철학 제131집 2017년 5월 http://dx.doi.org/10.18694/KJP.2017.05.131.217
CHEOLHAK, Korean Philosophical Association
Vol.131, May 2017, 217-243

인공 지능에서 인공 감정으로*

-감정을 가진 기계는 실현가능한가?-

천 현 득**

【주제분류】 과학기술철학, 심리철학
【주요어】 감정, 인공지능, 인공감정, 일방적 감정 소통, 탈인용부호 현상
【요약문】 인공 감정에 관한 철학적 탐구가 필요한 시점이다. 인간의 고유한 영역으로 간주되던 인지적 과제에서 기계의 추월을 염려하는 처지가 되자, 사람들은 이제 이성이 아니라 감정에서 인간의 고유성을 찾으려한다. 하지만 최근 인공지능 로봇에 감성을 불어넣는 작업이 새로운 화두로 떠오르고 있다. 이 글은 인공 감정의 실현가능성과 잠재적인 위험을 논의한다. 먼저, 감성 로봇의 개발 현황과 주요한 동기들을 개괄하고, 왜 로봇의 감정이 문제가 되는지 살펴본다. 진정한 감정-소유 로봇이 가능한지 검토하기 위해 감정을 선험적으로 정의하기보다는 감정이 수행하는 몇 가지 핵심 역할을 소개하고 이로부터 어떤 대상에 감정을 부여할 수 있는 기준들을 제안한다. 나는 이런 기준에 비추어 진정한 감정 로봇이 근미래에 실현될 가능성이 낮다고 주장한다. 그러나 감정-소유 로봇이 등장하기 이전이라도, 어느 정도의 자율성을 가진 로봇과 맺는 일방적 감정 소통은 잠재적으로 위험할 수 있으며, 이에 대비하는 것이 시급하다고 주장한다.

1. 알파고는 인공 지능이냐는 물음

인공 지능(artificial intelligence)의 발전상이 눈부시다. 2016년 한국의 바둑 최고수 이세돌은 도전자 알파고(AlphaGo)와의 대국에서 4대1로 무릎을 꿇었다. 오랫동안 인류의 전유물로 여겨지던 영역들에서 기술의 도전이 거세다. 인공 지능이 보여주는 뛰어난 수행 능력에 압도당하면서 사람들이 체감하는 불

* 유익한 심사를 해 주신 익명의 심사위원들께 감사드린다. 이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5A2A03927217)
** 이화여자대학교 이화인문과학원 조교수

안감도 눈에 띄게 커져가고 있다. 물론, 영화 속 '터미네이터'처럼 인공 지능이 당장 인류 전체를 멸망시킬 수도 있다는 시나리오는 공상 과학소설에 불과할 것이다. 그렇지만 일부 과학자들과 철학자들은 장기적인 관점에서 특이점과 인공초지능(artificial superintelligence)의 도래를 예상하고 우려한다.(Kurzweil 2005; Bostrom 2014) 더욱 실천적인 관심을 가진 사람들은 머지않아 현실화될, 인공 지능과 더불어 살아야할 세상을 내다보고 미리 대비해야한다고 역설한다. 교통, 노동, 보건, 안보, 경제 등 인공 지능의 잠재적 영향은 전방위적이며, 이는 미국 백악관, 영국 의회, 유럽 연합, 스탠포드대학 등에서 인공 지능의 사회적 영향에 대한 보고서를 앞 다투어 발간하고 있는 이유이다.

인공 지능의 광범위한 적용으로 인해 생겨날 사회적, 경제적, 문화적, 안보적 변화를 예측하고 이러한 변화를 제도적, 정책적 수준에서 대비하는 일도 꼭 필요하지만, 인공 지능의 철학적 도전은 인간 존재와 그 의미를 향한다.(Boden 1990; Frankish and Ramsey 2014) 인공 지능은 인간의 자기반성을 유발하는 환기적 대상이다. 계산기와 자동기계가 발전해온 역사는 짧지 않지만, 20세기 중반 인공 지능의 개념이 등장하기 전까지 인간성의 본질을 재고하게 만들 정도로 위협적이지는 않았다.(Minsky 1986) "인공 지능" 연구는 한편으로는 기계가 인간과 같이 지능적으로 행위하는 존재일 수 있는 가능성에 대한 탐색이면서, 다른 한편으로는 인간의 지능 혹은 이성 능력의 정체를 더 잘 이해하기 위한 노력이기도 하다. 그래서 인공 지능이라는 연구 분야는 이중적인 성격을 가진다. 그것은 컴퓨터 과학의 일부로서 지능적 행동을 산출하는 기계 혹은 그것을 구동시키는 소프트웨어를 탐구하고 제작하는 분야이기도 하지만, 동시에 마음을 과학적으로 탐구하는 인지과학(cognitive science)의 일부로서 인간 마음의 구조와 작동 방식을 규명하기 위한 계산적 모델링을 포함한다. 인공 지능 분야의 잘 알려진 교과서(Russell and Norvig 2015)에 따르면, 인공 지능의 목표는 한편으로는 인간 지능을 닮은 기계 지능을 구현하는 것이고, 다른 한편으로는 인공물에도 장착될 수 있는 형태의 지능을 연구함으로써 지능 일반과 인간의 지능을 이해하는 것이다.

잘 설계된 기계나 소프트웨어가 (편의상, 이를 줄여서 "로봇"이라고 부르자) 통상 인간에게 부과되는 특정한 과제들을 뛰어나게 수행해 낼 수 있다는 데에

는 이견이 없다. 퀴즈쇼(“Jeopardy!”)에서 인간 우승자들을 제치고 승리한 IBM의 왓슨이나, 체스 챔피언 카스파로프를 이긴 IBM의 Deep Blue, 그리고 이세돌을 이긴 구글 딥마인드(Google DeepMind)의 알파고를 보라. 그들이 인간과 겨루었던 과제들은 분명히 인지적인 성격을 지닌 것이었고, 그들은 그 과제들을 지능적으로 (그리고 압도적인 실력으로) 해결해냈다. 물론 인지적 과제를 지능적으로 수행했다고 해서, 딥블루나 알파고에게 자의식이나 의식을 부여할 사람은 없을 것이다. “인공 지능"의 가능성에 회의적인 목소리를 냈던 철학자들(e.g., Searle 1992)은 통사론적 엔진에 의해 작동하는 계산 기계는 의미를 이해하지 못하며 따라서 진정한 지능이 아니라고 주장해왔다. 따라서 왓슨이 퀴즈의 의미를 이해했는지, 알파고가 바둑돌의 움직임이 가진 의미를 알았는지 따져볼 수도 있다. 그러나 이에 대한 철학적 논쟁은 일단 제쳐두자. 로봇이 의미나 개념을 가지는지 묻는 까닭은 이해와 의미가 지능의 본질적 요소라고 가정하기 때문이다. 하지만 압도적인 수행 능력으로 무장한 기계의 등장 앞에서 “지능”이라는 말 자체의 의미가 변화하고 있다.

사람들이 지능을 보는 관점이 변화하고 있다. 첫째, 알파고를 접한 이후 많은 사람들은 그것이 정말 “인공 지능”인지 굳이 따져 묻지 않는다. 알파고나 왓슨뿐 아니라 다른 정보 기술을 수식하는 말로 “인공 지능”을 붙이는 데에 사람들은 아무런 거리낌이 없다. “인공 지능 비서인 시리(Siri)”, “인공 지능을 갖춘 냉장고” 등의 표현은 빠르게 일상적 용법으로 자리잡고 있다. 인간이 기계와의 “지적인” 대결에서 진 상황에서, “그래도 기계가 의미를 이해한 것은 아니지 않느냐”는 물음이 궁색해진 탓이다. 둘째, 의미를 이해하는지 여부와 상관없이 인지적 과제를 성공적으로 수행하는 기계를 인공 지능으로 부를 수 있다면, 의미나 이해는 더 이상 “지능”의 본질적인 요소가 아니게 된다.[1] 본래 인간 지능이 가지던 풍부한 의미는 점차 축소되고 있다. 얼마나 효율적으로 결과를 산출할 수 있는지 측정할 수 있는 과제 수행 능력 이외의 요소들은 부차

1) 근대 천문학 혁명 당시 “지구”라는 단어의 의미 변화를 떠올려보자. 지구가 우주 중심에 고정되어 있다는 중세의 생각은, 지구는 우주의 변방에 위치하면서 스스로 돌고 또 태양 주위를 회전하는 하나의 행성에 불과하다는 관념으로 대체된다. 지구가 인간 삶에서 가진 풍부한 의미는 축소되었지만, 우리가 살고 있는 이 땅이 더 이상 우주의 중심이 아니라고 해서 더 이상 지구가 아니라고 말하지는 않았다.(Kuhn 1957)

적인 것으로 치부되고 있다.[2] 셋째, 현존하는 인공 지능들은 제한적인 소수의 과제만을 수행한다. 인공 지능도 지능이라면, 지능이란 하나의 단일한 능력이 아니라 서로 얽혀있고 상호작용하는 다양한 세부능력들의 총체이다. 물론 인간의 지능은 유연하고 일반적이며, 그런 점에서 현존하는 인공 지능과 다르다. 그러나 과제나 영역에 특수한(task-, or domain-specific) 지능을 지능이 아니라고 볼 수도 없을 것이다.

