 목적에 해당된다(what and why of computation). 알고리즘 수준은 특정 계산 목적을 구현하는 방법(how of computation)을 지칭하고, 물리적/구현적 수준은 해당 정보처리 시스템이 물리적으로 구현되는 수준, 즉, 하드웨어를 지칭한다. 인간과 인공지능의 경우 하드웨어는 다르지만, 처리의 목적 및 방법/알고리즘에서 공통점을 찾을 수 있다. 인간과 인공지능은 처리 목적을 공유할 수 있는데, 예를 들어 인간의 시지각과 컴퓨터의 이미지 처리는 모두 입력 자극을 처리하여 환경의 주요한 사건을 탐지하는 계산 목적을 가지고 있다. 동일한 계산 목적에 대해서 이를 달성하는 방안은 다양하게 존재하고 기계와 인간이 동일한 알고리즘을 사용하는 것도 가능하다. 하지만 알고리즘의 사용

은 하드웨어에 의해 제약될 수밖에 없는데, 눈이 두 개인 인간과 그러한 제약이 존재하지 않는 기계의 경우 깊이와 움직임 지각에 사용하는 알고리즘이 동일할 수 없다.

하드웨어의 발달에 따라 인간과 기계가 사용하는 알고리즘 간의 유사점은 점점 줄어들고 있지만 그럼에도 매크로한 수준에서 정보처리 시스템으로서 인간과 기계가 가진 공통점이 존재한다. 이와 관련하여 인공지능의 역사는 유용한 통찰을 제공한다. 인공지능은 1950년대 시작된 이후 현재에 이르기까지 주기적인 융성기와 침체기를 겪으면서 발전하였다. 각 시기에 주도적으로 사용된 접근은 지능적 수행의 핵심이 무엇인지에 대한 당시의 이해를 반영하였는데, 해당 접근이 한계에 직면하면서 지능적 수행에 필요한 요소가 무엇인지에 대한 이해가 확장되었다. 이 절에서는 인공지능의 역사를 바탕으로 인간과 인공지능에 공통적인 작동 원리가 무엇인지를 살펴본다.

### 전략과 알고리즘

1950년대 인공지능 연구가 시작되면서 초기 성공은 1956년 뉴웰과 사이몬이 개발한 Logic Theorist(LT)의 성공에 크게 힘입었다. 앞에서 살펴보았듯이 LT는 러셀과 화이트헤드의 작업을 계산론적으로 구현한 것인데, 그 핵심은 수학 증명을 하는 계산 알고리즘의 개발이다. LT의 성공에 힘입어 다양한 영역에서 과제 수행을 가능하게 하는 알고리즘을 개발하려는 시도가 활발하게 이루어졌다. 일반문제해결자(General Problem Solver; GPS)는 논리적 증명에 특화된 LT와는 달리 수학 이외의 여러 문제에 대해서 해결책을 찾는 것을 목적으로 하였다

(Newell & Simon, 1972). 일련의 잘 정의된 문제(well-defined problems)에서 주로 적용이 이루어졌는데, 대표적으로 하노이 탑, 선교사와 식인종 문제, 애너그램(anagram) 문제를 들 수 있다. 이들은 문제에 대한 지식과 전략을 구분하고, 다양한 문제 상황에 공통적으로 사용할 수 있는 전략을 구현하고자 하였다. 문제를 푸는 것은 문제공간(problem space)을 구성하고 그 내에서 목표를 찾아가는 검색과정으로 간주되었는데, 이 과정에서 자주 사용되는 전략으로 수단목표분석(means-end analysis), 언덕오르기(hill-climbing) 등이 존재한다. 수단목표분석은 문제가 주어질 때 현재의 상태와 목표 상태를 확인하고 그 차이를 줄여가는 전략을 지칭한다. 언덕오르기 전략은, 최적의 해결책을 찾기 어려운 경우 시용되는데, 가용한 선택지 중 목표 상태와 현 상태의 차이를 조금씩 줄여가는 선택을 하는 전략이다. 초기 인공지능 시스템의 핵심은 이와 같은 전략과 알고리즘을 만드는 것에 있었다. 일견 복잡해 보이는 처리도, 일단 'if-then' 진술문의 형태로 구현되면 기계의 뛰어난 연산 능력의 도움을 받아서 인간과 유사한 또는 뛰어넘는 수행을 보이는 것이 가능하였다. 기계는 인간보다 뛰어난 연산 능력을 소지하고 있을뿐만 아니라 인간처럼 의미적인 요인에 의한 처리 방해를 경험하지 않기 때문에 복잡한 계산을 더 빠르고 정확하게 처리할 수 있었다.

전략과 알고리즘을 강조한 초기 접근이 일정 부분 성공을 거두었으나 곧 한계에 부딪히게 되었다. 논리적인 문제들은 잘 풀었지만, 일상적인 문제들을 잘 해결하지 못하였는데, 대표적으로 시지각의 문제를 들 수 있다. 인간에게는 매우 손쉬운 지각의 문제는 인공지능에게는 매우 복잡한 문제로 드러났는데,

1957년에 프랑크 로젠블라트(Frank Rosenblatt)는 사람의 뇌에 있는 신경조직을 인공적으로 재현한 퍼셉트론(Perceptron)을 제안하였다. 다수의 입력을 받아서 출력하는 일종의 인공신경망이었는데, 단순한 구조임에도 패턴 인식이 가능하여 초기에는 많은 주목을 받았다. 그러나 선형분리가 가능한 문제밖에 풀지 못한다는 한계가 지적되고 일련의 한계에 직면하면서 인공지능 연구는 인공지능의 겨울(Winter)이라고 불리는 첫 번째 침체기를 맞게 되었다(Lee, 2018; Russell, & Novig, 2016).

이 시기 인공지능 연구가 당면한 문제는 다양했지만 그 중 하나는 지능적 처리를 구현하는 데 있어 과도하게 전략과 알고리즘에 의존한 것이다. 알고리즘의 효과와 성능은 그 자체만으로 가능하지 않고 사용하는 자료의 크기와 질에 의존한다. 이는 인간의 정보처리에서도 분명하게 드러나는데, 적절한 처리 전략의 사용이 중요하지만 입력 자극을 어떻게 처리하는지 못지않게 과거 경험과 지식을 어떻게 활용하는가가 핵심적인 역할을 한다. 예를 들어, 주어진 정보를 기억할 때 어떤 부호화 전략을 사용하는가가 기억의 질에 영향을 주지만, 부호화 전략의 사용은 기억에 저장된 정보에 의해 영향을 받는다. 기억에 저장된 정보에 따라서 어느 정보가 부호화 될지, 어떤 방식으로 정교화되는 지가 달라지게 된다. 인지 연구자들은 일찍이 개념정보에 의해서 주도되는 하향처리(top-down processing)와 입력 자극에 의해서 주도되는 상향처리(bottom-up processing)를 구분하였는데, 인간의 정보처리는 장기기억에 저장된 다양한 형태의 지식, 경험에 적극적으로 의존한다. 온전히 자료에만 기반한 처리는 계산론적으로 상당한 처리 부담을 초래할 수 있는데, 기억에 저장된 정보를

바탕으로 한 하향처리는 입력정보에서 부족한 정보를 메꾸거나 효율화하는 것을 가능하게 한다.

### 지식과 경험

전문성(expertise)은 학습과 훈련 과정에서 습득된 지식과 경험이 일정 수준 이상으로 축적된 것을 지칭한다. 인지 연구의 초기에서부터 전문성의 획득과 영향 요인에 대한 연구가 활발하게 이루어졌는데, 초기 연구된 문제 영역 중의 하나는 체스였다. 체스는 제한된 문제 공간에서 일련의 규칙을 적용하여 목표를 이루는 게임으로 적절한 게임 전략을 구사하는 것이 중요하다. 체스 전문성에 대한 초기 가설은 체스 마스터와 같은 전문가들은 특별한 능력, 예를 들면 높은 지능이나 큰 작업 기억 용량, 빠른 정보처리 속도 등을 소지했을 것이라는 것이었다. 체스 전문가와 초보자를 비교한 일련의 연구들은 그러나 두 집단을 구별짓는 것은 선천적으로 타고나는 능력의 차이가 아니라 전문성을 획득하는 과정에서 습득한 체스에 대한 지식임을 보여주었다(Chase & Simon, 1973). 체스뿐만이 아니라 다양한 과제 영역에서 전문성의 핵심이 선천적인 능력보다는 해당 영역에서 대한 지식임이 드러나게 되었다. 전문성의 핵심은 일차적으로는 지식에 있는 것으로 보이는데, 전문가가 되는 과정에서 해당 영역과 관련된 다양한 지식이 습득하게 된다. 이렇게 습득된 지식은 입력 정보에 대한 효율적인 처리를 안내하는데, 그 결과 전문가와 초보자는 소지한 지식의 양뿐만 아니라 질적으로 다른 처리 전략을 사용하는데 된다(Chi, Glaser, & Farah, 1991).

