과학철학 12 (2009) pp.49~69

인공지능의 한계와 일반화된 지능의 가능성 : 포스트 휴머니즘적 맥락*†

이 상 욱‡

본 논문은 현대 인공지능 연구결과나 최근 등장하고 있는 포스트 휴머니즘의 가능성이 데카르트를 '난처하게' 할 수 있는 내용을 담고 있음을 지적하고 그것이 함축하는 바를 논의한다. 특히 데카르트를 난처하게 한 최근의 연구들이 단순히 인간과 기계의 구별의 불분명하게 한 것이 아니라 지능에 대한 우리의 생각을 바꾸도록 재촉하고 있다는 점에 주목한다. 이 점을 보다 자세히 논의하기 위해 현대적 지능 정의의 시초인 튜링 검사에 대해 살펴보고 튜링 검사가 실용적인 이유에서 흉내 내기로서의 지능 개념에 기초하고 있음을 보인다. 그리고 이처럼 인간 지능을 인공적인 장치가 어떻게 '흉내 낼' 것인지에 초점이 맞추어진 고전적 인공지능은 연구는 상당한 성과에도 불구하고 경험적 한계와 철학적 반론에 직면해 있음을 설명한다. 마지막으로 최근 인공지능 연구에서 나타나는 지능에 대한 탈인간적 정의의 흐름과 이를 바탕으로 한 연구들을 살펴보고 이러한 최근 연구경향이 보다 일반화된 형태의 지능의 가능성에 대한 철학적 분석에 어떻게 기여할 수 있을지를 논의한다.

【주요어】 인공지능, 포스트 휴머니즘, 데카르트, 튜링 검사, 인간, 기계, 흉내내기, 일반화된 지능

---

* 접수완료: 2008.6.17/ 심사게재확정: 2008.10.20/ 수정완성본 접수: 2009.6.2

† 본 논문의 초기 형태는 2006년 10월 15일 제19회 한국철학자대회 분석철학분과에서 발표되었다. 논평을 맡아주신 최훈 선생님과 유익한 토론을 해주신 여러 선생님들께 감사드린다. 두 심사위원 선생님께서 이 논문의 장단점을 정확히 지적해주셨고, 이 지적을 고려해서 제목을 수정하였다. 두 분께 감사드린다. 일반화된 지능에 대한 보다 세부적인 논의는 별도의 논문으로 발표할 계획이다. 이 논문은 2006년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2006-321-A00129).

‡ 한양대 철학과 교수

1. 머리말: '데카르트적 난처함'

저명한 뇌과학자인 안토니오 다마지오는 1994년에 인간의 뇌가 합리적 추론 능력과 감정표현 능력에 어떻게 관련되는지를 설명한 탁월한 저서를 발표하였는데 그 제목을 『데카르트의 오류』라고 했다. 다마지오 주장의 핵심은, 인간이 가지고 있는 지적인 능력 혹은 보다 거창한 이성(reason)을 떠올릴 때 전형적으로 내세우는 합리적 추론 능력만이 아니라 우리보다 지적으로 못한 동물과도 공유하는 감정표현 능력에 우리 뇌의 상당부분이 사용되고 있으며 실은 인간의 감정표현은 고도의 지적 활동이라는 것이었다.[1)]

이는 데카르트가 인간과 동물을 구별하면서 그 핵심적인 차이점을, 단지 기계에 불과한 동물과 인간이 공유하는 저급한 감정이 아닌 오직 인간만이 향유할 수 있는 이성적 사고에서 찾았던 것과 분명한 대조를 보인다. 데카르트 이후 수많은 비교행동학자들은 동물이 인간만큼 복잡한 양식의 추론 능력을 가지고 있는 것은 아니지만 나름대로 유추능력과 논리적 판단능력을 가지고 있음을 보여주었다. 이 점에 있어서 야생 침팬지를 대상으로 이루어진 제인 구달의 연구는 동물의 지능과 인간의 지능이 분명한 차이에도 불구하고 연속적인 지적 능력의 스펙트럼 상의 다른 지점으로 이해될 수 있음을 시사했다.[2)] 이와는 별도로 인공지능 연구자들은 인간의 다양한 지적 능력 중에서 상대적으로 인공적 구현이 손쉬운 영역은 데카르트가 인간 이성의 핵심으로 보았던 논리적, 합리적 추론능력이라는 점을 깨닫게 되었다. 물론 이는 인공지능 철학자의 가능성을 즉각적으로 함축하는 것은 아니지만, 현재 상당한 영역의 형식 추론에 있어서 인공지능의 추론이나 증명 능력은 눈부시게 발전하고 있다.

데카르트 이후 전개된 이상의 상황을 살펴볼 때 왜 다마지오가 『데카르트의 오류』라는 제목을 사용했는지 이해할 수 있다. 다마지오로서는 실은

---

1) Damasio 1994
2) Goodall 1988

인간과 다른 동물들 그리고 인공적으로 만들어진 지능 사이의 연속적인 흐름의 한 극단적인 형태에 불과한 인간의 합리적 추론능력을 데카르트가 그토록 강조한 것이 후속 연구에 의해 잘못으로 판명되었음을 부각시키고 싶었을 것이다. 그리고 그에 더해 자신이 연구한 내용, 즉 철학자들에게는 사소해 보이는 인간의 감정 표현 및 다른 사람의 표현을 해석하는 과정에 고도의 지적 능력이 요구되며, 인간 두뇌의 상당 부분이 이에 사용되고 있음을 이야기하고 싶었을 것이다. 다소 진부하게 들리는 '생각하는 인간'보다는 '감정적 인간'이 더 핵심이라는 것이다.

필자는 다마지오가 데카르트의 주장에 대해 가진 문제의식에 공감하지만 '데카르트의 오류'라는 표현에 대해서는 다소 불편하다. 우선 오류라는 말은 상당히 부정적인 어감을 갖는데 몇 백 년 전 철학자가 당시 이루어지지 않은 경험적, 이론적 연구를 미리 예언하지 못했던 것을 '오류'라고까지 말할 수 있는지 잘 모르겠기 때문이다. 그런 의미라면 오류를 범하지 않는 철학자는 아마 거의 없을 것이니 구태여 데카르트만 골라내어 '데카르트의 오류'라고 명명하는 것은 부당하다는 느낌이 든다.