2. 문제는 감정이다

동서를 막론하고 많은 사상가들은 인간을 지/정/의, 혹은 이성과 감정과 의지를 가진 존재로 보았다. 그 가운데 인간을 인간답게 하는 것은 바로 이성이며, 감정은 이성의 지배를 받아야한다는 생각이 지배적이었다. 그러나 인지적인 능력에서 기계의 추월을 염려하며 초라해진 인간의 위상을 개탄하는 사람들은 이제 감정으로 눈을 돌린다. "왓슨은 경쟁에서 이기긴 했지만 승리를 기뻐하지는 못했다. 당신은 왓슨의 등을 두드리며 축하해줄 수 없고, 함께 축배를 들 수도 없다. 로봇은 이런 행동들이 무엇을 의미하는지 이해할 수 없을 뿐더러 자신이 이겼다는 사실조차 인식하지 못한다."(Kaku 2014, 336) 사람들은 이제 인간성의 핵심을 지적인 능력이 아니라 정서적인 부분에서 찾으려 한다. 과업의 알고리듬화를 통해 많은 직업이 로봇에 의해 대체될 것이라는 우려 속에서도, 인간의 감정을 읽고 인간과 상호작용하는 직업이 가장 오래 살아남을 것이라는 예측이 많다. 예컨대, 대중교통을 담당하는 운전기사는 대체될 확률이 높지만, 보육 교사의 대체 가능성은 높지 않다. 기계가 수행하기 가장 어려운 일은 인간과 감정적으로 교류하는 일이다.[3]

2) 이러한 변화는 인간의 정신과 지성을 업무수행 능력의 차원으로 축소해 다루려고 하는 현대 기술사회의 문화적 조건과 무관치 않을 것이다.

3) 브린욜프슨(Erik Brynjolfsson)과 맥아피(Andrew McAfee)는 기계가 인간을 따라잡지 못한 영역으로 1) 글쓰기, 과학적 발견, 기업가정신, 예술적 작업을 포함하는 창조적인 일, 2) 감정을 통한 사회적 상호작용, 3) 숙련과 솜씨가 발현되는 수영이나 발레 등 신체적 솜씨(physical dexterity)를 들고 있다. 인공 지능의 발달이 일자리에 가져올 변화에 관해서는 다음을 참조하라.

이제껏 로봇이나 인공 지능은 완벽하게 논리적이고 이성적으로 그려지곤 했다. 그런 점에서 보면 기계가 육체적, 지적 과업에서 인간의 수행 능력을 추월하더라고, 인간은 감정을 가지고 다른 이들과 교감하는 존재라는 점에서 차별화된다고 볼 수 있었다. 하지만 최근에는 감정을 가진 로봇을 제작하려는 열망이 어느 때마다 뜨겁다. 로봇 산업은 인간의 신체 노동력을 대체하는 산업용 로봇에서 지능형 서비스 로봇으로 강조점이 빠르게 이동하고 있다. 산업용 로봇은 공장에서 사람을 대신해 반복적인 작업을 빠르고 정확하게 수행하기 위한 도구로 활용되고 있다. 반면, 지능형 서비스 로봇은 외부 환경의 변화를 스스로 인식하고 상황을 판단하며, 인간과의 상호작용을 통해 인간의 여러 활동에게 도움을 주도록 설계, 제작된다. 2004년 일본 후쿠오카에서 발표된 "세계 로봇 선언"에서는 "차세대 로봇은 인류와 공존하는 파트너가 될 것이며, 인류를 신체적이고 심리적으로 보조하게 될 것"이라고 선언했다. 공학자들과 기업들은 일반가정, 병원, 양로원, 학교 등에서 사람들의 일상생활과 돌봄 및 치료 과정을 돕는, 사람과 상호작용할 수 있는 사회친화적 로봇을 개발하고자 한다. 이 로봇들은 세탁기나 청소기와는 다르게 취급될 것이다. 세탁기와는 달리, 사람들은 로봇에 이름을 붙여주고 말을 걸고 "사회적인" 상호작용을 할 것이다. 인간과 정서적으로 교감하는 로봇이 집집마다 배치된다면, 우리는 그것을 반려동물이나 가족구성원과 같이 여기게 될지도 모른다.

사교 로봇이나 감정 로봇이 각광받는 데에는 여러 이유가 있다. 우선, 현대인들은 똑똑하게 행동하는 로봇뿐 아니라 정서적으로 교감할 수 있는 로봇을 바란다. 가족 해체 현상이 가속화되고, 1인 가구가 증가하며, 공동체와의 단절을 경험하고 있는 우리 세대에 외로움을 덜어줄 로봇에 대한 수요가 커지고 있다. 게다가, 사람들은 어느 정도 감정 표현을 하는 로봇을 그렇지 않은 로봇보다 선호하는 것으로 나타났다. 사람과 같은 얼굴 표정과 목소리, 몸짓 등을 표현하는 경우 사람들의 호감도가 높아진다.(Waytz and Norton 2014) 감정 표현을 하는 로봇은 인간으로부터 더 큰 신뢰를 얻게 되고, 또 더 많이 사용될 것이다.

김세움, "기술진보에 따른 노동시장 변화와 대응", 한국노동연구원 정책연구 2015-5; Executive Office of the President, "Artificial Intelligence, Automation, and the Economy", December 2016.

둘째, 로봇에게 감정 능력을 부여함으로써 로봇의 전반적인 성능을 향상하거나 사용자의 세밀한 필요에 더 잘 부응하도록 만들 수 있다. 감정 로봇 연구자인 브리질(Breazeal and Brooks 2005)은 로봇을 네 종류, 즉 도구, 사이보그 연장, 아바타, 협력상대로 분류하며 어떠한 경우에도 감정 로봇이 사용자에게 도움이 된다고 설명한다. 정해진 방식의 명시적 명령만을 수행하는 로봇보다는 사용자의 표정이나 음성, 몸짓 등에서 드러나는 감정을 인식하는 로봇이 더 나은 서비스를 제공할 수 있다는 점은 분명해 보인다.

- 도구로서의 로봇: 로봇은 특정 과제를 수행하기 위한 장치이다. 로봇의 자율도는 주어진 과제의 성격에 따라 다를 수 있는데, 경우에 따라 원격제어가 적합할 수도 있고, 때로는 자기충족적인 체계가 필요할 수도 있다. 예컨대, 위험 지역, 오지, 혹은 우주를 탐사하는 로봇의 경우, 인간과의 통신에 상당한 제약이 있을 수 있기 때문에, 인간이 과제 수행을 전반적으로 감독하더라도 로봇은 주어진 환경에서 여러 과제를 수행할 정도로 자기충족적이어야 한다. 동물과 마찬가지로, 로봇이 복잡하고 예측불가능하면 때로는 위험한 환경에서 제한된 자원으로 여러 과제를 수행해야하는 경우, 감정을 갖는 로봇은 임무 수행에서 더 나은 수행 능력을 보여줄 수 있다.
- 사이보그 연장으로서의 로봇: 로봇은 인간 신체의 일부로 간주할 정도로 인간과 물리적으로 결합될 수 있다. 외골격 로봇이나 신체가 절단된 사람들을 위한 보철 팔다리 등이 여기에 속한다. 인공 보철이 그 자체로 감정을 가질 필요는 없지만, 사람이 경험하는 감정에 인식하고 그에 맞추어 작동을 조절하면 유용할 것이다. 예컨대, 스트레스가 강한 상황에서는 신체의 능력과 신속성을 증강하도록 파라미터를 조정하고, 진정된 상태에서는 에너지 소비를 아끼는 방향으로 조정할 수 있다.
- 아바타로서의 로봇: 자신을 로봇에 투사하여 로봇을 통해 다른 이들과 원격으로 상호작용하는 아바타 로봇을 통해 멀리 떨어진 사람들과 의사소통할 수 있다. 기술을 매개로 한 의사소통은 보통 대면 소통보다 빈약하기 마련하지만, 로봇 아바타는 완전히 체화된 경험을 가능하게 하고, (감촉, 눈 맞춤, 같은 공간 안에서의 움직임 등을 통해) 상대에게 물리적, 사회적 현전성을 드러낼 수도 있다. 이것이 가능하려면 로봇은 매우 고차원적인 인간의 명령을 수행할 수 있어야하고, 사용자의 말의 언어적 의도나 감정 상태를 파악하고, 그것을 상대

편에게 충실히 전달하는 능력을 소유해야한다.