인지과학의 전문성 연구와 발맞추어 인공



지능 분야에서도 전문가 시스템(expert system)의 개발이 진행되었다. Dendral은 최초의 전문가 시스템으로, 분자의 기본 화학식(예,  $C_6H_{13}NO_2$ )과 질량 스펙트럼(예,  $m=15$ )을 주면 분자구조를 유추하였다. 이 문제를 풀기 위해서는 주어진 화학식에 대해서 가능한 분자구조를 유추하고 각각에 대한 질량 스펙트럼을 예측하고 이를 관측값과 비교하여 적합한 분자구조를 찾아내어야 한다. 분석화학자들에게 자문을 구한 결과 이들이 스펙트럼에서 분자의 공통 하위 구조를 암시하는 패턴에 대한 지식을 활용한다는 것이 드러났는데, 해당 분자에 대해 특정 하위 구조가 존재하는지 아는 경우 가능한 후보들의 수가 엄청나게 줄어들게 된다. 분석화학자들의 문제 풀이에 사용하는 지식을 특수 목적 규칙으로 만들어 Dendral에 반영하였고, 그 결과는 아주 성공적이었다. Dendral의 성공은 이후 혈액전염병 진단에 사용된 MYCIN 시스템의 개발로 이어졌다(Buchanan & Shortliffe, 1984). 이러한 성공에 힘입어 전문가 시스템은 산업으로 성장하게 되었고 1982년에는 최초의 상용 전문가 시스템이 사용된 후 여러 분야에서 상당한 비용 절감을 이루는 데 기여하였다. 또한 지식 획득과 표현의 문제가 인공지능 연구의 중심에 자리잡게 된 계기로 작용하게 되었다(Lee, 2018; Russell & Novig, 2016).

전문가 시스템이 보인 성공에도 불구하고, 전문가 시스템을 향상하고 유지하는 문제는 상당히 어려운 것으로 드러나게 되었다. 전문가 시스템의 핵심 구성 요소는 전문 지식과 이를 활용하는 추론 규칙이다(Hayes-Roth, Waterman, & Lenat, 1983). 인간 전문가가 소지한 지식이 전문가 시스템을 구성하는데 사용되게 되는데, 전문가들이 가진 지식은 크게

서술지식과 암묵지식(implicit knowledge)으로 구분될 수 있다. 서술지식은 명시적인 지식(explicit knowledge)으로 전문가를 인터뷰 하거나 또는 전문 서적을 통해 추출 가능하기 때문에 시스템에 구현하는 것이 상대적으로 용이하다. 그러나 전문가가 오랜 경험을 통해서 습득한 암묵적인 지식이나 표상 구조의 차이는 이러한 방법으로 포착이 쉽지않다. 이에 대한 해결책으로 전문성을 드러내는 과제(예, 체스 경기에서 다음 수를 결정하기)에서 이들의 수행을 관찰하고 이를 전문성과는 무관한 과제(예, 체스판의 배열을 기억하기)에서의 수행 또는 비전문가의 수행과 비교하는 방법이 사용되었다(Chi, 2006). 전문성이 심화될수록 암묵적 지식을 추출하는 것이 어려워졌고 프로그래머와 엔지니어들이 수행하기에는 버거운 과제가 되었다. 암묵지를 추출하는 데 성공해도, 해당 분야의 전문가의 수행을 따라가기에는 역부족인 경우가 종종 발생하였는데, 전문가 시스템의 경우 해당 영역에 대한 지식을 많이 가지고 있으나 상식(common sense)이라고 할 일상적인 지식은 가지고 있지 않았기 때문이다. 예를 들어 MYCIN에게 장에 콜라라균(*Cholerae Vibrio*)에 감염된 환자에 대해서 질문하자 MYCIN은 두 주동안의 테트라사이클린(tetracycline)이라는 약물을 처방하는데, 이러한 처방은 박테리아를 죽일 수 있지만 동시에 환자도 죽일 수 있는 양이었다. 상식의 결여는 전문가 시스템이 인간 전문가 수준의 수행을 하는 데 걸림돌이 되었다(Lee, 2018; Russell & Novig, 2016).

가장 큰 걸림돌로 작용한 것은 학습 능력의 부재였다. 전문가 시스템이 지속적으로 사용되기 위해서는 기반이 되는 지식이 끊임없이 업데이트 되지 않으면 안 된다. 새로운 바이

러스가 출현할 수도 있고, 새로운 연구 결과나 치료 기법이 도입되거나, 병원이나 정부의 정책과 제도가 바뀌면서 진단 기준을 바꾸어야 하는 경우도 존재한다. 당시의 전문가 시스템들은 스스로 지식 기반을 업데이트 할 수 있는 능력이 부재하였기 때문에 매년 엔지니어들이 다시 프로그램 하지 않으면 안되었다. 이러한 문제점들이 드러나면서 인공지능 연구는 두 번째의 겨울을 맞이하게 되었다(Lee, 2018; Russell, & Novig, 2016).

### 학습과 딥러닝

초기 전문가 시스템의 실패는 지능적인 시스템에서 학습 능력이 필수적임을 보여주었다. 학습 능력은 인간을 포함한 모든 유기체의 생존에 매우 중요한 능력인데, 인공지능 또한 환경 속에서 독립적인 존재로 살아가기 위해서는 물리적, 사회적 환경의 변화에 맞추어 스스로의 지식 기반을 업데이트 하고 행동을 조정하는 것이 필요하다. 기계학습은 기계가 스스로 학습을 통해서 지식을 축적하고 구조화하는 능력을 지칭한다. 현재 기계학습을 위해 사용되는 기법에는 크게 세 가지가 존재하는데, 강화학습(reinforcement learning), 지도학습(supervised learning), 비지도 학습(unsupervised learning)이다. 강화학습의 경우 기계가 특정 행동을 한 다음 그 결과물에 대한 강화(예, 맞고 틀리고의 피드백)를 제공하여 목표 행동을 학습하도록 하는 기법이다. 강화학습은 에이전트(agent) 학습에 유용한데, 에이전트가 환경을 탐색하면서 받는 피드백을 통해서 적응적인 행동을 학습하는 데 효과적이다. 예를 들어, 에이전트가 앞에서 굴러오는 공을 잡는 것을 학습한다고 할 때 에이전트는 공을 잡기 위해

서 다양한 시점에 다양한 방향으로 움직일 수 있다. 에이전트가 특정한 움직임을 선택한 다음 에이전트가 공을 잡거나 놓치는 결과가 발생하는데, 이 결과는 일종의 강화물로 이후 에이전트의 움직임을 수정하는 데 활용되게 된다. 에이전트는 초기에는 무선적으로 행동하지만 학습이 지속되면서 누적된 보상을 최대로 하는 움직임 행동의 알고리즘을 찾아가게 된다. 이는 전형적인 조작적 조건형성(operant conditioning)의 과정인데, 이 과정을 통해서 인공지능 에이전트는 스스로 환경을 탐색하고 상호작용하면서 최적의 행동 패턴(예, 움직이는 공을 잡거나 게임에서 상대 무찌르기)을 찾아낼 수 있게 된다(Gold & Scassellati, 2009).

지도학습과 비지도 학습 기법은 기계가 자료에 존재하는 패턴을 학습할 때 주로 사용되는 기법이다. 예를 들어 이미지를 보고 사람인지 아닌지를 판별하거나 필기한 내용을 보고 문자, 단어 인식을 학습하는 것 등을 들 수 있다. 지도학습과 비지도학습의 차이는 시스템을 훈련시킬 때 주어지는 자료의 정답 또는 값(예, 특정 이미지가 사람인지)이 주어지는지의 유무에 있다. 지도학습의 경우 사례의 값이 알려진 훈련 자료가 주어지기 때문에 기계는 자료와 자료의 값(어느 이미지가 사람이고 어느 이미지가 고양이인지)을 연결하는 규칙을 학습하면 되는 반면, 비지도 학습(자율학습이라고도 불린다)의 경우 훈련 사례만 제시되기 때문에 기계가 스스로 분류 규칙과 분류에 사용되는 범주를 학습해야 한다. 지도학습의 과정은 아동의 개념 학습 과정에 비유될 수 있는데, 아동이 주변 사물의 명칭을 학습할 때 주위의 어른이 특정 사례의 명칭(예, 개인지 고양이인지)을 알려주는 상황에서 일어

나는 학습에 상응한다. ‘정답’이 주어지면 대응 규칙을 찾는 것, 예를 들면 무엇이 고양이의 특징이고 개의 특징인지를 찾아내는 것이 수월해 진다. 하지만 항상 누군가가 정답을 알려주는 것이 가능하지 않은데, 비지도 학습은 인간이 새로운 환경이나 자극에 직면하여 스스로 분류 체계(예, 여행에서 찍은 사진을 몇 개의 주제로 묶을지)를 만들고 분류에 사용되는 규칙(예, 여행 준비 사진도 포함한다)을 정해야 하는 경우도 존재한다. 비지도 학습의 경우 정답이 주어지지 않기 때문에 자극의 규모와 복잡성에 따라 패턴 인식과 추출, 범주화 등 다양한 정보처리가 관여하며 지도 학습에 비해서 훨씬 더 많은 시간과 노력이 필요로 한다.