필자의 불편함의 보다 큰 이유는 필자가 데카르트를 심신 실체 이원론을 적극적으로 옹호한 철학자로 평가하는 것보다는 당시 널리 퍼져있던 '존재의 대사슬(Great Chain of Being)'의 생각에 기반한 중세적 세계관에서 탈피하여 적어도 인간을 제외한 모든 것이 기계적인 인과관계로 이해가능하다고 주장한 급진적인 자연철학자로 평가해야 한다고 생각하는 데서 찾을 수 있다. 실제로 우리에게는 『방법서설』로 익숙한 데카르트의 주저는 그가 자연세계와 인간세계 전반을 기계철학의 입장에서 설명한 4부작의 첫 부분에 해당된다. 필자는 데카르트의 자연철학을 (지금에 와서는 설득력이 많이 떨어진) 실체 이원론을 주장한 보수적인 것이었다기보다는, 자연세계가 정령이나 숨겨진 힘과 같은 불가해한 요인을 끌어들이지 않고서는 이해가능하지 않다는 기존의 견해를 반박하면서 자연세계를 기계적 인과작용으로 이해할 수 있다는 과감한 주장을 (물론 인간의 영혼은 제외하고) 제기했다는 점에서 그 의의를 갖는 것으로 보아야 한다고 생각한다.[3] 그런 의미에서 데카르트는 경험적 오류보다는 혁명적 생각으로 평가되어야 한

다는 것이 필자의 생각이다.

그러나 현대 인공지능의 발달이 데카르트를 (반사실적인 의미에서) '난처하게' 만들었음은 분명하다. 이는 단순히 현대 인공지능 연구자들이 상대적으로 기초적인 형태라도 인간의 지적 능력을 흉내 낼 수 있는 기계를 만들었다는 사실 때문만은 아니다. 그보다는 현대 인공지능이 다른 부분에서의 부분적 성공에도 불구하고 인간의 지적 능력에서 아직까지 흉내 내지 못하고 있는 영역이 인간의 복잡한 감정표현과 그것의 질적 속성이기 때문이다. 몇몇 철학자들은 인간의 감정까지 포함해서 의식의 질적 속성은 결코 인공지능이 성취할 수 없는 부분이라고 주장하기조차 한다. 이런 상황은 적어도 인공지능, 즉 기계가 성취할 수 없는 것을 인간의 본질적 속성으로 본다면 인간의 합리적 추론능력이 아니라 인간의 복합적인 감정 능력을 인간의 고유한 특징으로 생각하는 것이 적절하다고 시사하는 듯하다. 이는 기본적으로 기계인 동물과 인간의 절대적인 차이를 합리적 추론능력에서 찾았던 데카르트로서는 상당히 뜻밖의 결론이며 수용하기 어려운 사실일 것이다.

데카르트를 더욱 난처하게 만드는 것은 사이보그와 최근 부상하고 있는 포스트휴머니즘의 가능성이다. 인공지능이나 지능적인 동물이 기계에서 인간으로의 근접을 상징한다면 사이보그와 포스트휴머니즘은 인간에서 기계로의 근접을 상징한다. 여기서 주의할 점은 데카르트는 인간 몸의 기계성에 대해서는 누구보다도 강력한 옹호자였다는 사실이다. 그러므로 단순히 우리 몸의 기능을 향상시키기 위해 기계 팔을 붙이거나 인공 신장을 다는 방식으로 사이보그가 되는 것은 데카르트의 입장에서는 인간의 정체성에 아무런 위협을 가하지 않는 '자연스러운' 일이다. 하지만 두뇌는 다르다. 데카르트가 보기에 비록 영혼이 두뇌 어딘가에 존재하는 것은 아니지만 송과선을 비롯한 두뇌의 여러 부분이 영혼과 몸의 상호작용을 매개한다는 점은 분명하다. 그러므로 두뇌의 정신내용을 온라인상에 올려놓고 가상공간의 삶을 즐긴다든지, 인간의 지적 능력을 (기본적으로는 '기계적인' 방식으로)

3) 매즐리쉬 2001, 우드 2004

향상시키거나 변화시키는 새로운 뇌 신경회로를 사용하는 등의 행위는 인간과 인간이 아닌 기계 사이에 설정한 데카르트의 엄격한 구분을 무너뜨리는 것이다.[4]

물론 현재에도 인간과 기계의 구별은 분명히 존재하고 아마 미래에도 여전히 그럴 것이다. 그러나 이러한 구별 짓기의 가능성은 구별 짓기의 정당성과 혼동되어서는 안 된다. 매즐리쉬도 지적하듯이 인간과 기계의 구별은 서로 다른 시기에 서로 다른 방식으로 이루어져 왔다. 게다가 다른 기기의 구별 기준은 많은 경우 서로 상충되는 것이었다. 예를 들어, 기계의 복잡성이나 수행하는 작업이 단순하던 시기에는 기계는 인간과 달리 체스를 두는 것처럼 고도의 지적 능력을 발휘할 수 없다고 폄하되었다. 하지만, 대부분의 평균적인 인간보다 훨씬 체스를 잘 두는 기계가 나오자 기계는 인간적인 특징인 '실수'나 '당황스러움' 등을 보여주지 않는다는 점에서 '기계적이고 냉정하다'는 평가를 받아야 했다. 그러므로 인간과 기계가 구별되어야 하는 사회적, 문화적 필요성이 있는 한 구별은 지속적으로 이루어질 것이고 구별기준도 매 시기마다 적절하게 선택될 수는 있을 것이다. 하지만 매번 새롭게 제시되는 인간과 기계의 구별 기준은 동시에 무엇을 인간과 기계의 본질적 속성으로 규정하는지도 바꾸게 될 것이다.

이처럼 현대 경험과학의 연구결과나 최근 등장하고 있는 새로운 경험적 가능성은 데카르트를 '난처하게' 만들 것임은 분명하다. 본 논문은 이러한 데카르트적 난처함의 내용을 현대 인공지능 연구와 포스트 휴머니즘을 배경으로 살펴보고 그것이 함축하는 바를 논의한다. 특히 필자가 주목하고자 하는 바는 데카르트를 난처하게 한 최근의 연구들이 단순히 인간과 기계의 구별을 불분명하게 한 것이 아니라 지능에 대한 우리의 생각을 바꾸도록 재촉하고 있다는 점이다. 이 점을 살펴보기 위해 다음 절에서는 현대적 지능 정의의 시초인 튜링 검사에 대해 살펴보고 튜링 검사가 실용적인 이유에서 흉내 내기로서의 지능 개념에 기초하고 있음을 보인다. 그리고 3절에서 이처럼 인간 지능을 인공적인 장치가 어떻게 '흉내 낼' 것인지에 초점이

4) *Cf.* Brooks 1991, 브룩스 2005

맞추어진 고전적 인공지능 연구는 상당한 성과에도 불구하고 경험적 한계와 철학적 반론에 직면해 있음을 설명한다. 이어 마지막 절에서는 최근 인공지능 연구에서 나타나는 지능에 대한 탈인간적 정의의 흐름과 이를 바탕으로 한 연구들을 살펴본다. 그리고 이러한 인공지능의 최근 경향이 보다 일반화된 형태의 지능의 가능성에 대한 철학적 분석에 어떻게 기여할 수 있을지를 논의한다.