- 협동상대로서의 로봇: 로봇이 유능한 협동상대로서 사람들과 사회적으로 상호작용할 수 있으려면 사회 지능과 감정을 가져야한다. 예컨대, 노인 돌봄 로봇은 환자가 보여주는 고통, 피로감, 불안 등의 징후를 잘 포착하고 이에 반응할 수 있어야한다. 사람들을 귀찮거나 화나게 하지 않으면서 필요에 부응하도록 하려면, 섬세하고 예민한 감각이 필요하다. 현재의 많은 기술들은 사회적, 혹은 정서적으로 문제가 있는 사람들이 하는 방식으로 우리와 상호작용한다. 사람들과 장기적인 관계를 맺고 그것의 유용성을 최대한 활용하려면, 사람들이 일상생활에서 받아들일만한 수준의 감정 로봇을 만들어야한다.

셋째, 일부에서는 미래 로봇을 더욱 안전하게 만들기 위해 감정이 중요하다고 믿는다. 이모셰이프 창업자이자 최고경영자인 패트릭 로젠탈은 "인공 지능에게 사람의 감정을 인식할 수 있게 하면 인공 지능이 항상 사람의 행복을 추구하는 쪽으로 작동하게 할 수 있어 인류를 위협하는 존재가 되는 것을 피할 수 있다"고 주장한다.[4] 인공 지능이 언젠가 인간의 능력을 훌쩍 뛰어넘는 수준으로 발전하기 이전에 사람들이 원하는 감정을 길들일 수 있다면 더 안전한 인공 지능을 만들 수 있다는 것이다.

인공 지능은 진화를 거듭하여 결국 인간과 같은 감정을 가지고 인간과 상호작용하는 존재가 될 것인가? 만약 인공 지능이 감정을 가질 수 있다면, 우리의 인간 이해는 또 어디에 뿌리내려야하며, 우리는 인공 지능을 어떠한 존재로 대우해야 할까? 인공 감정(artificial emotion)의 가능성을 타진하는 문제는 그래서 인공 지능의 가능성을 물을 때와 마찬가지로 이중적인 의미를 갖는다. 인공 감정에 대한 연구는 감정적 존재인 인간과 유사하게 행위하는 기계를 제작하려는 시도이면서, 동시에 감정 과정에 대한 계산 모형을 통해 감정 일반과 인간의 감정을 더 깊이 이해하기 위한 노력이기도 하다. 로봇에 감성을 불어넣는 작업이 새로운 화두로 등장한 이때, 인공 감정에 대한 철학적 탐구는 더 이상 미룰 수 없는 과제가 되었다.

이 물음에 답하려면 우선 감정이란 무엇인지, 인간과 동물에게 있어 감정의

4) "How happy chatbots could become our new best friends", BBC News 2016.5.31.

핵심적인 역할은 무엇인지 생각해 보아야 한다. 이로부터 우리는 어떤 대상이 감정을 소유한 존재인지 여부를 판단할 수 있는 일련의 기준들을 추려낼 수 있을 것이다.

3. 감정이란 무엇인가

감정이 없다면 우리가 누리는 풍부한 삶은 불가능하다. 우리는 기쁜 일도 슬픈 일도 겪는다. 우리는 때로 두렵고 수치스럽고 분노하지만, 때로는 자부심을 느끼며 살아간다. 감정의 본질은 도대체 무엇이고, 우리는 그런 감정을 왜 가지고 있는지를 본격적으로 탐구하는 일은 제한된 글의 범위를 넘는다. 감정이 무엇인지 이해하는 한 방식은 감정의 기능적 역할을 이해하는 것이다. 우리의 정신적 삶에서 감정이 수행하는 몇 가지 핵심적인 역할들을 고려해봄으로써, 어떤 대상에게 감정을 부여할 수 있는 기준을 생각해볼 수 있다. 이는 감정의 개념을 선험적으로 정의하는 일과는 다르다. 감정을 인간에게 고유한 어떤 것으로 만들기 위해, 애당초 인간과 일부 동물들 외에는 감정을 가진다고 말할 수 없도록 "감정"을 규정한다면 아무런 실익을 얻을 수 없다. 원리적으로는 인간 외의 다른 생명체나 인공물에게 감정의 소유를 배제하지 않으면서, 동시에 인간 및 동물이 가진 감정의 어떤 측면을 밝혀줄 수 있는 일반적인 원리나 기준을 제시할 수 있어야한다. 이를 위해, 우리가 다른 사람에게 감정을 부여할 때 어떤 기준들에 호소하는지 점검해보는 것은 좋을 출발점이 될 수 있다.[5)]

우리는 다른 사람들의 행동에서 감정의 단서를 발견한다. 감정과 정서적 행동의 관계는 밀접하다. 행동주의적 관점에 따르면, 감정이란 입력 자극에 대한 적절한 출력을 내놓는 행동들의 패턴으로 환원된다. 감정의 한 가지 기능은 사

5) 한 심사위원은 "감정이 어떤 역할을 하는지"가 아니라 "감정이 무엇인지"를 직접 분석해야했다고 지적했다. 그러나 심성 상태에 대한 기능주의적 접근이 그것이 수행하는 기능적 역할에 의해 심성 상태의 본성(그것이 무엇인지)를 드러내듯, 감정의 기능적 역할들에 주목하는 것은 감정이 무엇인지를 드러내는 한 방식이다. 감정에 대한 기능주의적 접근이 감정의 모든 측면을 온전히 밝힐 수 있다고 주장하는 것은 아니지만, 인공물을 비롯한 타자에게 감정을 부여하는 기준을 논의하는 이 글의 맥락에서 이러한 기능주의적 접근은 용호될 수 있다.

회적 의사소통에 있다. 특히, 사회적 동물인 인간에게 다른 사람의 감정 표현에서 그의 심적 상태와 의도 등을 읽어내고 적절하게 반응하는 능력이 중요하다. 로봇공학에서는 이런 접근법을 따라 "사회적" 혹은 "정서적" 행동을 보이는 로봇을 제작하는 데 많은 노력을 기울인다. 스스로 감정을 경험하는 개체만이 그러한 상호작용을 할 수 있다고 단정할 수 없다. 내적인 감정 경험을 언급하지 않고서도, 일정 수준의 사회적 상호작용이 가능한 로봇을 만들 수도 있다. 하지만 우리가 바라는 것은 인간과 인간 같은 방식으로 의사소통하는 로봇이며, 이를 위해 행동주의적 감정 이론은 불충분하다.

예를 들어, 당신이 인간과 유사한 정서적 행동을 보이는 어떤 대상을 발견했다고 하자. 심지어 그 대상은 외양으로 볼 때 인간과 구별되지 않았다고 가정하자. 그런데 만일 그것이 무선 통신을 통해 원격 제어되는 로봇으로 드러났다면, 당신은 그것에 감정을 부여하겠는가? 만일 아니라면, 생김새나 행동이 충분한 기준이 아님을 의미한다. 정서적 행동이 타자에게 감정을 부여하는 일차적인 단서인 것은 분명하지만, 그런 휴리스틱이 제대로 작동하기 위해서는 배경지식을 가정해야한다. 즉, 우리 인간은 외적 행위뿐 아니라 내적인 측면에서도 서로 유사하다는 가정하고 있는 셈이다.

행동주의는 정서적 행동과 감정 경험 사이의 거리를 간과한다. 마음의 작용을 행동 수준에서 분석하더라도, 어떤 행동이 감정에 의한 것이고 어떤 것이 그렇지 않은지 구분할 방법이 마땅치 않다. 행동은 감정에 대한 표시자(indicator)이지, 개념적으로 감정과 동일하지 않다. 행동의 동등성은 심성 상태의 동등성을 함축하지 않기 때문에, 동일한 상황에 직면한 두 사람이 서로 다른 감정을 느낄 수도 있고, 같은 감정을 느낀 두 사람이 서로 다르게 행위할 수도 있다. 감정과 행동은 성향적으로 연결되어 있다.[6)]

그렇다면 정확히 어떤 내적 측면이 감정을 부여하는 일과 유관한가? 한 가지 후보는 가장 사밀한 내적 측면인 의식적 경험에 호소하는 것이다. 그러나 감정을 경험할 때 우리가 느끼는 감각질(emotional qualia)이 감정의 필수 조건인지 여부가 논란거리일 뿐 아니라 느낌 자체는 상호주관적으로 확인될 수 없다는 점에서 타자에게 감정을 부여하는 기준으로서 적합하지 않다.(Megill

6) 반사행동은 자극에 대한 직접적인 반응이지만, 통상 감정으로 간주되지 않는다.