현재 많은 관심을 받는 딥러닝은 비지도 학습 상황에서 주로 사용되지만 그 핵심 특징은 학습을 구현하는데 다층의 구조를 가진 인공신경망을 사용한다는 것이다. 초기 퍼셉트론 연구에서 시작된 인공신경망은 뇌의 신경원이 작동하는 방식으로 계산을 수행하는 시스템이다. 인공신경망을 사용하는 경우 학습의 결과는 (특정 노드에 개념적으로 저장되지 않고) 신경망 내의 다양한 단위와 연결에 분산되어 저장된다. 학습의 결과 입력과 출력의 함수 관계가 변화하게 되는데, 이때 입력층(input layer)과 출력층(output layer) 사이의 은닉층(hidden layer)의 존재유무 및 수에 따라 망 전체의 계산 능력이 달라진다. 딥러닝은 이러한 은닉층을 활용한 인공신경망을 지칭하는데, 은닉층이 늘어날수록 처리해야 하는 변수가 증가하기 때문에, 신경망을 훈련시키기 위해서는 방대한 양의 데이터와 연산 능력이 요구된다. 초기의 신경망에서 1-2개 정도의 은닉층을 가진 신경망이 훈련 가능하였으나, 최근에

는 10개층 이상의 은닉층을 가진 신경망이 사용되고 있다. 알파고의 경우 12개의 은닉층을 가지고 있었다. 딥러닝은 불완전한 데이터, 모호한 자극을 처리하는 데 유용한데, 초기에 시각 이미지 인식에 활발하게 사용되어 많은 성과를 내었다. 딥러닝 기법은 GPU(Graphical Processing Unit)와 빅데이터와 결합하여 다양한 영역에서 많은 성과를 보이고 있으나, 인지를 다시 블랙박스화시킨다는 문제점을 가지고 있다. 즉, 주어진 자료를 성공적으로 분류하지만 왜 그러한 분류를 하였는지를 설명하지 못한다는 것이 주요한 단점으로 지적되고 있다.

딥러닝을 포함한 기계학습 기법의 발달과 성공은 지능적인 시스템의 구현에 학습 능력이 핵심적임을 보여준다. 학습은 심리학의 오랜 연구 주제인데, 행동주의의 영향으로 심리학 연구 초기부터 학습에 대한 연구가 시작되었다. 당시 행동주의는 인간의 내적인 활동과 내적으로 일어나는 처리에 대한 고려를 금기시하고 외적으로 발생하는 행동에 초점을 맞추었는데, 그 결과 지각(perception)은 변별(discrimination)이 되었고, 언어(language)는 언어 행동(verbal behavior)이 되었다(Miller, 2003). 인지혁명(cognitive revolution) 이후 내적으로 일어나는 처리에 대한 연구가 시작되었으나 인지 연구자들은 학습의 문제를 기본적으로 기억의 문제로 보았다. 정보를 습득해서 이를 장기기억에 저장했다가 이후 필요한 상황에서 인출하는 것이 학습이라고 보았는데, 따라서 입력 자극에 대한 적절한 주의, 효과적인 부호화 전략의 사용, 및 적절한 인출단서의 사용이 중요하게 간주되었다. 이러한 접근은 정보의 습득에 관여하는 다양한 기억 기계와 전략의 효과를 밝히는 데 공헌하였으나, 동시에 학습의 문제를 기억의 문제로만 치환하였다는 문

제점을 낳았다. 학습은 단순히 정보 저장의 문제가 아니라 환경에 존재하는 규칙성에 대한 탐색과 추론 및 이를 바탕으로 한 행동 조절을 포함한다. 학습을 위해서는 기억뿐만 아니라 패턴 인식, 관계 추론, 의사결정 등 다양한 인지과정이 관여하며 또한 유기체가 환경과 상호작용 하는 과정에서 조건형성, 행동조성, 동기 등 다양한 요인이 영향을 미친다. 기존의 인지적 접근에서는 학습의 문제를 너무 좁게 접근하였다는 아쉬움이 존재하는데, 기계학습의 발전은 역사적으로 인간의 학습에 대한 접근이 얼마나 제한적이었는지를 보여준다.

환경에 존재하는 규칙성과 패턴을 학습하고 환경과 상호작용하면서 스스로의 행동을 수정할 수 있기 위해서 학습 능력은 필수적이다. 딥러닝의 발전에 힘입어 인공지능의 학습 능력은 지속적으로 개선되고 있고 다양한 영역에서 성공이 보고되고 있다. 딥러닝에 의해서 주도되는 현재의 세 번째의 인공지능 융성기가 과거의 인공지능 연구처럼 또 다른 거울을 맞이하게 될지, 만약 또 다른 침체기를 맞이하게 된다면 어떠한 한계점 때문인지는 아직 불분명하다. 그 한계점이 무엇이든지 인공지능에 대한 연구는 지능과 지능적 수행에 대한 우리의 이해를 지속적으로 확장할 것으로 기대된다. 이와 동시에 정보처리 시스템으로서의 인간과 인공지능의 공통점과 차이점 또한 더욱 분명하게 드러날 것으로 보인다.

### 공생의 역사: ITS를 중심으로

인공지능 개발에 있어 인지과학과 컴퓨터 과학은 오래 전부터 공생의 관계를 맺어 왔다.

대표적인 인공지능 시스템 중의 하나인 ITS (Intelligent Tutoring System)는 컴퓨터를 사용하여 맞춤형 학습을 제공하려는 시도로 컴퓨터의 출현 초기부터 그 개발이 시작되었다(Jeong, 2009). ITS의 개발은 여전히 진행 중이지만 다른 어떤 시스템보다 인지과학과 밀접하게 상호작용하면서 진행되었다(Aleven, & Koedinger, 2002; Anderson et al., 1987; Chi et al., 2001; Fox, 1993; Ohlsson, 1986; Ritters et al., 2007; Merrill et al., 1992; VanLehn, 2011). ITS 개발은 인간 튜터링에 대한 연구에 의해서 영향을 받았을 뿐만 아니라 인간 튜터링에 대한 연구를 촉발시켰는데, 다른 영역에서도 인지과학과 인공지능 간의 생산적 협력 관계의 모형이 될 수 있을 것으로 보인다. 이 절에서는 ITS의 구성 요소 별로 인간 학습자 및 교수자에 대한 연구가 어떻게 ITS 개발에 반영되고 기술과 접목되었는지를 살펴보는데, ITS의 작동은 세 가지 핵심 모형에 의존한다. 교수하고자 하는 지식 영역에 대한 도메인 모형(domain model), 교수 모델(pedagogical model) 및 학습자 모델(student model)이 그것이다.

### 도메인 모형

도메인 모형(domain model)은 교수해야 할 대상이 되는 영역 지식에 대한 모형을 지칭한다. 물리학을 교수하는 경우 물리학 지식이, 수학을 교수하는 경우 수학 문제 풀이의 방법 등이 해당된다. 도메인 모형은 ITS가 학습자에게 제시할 학습 내용/문제를 선정하고 학생들의 수행을 평가하는 데 관여한다. 초기 ITS는 주로 수학, 물리학, 프로그래밍 등의 영역에서 주로 개발되었는데(Ritter et al., 2007; VanLehn et al., 2005), 이들 영역의 지식이 상대적으로

잘 구조화되어 있고 정의될 수 있기 때문이다. 교수해야 하는 지식에 대한 명확한 정의에 대한 구조가 존재하지 않으면, 컴퓨터가 효과적으로 학습 과정을 안내하거나 피드백을 제공하기 어렵기 때문이다(Alkhatlan & Kalita, 2018).

도메인 모델이 교수 과정에서 어떻게 사용되는지에 따라 의미망(semantic networks), 생산 규칙(production rule), 또는 제약(constraints) 등의 다양한 형태로 표상된다(Martin, 2002). 초기에는 영역 전문가들의 도움을 받아 해당 분야의 대표적인 교재 및 핵심 자료를 바탕으로 도메인 모델이 만들어졌다. 영역 전문가의 도움이 여전히 필수적이지만 기계학습 기법이 발달하기 시작하면서 인공지능이 이 과정을 지원하기 시작하고 있다. 예를 들어 최근 개발되고 있는 대화형 인공지능 튜터 중의 하나인 WDBT(Watson Dialogue-Based Tutor)에서는 전문가가 핵심 교재를 선정하는 역할을 하고 나면, 시스템이 교수되어야 하는 목표 지식을 교재에서 자체적으로 추출하여 도메인 모델을 만든다(Chang et al., 2018).

초기 ITS에서 도메인 모델은 학생이 반드시 습득해야 할 목표 지식으로 간주되었다. 그러나 최근에는 해당 영역의 교수 활동을 안내하기 위한 참조의 틀로 사용되기도 하는데, 이러한 추세는 대화형 ITS에서 두드러진다. 예를 들어, WDBT에서는 도메인 모델 내의 목표 지식을 활용하여 학생들에게 질문을 던지거나 피드백을 제시한다. 이러한 교수법을 채택하는 경우 도메인 모델이 엄격하게 정의되지 않아도 되는데, 최근 심리학, 사회학, 초인지 기술(meta-cognitive skills) 같은 잘 정의되지 않은 영역에서도 ITS의 사용이 시도되는 배경이 되고 있다(Holmes, Bialik, & Fadel, 2019/2020; Lane

et al., 2013; Roll et al., 2007).