2. 튜링 검사와 흉내 내기로서의 지능

알란 튜링이 1950년에 발표한 「계산 기계와 지능」이라는 논문은 현대 인공지능학의 개념적 기초를 놓았다고 할 수 있다.[5] 이 논문에서 튜링은 우선 지능의 본성이 무엇이고 지능을 어떻게 이해해야 하는지에 대한 철학자들의 논쟁은 당분간 끝이 날 것 같지 않다고 불평한다. 그래서 철학자들이 지능의 본질에 대해 합의에 도달하기를 기다리기 보다는 현실적인 상황에서 지능을 가졌는지의 여부를 확인할 수 있는 검사를 제안하기로 한다. 얼핏 생각하면 튜링의 이와 같은 실용주의적 접근법은 처음부터 실패를 내재하고 있는 것으로 생각될 수 있다. 지능이 무엇인지도 모르면서 어떻게 특정 지능후보자가 지능을 가지고 있는지 여부를 검사하겠다는 것인가?

튜링의 해법은 검사를 지능에 대한 절대평가가 아니라 비교평가로 만드는 것이었다. 튜링은 지능에 대해 의견이 분분한 여러 학자들이 적어도 한 가지에 있어서는 견해가 일치하고 있음에 주목했다. 그것은 인간이 지능을 가진다는 사실이다. 그리고 인간이 가진 지능은 적어도 지능이라는 개념으로 우리가 의미하는 바의 대부분을 포함하는 것으로 여겨질 수 있었다. 그러므로 어떤 존재든 인간과 비교해서 손색이 없는 정도로 지적인 행위를 보여준다면 그 존재 역시 인간과 비슷한 수준의 지능을 가지고 있다고 추

5) Turing 1950, Crane 2003은 튜링 검사에 대한 철학적 후속 논의를 잘 정리하고 있다.

론할 수 있다. 물론 이러한 추론은 방법론적 행태주의의 가정을 포함하고 있다. 지능을 가졌는지의 여부에 대한 판단을 외부로 드러나 행태로만 판단하겠다는 생각이 그것이다. 그러나 전통적인 내성론이 인간의 마음 상태를 객관적으로 조사하기에 적절하지 않다는 점은 이미 19세기 경험 심리학의 등장부터 인정되던 바였고 지능 검사의 결과를 상호주관적으로 합의가능한 방식으로 만들기 위해서는 행태 이외에 호소할 수 있는 근거를 찾기는 쉽지 않다. 이러한 고려를 거쳐 튜링은 지금은 잘 알려진 튜링 검사를 제안하게 된다.

여기서 우리는 튜링 검사의 구체적인 형식에 주목해야 한다. 첫째 특징은 튜링 검사는 지능이 무엇인지에 대한 '정의'로서 제안된 것이 아니라 누구나 인정할 수 있는 인간의 지능을 특정 대상에게 '확장'시킬 수 있는지를 판가름하려는 목적으로 제안된 것이라는 점이다. 그래서 검사방식은 검사대상에게 여러 질문을 던져 그 대상이 질문에 대답하는 방식이나 답의 내용에서 지능적인 면모를 찾아볼 수 있는지를 판단하는 절대적인 방식이 아니라, 검사자가 피검사자의 정체를 모르는 상황에서 피검사자인 인간과 검사대상(예를 들어 기계)의 답변을 상호비교하여 인간과 동일한 정도의 지능을 기계(검사대상)에게 부여할 수 있을 때 기계에게 지능을 인정해주자는 비교적인 방식으로 이루어진다.

둘째 특징은 튜링 검사가 통계적인 추론을 사용하고 있다는 점이다. 즉, 튜링은 단 한차례의 검사를 통해 기계가 검사자를 속이고 다른 피검사자인 진짜 인간보다 자신이 더 인간적인 답변을 할 수 있다고 믿게 하는데 성공했을 때 바로 즉시 지능을 가진 것으로 대우해주자고 제안한 적은 없다. 이런 식의 검사는 특정 검사자의 지적, 문화적 배경에 의해 영향을 받을 것이고 당연히 매우 우발적인 방식으로 지능을 부여하거나 부여하지 않거나 할 가능성이 있다. 이는 가상적으로 피검사자를 모두 인간으로 하거나 모두 기계로 했을 경우를 생각해보면 분명하다. 누군가를 분명하게 인간이 아닌 것으로 판결해야 한다는 전제와 '인간적인 대응'이라는 표현이 가진 의미의 모호함이 결합하면 일회적 검사의 신뢰성은 높지 않다고 보아야 한다. 그래서 튜링이 제안하는 것은 지능 부여에 대한 통계적 추론이다. 즉

만약 미래의 인공지능을 갖춘 기계가 수많은 튜링 검사를 통해 인간과 비교하여 평균적으로 못하지 않는 승률을 보인다면 그것이 인간의 지능과 필적할만한 지능을 가지지 않았다고 보기 어렵다는 점을 강조하는 것이다.

이제 튜링 검사가 지능에 대해 함축하는 바를 살펴보자. 우리는 우선 튜링 검사가 지능에 대한 거대한 존재론적 물음에 답하기 위해 마련된 것이 아니라 인공지능을 갖춘 미래의 기계가 언제 지능을 가졌다고 인식론적으로 안전한 방식으로 확인할 수 있는가라는 보다 제한된 실용적 질문에 대답하기 위해 제안되었다는 점을 분명히 할 필요가 있다. 논문 어디에서도 튜링은 튜링검사를 통과하는 것이 바로 지능이라는 식의 주장을 한 적이 없다.

이 점을 고려할 때 튜링 검사가 방법론적 행태주의를 따르고 있다는 점도 그다지 문제삼을 일은 아니라고 생각할 수 있다. 튜링은 지능의 본질적 속성이 행태라고 주장하거나, 아예 철학적 행태주의자처럼 지능과 같은 마음상태는 곧 행태로 환원되어 정의될 수 있다고 주장하지 않는다. 튜링은 그와 같은 논쟁에 휘말리는 것을 바람직하다고 여기지도 않았고 그러한 논쟁이 언젠가 끝이 날 것이라고 생각하지도 않았다. 그보다 튜링은 만약 우리가 인간의 지능과 비교하여 다른 존재의 지능 여부를 판단하려면, 인간에게만 고유할 수도 있는 내성의 능력이나 겉모습이나 예의범절과 같은 사회문화적 요인들을 제거할 필요가 있다는 점을 강조했을 뿐이다. 이런 인간의 종적 특징을 지능과 같은 추상적 능력을 평가하는 데 동원하는 것은 적절하지 않기 때문이다.