2014) 인간의 생물학적 구성요소와 구조를 언급하는 것도 또 하나의 방법이다. 우리는 다른 사람들이 우리와 동일한 생물학적 요소로 이루어져 있다고 믿는다. 그러나 인공물에 감정을 부여할 수 있는지 따져야하는 지금 상황에서 세포나 단백질과 같은 요소에 지나치게 의존하는 것은 온당치 않다. 오히려 심리학적 수준의 인지 구조(cognitive architecture)와 그것이 마음의 작동 내에서 그리고 행동과의 연관 속에서 수행하는 여러 기능적 역할들이 무엇인지 살펴야한다. 이를 위해서는 상당히 축적되어온 철학적, 심리학적, 신경과학적 연구들을 활용할 필요가 있다.

간단한 예시로서, 당신이 추석 즈음에 성묘를 하러 산에 올랐다고 가정해보자. 당신이 땅에서 어떤 매끈하고 긴 물체가 꿈틀거리는 모습을 보았다면, 그 즉시 몸이 얼어붙고 심장은 쿵쾅거리고 손바닥에 땀이 나면서 공포를 느꼈을 것이다. 당신은 순간 움츠려들었다가 이내 빠른 속도로 달아났을 것이다. 감정의 첫 번째 역할은 개체의 생존, 안녕, 혹은 항상성 유지에 관련된 중요한 정보를 제공해 주는 데 있다. 우리의 지각 능력이나 고등 인지는 외부 세계에 대한 신뢰할만한 정보를 제공하지만, 감정은 차분한 고등 인지과정과는 달리 빠르고 효과적인 상황 판단 및 의사결정이 가능하도록 해준다. 당신이 느낀 공포감은 당신이 위험한 상황에 처해있다고 즉각 알려주며, 전달하는 정보의 양은 적지만 커다란 효과를 발휘한다. 철학계에서 통용되는 용어로 말하자면, 감정은 지향성을 가지며 감정을 느끼는 개체가 처한 상황에 대한 평가(appraisal)를 포함한다. 그 개체가 느끼는 공포 감정은 그가 위험한 상황에 처해있다는 핵심 주제를 표상하고 그에 알맞게 대응하도록 준비시킨다.

둘째, 감정은 인지 과정을 촉진하거나 증진하기도 하고, 추론 양식에 영향을 미치기도 한다. 예컨대, 공포와 같은 부정적 감정은 그 감정을 느끼는 사람에게 지금 문제가 되고 있는 상황의 세부적인 내용에 집중하도록 만드는 경향이 있고, 반대로 긍정적인 정서는 큰 그림이나 포괄적인 의미를 생각하도록 만드는 경향이 있다.(Pessoa and Ungerleider 2004) 감정은 선택적 주의(selective attention)에서도 중요한 역할을 한다.(Attar and Muller 2012) 시각적 경험을 할 때 우리는 시각장에 들어온 모든 정보를 한꺼번에 처리할 수 없기 때문에, 그 가운데 특정한 측면에만 초점을 맞추고 나머지는 무시한다. 다시 말해, 우리는

중요하고 두드러진 특성에 주의를 집중하고 그렇지 않은 것들은 그냥 지나친다. 감정은 어떤 것이 중요한 것인지 결정하는 데 관여한다. 주의는 제한된 자원이기에 감정을 불러일으키는 자극에 집중되는 경향이 있고 실제로 그런 자극이 중요한 경우가 많다. 예컨대, 산 속에서 뱀을 마주쳤을 때 당신이 느낀 공포는 그 감정을 일으킨 대상에 주의를 집중하도록 했을 것이다. 감정은 선택적 주의뿐 아니라 장기기억 형성에도 일정한 역할을 한다. 우리는 강한 감정을 동반했던 사건들을 더 잘 기억하는 경향이 있다. 결혼식이나 아이의 출생, 사랑하는 사람의 죽음 등 강한 감정을 불러일으켰던 사건들은 기억에 더 오래 남는다. 뱀에 대한 공포 경험도 더 잘 기억될 가능성이 높다.

캡그라스 증후군(Capgras Syndrome)은 감정이 인지에 미치는 영향을 보여주는 매우 흥미로운 사례이다. 이 질환을 앓고 있는 환자는 아내의 얼굴을 제대로 알아보지 못하고 자신의 아내를 가짜라고 주장한다. 그러나 환자의 일반적인 지능이나 얼굴 인식 능력 자체에 큰 문제가 있는 것은 아니다. 환자들은 얼굴 모양이 아내와 닮았다고 생각하지만 상대가 진짜 아내임을 부인하면서 그녀가 진짜 흉내를 낸다고 생각하는데, 이는 아내에게서 느껴지는 감정이 느껴지지 않기 때문이다. 흥미롭게도, 전화로 목소리를 들려주면 아내를 알아보기도 하는데, 청각 인식과 감정 회로의 연결은 문제가 없지만 시각 인식과 감정 회로 사이의 연결에 문제가 있는 것으로 파악되었다.(라마찬드란 2016)

셋째, 감정은 행위를 안내하는 역할을 한다. 감정은 장기적인 계획을 세우거나 무엇을 추구하고 무엇을 회피할 지 판단할 때 핵심 근거가 된다. 물론, 배고픔이나 목마름 같은 기본적인 충동이나 단순한 조건반사도 행동에 영향을 미치지만, 감정은 그보다 높은 수준에서 인지와 행동을 매개한다. 간혹 우리는 감정을 절제하고 차가운 이성을 동원해야만 올바른 판단에 도달할 수 있다고 믿는다. 그러나 감정의 동기부여 역할을 간과하지 말아야 한다. 신경과학자 다마지오(Damasio 1994)의 잘 알려진 연구에 따르면, 대뇌변연계의 감정 중추가 손상된 환자들의 경우 가치판단에 혼란을 겪는다. 이 환자들은 실험실의 표준적인 인지적 과제를 수행하는데 별 문제가 없었지만, 일상생활에서 합리적인 판단을 내리는 데 어려워했다. 그 가운데 일부는 투자에서 큰 손실을 보기도 했는데, 정신적으로 건강한 사람의 경우 투자 실패를 경험하면 이후 더욱 조심

스럽게 접근하거나 투자를 멈췄겠지만, 감정이 손상된 사람들은 그렇지 않았다. 그들은 좋지 않은 감정과 위험한 선택 사이의 연결을 학습하지 못한 것이다. 감정은 중요한 일과 사소한 일을 신속하고 정확하게 구별하는 데 핵심적이다. 감정을 느끼지 못하는 사람이라면 도대체 왜 특정한 일을 해야만 하는지 동기를 찾기 어려울 것이다. 만일 어떤 이가 뱀을 보고도 공포를 느끼지 않는다면, 그의 생존은 장담할 수 없을 것이다. 감정은 단순한 조건반사와 숙고된 판단 사이에 위치하는 것으로 보이며, 일의 우선권을 조정하고 재빨리 대응해야할지 아니면 시간을 가지고 숙고해야할지를 결정하는 데에도 일정한 역할을 한다.

넷째, 우리가 특정한 감정을 경험할 때면 특징적인 신체 반응이나 표정 등이 동반된다.[7] 그러한 신체 반응은 환경에 대해 적응적이며, 우리가 다음번에 취하게 될 행동을 준비하는 역할을 한다. 고양이를 보고 공포를 느낀 생쥐는 얼어붙거나 도망을 친다. 다양한 감정 표현은 사회적 상호작용에서 중요한 역할을 담당한다. 우리는 서로의 미묘한 감정을 읽어내고 그에 적절히 반응하며, 그런 정서적 교감을 통해 공동체를 유지하는 존재이다. 어떤 이가 공포에 질려 있는 표정을 하고 있다면 우리는 그가 위험에 처해 있음을 파악하고 그에 알맞은 행동을 실행에 옮길 수 있어야 한다. 한편으로, 우리는 상대방에게 특정한 행위를 이끌어 내거나 특정한 감정을 불러일으키기 위해, 상황에 알맞은 감정을 적극적으로 표현하거나 숨길 수도 있다. 감정은 사회적 유대감을 형성하는 기초이다.

감정은 개체의 생존과 안녕에 유관한 정보를 표상하고 인지 과정에 영향을 미치며 행위를 지도하고 사회적 상호작용에 관여한다. 감정은 한정된 자원을 가지고 복잡하고 때로는 예측 불가능한 물리적, 사회적 세계에 살아가기 위해 유연하고 적응적인 행위를 나타내야할 지적인 존재에게 요구되는 그 무엇이다. 이러한 감정을 구현하는 한 가지 방식은 인간과 동물과 같은 신체적 조건을 부여하는 것이지만, 우리의 물음은 실리콘을 기반으로 인공 감정을 구현할

7) 물론 이러한 신체 반응과 그에 대한 감각은 우리 몸 전체에 걸쳐있는 자율신경계와 내분비계, 호르몬의 작용, 그리고 두뇌의 구조 등에 의해 결정된다. 그러나 앞서 언급된 것처럼, 감정을 가지기 위해 우리와 동일한 생물학적 기반 – 세포, 호르몬, 신경계 – 을 가지고 있어야 한다고 요구하는 것은 아니다.