#### 교수 모형

교수 모델(pedagogical model)은 도메인 모델 내의 지식을 어떠한 방식으로 학생들에게 교수할 것인지에 대한 것이다. 예를 들어, 수학 문제풀이를 교수할 때, 먼저 배경 지식과 풀이 방법을 설명하고 문제를 풀게 할지, 문제를 먼저 풀게 하고 설명을 할지, 아니면 많은 문제 풀이를 통해 스스로 관련 규칙을 깨닫게 할지 등의 다양한 교수 전략이 존재한다. 초기의 ITS 교수 모형은 검증된 지식과 절차의 전달을 강조하였다. 그러나 구성적인 학습 활동의 효과성이 알려지면서 자기에게 설명하기(self-explanation)와 같은 학습자의 적극적 학습 활동을 유도하는 교수 방식이 사용되고 있다(Rosé et al., 2001). 예를 들어, 정답을 바로 학생들에게 제시하기보다는 학생들이 스스로 문제를 다양한 방면에서 검토하고 해법을 찾아가도록 유도하거나, 문제를 풀기 전에 학생들에게 어떠한 방법을 통해서 풀지를 설명하도록 하는 것을 들 수 있다(Rau, Aleven, & Rummel, 2009). 최근 들어서 많이 사용되는 대화형 ITS에서는 지식이나 풀이 방법을 알려주기 보다는 대화를 통해서 학생들이 스스로 학습하도록 유도하기도 한다.

ITS 교수 모형의 개발에는 인간 튜터와 학습자에 대한 연구가 중요한 역할을 하였다. 초기 많은 주목을 받은 것은 소크라테스식 교수법(Socratic tutoring)이었는데, 소크라테스식 교수법의 특징은 질문을 통해서 학습자가 답을 하는 과정에서 스스로 정답을 찾아가도록 하는 것이다. Collins와 Stevens(1982)는 최고의 교사들이 소크라테스식 교수법을 사용함을 보고하였고, 이는 소크라테스식 교수법을 사용

하는 ITS의 개발 시도로 이어졌다(Rosé et al., 2001). 그러나 소크라틱 소크라테스식 교수법 같은 정교한 교수 기법은 그 자체로 실시의 난이도가 높고 일상적인 일대일 교수 상황에서 자주 사용되지 않음이 보고되었고, 이후의 ITS 시스템에서는 보다 통상적인 튜터링 대화의 패턴을 구현하는데 집중하였다(VanLehn, 2011). Graesser, Person, 및 Magliano (1995)는 인간의 튜터링 대화를 분석하여 튜터와 학습자의 대화가 다음의 다섯 단계를 따라 진행됨을 제안하였다: (1) 튜터가 질문을 하거나 문제를 제시한다; (2) 학습자는 답변을 시도한다; (3) 튜터는 학습자의 답변을 간략하게 평가한다; (4) 튜터와 학생이 대화를 통해 함께 답변을 향상시킨다; (5) 튜터는 학생들이 이해했는지를 점검한다. 이를 바탕으로 대화형 ITS의 초기 시스템인 오토튜터(Auto-tutor)가 개발되었다. 초기에는 텍스트를 기반으로 학습자와 상호작용하였으나, 자연어처리의 발달과 함께 음성인식 기술을 사용한 다양한 대화형 ITS들이 최근 개발되고 있다(Chang et al., 2018; VanLehn, Jordan, & Litman, 2007).

#### 학습자 모형

학습자 모델(student model)은 학습자에 대한 모델로 학습자에 대한 다양한 정보를 포함한다. 예를 들어, 학습자가 이제까지 학습한 내용과 성취 수준, 아직 모르는 내용, 과거에 범한 오류, 자주 틀리는 문제 유형, 자주 사용하는 방략 등이 그 안에 포함될 수 있다. 또한 학습자의 과거 학습 경력, 학습자의 주의 집중 능력 등 행동 특성에 대한 정보가 포함될 수도 있다. 학습자 모형은 ITS에서 추구하는 개인화된 학습의 구현에 핵심적이다. 학생에

게 부족한 지식을 맞춤으로 교수하기 위해서는 학생이 무엇을 알고, 어떤 부분의 지식과 문제 해결 능력이 부족한지를 파악하고 이를 바탕으로 학생이 잘못 알고 있는 부분에 대해서 피드백을 또는 추가 설명을 제공할 수 있기 때문이다. 또한 학습자의 다음 행동을 예측한다거나, 새로운 진도를 시작해야 할지, 대안적인 학습 방법을 학습할지 등의 다양한 교수 관련된 결정을 내리는 기반이 된다.

학습자를 진단하는 문제는 측정의 문제이기도 한데, 전통적인 심리측정이나 교육측정에서는 시험, 검사 등에서 참가자들의 답변을 바탕으로 이들의 성취 수준과 개인 특성을 파악하는 접근을 취한다. 그러나 학습 시작 전후가 아닌 교수과정에서 이러한 검사를 실시하는 것은 학습과 교수에 방해가 될 수 있기 때문에, 학습 과정에서 자연스럽게 출현하는 행동을 바탕으로 학습자의 이해 수준과 오개념을 진단하고 대응하는 것이 필요하게 된다. 교수 동안에 학습자가 보이는 행동에 기반하여 정확한 학습자 모형을 구성하는 것은 상당히 어려운 작업임이 드러났는데, ITS 개발 초기부터 어려움이 지적되었고 이를 극복하기 위한 다양한 전략들이 제안되었다(Self, 1990). 가장 많이 사용되는 해결책은 오버레이 모델(overlay model)로, 도메인 모형을 바탕으로 학습자의 지식 상태를 판별하는 것이다. 예를 들어 시스템이 특정 개념을 설명했으니 학습자가 알 것이라고 추정하는 것이다. 오버레이 모델이 학습자 모형을 구성하는 데 종종 사용되지만, 일련의 문제점이 존재한다. 도메인 모델을 기반으로 하기 때문에 학습자가 도메인 모델에 포함되지 않은 오개념(misconception)이나 지식을 학습자가 가지고 있는 경우 이를 정확하게 판별할 수 없게 된다. 또한 학습 경

로 같은 역동적인 상태나 지식 간의 관계 정보를 표상하는 데에 한계가 존재한다.

ITS 연구 초기에 인간 튜터의 경우 학습자에 대한 정확한 진단을 수행할 수 있을 것이라는 믿음이 존재하였다. 이는 인간 튜터의 정확한 진단 능력이 일대일 튜터링의 성공에 중요한 역할을 한다는 가설의 기반이 되기도 하였다. 그러나 학습자의 이해를 정확하게 진단하는 것은 인간 튜터에게도 매우 어려운 과정인데(Chi, Siler, & Jeong, 2007), 인간 튜터들은 진단을 위한 질문 자체를 자주 하지 않았을 뿐만 아니라 학생들의 수준에 대한 정보를 제공받고도 이를 효과적으로 활용하지 못하였다(VanLehn, 2011). 최근에는 잠재의미분석(Latent Semantic Analysis; LSA)이 학습자 모형을 구성하는데 사용되는데, 특정 지식 요소를 학습자가 소지하고 있는지, 몇 번 적용했는지 보다는 학생들의 답변, 말뭉치를 대상으로 도메인 모형과의 의미가 일치하는지에 대한 적합도(goodness of fit)를 계산하여 이루어진다(Fotlz, 2016; Graesser et al., 2005; Nye et al., 2014).

### 인공지능의 활용

인공지능이 독자적인 알고리즘을 바탕으로 진화하기 시작하였고 인간의 수행을 뛰어넘고 있지만, 인공지능은 여전히 인간을 바라보고 영감을 얻을 수밖에 없다. 인간이 필요로 하는 서비스와 기능을 개발하고 인간과 효과적으로 상호작용하기 위해서는 인간이 무엇을 필요로 하는지, 무엇을 잘하고 못하는지, 어떤 방식으로 도와줄 수 있는지를 이해하는 것이 중요하기 때문이다. 앞으로 다양한 영역에서

인공지능이 인간의 활동을 지원할 것으로 기대되는데, 이 절에서는 그 중에서도 인지과학 연구 및 인간 정보처리에 대한 지원을 살펴본다.

#### 인지에 대한 이해의 심화

인공지능은 인간의 인지과정을 이해하는 작업, 즉 인지에 대한 연구를 지원하는데 중요한 역할을 할 수 있을 것으로 보인다. 크게 두 영역에서의 지원을 생각해 볼 수 있다. 인지 연구에 필요한 실증적 자료수집과 분석 과정에 대한 지원, 그리고 인지 이론을 정교화하고 검증하는 작업에 대한 지원을 들 수 있다.