이런 의미에서 튜링은 인간 아닌 다른 존재의 지능에 대해 연구하고 그 지능을 확인하는 과정에서 인간중심주의를 기능적으로 배제할 수 있는 방법론적 원칙을 제시한 것이라 볼 수 있다. 적어도 써얼의 중국어방 논제처럼 튜링의 이 방법론적 원칙의 핵심에 도전하는 논제가 아닌 이상은 튜링의 접근방법은 지능을 행태로 환원시켰다는 식의 비판을 받을 이유가 없음은 분명하다.

다음으로 주목해야 하는 튜링 검사의 특징은 이 검사가 앞서 지적했던 이유로 지능 자체를 검사하기 보다는 우리가 암묵적으로 지능적이라고 생

각하는 인간의 행위를 기계나 다른 지능소유 후보자들이 얼마나 잘 '흉내' 내는지를 판별하도록 설계되었다는 점이다. 이는 지능에 대한 일반이론에서 튜링검사가 출발하지 않고 논쟁의 여지가 거의 없는 명제 '모든 인간은 지능을 갖는다'에서 출발했기 때문에 갖는 근본적인 특징이다. 즉, 튜링 검사는 미래의 모든 지능소유 후보자들을 오직 인간의 지능을 기준으로 그것에 얼마나 근접하는지를 평가하게 되었던 것이다. 그리고 여기에 방법론적 행태주의가 더해지면 인간 지능에 대한 근접성은 당연히 인간의 (언어적 형태로 한정된) 지적인 행태를 얼마나 지능소유 후보자들이 잘 '흉내' 내는가에 의해 평가될 수밖에 없게 된다. 이처럼 튜링이 직접적으로 지능을 '흉내내기 능력'으로 정의하지 않았으면서도 결국에는 '흉내내기 능력'이 튜링 검사에 핵심적인 요소로 자리 잡게 된 것은 이후 인공지능 연구의 전개과정이나 최근의 변화에 지대한 영향을 끼치고 있다.

튜링은 논문이 발표된 1950년을 기준으로 수십 년 내에 튜링 검사를 통과할 수 있는 기계가 등장할 것이라고 예측했다. 하지만 2009년 현재에도 튜링검사를 통과한 기계는 등장하지 않았고 잘 알려진 여러 계산적 문제 때문에 당분간 그런 기계가 등장할 가능성도 희박하다.[6] 그럼 튜링의 직관에 어떤 문제가 있었기에 이런 상황이 도래한 것일까? 다음 절에서는 그 점에 대해 논의해 보자.

3. 고전적 인공지능의 전망과 한계

튜링 검사는 미래에 등장할 수 있는 인공지능 후보자들이 인간에 필적할 만한 지능을 가졌는지를 통계적으로 알려준다. 고전적 인공지능은 이러한 지능 확인 검사와 함께 흔히 계산주의(computationalism)라 불리는 입장에 기반한다. 이는 폰 노이만 설계를 따르는 보편 튜링 기계로서의 현대 컴퓨터와 우리 뇌가 작동하는 방식 모두를 기능적 관점에서 통합적으로 이해할

6) *Cf.* Boden 1989, 1990, Fogel 1995

수 있다는 입장이다. 우리 두뇌가 마음상태를 처리하는 방식이 진정으로 폰 노이만 설계를 따르는 보편 튜링 기계인지에 대해서는 최근 뇌과학의 연구자들 사이에서 논란이 많다. 대체적인 견해는 우리 뇌가 인지과정을 수행하는 방식은 전통적인 기호처리 과정이기 보다는 기호보다 낮은 수준에서 작동하는 연결주의적 방식을 따른다는 것이다. 그러므로 고전적인 인공지능과 비교적 새로운 연결주의 사이에 미래 인공지능 연구를 두고 최근 일종의 주도권 싸움이 벌어지고 있다.[7]

하지만 인간의 마음이 작동하는 방식이 어떠한지에 대한 논쟁에서 벗어나면 고전적 인공지능과 연결주의 모두 비교적 최근까지 '흉내내기'로서의 지능 개념에 입각해 있었다는 공통점을 갖는다. 간단하게 말하자면 인간이 이러이러한 능력을 가진다는 점은 분명하니 우리가 한 번 이런 능력을 기계로 구현해보자는 방법론적 목표를 공유하고 있다고 할 수 있다. 이 점에 국한하면 고전적 인공지능과 연결주의는 자신의 구현방식이 인간의 지능을 '흉내내기'에 더 적합하다고 주장하는 것에 지나지 않는다고 볼 수 있다. 그러므로 우리의 논의에서 중요한 점은, 특정한 인지능력마다 고전주의와 연결주의 모두 상대적 우위를 점하기도 하지만 결국에는 어떤 접근도 인간의 지능을 튜링 검사를 통과할 정도로 모방하는 인공지능을 만드는 데 성공하지 못했다는 점에 있어서는 공통점을 가진다고 할 수 있다.

물론 인공지능 연구가들은 전형적인 과학자의 방법론적 낙관주의에 근거하여 이러한 현재 상황이 미래에는 개선되거나 타개될 것이라고 전망한다. 예를 들어, 두뇌가 진정으로 원리상 연결주의적으로 작동한다면 충분히 많은 수의 연결점(node)을 가진 충분히 복잡한 연결망은 두뇌의 작용을 상당한 정도로 흉내낼 수 있을 것이다. 하지만 이런 '큰 가정'에 입각한 전망은 수많은 원리적, 실천적 어려움을 내포하고 있다. 게다가 몇몇 학자들은 인공지능이 인간의 지능을 흉내내는 데는 원리적인 문제점이 존재하며 그러한 문제점은 아마도 미래에도 해결되기 어려울 것이라 전망한다. 이 절에서는 이러한 입장을 개진하는 학자 중에서 상이한 이유를 대표하는 사

7) Cole, Fetzer and Rankin 1990, Churchland 1995

회학자 콜린스와 철학자 써얼의 견해를 살펴본다.

'실험자의 회귀(Experimenters' Regress)' 개념을 사용하여 과학지식이 형성되는 과정에서 암묵적으로 획득되는 지식의 중요성과 인식론적 문제점을 지적한 과학사회학자 해리 콜린스는 인공지능의 불가능성에 대해 독특한 근거를 제시한다.[8] 콜린스 주장의 핵심은 인간의 자연스러운, 즉 '인간적인' 행위가 너무나 비알고리즘적이어서 근본적으로 알고리즘적인 인공지능은 구현할 수 없다는 것이다. 이러한 비판은 유명한 괴델정리를 사용한 마음의 비알고리즘적 속성을 지적한 루카스-펜로즈 논점과는 다르다. 왜냐하면 콜린스 주장은 핵심은 인간은 태어난 후 사회화 과정을 거치면서 인간에게 고유한 여러 독특한 관습, 신념, 기대의 조합을 획득하게 되는데 이러한 조합의 특수함은 몇 가지 규칙의 조합을 통해서 알고리즘적으로 정의되거나 학습될 수 있는 것이 아니라 마치 도예가가 오랜 경험을 통해 좋은 도자기 만드는 기술을 배우듯이 여러 다양한 시행착오적 경험을 통해 배우면서 비알고리즘적으로만 획득될 수 있다는 것이기 때문이다.[9]