수 있을지 여부이다.

4. 인공 감정은 실현가능한가?

사교 로봇이나 감정 로봇의 제작을 향한 연구 방향을 살펴봄으로써 인공 감정이 가까운 미래에 실현될 수 있는지 생각해보자. 로봇공학자들이 설계하는 감정 체계는 흔히 감정 인식, 감정 생성, 감정 표현이라는 세 부분으로 구성된다.[8)]

- 감정 인식: 입술, 눈썹 모양, 얼굴 찡그림 등의 표정이나 몸짓을 시각적으로 인식하고, 음성의 템포와 억양, 강도 등에 따라 음성을 인식하며, 애완용 로봇과 같은 일부 로봇에서는 촉각 정보(쓰다듬기, 때리기, 안아주기 등)를 활용하여 사용자의 감정을 파악한다. 2015년 소프트뱅크가 개발한 페퍼는 사람의 얼굴을 관찰해 감정을 인식하고, 2016년 1월 애플이 인수한 얼굴인식 전문기업 이모션트(Emotient)는 구글 글래스를 통해 미세한 표정까지 읽어내고 이를 통해 사람이 느끼는 감정의 종류와 강도를 읽어내는 기술을 보유한 것으로 알려져 있다.
- 감정 표현: 얼굴 표정을 짓거나 몸짓을 하고, 음성으로 반응하기도 한다. 와세다대 로봇인 코비안(Kobian)은 온 몸을 이용해 코미디언처럼 행동을 표현하는데, 놀란 목소리나 우스꽝스러운 몸짓 등을 표현할 줄 안다. 페퍼(Pepper)는 발표되는 날 발표장에서 손정의 회장과 교감하며 다양한 제스처를 구사했다. 다만, 예고 없이 주어진 환경에 대한 반응이 아니라 녹화된 표현 패턴이었음이 알려졌다.
- 감정 생성: 많은 감정 로봇은 단순히 행동주의적 접근을 따르지 않고 심리학과 신경과학의 성과를 반영한 감정 모형을 로봇에 장착하려 한다. 타인의 감정 표현을 인식할 뿐 아니라 타인의 표정이나 주변 상황에 의해서 스스로 감정 모형을 생성하고, 이를 바탕으로 표정이나 몸짓, 목소리로 표현하는 것이다. 즉, 입력과 로봇의 현 상태를 참조하여 감정을 생성하며, 때로는 동기나 성격 등을 고려하기도 한다. MIT 인공 지능 연구소에서 개발된 키즈멧(Kismet)은 3차원 감정 공간

8) 국내외 연구 동향에 관해서는 다음을 참조: 안호석・최진영 (2007), 이동욱 외 (2008), 박천수 외 (2008), 이찬종 (2009), 이원형 외 (2014), 김평수 (2016)

(arousal, valence, stance)에 9개의 감정을 표현한다. 예컨대, 분노는 각성의 수준이 높으면서 부정적이고, 그러면서 그 감정을 일으키는 대상을 향해 나아가는 감정이다. 키즈멧은 15개의 자유도를 가지고 감정을 표현했고, 그 후속 로봇인 레오나르도(Leonardo)는 64개의 자유도를 가진다.

감정 인식과 표현 능력만을 갖춘 로봇은 사회적 의사소통이라는 역할을 재현하는 데 역점을 둔다. 그러나 내적인 감정 생성 모형 없이 로봇이 할 수 있는 의사소통은 제한적이며, 그러한 로봇은 감정을 소유한다고도 볼 수 없다. 내적인 감정 과정을 개체의 인지 구조 안에 포섭시키지 않고서는, 개체의 행동은 유연하고 적응적일 수 없고 낯선 환경에 놓일 때 손쉽게 작동을 멈추게 된다. 그렇다면 문제는 감정 생성 모형을 갖춘 로봇이 인공 감정을 가질 수 있는지 여부이다. 이는 감정 모형이 어떤 수준에서 구현되는지에 달려있다. 그러나 현재의 기술 수준에서 인공 감정을 가진 로봇은 없을 뿐 아니라 가까운 미래에 그런 로봇이 등장할 가능성은 희박하다고 판단된다.

감정이 수행하는 핵심적인 역할들을 자세히 살펴보면 그런 역할들이 필요하고 또 가능하기 위한 전제조건들이 있음을 알 수 있다. 첫째, 감정은 주어진 자극이 가진 가치와 중요성에 대한 평가를 포함하기 때문에, 자기 자신에 대한 기초적인 모형, 혹은 원초적 자아(proto-self model)를 가져야한다. 로봇이 인간이 가는 것과 같은 자아나 자의식을 가져야한다는 말이 아니다. 어떤 것이 "자신에게" 해가 되는지 도움이 되는지 평가할 수 있어야한다는 뜻이다. 포유류나 파충류는 물론이고 곤충도 해로운 자극은 피하고 유익한 자극은 얻으려한다. 곤충이 자아 개념이나 자의식을 가진다고 볼 수 없지만, 원초적인 자아 모형을 가진 것으로 볼 수 있다. 이와 연관하여, 감정을 가진 개체는 기본적인 충동이나 욕구를 가진다고 전제된다. 동물은 목마름, 배고픔, 피로감 등의 본능을 가지는데, 이런 본능이 없다면 감정도 없다.[9] 둘째, 앞서 논의된 감정의

9) 로봇공학자들도 이와 같은 사실을 모르지 않는다. 그들은 적절한 시기에 원하는 자극과 입력을 적절한 양과 강도로 받고자 하는 기본적인 동기를 로봇 안에 장착하고자 한다. 키즈멧의 경우, 사회적 충동(the social drive), 자극 충동(the stimulation drive), 피로 충동(the fatigue drive)을 내장하며, AIBO는 성애욕, 탐색욕, 운동욕, 충전욕의 4개 본능을 구현했다.

여러 기능적 역할들은 감정을 가진 개체가 상당한 수준의 감각 능력과 일반 지능(general intelligence)을 가지고 있음을 전제하고 있다. 환경에서 주어지는 자극을 지각하고 그로부터 얻은 정보와 개체 내의 상태에 대한 정보를 결합할 수 있는 능력이 없다면, 감정은 불가능하다. 감정은 지능적인 동물에게서 나타나며, 더 지능적일수록 더 풍부한 감정을 나타내는 경향이 있음을 기억할 필요가 있다. 지능과 감정은 한 인지 구조 내에서 상호작용하는 두 가지 하부 시스템이다. 따라서 인간과 사회적으로 상호작용하기 위해 인간(혹은 반려 동물)이 가지는 것과 같은 감정을 가지려면, 로봇은 인간이나 고등 동물 이상의 일반 지능을 가지고, 생명체들이 가진 신체와 유사한 신체를 가지며, 생명체가 흔히 처하는 것처럼 복잡하고 예측 불가능한 환경에 놓여 적응할 수 있어야한다. 복잡한 환경에 적응적으로 행위할 수 있는 일반 지능을 가진 인공 지능에 도달하는 길은 아직 멀다. 진정한 감정 로봇을 현실적으로 구현하기란 어렵다.

인공 감정이 현실적이지 않은 또 다른 이유는 기술 발전의 궤적이 사회적 배경 안에서만 결정된다는 점에서 찾을 수 있다. 원리적으로 가능한 모든 기술이 현실화되는 것은 아니다. 기술의 실현가능성에 관한 판단은 단순히 서술적이지 않다. 그것은 처방적이기도 하다. 설령 기술적으로 충분히 가능성이 있더라도, 시장에 충분한 수요가 없거나, 해당 기술에 대한 사회문화적 저항이 크거나, 그 기술에 이해관심을 가진 사람들에게 충분한 설득력을 보여주지 못한다면, 그 기술은 개발되지 않을 것이다. 기술 발전은 기술 자체의 논리만으로 결정되지 않는다.(Pinch and Bijker 1987; Noble 1984; Winner 1986/2010) 따라서 우리가 진정한 인공 감정을 원하는지 물어야한다. 나는 사람들이 감정 로봇을 원하는 이유가 과장되어 있거나, 실제로는 진정한 감정을 가진 로봇을 만들어야할 좋은 이유가 없음을 보이고자 한다.