#### 실증 연구 지원

인지 연구는 전통적으로 실증적 자료수집을 중심으로 진행되었다. 주로 인간 참여자를 대상으로 자료가 수집되는데, 자료 수집 방법에 따라 실험법, 관찰, 또는 설문조사가 사용된다. 자료 수집 방법이 무엇이든지 실증적 연구를 진행하는 것은 간단하지 않은 작업이다. 연구 주제에 대한 전문 지식을 갖추는 것은 물론, 실험 과제와 질문을 구성하고 참여자를 모집해서 자료를 수집하고 분석하는 데 상당히 복잡한 지식과 절차가 요구된다. 자료수집과 분석 과정은 이미 오래 전부터 컴퓨터와 프로그램의 도움을 받아서 이루어지고 있다. 컴퓨터를 포함한 다양한 기기들이 자극/과제/문항을 제시하고 반응을 수집하는 데 활용되고 있으며, 통계 프로그램은 수집된 자료의 분석에 필수적인 역할을 한다. 그럼에도 연구의 많은 과정은 여전히 연구자에 의해서 직접 수행되는데, 인공지능의 발달과 함께 자료 수집과

분석 과정에 대한 보다 광범위한 지원이 가능해질 것으로 보인다.

심리학 이외의 영역에서는 이미 인공지능이 가설을 만들고 연구를 설계하고 수행하는 작업에 참여하는 것이 시작되었다. King et al. (2009)은 기능유전체학(functional genomics) 영역에서 스스로 가설을 제안하고 실험을 설계하는 로봇 과학자 Adam을 개발하였다. Adam은 이스트에 대한 가설을 만들고, 이를 검증하는 실험을 설계하는 것은 물론 자동화된 실험실에서 자료 수집을 통해 가설 검증을 실시하였다. 심리학에서는 사람들 대상으로 자료를 수집하는 것이 중요하기 때문에 인공지능이 어디까지 심리학 연구 진행을 보조할 수 있을지 아직 불분명하지만 이와 관련하여 최근 주목할 만한 추세로 챗봇(chatbot)의 사용을 들 수 있다.

다양한 영역에서 봇(bot), 챗봇(chatbot)의 사용이 증가하고 있고 연구에 필요한 자료를 수집하는 과정도 예외가 아니다. 챗봇은 실험자의 수행하는 작업을 대신하는 것뿐만 아니라 실험자로 인해 야기되는 실험자 효과(experimenter effect)를 줄이는 방안으로 고려되고 있다. 자료수집 과정에서 연구자는 다양한 방식으로 결과에 영향을 준다. 실험자의 종교, 인종, 성별 등이 연구 참가자들에게 의도하지 않은 영향을 주는 것은 물론 실험자가 연구 가설을 알고 있는 경우 이는 요구특성(demand characteristics)으로 작용하여 참여자의 반응과 행동에 영향을 줄 수 있다. 전통적으로 실험 심리학에서는 이러한 문제에 대한 해결책으로 실험 절차의 표준화, 실험자 훈련 및 이중암맹(double blind) 처치를 강조해왔다. 그러나 이중암맹 처치가 실용적으로 가능하지 않은 경우도 존재하고 여러 명의 연구자가 동시에 자

료수집에 관여하는 경우 절차의 표준화와 훈련에도 불구하고 실험자 간 차이가 발생할 수 밖에 없다.

Hasler, Tuchman, 및 Friedman(2013)은 연구 참여자를 모집하는 과정에서 챗봇을 사용하고 인간 인터뷰어(interviewer)의 참여자 모집과 비교하였다. 이들은 가상공간인 세컨드라이프(second life)에서 설문 참여자를 모집하고 종교에 대한 간단한 질문에 답하도록 하였는데, 참여자들이 인터뷰에 응한 비율은 인간 설문자가 요청했을 때 더 높았으나 인터뷰에 응한 후 실제 종교 질문에 대한 답변을 하는 비율은 두 조건에서 비슷하였다. 반면 설문자에 대한 부정적인 반응(예, “다시 연락하지 마세요”)은 인간 인터뷰어일 때 더 많이 발생하였다. 즉, 참여자를 모집하는 것은 인간 인터뷰어가 더 효과적이었지만 이미 모집된 참여자에게 답변을 이끌어내는 능력은 두 집단이 유사하였고, 전반적인 반응은 챗봇일 때 더 긍정적이었다. 이러한 결과는 챗봇이 인간 실험자보다 민감한 주제에 대한 참여자의 반응을 이끌어 내는데 더 효과적일 수 있음을 시사한다(Wijenayake, Berkel & Gonclave, 2020). 아직까지는 챗봇의 자연어 대화능력의 부족하여 활용이 제한적이나, 인공지능의 발달과 함께 그 역할이 확대될 것으로 보인다.

자료수집과 관련하여 빅데이터의 활용 가능성에도 주목할 필요가 있다. 빅데이터는 내적인 심리 모형을 만들고 이를 검증하는 방식으로 이루어지는 심리학 연구에 적합하지는 않다(Kim, 2019). 전통적인 인지 연구의 방법론은 이론에 기반하여 도출된 변인을 실험적으로 조작하여 독립변인과 종속변인 간의 인과 관계를 명확하게 규명하는 것을 강조한다. 자료 수집에 있어서 엄밀한 변인 조작과 가외



변인의 통제가 실험이 제공하는 설명력의 근간이다. 실험법은 이론에 대한 검증을 가능하게 하고 변인 간의 관계를 명확하게 밝힌다는 장점을 가지고 있지만, 생태학적 타당도가 종종 문제로 지적된다. 통제의 필요성으로 인해 인위적인 과제가 사용되기 때문에 결과의 타당성이 문제시되는 것은 물론, 실험실 내에서 구현될 수 없는 인지과정이나 현상을 적절하게 연구하지 못한다는 제한점이 존재한다. 빅데이터는 일상의 현장에서 수집되는 데이터라는 면에서 또한 그 규모에서 실험을 통한 자료수집에 대한 보완책이 될 수 있다. 디지털 기술의 발달로 인간 생활의 많은 부분은 온라인으로 옮겨오고 있고, 그 결과 인간 행동의 많은 부분이 디지털 흔적을 남기게 되었다. 이를 활용하여 실험 연구에서 발견된 현상에 대한 검증은 물론 실생활에서 인간의 기억이나 의사결정, 문제해결이 어떻게 일어나는지를 연구할 수 있을 것으로 보인다. 빅데이터가 인지 및 심리학 연구에 보다 적극적으로 사용되기 위해서는 분석 기법의 발전과 함께 기존의 연구법들과 빅데이터가 어떻게 서로 보완적으로 사용될 수 있는지에 대한 모색이 필요한 것으로 보인다(Kim, 2019; Jones, 2016; Siemens & Baker, 2012).

### 인지 이론 개발과 검증

인간의 마음은 상당히 복잡한 방식으로 작동하며, 마음이 행동으로 표현되는 과정에서 수많은 변인들과 기제들이 관여한다. 이를 온전히 언어적인 기술에 의존하여 기술하고 이론화 하는 데는 한계가 존재한다. 인지 이론을 계산론적으로 명세하는 것은 복잡하고 모호한 개념을 정의하고 구체화시키는 작업을 도와주는데, 일찍이 계산 심리학(computational

psychology), 계산론적 인지 모델링(computational cognitive modeling)에서는 인지와 인지 기능의 본질을 탐구하는데 있어 해당 인지 과정에 대한 계산론적 모형을 구축하는 방법을 사용해 왔다. 계산론적 모형을 구성하기 위해서는 이론의 가정을 분명하게 하고 변인 간의 관계를 정밀하게 명세하는 것이 필요하기 때문에, 이론의 내적 정합성과 완결성을 높이는데 유용한 도구가 된다(Wilson & Collins, 2019).

계산론적 모형의 검증에 시뮬레이션(simulation)이 종종 사용된다(Berends & Romme, 1999; David, Eisenhardt, & Bingham, 2007; Sun, 2008). 기억에 대한 계산 모형이 망각곡선, 계열위치 효과 같은 기억관련 현상들을 얼마나 잘 묘사하는지를 검증하거나 또는 경쟁 모델 가운데 인간 행동을 더 잘 설명하는 모델을 선택하기 위해서 시뮬레이션이 사용되기도 한다. 시간의 흐름에 따라 모형의 구성 요소들이 어떻게 작용하고 변화하는지를 관찰하거나 현실에서 조작이 불가능하거나 어려운 변인의 효과를 검증하는 것도 가능하다(Davis et al., 2007; Gorniak & Roy, 2007).

최근에는 매크로 모델링(macro-modeling)이 활발하게 이루어지고 있다(Forbus, 2010). 기존에 개발된 여러 모델들을 통합하여 더 큰 단위의 현상을 설명하려는 시도인데, 예를 들어 어휘처리, 단어재인에 대한 모형을 통합하여 상위의 문장 처리에 대한 모델을 만드는 것이 이에 해당된다. 인공지능의 발달은 심리 현상에 대한 계산이론의 발달과 검증을 더욱 활발하게 만들 것으로 기대되고, 대상이 되는 인지 기제와 현상의 범위가 확장될 것으로 보인다.