간단하게 정리하자면 콜린스는 인간의 지적 능력의 상당 부분이 명제화되거나 알고리즘화 될 수 없는 형태로 존재하며 이것의 획득한 오랜 기간의 복잡한 사회화 과정을 거쳐 이루어진다는 것이다. 이런 이유로 인간에 의해 만들어진 인공지능은 아무리 우리가 사회적 지식을 알고리즘적으로 미리 주입시키더라도 튜링 검사를 통과할 수 없게 된다. 튜링 검사에서 결정적으로 인간과 기계를 구별하는 질문들을 생각해보면 이 점이 명확하게 드러난다. 튜링 검사에서 인간이 아님을 확인할 수 있는 질문은 인간이 수행한다고 보기에는 너무 뛰어난 능력을 보여주거나 (예를 들어, 엄청나게 복잡한 계산을 순식간에 해치우는), 아니면 특정한 사회문화적 배경에 익숙한 인간이 보이기 어려운 반응 즉, 아시모프의 과학소설 『바이센테니얼 맨』에 나오는 앤드류가 자신의 몸을 보다 불완전하게 바꾸면서까지 닮고 싶어했던 '인간적인' 특이적 반응을 유도하는 질문이다.[10]

---

8) '실험자 회귀'에 대한 비판적 분석은 이상욱 2006 참조.

9) Collins 1990, Collins and Kusch 1998

10) Asimov 1976

물론 기계는 이런 사실에 대해 미리 대비하도록 프로그램되어질 수 있지만 일반적으로 프로그래머가 미리 예상할 수 있는 이런 '인간적인' 질문의 종류와 수보다는 튜링 검사의 검사자가 생각할 수 있는 상황과 수가 더 많기 마련이다. 게다가 매 질문마다 존재하는 '인간적인' 답변은 하나 이상이고 마찬가지로 '기계적인' 답변도 하나 이상이다. 그리고 연이은 질문은 서로 연결된 방식으로 인간적이거나 기계적인 답변을 요구할 수도 있다. 이 경우 인공지능에게 매번 '기계적인 답변'의 함정을 피해가도록 미리 프로그램해 넣은 일은 가능한 경우의 수가 기하급수적으로 불어나기 때문에 실질적으로 거의 불가능한 일이다. 그러므로 콜린스의 비판은 고전적 인공지능의 유명한 구성문제(frame problem)와 연결 지을 수도 있다.

이렇게 생각하고 보면 콜린스의 비판의 한계는 비교적 분명하게 드러난다.[11] 콜린스는 인공지능 연구자들에게 잘 알려진 구성의 문제를 사회학적 시각에서 새롭게 제기한 것이라고 할 수 있다. 비록 그가 제기한 구체적인 논점은 지능의 암묵지(implicit knowledge) 의존성을 부각시키는 새로운 측면은 있지만 기본적으로 고전적 인공지능 연구자들이 직면한 문제를 새로 진술한 것으로 볼 수 있다. 이 점에 있어서 콜린스의 비판이 연결주의에도 적용될 수 있을지는 분명하지 않다. 연결주의의 특징은 인간의 사회화 과정과 비슷한 '훈련' 과정을 연결망에게 부여할 수 있다는 점이다. 즉, 연결망 사이의 연결 강도가 초기값에서 전형적인 문제들을 풀어나가면서 특정한 방식으로 조절될 수 있고 결국에는 유사한 문제가 주어졌을 때 풀어낼 수 있는 성공확률이 높아지게 된다. 이러한 작동방식은 어린아이가 수많은 사회적 훈련과정을 거쳐 특정 사회문화적 배경을 암묵지의 형태로 습득하게 되는 과정과 유사하다고 할 수 있다. 그러므로 콜린스의 인공지능에 대한 비판은 고전주의에 국한한 것으로 여겨져야 할 것 같다.

---

11) 최훈 선생님은 논평문에서 콜린스의 비판은 고전적 인공지능 구현 노력에 대해 이미 오래전에 드레퓌스 등에 의해 제기된 현상학적 비판과 별 다를 바 없다고 주장했다. 필자는 콜린스의 사회학적 분석이 드레퓌스의 논의보다 사회공동체의 규약적 성격에 더 주목하고 있는 분석이라고 생각하지만 둘 사이의 차이점은 이 논문의 핵심에서 벗어나 있으므로 더 논의하지 않는다. 드레퓌스의 논의는 Dreyfus 1992 참조.

이러한 콜린스 견해의 한계에도 불구하고 필자는 콜린스의 지적에서 중요한 시사점을 얻을 수 있다고 생각한다. 그것은 바로 인간 지능이 가지는 독특함과 흉내 내기 검사로서의 튜링 검사가 지니는 내재적 한계이다. 앞서 지적했듯이 인간의 지적 능력을 흉내 내려는 고전적 시도가 실패한 영역은 연역추론과 같이 전통적으로 지성의 대표적인 능력이 발휘되는 분야로 여겨져 왔던 곳이 아니었다. 튜링 검사를 통과한 기계가 없는 이유는 주어진 상황에서 인간이 적절하게 '인간적인' 방식으로 대응할 수 있는 방식이 연역추론이 허용하는 범위를 훨씬 넘어서도록 많다는 사실과 '인간적이지 않은' 대응방식 역시 상당히 많다는 사실, 그리고 가장 결정적으로 이 두 방식 사이의 차이를 몇 가지 규칙으로 간단하게 규정할 수 없다는 사실이다. 아직 우리가 충분히 이해하지 못하고 있는 방식으로 인간은 동료 인간이 자신과 비슷한 방식으로 지능을 발휘하고 있는 방식을 알아채는 것 같다. 그런데 그런 알아챔을 알고리즘적 방식, 즉 규칙의 결합으로 유한하게 포착해내기는 거의 불가능해 보이는 것이다. 그래서 콜린스는 이러한 차이점이 결국에는 오랜 사회화 과정을 통해 점진적으로 획득될 수밖에 없다고 보는 것이다.