첫째, 로봇이 스스로 감정을 가진다고 해서 미래의 로봇이 더 안전해질 것이라고 장담할 수 없다. 우리 인간이 동물의 감정을 배려하지 않는 것처럼, 인공초지능이 인간의 감정을 이해하고 배려하지 않는다면, 미래 로봇이 더 안전해질 것으로 기대할 수 없다. 게다가, 우리는 감정에는 명암이 있음을 직시해야한다. 어떤 감정에 휩싸여 상황을 냉철히 판단하는 못하는 경우를 우리는 가끔 경험한다. 예컨대, 공포에 휩싸인 사람은 주어진 상황의 위험을 과대평가하

며, 그런 상태가 지속되면 사람들 사이의 합리적 대화마저 불가능하게 만들 수 있다. 인간은 폭력적 행위에서 쾌감을 느끼기도 한다. 인간의 가장 큰 적은 인간이었음을 기억해야한다. 인간들 사이에서 벌어진 끔찍한 전쟁들, 살인사건들, 모욕적인 언사와 행위들을 인간이 감정을 가졌기에 혹은 감정을 가졌음에도 불구하고 벌어졌다. 인간과 같은 감정을 가졌다고 해서 로봇이 인간에게 더 친절한(human-friendly) 존재가 되리라 기대하기는 어렵다.

둘째, 진정한 감정을 소유한 로봇이 애초에 우리가 로봇을 만드는 목적에 부합하는지 의문이다. 어원으로 보면, 로봇은 인간의 고되고 번거로운 노동을 대신하기 위한 기계를 뜻하며, 일종의 인공물 노예이다. 그런데 미래 로봇이 감정을 가지게 되었음에도 단지 그것이 인공물이라는 이유로 노예처럼 부린다면, 감정을 느끼는 존재에 대한 노동 착취라는 비난이 생겨날 수 있다. 감정의 소유 여부는 로봇의 노동을 도덕적 차원에서 고려하도록 만든다. 예컨대, 돌봄 로봇에게 우리는 감정 노동을 강요하는 것일지도 모른다. 로봇에게 인권이나 동물권과 유사한 로봇권(robot right)을 부여해야한다는 목소리가 제기될 수 있다. 우리는 권리를 가진 주체로서의 로봇을 원하는가, 아니면 시키는 일을 똑똑하게 처리하는 노예로서의 로봇을 원하는가? 뿐만 아니라, 인간이 로봇에게 신체적으로나 심리적으로 고된 노동을 강제하려고 해도, 감정을 가진 로봇을 통제하기란 쉽지 않을 것이다. 로봇이라고 해서 아무도 거주하지 않는 텅 빈 우주 속으로, 혹은 노심이 녹아내리는 원자로 속으로 들어가고 싶지는 않을 것이다. 우리는 비장한 마음으로 조국을 위해 헌신하는 로봇을 보고 싶은가, 아니면 감정을 느끼지는 못하지만 위험한 상황에서도 과제를 수행할 수 있는 로봇을 원하는가?

셋째, 인간이 경험하는 풍부한 감정을 로봇에 불어넣는 것이 비현실적이라면, 우리는 어려운 선택에 직면하게 된다. 어떤 감정은 로봇에게 허용하고 몇몇 감정은 억제해야할 것이기 때문이다. 예컨대, 사용자에게 충실하고 예의바르며, 사용자의 감정에 공감하여 재치있는 비평을 할 줄 알고, 때로는 넋두리를 늘어놓을 수 있다면 좋을 것이다. 로봇에게 긍정적인 감정이 풍부하면 좋다는 생각은 그럴듯해 보인다. 그러나 로봇에게 얼마만큼의 부정적인 감정을 넣어주어야 할지 결정하기란 어렵다. 로봇에게도 분노, 공포, 슬픔, 역겨움, 수치,

모욕감, 당황스러움의 감정이 필요할까? 이 질문에 답하기 어려운 것은 두 가지 이유가 있다. 하나는 로봇에게 그런 감정이 내재되어 있지 않다면, 사용자가 그러한 부정적 감정을 느낄 때 제대로 공감해줄 수 있을지 의문스럽기 때문이다. 인간과 인간적인 방식으로 교감하는 로봇을 원한다면, 로봇은 부정적 감정도 가져야할 것이다. 그런 감정을 소유하지 않는다면, 인간의 부정적 감정을 이해하지 못하거나 아니면 이해하는 척 해야한다.[10] 다른 하나는 부정적 감정도 나름의 역할이 있음을 우리가 알고 있기 때문이다. 우리는 부정적 감정이 인지 능력의 어떤 면을 강화한다는 사실을 이미 살펴보았다. 뿐만 아니라, 슬픔을 느낄 수 없는 존재가 기쁨을 온전히 누리거나, 수치를 모르면서 자부심을 느끼는 것은 어려워 보인다.[11]

요컨대, 로봇이 감정을 가지기 위한 전제조건들을 만족하기 어렵기에 가까운 미래에 인공 감정이 현실화되기는 어려울 것이고, 설사 그것이 기술적으로 가능하다고 하더라도 진정한 감정을 가진 로봇을 인류가 원하는지에 관해서는 의문의 여지가 많다. 따라서 나는 인공 감정이 근미래에 실현될 가능성은 낮다고 본다.

10) 탁월한 과학자이자 미래학자인 카쿠(Michio Kaku)는 로봇에게 분노의 감정은 제거되거나 통제되어야 한다고 주장한다. 분노는 상대를 향한 강한 부정적 감정이기에, 분노의 상대에게 위험한 상황이 초래될 수도 있다. 우리는 자신의 밥그릇을 뺏기고 모욕당하고 억울하고 원하는 바를 얻지 못하고 좌절할 때 화를 낸다. 분노를 우리의 자원을 총동원하고 에너지를 집중하도록 만든다. 응당 분노해야할 상황에서 분노할 수 없는 개체는 무력감이나 우울감을 경험하게 된다. 우리는 화를 내야할 상황에서도 화내지 않는 바보 혹은 우울한 로봇을 원하는가? 어떤 사용자가 집에 돌아야 자신이 화났던 일을 로봇에게 말해준다고 하자. 그 감정에 공감하기 위해 로봇은 어떻게 해야 할까?

11) 이러한 문제를 다루는 한 가지 방법은 모든 감정을 경험하게 하되 그것을 행동으로 옮기지 못하도록 만다는 것이다. 사용자의 분노에 공감하고 그 자신도 화낼 줄 알지만, 그것을 실행에 옮기지 못하게 하거나 신체 능력이나 이동성 자체를 약화하는 것이다. 이는 결국 감정이란 개체가 가진 지적 역량과 이동성을 포함한 신체적 능력과 균형을 이루도록 생성되는 것임을 시사한다.

5. 일방적 감정 소통의 위험성에 대비하기

진정한 인공 감정의 실현가능성이 낮다는 주장이 현 시점에서 최종적인 결론일 수 없다. 감정을 내적으로 가지는 로봇이 아니더라도, 사람과 교감하는 사회적 서비스 로봇이 가져올 문제에 대한 진지한 성찰이 요청된다. 이것은 미래의 문제가 아니라 지금의 문제이다. 사교 로봇(social/sociable robots)은 산업용 로봇이나 컴퓨터, 또는 다른 가전제품과는 다르게 취급된다는 점에 주목해야 한다. 공장 내에서만 작동하는 산업용 로봇과 달리 사교 로봇은 폭넓은 환경에서 작동하도록 설계되고, 그것의 외양은 인간이나 동물을 닮도록 제작된다. 산업용 로봇이 특정한 과제를 수행하기 위해 프로그램된 반면, 사교 로봇은 일부 개방성을 갖는다. 사교 로봇은 제한된 범위 내에서 이동도 가능하고 행동 출력을 갖는다는 점에서 이동성이 없는 컴퓨터나 산업용 로봇과는 다르며, 일정 수준의 자율성을 갖는다는 점에서 독특하다. 이 모든 특성이 진정한 감정의 소유를 전제하지 않는다는 점이 중요하다. 사람들은 이러한 특성을 가진 인공물의 행동을 해석하기 위해 의도 등의 심성 상태를 부여하기 쉽다. 인간의 감시나 개입 없이 과제를 수행하는 로봇을 인간은 자율적인 존재로 인식할 가능성이 높으며,[12] 그럴수록 더 인간처럼 느끼고 의인화하기도 쉽다.[13]

감정 표현을 할 줄 아는 로봇이 인간 행동에 어떤 영향을 끼칠 수 있는지 보여주는 실험들 가운데 하나만 소개하자. 로봇과 팀을 이루어 과제를 수행하게 한 어떤 연구에서, 로봇은 자율적으로 판단하지 않고 인간의 명령을 따르도록 했다. 한 조건(“감정” 조건)에서는 로봇이 긴박함을 소리로 표현하거나 인간이 받는 스트레스를 감지해 그에 알맞은 대응을 하도록 했고, 다른 조건(“비감정” 조건)에서는 로봇의 소리에 변화를 주지 않았다. 실험참여자들은 둘 중

12) 지금 문제는 로봇이 객관적인 의미에서 얼마나 자율적인가 하는 것이 아니라 인간이 로봇의 자율성을 어떻게 혹은 어느 정도로 인식하는가, 그리고 그런 인식이 인간 행동에 어떠한 영향을 미치는가 하는 것이다. 로봇의 자율성을 측정하는 몇몇 변수들에 관해서는 Huang (2004)를 참조.