인간인지에 대한 연구가 지난 70년 동안 활발하게 진행되었으나 인지 작동의 많은 부분

은 아직 충분하게 이해되지 못한 채 미지의 영역으로 남아있다. 의식(consciousness)의 중요성에도 불구하고 의식이 무엇인지는 아직 분명하게 이해되고 있지 않다(Dennett, 1991). 강한 인공지능(strong AI)에 대한 논의는 의식에 대한 논쟁을 다시 촉발시키고 있는데, 의식에 대한 설명 중의 하나는 지능적 시스템이 보이는 일종의 부수현상이라는 것이다(McDermott, 2007). 의식을 가진 인공지능이 과연 출현할 수 있을지는 아직 분명하지 않으나, 인공지능의 발달은 의식에 대한 일련의 가설이 검증될 수 있는 기회가 될 수 있다. 인지과학의 또 다른 오랜 논쟁 중의 하나는 정보처리 시스템이 세상을 어떻게 부호화하고 표상하는가이다(Paivio, 1991). 단순히 명제적 표상을 사용하는가, 이미지를 사용하는가의 논쟁을 넘어서, 마음에서 사용하는 언어로 대표되는 상징 표상과 대뇌의 신경원의 흥분과 휴지 상태로 대표되는 비상징 표상 간의 관계에 대한 논쟁이 존재한다(Barsalou, 2008). 이러한 긴장은 인공지능 연구에서도 존재하는데, 인공지능 연구 초기 Simon 같은 학자는 기호(symbols)를 처리하는 능력을 강조하면서 상징적 또는 기호적 모델링(symbolic modeling)을 중시하였으나, 최근의 딥러닝 기법의 발달은 비기호적(sub-symbolic) 모델링에 의해 전적으로 주도되고 있다. 인공지능에서 이 두 방식의 표상 또는 모델링 방식 간에 존재하는 긴장이 어떤 식으로 통합되게 될지는 아직 불분명하지만, 그 결과는 인간인지에도 중요한 함의와 통찰을 제공할 것으로 보인다.

#### 인간 정보처리의 지원

3차 산업혁명 시대의 기계들이 인간의 물리

적인 노동력을 대체하는 것들이었다면, 4차 산업혁명 시대의 기계들은 인간의 정보처리를 대체한다. 인간의 물리적 세계에 대한 탐색과 조작이 기계의 발달로 가속화되었듯이, 인간이 세상을 이해하고 파악하는 능력은 인공지능의 발달로 가속화될 것으로 기대된다. 디지털 기기는 이미 업무와 일상의 많은 활동을 수행하는 데 필수 불가결한 존재가 되었다. 이제까지 이들 기기의 역할이 주로 정보를 검색하거나 저장하고 제시하는 것에 국한되어 있었지만 앞으로는 더 정교하고 복잡한 처리를 지원할 것으로 예상된다. 자율운전부터 의료영상판독, 법률지원까지 다양한 영역에서 다양한 처리를 지원하는 도구와 기기들이 개발되고 있는데, 이 절에서는 가상/증강현실 기술 및 정보에 대한 조직과 추론을 돕는 도구들에 대해서 살펴본다.

#### 가상 및 증강 현실

인간의 환경 지각과 현실 인식에 영향을 주는 다양한 테크놀로지들이 개발되고 있다. 가상 및 증강현실 기술은 현실 및 온라인 공간에서 인간의 지각과 경험을 변화시킨다. 별도의 기기를 요하는 몰입형 가상현실(immersive virtual reality)은 현실과 분리된 대안적인 가상 공간, 현실을 경험할 수 있도록 한다. 증강현실(augmented reality)의 경우 모바일 기기(예, 스마트폰, 스마트 안경)를 사용하여 현실에 대한 증강된 경험을 제공하는데, 현실과의 상호작용은 유지되지만 현실에서 보이지 않는 정보를 추가적으로 제공하여 현실에 대한 인식과 체험을 증가시킨다.

가상현실과 증강현실을 활용하려는 시도가 심리학에서 활발하게 이루어지고 있는데, 대표적으로 가상현실을 사용한 불안, 공포, 외상

후스트레스에 대한 치료 프로그램을 들 수 있다(Rothbaum et al., 1999). 인지와 관련하여 가상현실과 증강현실은 이미 다양한 교육과 훈련 프로그램들에서 사용되고 있는데, 생생한 경험을 제공할 수 있기 때문에 현장감있는 실습과 훈련이 중요한 분야에서 특히 유용하다. 예를 들어, 치과 진료 중 흔하게 행해지는 발치(extraction)를 제대로 학습하기 위해서는 치아의 구조에 대한 지식은 물론 발치의 절차 및 도구 사용에 대한 실습이 필수적이다. 임상장면에서의 실습에 대한 사전 단계로 가상현실에서 발치에 관련된 절차를 학습하고 실습하는 것이 시도되고 있다(Park et al., 2019). 외국어 회화 학습 또한 가상현실이 유용한 학습 경험을 제공할 수 있는 영역이다. 실제 원어민과 대화를 연습하는 것이 중요하나, 경제적인, 시간적인 제약이 존재한다. 가상현실을 사용한 회화 학습은 이에 대한 대안으로 활용될 수 있는데, 가상 환경이 주는 몰입감을 활용하여 실제 현지에서 외국인과의 대화하는 것과 같은 효과를 낼 수 있다(Jeong & Jeong, in press). 증강현실 기술의 경우 현장 학습에서의 증강현실 기기는 현장 학습의 효율성을 증가시키는데 기여할 수 있다. 예를 들어 현장이나 유적 등을 방문할 때 해당 공간의 역사에 대한 정보를 제공하거나 주변에 있는 건물, 도구, 동식물들에 대한 정보를 제공하는 것이 가능한데, 이를 통해 보다 생생한 학습경험을 만들어 낼 수 있다.

가상현실과 증강현실은 모두 인간의 현실 인식과 지각에 영향을 주는 기술로, 인간의 환경 지각과 인식 경험을 증대하고 현실의 제약을 뛰어넘는 경험을 가능하게 한다는 장점을 가지고 있다. 그러나 이들 기술들이 제공하는 기능이 제대로 활용되기 위해서 선결되

어야 하는 문제점들이 존재한다. 일부는 기술적인 문제들로, 예를 들어 몰입형 가상 현실을 체험하기 위해서 HMD 기기를 사용할 때 사이버 멀미가 발생한다(Jung et al., 2007). 가상현실의 구현을 위해서 사용자의 머리 위치를 추적해서 그에 맞게 영상을 제시해 주어야 하는데, 영상의 움직임과 머리 움직임 간에 시차가 발생해서 멀미가 유발된다. 디스플레이 시야각 조절, 운동감각 부여 등이 해결책으로 시도되고 있으나, 현재까지는 광범위한 가상현실 사용에 걸림돌로 작용하고 있다.

기술적인 문제를 해결하는 것과 동시에 기술이 가져오는 의도치 않은 효과에 대해서도 주의를 기울일 필요가 존재한다. 가상현실의 경우 가상 공간에 대한 생생한 경험을 가능하게 하지만 학습자는 동시에 현실과는 다른 방식으로 환경과 상호작용하지 않으면 안되고 이는 정보처리의 과부하를 야기할 수 있다. 증강현실의 경우 실제 현실이 주는 정보와 증강현실 기기가 제공하는 정보를 동시에 처리하고 통합해야 하는 것이 인지 부하의 원천으로 작용하는 것으로 보인다(Wu et al., 2013). 가상현실이나 증강현실이 주는 경험이 의도된 학습 효과를 만들어 내지 못하고 학습자의 주의가 유혹적인 세부 정보(seductive detail)에 머물게 될 위험도 존재한다. 이 경우 가상현실과 증강현실이 학습자의 호기심과 동기를 야기하는 데는 성공하지만, 학습과 무관한 세부 정보에만 관심을 기울이도록 하여 학습에 필요한 처리와 활동을 촉진하지 못할 수도 있다. 이들 기술이 약속하는 치료적, 교육적 효과가 존재하지만, 그 가능성이 제대로 실현되기 위해서는 이들 기술이 인간의 정보처리 및 기타 심리과정과 어떻게 상호작용할지에 대한 세심한 주의와 연구가 필요하다.

### 정보의 조직화와 추론

정보처리의 핵심은 정보에 대한 조직화와 처리이다. 기존에 인간의 정보처리를 지원하는 도구들의 대부분이 정보의 검색과 저장에 대한 지원에 치중하였다면(Lajoie & Derry, 2013), 인공지능의 발달은 더 고급의 정보처리를 지원하는 도구의 개발을 가능하게 할 것으로 기대된다. 불필요한 자극을 무시하고 중요한 자극을 선별하는 것부터, 정보들 간의 관련성을 탐지하고, 유사한 것들을 함께 묶거나, 주어진 정보로부터 주어지지 않은 정보를 추리하고 해결책을 찾는 등 다양한 활동을 지원하는 도구들이 개발될 것이다. 이미 영상 판독, 질병 진단, 법률, 번역, 심리 평가, 글쓰기 등의 영역에서 인간 전문가가 수행하던 작업을 대신하는 시스템이 개발되거나 논의되고 있다. 이러한 도구의 개발은 정상인의 처리를 지원하는 것은 물론 장애를 지닌 사람들을 대상으로 장애의 정도를 완화하거나 극복하는데 사용될 수 있다(Kim, Song, & Jeong, 2019).