이러한 점을 생각해볼 때 우리는 인간이 지능을 가졌다는 튜링의 직관적 전제에 동의하면서도, 인간의 지능이 다른 모든 가능한 지능이 '흉내 냄으로써' 지능의 지위를 획득할 수 있는 전형적인 것으로 여겨져야 할 것인지에 대해 의문을 갖게 된다. 만약 콜린스가 지적하는 것처럼 인간의 지능의 상당 부분이 구체적이고 우발적인 사회화 과정의 결과물이라면 인간과 사회화 과정을 공유하지 않는 침팬지나 인공지능 로봇이 인간과 구별되는 방식으로 행위하는 것은 너무나 당연한 일이다. 비록 튜링이 인간의 지능만이 유일한 지능의 척도이고 그렇기에 인간 지능을 기준으로 다른 지능을 평가해야 한다는 원론적인 입장을 견지한 것은 아니지만 그럼에도 불구하고 지능 검사를 '흉내내기'의 형태로 제작함으로써 결과적으로 그러한 입장을 전제한 셈이 되었다고도 할 수 있다. 정리하자면 콜린스가 지적한 인간지능의 사회화 의존성은 인간지능이 분명히 지능이 취할 수 있는 한 가지 형태이긴 하지만 다른 지능이 모방해야만 하는 유일하거나 표준적인 형

태가 아닐 가능성에 대해 시사한다고 할 수 있다.

다음으로 써얼의 중국어방 논제에 대해 살펴보자. 이 논제의 내용은 잘 알려져 있고 그간 수많은 논의가 있어왔으므로[12], 필자는 우리의 논의에 관련된 부분으로 바로 들어가기로 한다. 써얼 주장의 핵심을 튜링 검사와 연결지어 생각해보면 다음과 같다. 설사 튜링 검사를 통과한 기계가 존재하더라도 그것을 그 기계가 진정한 지능 혹은 마음을 가지고 있는 증거로 해석할 수는 없다. 중국어방에 있는 사람이 중국어를 전혀 이해하지 못한 상태에서도 충분히 만족스러운 중국어 관련 행태(behavior)를 보일 수 있듯이 튜링 검사를 통과한 기계도 검사 내용에 대한 '진정한' 이해 없이 순전히 전형적인 인간의 행태적 특징을 흉내 낼 수도 있기 때문이다. 보다 중립적인 방식으로 써얼 주장의 함의를 정리하면 다음과 같을 것이다. 우리와 인공지능을 갖춘 기계는 튜링 검사의 의미에서는 서로 동등할 수 있겠지만 관련된 내용을 진정으로 '이해'라고 있는지에 있어서는 서로 다른 정도의 지능을 발현하고 있다고 할 수 있다.

써얼의 논증에서 가정되고 있는 것은 튜링 검사를 통과한 기계가 중국어방과 유사한 상황에 놓여있다는 것이다. 이는 단순한 컴퓨터 소자의 집합인 기계가 '이해'라는 독특한 마음의 상태를 향유할 수 없다는 써얼의 직관에 근거한다. 그러나 국내외 저자들에 의해 잘 지적되었듯이 이러한 전제가 별다른 논증없이 그대로 수용되기는 힘들다. 단순한 세포의 집합체인 우리 인간도 '이해'라는 독특한 특성을 향유하고 있기 때문이다. 이에 대해 써얼은 우리 뇌의 복잡도가 현재 인공지능을 갖춘 기계의 복잡도에 비해 훨씬 크다는 점을 지적할 것이다. 써얼의 생물학적 물질주의의 핵심 주장이 바로 그것이다. 마음이라는 질적인 속성이 물질로부터 발현하는 것은 사실이지만 적어도 인간의 두뇌 정도의 복잡도가 필요하다는 것이다. 그러나 써얼조차도 어떻게 '단순한' 물질에서 '이해'와 같은 전형적인 마음의 속성이 나타날 수 있는지에 대해서는 설명하지 못하고 있다. 필자로서는 이 점에 대한 해명 없이 현재 기계는 이러한 '이해'를 가지고 있다고 보기

12) 소흥렬 1994, 손병홈, 송하석, 심철호 2002, 송하석 2000, Searle 1980, 1992

어렵다는 직관과 몇 가지 간접적인 경험적 근거에 근거하여 어떠한 종류의 기계도 지향성이나 이해와 같은 고도의 정신적 속성을 구현할 수 없다는 주장에 이를 수 있다고 생각하지 않는다.

인간은 오랜 진화의 역사에서 두뇌 구조를 발달시켜왔고 그 과정에서 향유할 수 있는 정신적 속성의 종류와 깊이에도 상당한 변화가 있어왔을 것이다. 인공지능을 갖춘 기계가 인간의 진화과정과 유사한 학습과정이나 연구의 대상이 되는 과정을 거쳐 이러한 인간의 경험과 유사한 경험을 축적하지 못할 원리적 이유는 없다. 만약 인간이 겪은 수백만 년의 역사적 경험과 질적으로 동일한 경험을 기계에게도 요구해야 한다고 주장한다면 이는 결국 콜린스의 논점으로 회귀하는 것이다. 데넷을 따른다면 다양한 수준의 복잡도를 가지는 물질적 구성은 다양한 수준의 지적인 능력을 발휘할 수 있다. 생명체가 다윈 생물, 스키너 생물, 포퍼 생물, 그레고리 생물로 진화를 거듭할 때 핵심적인 변화는 현재 인공지능에서 익숙한 디자인 원리들로 포착될 수 있다. 이 사실은 지능의 다양한 수준을 인공적으로 얻는 데에 원리적인 장애물이 있다는 주장의 직관적 호소력을 상당히 위축시킨다.

물론 그럼에도 불구하고 여전히 인간 지성의 '질적 특징'은 여전히 미해결의 가능성으로 남을 가능성이 있다. 찰머스가 지적하듯이 현대 물질주의조차 인간이 왜 좀비가 아닌지를 충분히 설득력 있게 설명해주지 못하다고 평가할 수도 있기 때문이다. 아마도 써얼이 중국어방 논제를 통해 호소하려는 직관도 이러한 마음의 질적 특성일 것이다. 그러나 기능적이고 인지적인 마음의 능력을 여러 단계에서 인공적으로 구현하는 것에 원리적인 문제가 없다는 점에 동의한다면, 인간의 마음 및 지성이 가진 독특한 질적 특성을 기계가 구현하지 못한다고 기계에게 지적인 능력을 부여하는 것을 주저하는 것은 인간중심주의의 옹호하기 어려운 형태라고 볼 수밖에 없다.

앞서 콜린스의 견해를 논의하면서도 지적했듯이, 우리가 현재 비교적 자세히 알고 있는 지능이 인간이 구현하고 있는 지능이기에 우리의 지능 논의는 인간 지능에 대한 우리의 선지식에 의존할 수밖에 없다. 그러나 이는 앞으로 등장할 수 있는 새로운 지능이 항상 인간 지능의 모든 면을 그대로 모방하는 방식으로 평가되어야 함을 의미하지는 않는다. 써얼의 주장이 힘

을 얻기 위해서는 인간 지능이 갖는 순수하게 질적인 특성이 인간인 우리에게는 매우 절실하고 중요한 특징인 동시에 전체 가능한 지능의 영역에서도 핵심적인 특징이라는 일반적인 논증이 필요하다. 그렇지 않다면 인간에서 고유한 한 속성, 하지만 인간의 존엄성과는 본질적인 연관이 없었던 피부색을 근거로 다른 인간을 차별했던 인종차별주의와 마찬가지로 지성 일반에서 본질적이지 않은 '이해하는 느낌'이라는 속성을 가지고 인간 이외의 지적 대상을 차별하는 사태가 벌어질 수도 있을 것이다.