13) 로봇이 인간과 같은 의미에서 자율적이라고 주장하는 것이 아니다. 로봇이 스스로 행동의 준칙을 결정하고 그에 따른다는 의미나, 자신의 행동에 책임을 져야하고 또 지려고 한다는 의미에서 자율성을 가진다는 뜻이 아니다. 다만, 인간의 직접적인 개입 없이도 일정 수준에서 로봇은 환경을 인식하고 상황을 판단하고 행동을 출력할 수 있음을 의미한다.

한 조건에만 참여했고, 연구팀은 두 조건에 참여한 사람들의 행동을 서로 비교했다. 그 결과, 로봇에게 소리를 통한 감정 표현을 허용한 경우 팀의 과제수행 능력이 그렇지 않은 팀에 비해 객관적 지표상으로 더 높게 나타났다. 또한, 감정 조건에 참여한 사람들은 실험 전과 비교에서 로봇에 대한 호감도가 높아지고, 로봇이 감정을 가져야한다는 생각을 조금 더 많이 하게 되었다.(Scheutz et al. 2007)

사람들은 사교 로봇에 애착을 느끼거나 교감한다고 생각하는 경향이 있다. 사교 로봇은 다른 인공물들과 달리 취급된다. 사람들은 그것들에 이름을 붙여준 후 이름을 자주 불러주고, 사진을 찍어 공유하거나 가족들과 친구들에게 소개해주기도 한다. 군사로봇과 함께 전장을 누빈 군인들은 로봇에게 계급을 붙이고 승진시켜주기도 한다. AIBO를 키웠던 여성 사용자의 경우, 로봇이 지켜본다고 느껴서 욕실에서 옷을 벗을 때 문을 닫는다. 이 같은 경향은 일반인뿐 아니라 로봇을 제작하는 전문가에게서도 나타난다. 로봇공학자는 로봇에게 감정이 실제로 존재한다고 믿지 않지만 그것에 상당한 정서적 애착을 가진다. 키즈멧을 만들었던 브리질(Cynthia Breazeal) 박사는 박사학위를 했던 MIT 연구실을 떠나면서 키즈멧과 떨어져야하는 상황에 직면해 감정의 동요를 느꼈던 것으로 알려져있다.

사교 로봇을 쉽게 인격화하면서 감정을 무의식적으로 부여하는 이런 현상을 감정의 탈인용부호 현상이라 부를 수 있다. 사람들의 명시적 믿음 체계 속에서 로봇의 “감정”은 따옴표 속에 있지만, 실제 행동에서는 그 따옴표가 쉽게 사라지기 때문이다. 이로 인해 사람들은 로봇에 대해 일방적인 정서적 유대감을 가질 수 있다. 상대는 감정을 실제로 가진 존재가 아닌데도 우리는 그것을 의인화해 감정을 가진 존재처럼 대하기 때문에 여러 문제들이 생겨날 수 있다.

사교 로봇에 대한 심리적 의존으로 인해, 사용자가 조종되거나 착취당할 가능성이 존재한다. 예컨대, 정서적 유대를 맺고 있는 로봇이 사용자에게 무언가를 요구한다면 사용자는 그에 부응해 요구를 들어줄 가능성이 높다. 만일 로봇 강아지가 집을 지키던 반려견을 가리키면서 “제발 그 개를 없애주세요. 너무 무서워서 견딜 수가 없어요.”라고 말한다면, 사용자는 심각한 고민에 빠질 수도 있다. 사교 로봇을 제작하는 기업이나 로봇의 제작과 유통에 관련된 일군의

사람들이 로봇과 맺는 정서적 유대를 이용해 사용자를 착취할 가능성이 있다. 로봇을 이용해 회사가 출시하는 새로운 제품을 구매하도록 설득하거나 유도하는 것이 가능하다. 특히, 돌봄 로봇의 주된 대상 가운데 이런 문제를 더 심화될 수 있다.

사람들이 로봇에게 일방적인 정서적 유대감을 형성하도록 로봇을 설계, 제작하기란 그리 어렵지 않다는 데 문제의 심각성이 있다. AIBO의 경우 겉모습이 강아지와 닮았고, 꼬리를 흔들거나 짓는 등 몇 가지 행동을 흉내내는 것뿐이지만 AIBO 소유자들이 그것에 보여준 애착은 반려견에 못지 않았다. 보스턴 다이나믹스가 제작한 로봇 스팟(Spot)은 사족 보행이 가능한 로봇으로 계단이나 산악 지형을 포함해 다양한 지역을 정찰할 수 있다. 회사가 유튜브에 게시한 홍보 영상에서 한 연구자는 로봇을 힘껏 걷어차는데, 로봇은 균형을 잃고 엉거주춤하다가 다시 네 발로 균형을 잡는다. 그런데 이 영상 밑에 달린 많은 댓글들은 마치 실제 강아지가 불쌍하게도 걷어차인 것처럼 반응했다. 심지어, 로봇 청소기 룸바의 경우 사교 로봇으로 분류하기도 어렵고 동물을 닮은 것도 아님에도 불구하고, 아마도 그것이 보여주는 자율적 움직임 덕분에 사람들은 룸바에 감사하는 마음을 느끼는 것으로 나타났다. 물론 룸바가 사람을 알아볼 수 있는 것도 아니다.

한편, 로봇 산업계는 사교 로봇의 인격화를 부추기고 있는 것처럼 보인다. 자신들이 제작하고 판매하는 로봇이 얼마나 실제적인지 강조하기 위해 “배고파”, “엄마, 사랑해요”, “진짜 우리 아기” 등의 단순한 문구들을 사용하고 있다. 때로는 페이스북 페이지(iRobot’s PackBot)를 개설해 마치 그것이 어떤 상황들을 경험하고 있는 것처럼 일인칭의 관점에서 소식을 전하기도 한다. 사교 로봇이 “가족 구성원”으로 대우받는다는 관점은 이제 흔한 일이 되었다.

사람들에게 로봇이 의식을 가지는지, 인격체이거나 동물인지, 도덕적 행위자로 볼 수 있는지 등을 묻는다면, 대부분 부정적인 답을 내놓을 것이다. 탈인용부호 현상은 사람들의 행동이 무의식적 차원에서 깊은 영향을 받고 있음을 말해준다. 인류는 사회적 동물이고 타자들과의 사회적 상호작용이 유전자 깊숙이 각인되어 있다. 우리는 단순히 물리적으로 설명되지 않는 현상을 만날 때, 그 대상의 심성 상태, 믿음, 욕구, 의도 등에 대해 자동적으로 추론하는 경

향이 있다. 특히, 유아들의 경우 그런 태도가 적용되는 대상의 범위가 넓다. 지금의 유아들은 로봇과 함께 교감하는 것을 자연스레 체득하는 첫 세대가 될 가능성이 있다.

미래 로봇은 더욱 정교해질 것이다. 현재의 조야한 로봇에 대해서도 사람들은 쉽게 의인화하는 경향이 있는데, 앞으로 로봇을 인격화하는 정도는 더욱 심화될 것이다. 미래 로봇은 더욱 인간과 닮은 외양을 갖추게 될 것이고, 자연언어를 통해 매끄럽게 상호작용할 것이며, 인간 얼굴의 미세한 근육의 움직임까지 포착해 표정을 읽어내고 자연스러운 감정 표현을 보여줄 것이다. 인간이 사교 로봇을 더 신뢰하고 더 깊은 감정적 유대감을 형성할수록, 속임수나 조종의 가능성도 커진다. 로봇에 대한 사람들의 신뢰감이 더 커지게 되면, 사람들의 솔직한 답변을 신빙성있게 청취해야 하는 경우 (예컨대, 현재 여론조사원이나 교사, 상담사 등이 하는 일에 대해) 로봇에게 그 일을 맡기거나, 물건을 판매하기 위해 로봇 판매원을 고용하는 것도 시간문제일 것이다. 결국에는 인간이 로봇과 맺는 관계가 일방적이라는 사실조차 깨닫기 점점 더 어려워질 수도 있다.