영역에 따라서 인간의 정보처리를 지원하는 수많은 도구들의 개발이 가능하나, 이 과정에서 중요하게 고려되어야 하는 것은 인간의 정보처리는 개인 내적으로만 일어나지 않는다는 것이다. 개인이 정보를 처리하고 내면화하지만, 개인만이 지식을 사용하고 활용하는 주체는 아니다. 사람 간의 상호작용 속에서도 많은 정보처리가 일어난다. 정보는 소집단과 공동체 내에서 공유되고 토론되고 기록되고, 변형된다. 인간 정보처리에 대한 지원은 개인 혼자서 하는 작업뿐만 아니라 서로 협력해서 팀으로, 집단으로, 공동체 차원에서 일어나는 정보처리에 대한 지원이 되어야 한다(Jeong, 2019). 이 작업은 소규모의 집단 작업에 대한 지원일 수도 있고 공동체 차원의 협업에 대한

지원일 수도 있다. 소집단 작업과 협업에 대한 지원으로 통상 비대면 의사소통에 대한 지원을 드나, 협업을 효과적으로 지원하기 위해서는 보다 다양한 처리가 지원되지 않으면 안된다. 협업의 파트너를 찾는 것부터, 서로의 작업 내용을 공유하고 합일점을 찾는, 그리하여 공동으로 새로운 지식과 정보를 만들어 내는 과정들이 지원되지 않으면 안된다(Jeong, 2019). 최근 들어 소집단 작업을 모니터링하고 피드백을 제공하는 것은 물론 챗봇이나 인공지능과 협업하는 것이 시도되고 있다(Wang et al., 2011; Wang et al., 2020). 인공지능의 발달과 함께 협업에 대한 보다 포괄적이고 효과적인 지원이 이루어질 것으로 기대된다.

인간의 정보처리를 지원하는 도구를 개발하는 것 못지않게 중요한 것은 해당 도구가 인간에 의해서 어떻게 사용되는지, 어떠한 영향을 미칠지를 이해하는 것이다. 어떠한 도구가 가능한지의 문제와 바람직한지의 문제는 동일하지 않다. 예를 들어, 최근 인간의 생활의 많은 부분이 온라인으로 옮겨 오면서 온라인에서 인간의 활동에 대한 모니터링이 증가하고 있다. 학생들의 경우 온라인 학습 환경에 언제 로그인하고 얼마나 오래 머물면서 무엇을 했는지, 같은 팀 구성원과 어떤 메시지를 주고 받았으며 같이 작업하는 문서의 어느 부분을 수정하였는지 등의 다양한 정보가 기록된다. 그 내용은 종종 교육적인 목적으로 교사에게 또는 같은 팀원 또는 수강생들과 공유된다. 모니터링은 수행을 객관화하고 점검하고 개선할 수 있는 바탕이 되지만 무분별한 모니터링과 공유는 부정적인 효과를 만들어 낸다. 과도한 모니터링은 수행에 대한 전념을 방해하는데, 학습에 사용해야 할 시간과 에너지의 일부가 활동을 기록하고 확인하는데 쓰이기

때문이다. 또한 손쉽게 양화가능한 지표들을 대상으로 모니터링이 이루어지다 보니 가시적인 행동에만 주의를 기울이고 학습 내용은 방기하는 의도하지 않은 결과가 발생하기도 한다. 게다가 모니터링의 결과가 무분별하게 공유될 때 해당 학습자는 물론 학습자가 속한 소집단이나 공동체에 부정적인 효과를 가져올 수 있는데, 동조나 무분별한 모방이 일어나기도 하고, 건강하지 않은 비교와 평가가 촉발될 수도 있다. 따라서 인간 활동을 지원하는 기기와 도구들의 개발과 사용에 있어 인간의 내적, 외적 활동에 미치는 영향이 세심하게 고려되어야 하는데, 이를 통해서 이들 도구와 기기들이 인간의 활동을 보다 효과적으로 지원하는 것이 가능해질 것이다.

## 결론

인간과 인공지능이 서로 경쟁 관계에 있는 것으로 보는 관점이 존재하지만, 인간과 인공지능이 협업 관계에 있는 것에도 주목할 필요가 있다. 분산 인지(distributed cognition)의 관점에서 인간과 인공지능은 과제가 요구하는 작업의 각기 다른 고리를 담당하는 것으로 볼 수 있다. 이 협업 관계에서 기계가 담당하는 역할이 변화를 거듭해 왔는데, 초기에는 물리적인, 또는 단순하고 반복적인 작업을 도와주는 데 머물러 있었으나 점차 복잡하고 어려운 작업까지 수행하고 있다. 이전에는 인간이 거의 전적으로 수행했던 글쓰기, 의사결정, 번역, 디자인 등의 지적인 작업들도 기계와 나누어서 수행하는 단계로 접어들고 있는데, 그 과정에서 이전에 인간이 해왔던 정보처리 작업의 일부, 경우에 따라서는 전부를 인공지능에

게 위임하게 될 수도 있을 것으로 보인다. 중국에는 인공지능과의 협업을 통해서 인간과 인간사회는 이전보다 훨씬 더 많은 일을 더 정확하고 빠르게 처리할 수 있게 될 것이다. 또한 그 과정에서 반복적이고 과도한 정신 노동에서 벗어나 창조적인 활동에 힘을 쏟는 것도 가능할 것으로 보인다.

그러나 인공지능과 협업하는 미래가 이러한 장미빛 전망만을 제공하는 것은 아니다. 기계의 항상적인 감시 속에서 스마트 기기에 중독된 채 관계 맺기를 기기를 통한 비대면의 소통으로 대체하면서 개인주의를 추구하지만 정작 개인성은 상실할 지도 모른다는 우려가 존재한다(Song, 2019; Turkle, 2016, 2017; Zoboff, 2019). 인간성과 인간관계에 미치는 영향뿐만 아니라 인간의 인지 역량에도 중요한 변화를 가져올 것으로 보인다. 기억이 중요한 것은 기억에 저장된 정보가 중요하기 때문일 수도 있지만 동시에 기억을 기반으로 상위 수준의 정보처리가 일어나기 때문이다. 글쓰기가 중요한 것은 작가가 되기 위해서가 아니라 글을 쓰는 과정에서 복잡한 주제를 정리하고 다듬는 활동이 일어나기 때문이다. 번역을 하는 것은 단순히 다른 언어로 문장을 변환하는 것이 아니라 글 배후에 있는 문화와 인간을 이해하는 것이기도 하다(Cronin, 2012). 인간이 글을 쓰거나 과제 수행에 필요한 인지과정의 일부를 기계에게 외주하게 될 때 이는 필연적으로 인간의 인지 능력을 변화시킬 수밖에 없다. 이러한 변화는 이미 감지되기 시작하였다(Carr, 2010). 전통적으로 매체(media)의 변화는 인간이 정보를 처리하는 방식에 직간접적으로 큰 영향을 주었다(Postman, 2006). 인공지능이 가져올 변화의 성격이 아직 분명하게 드러나고 있지는 않으나, 어떤 정보처리를 인공지능에

게 외주하는가에 따라서 미래의 인간은 검색은 잘 하지만 찾아낸 정보의 의미를 성찰하는 능력은 부족한, 인공지능의 지시를 따르는 것은 잘 하지만 정해진 경로에서 벗어나는 주도성과 자기 결정 능력은 부족한 모습으로 변화할 수도 있다. 인공지능이 가져올 이러한 도전과 위협에 대해서 성찰하고 탐구하는 것 또한 인공지능 시대에 인지과학자들이 해야 할 역할 중의 하나이다.

### 참고문헌

- Alven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26(2), 147-179.  
[https://doi.org/10.1207/s15516709cog2602\\_1](https://doi.org/10.1207/s15516709cog2602_1)
- Alkhatlan, A., & Kalita, J. (2018). Intelligent tutoring systems: A comprehensive historical survey with recent developments. arXiv preprint arXiv:1812.09628.
- Anderson, J. R., Boyle, F. C., Farrell, R., & Reiser, B. J. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modelling Cognition* (pp. 93-133). John Wiley & Sons.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.  
<https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Buchanan, B. G., & Shortliffe, E. H. (Eds.). (1984). Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley.
- Carr, N. (2010). *The shallows: How the internet is changing the way we think, read and remember*. Atlantic Books.
- Chang, M., Ventura, M., Ahn, J. W., Foltz, P., Ma, T., Dhamecha, T. I., ... & Haas, A. P. (2018). Dialogue-based tutoring at scale: Design and Challenges. In Practitioner and Industrial Track, Proceedings of the 13th International Conference of the Learning Sciences.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.  
[https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Chi, M. T. H., Siler, S., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction*, 22, 363-387.  
[https://doi.org/10.1207/s15326909xi2203\\_4](https://doi.org/10.1207/s15326909xi2203_4)
- Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471-533.  
[https://doi.org/10.1207/s15516709cog2504\\_1](https://doi.org/10.1207/s15516709cog2504_1)
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Lawrence Erlbaum Associates.
- Chi, M. T. H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, & P. Feltovich (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 167-184). Cambridge University Press.
- Collins, A. and Stevens, A. (1982). Goals and methods for inquiry teachers. In R. Glaser (Ed.), *Advances in Instructional Psychology*, Vol.