이런 차별이 정당화되지 못함은 일종의 '거꾸로 튜링 검사'를 생각해보면 쉽게 이해할 수 있다. 즉, 기계의 지능을 당연한 것으로 생각하고 기계의 지능과 비교하여 인간의 지능이 충분히 '기계적이지 못하다고' 인간에게 지능을 부여하지 않는 상황 말이다. 물론 이 과정에서 '기계보다 더 기계 같은 인간'이나 '인간보다 더 인간적인 기계' 등이 등장할 수도 있다. 거꾸로 튜링 검사 역시 원래 튜링 검사와 마찬가지로 통계적으로 결론을 도출할 것이기 때문이다. 이렇게 튜링 검사의 결과가 매 시행시기 마다에는 다양하게 나올 수 있다는 사실은 인간의 지능조차 실은 다양한 인지요소의 적절한 혼합이고 그러한 혼합은 사람마다 상당히 다를 수 있음을 시사한다. 이는 기계도 마찬가지일 것이다. 전형적으로 '기계적인' 지능이 한 가지로 규정될 수 있기보다는 다양한 기계 지능이 가능할 것이고 그들 사이를 하나로 묶어주는 유일한 특징은 인간과는 다르다는 정도일 것이다.

4. 포스트 휴머니즘 시대의 인공지능: 일반화된 지능에 대한 이해를 지향하며

최근 인공지능 연구는 새로운 경향을 보여주고 있다. 여기서 새로운 경향이란 고전주의와 연결주의처럼 인공지능을 구현하는 방식에 대한 새로운 경향이 아니라 인공지능 연구의 방향과 학문으로서의 인공지능 연구에 대한 자기정체성에서의 새로운 방향을 의미한다. 보덴은 인공지능의 철학

을 다룬 논문모음집 서문에서 인공지능 연구가 원칙적으로 두 가지 방식으로 전개될 수 있음을 지적했다. 첫째는 튜링 검사를 기본으로 하여 인공지능이 인간의 지능을 어디까지 흉내 낼 수 있을 것인지, 그리고 그러한 흉내내기가 지능을 발현하는 충분조건이 될 수 있을 것인지에 대한 논의이다. 이를 보덴은 좁은 의미의 인공지능 연구라고 정의한다. 그에 비해 인공지능 연구자들은 원칙적으로는 인간 지능에 구속받지 않고 자유롭게 지능이 가질 수 있는 여러 특징들을 조사하고 그 중에서 핵심적인 속성이 무엇인지를 분석하려는 노력을 할 수도 있는데 이것이 곧 넓은 의미에서의 인공지능 연구이다.[13] 전통적인 인공지능 연구는 당연히 좁은 의미의 연구에 집중되어 있었고 NASA처럼 지적 외계 생명체에 관심을 가진 특수집단을 제외하고는 넓은 의미의 인공지능 연구는 그다지 활성화되지 않았다.

최근 인공지능 연구경향에서 재미있는 점은 넓은 의미의 인공지능 연구가 활성화되고 있다는 점이다. 이러한 변화는 지능에 대한 일반 이론이 확립되어 이 이론에 입각한 통합적 연구가 가능해졌기 때문은 아니다. 그보다는 인공지능 연구자들이 자동항법장치와 같은 구체적인 인공지능의 문제들을 풀어나가다가 전통적인 인간중심주의적 인공지능 연구가 더 이상 효율적이지 않다는 점을 깨달았기 때문이다. 예를 들어, 최근 인공지능 연구자들은 자신들이 원하는 인공적인 지능의 구현을 위해서는 인간의 지능을 모방하는 것보다는 경우에 따라 다른 고등 동물이나 심지어 벌레 등의 지능을 모방하는 것이 더 유효하다는 점을 깨달았다. 물론 여기서 지능은 철저하게 행태적으로 이해되고 써얼이 강조하는 '이해'의 차원은 논외이다. 하지만 앞서 지적했듯이 과학자의 입장에서는 설사 원칙적으로 철학적 행태주의를 받아들이지 않더라도 방법론적 행태주의를 수용하지 않는다는 것은 연구를 포기하지 않고서는 거의 불가능하다. 인공지능 청소기가 청소도 말끔히 하는 한편 '집안을 매일 치우는 일은 정말 지겨워!'라는 메타적 성찰을 하는 것도 재미있는 가능성이이다. 하지만 청소기 설계과정에서 그런 요소를 고려하지는 않을 것이다. 그에 비해 청소를 잘 못하는 인공지능

13) Boden 1990 서문

청소기는 무용지물일 것이다.

그러므로 최근 인공지능 연구자들이 발견한 것은 자신들이 인간 지능의 흉내내기로 지능연구를 좁게 해왔다는 다소 새삼스러운 자각이 아니다. 그보다는 실제로 인간 지능을 지능의 표준적인 방식으로 여기고 이를 어떤 방식으로든 흉내 내려 노력하는 것이 전략적으로 유용하지 않다는 사실이다. 이렇게 생각하고 보면 튜링 검사를 통과한 인공지능을 단 기계가 아직 등장하지 않았다는 점은 유감스럽기는 하지만 인공지능의 미래를 뒤흔들 수 있는 일은 아니다. 다른 말로 표현하자면 써얼이 공격하는 강한 인공지능 논제가 거짓이더라도 인공지능 연구는 별 다른 지장을 받지 않을 것이라는 말이다. 그러나 이렇게 별다른 지장을 받지 않는 인공지능 연구는 꿩 대신 닭을 택하는 식으로 실패를 인정하고 인간 지능보다는 못하지만 그저 흉내 내기나 할 수 있는 그보다 못한 지능으로 연구주제를 바꾼 것이 아니다. 그보다는 인간 지능이 가진 독특함이 너무나 분명해서 상당히 다른 물질적 기반을 가진 인공지능을 단 기계로 그것을 모방하려는 시도는 그다지 생산적이지 않다는 점을 경험적으로 깨달은 것뿐이다. 하지만 인간의 지능이 다양하게 존재할 수 있는 수많은 지능 중 한 가지라는 관점을 취하면 이 과정에서 잃은 것은 많지 않다.[14] 기껏해야 과학소설에서 등장하는 인간과 구별되기 어려운 완벽한 휴머노이드는 당분간 기대하기 어려울 것이라는 점 정도이다.