2. Lawrence Erlbaum Associates.
- Cronin, M. (2012). *Translation in the digital age*. Routledge.
- Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2007). Developing theory through simulation methods. *Academy of Management Review*, 32(2), 480-499.
- Dennett, D. C. (1993). *Consciousness explained*. Penguin.
- Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7, 10. <https://doi.org/10.1057/s41599-020-0494-4>
- Foltz, P. W. (2016). Advances in automated scoring of writing for performance assessment. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 659-678). IGI Global. <https://doi.org/10.1111/j.1756-8765.2010.01083.x>
- Forbus, K. D. (2010). AI and cognitive science: The past and next 30 years. *Topics in Cognitive Science*, 2, 345-356. <https://doi.org/10.1111/j.1756-8765.2010.01083.x>
- Fox, B. A. (1993). Correction in tutoring. In Proceedings of the fifteenth annual conference of the cognitive science society (pp. 121-124).
- Gold, K., & Scassellati, B. (2009). Using probabilistic reasoning over time to self-recognize. *Robotics and Autonomous Systems*, 57(4), 384-392.
- Gorniak, P., & Roy, D. (2007). Situated language understanding as filtering perceived affordances. *Cognitive Science*, 31(2), 197-231.
- Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9(6), 495-522.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612-618.
- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4), 514.
- Hasler, B. S., Tuchman, P., & Friedman, D. (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior*, 29(4), 1608-1616.
- Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). *Building expert systems*. Addison-Wesley Longman Publishing.
- Hildt, E. (2019). Artificial intelligence: Does consciousness matter? *Frontiers in Psychology*, 10, 1535. <https://doi.org/10.3389/fpsyg.2019.01535>
- Holmes, W., Bialik, M., & Fadel, C. (2020). Artificial intelligence in Education: Promises and implications for teaching & learning (인공지능 시대의 미래 교육: 가르침과 배움의 함의; 정제영 이선복 옮김). Pakyoun story. (Original work published 2019)
- Jeong, H. (2009). Tutoring. In E. M. Anderman (Ed.), *Psychology of classroom learning: An encyclopedia*. Macmillan Reference.
- Jeong, H. (2019). Cognitive mechanisms of collaborative learning and technology supports. *The Korean Society for Cognitive Science*, 3, 1-30.

- Jeong, J. & Jeong, H. (in press). Effects of immersive virtual reality English conversations on language anxiety and learning achievement, *The Journal of the Korea Contents Association*.
- Jones, M. N. (Ed.). (2016). *Big data in cognitive science*. Psychology Press.
- Jung, J. Y., Cho, K. S., Choi, J., & Choi, J. (2017). Causes of cybersickness of VR contents: An experimental study on the viewpoint and movement. *The Journal of the Korea Contents Association*, 17(4), 200-208.
- Kim, C. (2019). Studying psychology using big data, *Korean Journal of Psychology: General*, 38(4), 519-548.  
<http://dx.doi.org/10.22257/kjp.2019.12.38.4.519>
- Kim, Y., Song, H., Jeong, Y-K. (2019). Smart technology and child development in the fourth industrial revolution era, *Korean Journal of Psychology: General*, 38(4), 487-517.  
<http://dx.doi.org/10.22257/kjp.2019.12.38.4.487>
- King, R., Rowland, J., Oliver, S., Young, M., Aubrey, W., Byrn, E., Liakata, M., Markham, M., Pir, P., Soldatova, L., Sparkes, A., Whelan, K., & Clare, A. (2009). The automation of science. *Science*, 324(5923), 85-89.
- Lajoie, S. P. & Derry S. J. (Eds.). (2013). *Computers as cognitive tools*. Lawrence Erlbaum Associates.
- Lane, H. C., Hays, M. J., Core, M. G., & Auerbach, D. (2013). Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology*, 105(4), 1026-1035.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Lee, K. H. & Lee, B. R. (2018). *Artificial intelligence*. KNOU Press.
- Lee, K. M. (2018). *Artificial intelligence: From Turing test to deep learning*. Life and Power Press.
- McDermott, D. (2007). Artificial intelligence and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 117-150). Cambridge University Press.
- Merrill, D. C., Reiser, B. J., Ranney, M., & Traflet, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences*, 2(3), 277-305.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141-144.
- Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Prentice-Hall.
- Nye, B. D., Graesser, A. C. & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, 24, 427-469.
- Ohlsson, S. (1986). Some principles of intelligent tutoring. *Instructional Science*, 14(34), 293-326.  
<https://doi.org/10.1007/BF00051825>.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/ Revue Canadienne de Psychologie*, 45(3), 255-287.
- Park, J-T., Kim, J. H., Kim, M. Y., Lee, J. H.



- (2019). Effects of educational content for dental extraction using virtual reality technology on dental extraction knowledge, skill and class satisfaction, *The Journal of the Korea Contents Association*, 19(2), 650-660.
- Postman, N. (2006). *Amusing ourselves to death: Public discourse in the age of show business*. Penguin.
- Rau, M. A., Alevan, V., & Rummel, N. (2009). Intelligent Tutoring Systems with multiple representations and self-explanation prompts support learning of fractions. In V. Dimitrova, R. Mizoguchi, & B. du Boulay (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 441-448). IOS Press.
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249-255.
- Rosé, C. P., Moore, J. D., VanLehn, K., & Allbritton, D. (2001). A comparative evaluation of socratic versus didactic tutoring. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Retrieved from <https://escholarship.org/uc/item/98j4479r>
- Rothbaum, B. O., Hodges, L., Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Hebert, P., Dave, G., Wills, B., & Baltzell, D. (1999). Virtual reality exposure therapy for PTSD Vietnam veterans: A case study. *Journal of Traumatic Stress*, 12(2), 263-271.
- Roll, I., Alevan, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition-applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, 2(2-3), 125-140.
- Russell, S. & Novig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Jpub.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* 3(3), 417-457.
- Siemens, G., & Baker, R. S. D. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254).
- Song, H. (2019). Clinical Psychology in the Age of in the new technology: Focusing on the recent studies trends, *Korean Journal of Psychology: General*, 38(4), 549-578. <http://dx.doi.org/10.22257/kjp.2019.12.38.4.549>
- Sun, R. (2008). *The Cambridge handbook of computational psychology*. Cambridge University Press.
- Turkle, S. (2016). *Reclaiming conversation: The power of talk in a digital age*. Penguin.
- Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., R, S., Taylor, L., Wingersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3), 147-203.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.

- VanLehn, K., Jordan, P., & Litman, D. (2007). Developing pedagogically effective tutorial dialogue tactics: Experiments and a testbed. In *Proceedings of Workshop on Speech and Language Technology in Education*, 17-20.
- Wang, Y., Murray, R. C., Bao, H., & Rosé, C. (2020). Agent-based dynamic collaboration support in a smart office space. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 257-260.
- Wang, H. C., Rosé, C. P., & Chang, C. Y. (2011). Agent-based dynamic support for learning from collaborative brainstorming in scientific inquiry. *International Journal of Computer-Supported Collaborative Learning*, 6(3), 371-395. <https://doi.org/10.1007/s11412-011-9124-x>
- Wijenayake, S., van Berkel, N., & Goncalves, J. (2020). Bots for Research: Minimising the Experimenter Effect. In *Proceedings of the CHI 2020 Workshop on Detection and Design for Cognitive Biases in People and Computing Systems*.
- Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, 8, e49547.
- Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, 41-49.
- Zuboff, S. (2019). *The age of surveillance Capitalism: The fight for a human future at the new frontier of power*. Profile Books.

1차원고접수 : 2020. 11. 08.

2차원고접수 : 2020. 12. 27.

최종게재결정 : 2020. 12. 29.

**Special issue: Psychology and Fourth Industrial Revolution 2**

## **Artificial Intelligence and Cognitive Science: Opportunities and Challenges**

**Heisawn Jeong**

Department of Psychology, Hallym University

Cognitive science and artificial intelligence have closely interacted with each other as they engaged in the studies of human and machine intelligence respectively. This relationship is likely to change in the near future with the rapid developments of artificial intelligence. This paper reflects on how the nature of the relationship between the two fields might change in the future. The developments of artificial intelligence presents both opportunities and challenges to cognitive science. First, the developments of artificial intelligence can lead to the deepening of our understandings of human intelligence by assisting cognitive science research. In addition, artificial intelligence can assist human intelligence by providing smart tools with which humans can perform with greater accuracy and efficiency. At the same time, artificial intelligence poses challenges to human intelligence as it is likely to change the information environments in which humans operate and alter the cognitive profiles of human intelligence. Active participation from cognitive scientists are needed in understanding and addressing these opportunities and challenges.

*Key words* : *Human intelligence, artificial intelligence, algorithm, knowledge, learning, Intelligent tutoring system, research supports, theory development, cognitive tools*