특히 인간과 기계의 경계가 점점 더 불분명해져가는 포스트휴머니즘 시대에는 지능에 대해 튜링 검사가 전제하는 간접적 인간중심주의적 태도를 취할 수는 없을 것이다. 인간의 마음이나 지성이 가진 질적 속성이 독특함을 인정하고 이것을 기계적으로 구현하는 일이 어렵다는 점을 인정하는 것은 그럼에도 불구하고 기계가 지능을, 우리 인간과는 다르지만 동등하게

14) 실은 인간의 지능조차 단일하게 규정될 수 없다. 자폐인의 지능은 특정 측면에서는 다른 사람들을 압도하지만 다른 측면에서는 떨어지는 특징을 보인다(Sacks 1995). 이러한 특징은 '장애'가 아니라 '차이'로 보는 것이 적절함은 Moon 2003에 잘 묘사되어 있다. 혹은 시공간적으로 보다 다양한 가능성을 염두에 둔다면 우리의 자손들은 우리와 비교할 수 없는 양극적인 '지적 능력'을 향유할 지도 모른다(Zelazny 1967).

훌륭한 지능을, 가질 가능성과 양립가능하다.[15] 그리고 현대 인공지능 연구는 보다 실용적인 이유에서 이러한 방향으로 나아가고 있다. 이제는 지능에 대한 보다 일반화된 이론이 필요할 때이다.

참고문헌

소홍렬 1994, 「인공지능의 조건」, 『철학연구(철학연구회)』 34: 15-26.

손병홍, 송하석, 심철호 2002, 「인공지능과 의식: 강한 인공지능의 존재론적 및 의미론적 문제」, 『철학적 분석』 5: 1-33.

송하석 2000, 「중국어 방 논변과 인공지능」, 『철학적 분석』 2: 145-68.

이상욱 2006, 「웨버 막대와 탐침 현미경: 실험자 회귀에서 탈출하기」, 『과학철학』 9(2): 71-100.

게이비 우드 2004, 『살아있는 인형: 인공생명의 창조, 그 욕망에 관한 이야기』, 김정주 옮김, 이제이북스. (원서: Gaby Wood 2002, *Living Dolls: A Magical History of the Quest for Mechanical Life*, Faber and Faber.)

로드니 A. 브룩스 2005, 『로드니 브룩스의 로봇 만들기』, 박우석 옮김, 바다출판사. (원서: Rodney A. Brooks 2002, *Flesh and Machines*, Vintage Books.)

브루스 매즐리시 2001, 『네번째 불연속: 인간과 기계의 공진화』, 사이언스북스. (원제: Bruce Mazlish 1993, *The Fourth Discontinuity: The Co-evolution of Humans and Machines*, Yale University Press.)

15) 혹은 설사 인간지능의 '이해적 특징'을 기계가 결코 구현할 수 없을 것이라는 써얼의 비관적 전망이 옳더라도 그것은 지능에 대한 일반적 논의와는 별 관련이 없는, '독특한' 인간 지능에 대한 특수 논의일 뿐이라는 것이다.

Asimov, Isaac 1976, *The Bicentennial Man and Other Stories,* New York Doubleday.

Boden, Margaret A. (ed.) 1989, *Artificial Intelligence in Psychology: Interdisciplinary Essays,* Cambridge, MA: The MIT Press.

Boden, Margaret A. (ed.) 1990, *The Philosophy of Artificial Intelligence,* Oxford: Oxford University Press.

Churchland, Paul 1995, *The Engine of Reason, the Seat of the Soul,* Cambridge, MA: The MIT Press.

Brooks, Rodney A. 1991, 'Intelligence without Representation', *Artificial Intelligence* 47: 139-59, reprinted in Haugeland 1997, pp. 395-420.

Cole, David J., Fetzer, James H. and Rankin, Terry L. 1990, *Philosophy, Mind, and Cognitive Inquiry,* Dordrecht: Kluwer Academic Publishing.

Collins, Harry M. 1990, *Artificial Experts: Social Knowledge and Intelligent Machine,* Cambridge, MA: The MIT Press.

Collins, Harry and Kusch, Martin 1998, *The Shape of Action: What Humans and Machines Can Do,* Cambridge, MA: The MIT Press.

Crane, Tim 2003, *The Mechanical Mind,* 2nd edition, London: Routledge.

Damasio, Antonio R. 1994, *Descartes' Error: Emotion, Reason, and the Human Brain,* New York: Penguin Putnam.

Dennett, Daniel C. 1991, *Consciousness Explained,* Boston: Little&Brown.

Dennett, Daniel C. 1995, *Kinds of Minds: Towards an Understanding of Consciousness,* London: Weidenfeld and Nicolson.

Dennett, Daniel C. 1998, *Brainchildren: Essays on Designing Minds,* Cambridge, MA: The MIT Press.

Dreyfus, Hubert L. 1992, *What Computers Still Can't Do,* Cambridge, MA: The MIT Press.

Fogel, David B. 1995, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence,* New York: IEEE Press.

Goodall, Jane 1988, *In the Shadow of Man*, revised edition, Boston: Houghton Mifflin Company.

Haugeland, John (ed.) 1997, *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, Cambridge, MA: The MIT Press.

Moon, Elizabeth 2003, *The Speed of Dark*, New York: Ballantine Books.

Searle, John R. 1980, 'Minds, Brains, and Programs', *The Behavioral and Brain Sciences* 3: 417-24, reprinted in Boden 1990, pp. 67-88.

Searle, John R. 1992, *The Rediscovery of the Mind*, Cambridge, MA: The MIT Press.

Sacks, Oliver 1995, *An Anthropologist on Mars: Seven Paradoxial Tales*, New York: Vintage Books.

Turing, Alan M. 1950, 'Computing Machinery and Intelligence', *Mind* LIX: 433-60, reprinted in Boden 1990, pp. 40-66.

Zelazny, Roger 1967, *Lord of Light*, New York: Eos.

The Limits of Artificial Intelligence and the Possibility of Generalized Intelligence: A Posthuman Context

Sangwook Lee

I point out first that both current artificial intelligence research and the 'realistic' possibility of posthumanism have revealed to us some 'awkward' facts for Descartes as regards his view on the essence of human intelligence, and then investigate its implication. Especially, I claim that recent developments in AI and posthumanism are not just blurring the human-machine boundary, but rather urging us to change our fundamental thoughts about the nature of intelligence in more radical ways. I discuss Turing Test, and argue that it is based on the conception of intelligence as 'mimicking' human intelligence, and therefore inherently problematic in its philosophical implications. Given the limited nature of Turing's practical but limited definition of intelligence, it is no wonder why classical AI research program is widely regarded as having great difficulties in achieving its 'mimicking' goal. In conclusion, I introduce some of recent efforts to understand intelligence in post-human ways, and examine the possibility of more generalized definition of intelligence.

[Key Word] Artificial Intelligence, Generalized Intelligence, Posthumanism, Descartes, Turing Test, human, machine, mimicking