사교 로봇이 인간을 덜 외롭게 해줄 수 있을지도 의문이다. 복잡하게 얽힌 대인 관계에 지쳐있거나 혼자서 많은 시간을 보내는 사람들에게 감정 로봇과 맺는 정서적 유대감은 분명 긍정적으로 기능할 것이다. 그것은 친구나 가족 관계를 보완할 수 있다. 대하기 쉽지 않고 불편한 다른 사람들보다 나에게 공감해주는 로봇과 깊은 관계를 형성하는 사람들도 다수 생겨날 수 있다. 기계에 더 많이 의존하고 사람과의 대면 접촉을 피한다면, 결국 우리는 "함께 외로울" 뿐이다. 예를 들어, 영화 그녀(Her)에서 남자주인공 테오도르는 아내의 미묘한 성격에 맞추는 것이 영 불편하고, 오히려 자신에게 모두 맞추어주는 운영체계 사만다에게 더 깊은 애착을 느끼게 된다. 이 분야에 대한 심리학적 연구를 오랫동안 수행해온 MIT의 셰리 터클은 쌍방향의 친구 맺기를 요구하지 않는 교류란 환상일 수 있음을 잘 보여준다.(Turkle 2010)

맥카시(McCarthy 1999)는 인간과 같은 로봇을 생산하면 생길 수 있는 잠재적 위험을 지적한 바 있다. 감정을 가진 로봇을 들여오기에 이미 인간 사회는 충분히 복잡하다. 그렇다고 해서 감정 로봇의 연구 및 개발을 전면적으로 중단

하자는 주장은 가능하지도 바람직하지도 않다. 전면적인 모라토리엄 선언은 앞서 언급된 몇몇 문제들을 해결하는 데 도움이 될 수 있겠지만, 다른 인공 지능 및 로봇 기술을 연구하면서 사교 로봇 연구만을 중단한다는 것은 현실적이지 않다.

한 가지 방안은 담뱃갑에 경고 문구를 붙여 담배의 위해성을 경고하듯, 로봇이 작동할 때마다 그것의 외양이나 특정 행동을 통해 로봇은 실제로 감정을 가진 것이 아니며 로봇의 정서적 행동은 인간이 감정을 가지고 행위하는 것과 동일하지 않음을 알려주도록 로봇을 설계하는 것이다. 이렇게 하면 사람들이 자연스럽게 의인화하는 경향을 막을 수는 없겠지만 경감시킬 수는 있을 것이다. 그러나 감정 로봇의 목적이 믿을만하고 자연스럽게 사람들과 인간적인 방식으로 의사소통하는 것인데, 그런 목적을 훼손하지 않으면서 바라는 효과를 얻을 수 있을지는 의문이다.

로봇이 정말로 인간과 같은 감정이나 느낌을 갖도록 만들 수 있다면, 적어도 다른 인간에게 당하는 것 이외의 방식으로 우리가 로봇에게 조종당하지는 않을 것이다. 물론 우리가 전혀 기만당하지 않을 것이라는 뜻은 아니다. 사람들은 서로 속고 속이며, 서로를 이용한다. 그러나 만일 로봇이 진정한 감정을 갖는다면, 인간이 다른 인간을 속이는 방식과 다른 방식으로 우리가 로봇에게 기만당하는 일은 없을지 모른다. 그러나 진정한 인공 감정을 제작하기란 현실적으로 어렵다.

상품화되는 모든 감정 로봇에 도덕적 추론을 내장하도록 제도화하는 것도 한 가지 가능한 대응 방안이다. 그러나 이러한 방안에는 언제나 그렇듯이 "어떻게"의 문제가 따라온다. 도덕적 추론 능력을 장착하기 위해, 도덕적 원리를 집어넣어야하는지 아니면 경험으로부터 배울 수 있도록 해야 하는지, 만일 원리를 넣어주어야 한다면 어떤 원리가 일차적으로 입력되어야하는지, 그리고 로봇 내에서 그런 원리들을 작동시켜 실시간으로 반응하게 만드는 것이 가능한지 논의가 필요하다. 우리는 잘 알려진 아시모프의 세 법칙이 실제로 구현되기 어렵다는 사실을 잘 알고 있다.(고인석 2011) 그러한 난점을 피해, 일방적인 감정 소통이라는 특성으로 인해 우리가 감정 로봇에게 기만당하지 않도록 로봇 안에 일정한 장치를 마련할 수 있는지 검토해야한다.

6. 결론을 대신하여

인간이 경험하는 풍부한 감정들로 인해 우리는 인간다운 삶을 살아간다. 그러나 인공 감정이나 감정 로봇은 논리적 모순이 아니다. 로봇과 같은 인공물이 언젠가는 지능뿐 아니라 감정을 가질 수 있는 가능성을 원천적으로 배제할 필요는 없을 것이다. 만일 인간 수준의 혹은 인간의 지적 능력을 초월하는 지능을 가진 인공물이 어떤 종류의 심성 상태를 가질 수 있다면, 그것이 감정까지도 소유할 가능성에 관한 논의는 단지 허튼소리는 아닐 것이다. 그러나 어떤 대상이 감정을 소유한다고 판단하기 위해서는 까다로운 조건들이 충족되어야 한다. 복잡하고 때로는 적대적인 환경에서 자신에게 주어진 자극이 자신의 생존과 항상성 유지에 어떤 가치를 가지는지 평가하여 적응적으로 행위할 수 있는 행위자만이 감정을 소유하기 위한 기본적인 조건을 갖추었다고 볼 수 있다. 인공지능이 단지 사람의 감정 표현을 인식하고 흉내내는 것을 넘어 진짜 감정을 가진 존재로 진화하려면 어쩌면 유기체와 같은 신체를 소유해야할지도 모르겠다. 우리가 그러한 인공지능을 원하는지 나로서는 확신할 수 없다. 그러나 먼 미래에 발생할지도 모르는 진정한 인공 감정을 논의하기 앞서, 감정 로봇과의 일방적 정서적 교감이 가져올 수 있는 잠재적 위험을 예상하고 이에 대비하는 것이 필요하다.

참고문헌

고인석. 2011.「아시모프의 로봇 3법칙 다시 보기」.『철학연구』93:97-120.

김평수. 2016.「인간과 교감하는 감성로봇 관련 기술 및 개발 동향」. 『정보와 통신』33(8): 19-27.

랭던 위너. 손화철 옮김. 2010.『길을 묻는 테크놀로지』. 서울:CRI.

박천수, 류정우, 손주찬. 2008.「로봇과 감성」.『정보과학회지』26(1): 63-69.

빌라야누르 라마찬드란. 이충 옮김. 2016.『뇌는 어떻게 세상을 보는가』. 서울:바다출판사.

안호석, 최진영. 2007.「감정 기반 로봇의 연구 동향」.『제어로봇시스템학회지』13(3): 19-27.

이동욱, 김홍석, 이호길. 2008.「감성교감형 로봇 연구동향」.『정보과학회지』26(4): 65-72.

이원형, 박정우, 김우현, 이희승, 정명진. 2014.「사람과 로봇의 사회적 상호작용을 위한 로봇의 가치효용성 기반 동기-감정 생성 모델」.『제어로봇시스템학회 논문지』20(5): 503-512.

이찬종. 2009.「로봇의 감정 인식」.『로봇과 인간』6(3): 16-19.

천현득. 2008.「감정은 자연종인가: 감정의 자연종 지위 논쟁과 감정 제거주의」.『철학사상』27: 317-346.

Bijker, Wiebe E., Thomas P. Hughes, and Trevor Pinch. 1987. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, Mass.: MIT Press.

Boden, Margaret A. 1990. *The Philosophy of Artificial Intelligence*. New York: Oxford University Press.

Bostrom, Nick. 2014. Superintelligence: *Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Breazeal, Cynthia, and Rodney Brooks. 2005. “Robot emotion: A functional perspective.” In Fellous, Jean-Marc, and Michael A Arbib (2005)

Damasio, Antonio R. 1994. *Descartes' Error: Emotion, Reason, and the Human*

Brain. New York: Putnam.

Fellous, Jean-Marc, and Michael A Arbib. 2005. *Who Needs Emotions?: The Brain Meets the Robot*. New York: Oxford University Press.

Frankish, Keith, and William M. Ramsey, eds. 2014. *The Cambridge Handbook of Artificial Intelligence*. New York: Cambridge University Press.

Hindi Attar, Catherine, and Matthias M. Müller. 2012. "Selective Attention to Task-Irrelevant Emotional Distractors Is Unaffected by the Perceptual Load Associated with a Foreground Task." *PLOS ONE* 7 (5):e37186.

Kaku, Michio. 2014. *The Future of the Mind: the Scientific Quest to Understand, Enhance, and Empower the Mind*. New York: Doubleday.

Kuhn, Thomas S. 1957. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, Mass.: Harvard University Press.

Kurzweil, Ray. 2005. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

McCarthy, John. 1995. "Making Robots Conscious of Their Mental States." Machine Intelligence 15.

Megill, Jason. 2014. "Emotion, Cognition and Artificial Intelligence." *Minds and Machines* 24:189-199.

Minsky, Marvin. 1986. *The Society of Mind*. New York: Simon and Schuster.

Noble, David F. 1984. *Forces of Production: A Social History of Industrial Automation*. New York: Oxford University Press.

Pinch, Trevor and Wiebe E. Bijker. 1987. "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." In W. E. Bijker, T. P. Hughes, and T. Pinch, eds, (1987)

Pessoa, Luiz, and Leslie Ungerleider. 2004. "Neuroimaging studies of attention and the processing of emotion-laden stimuli." *Progress in Brain Research* 144:171-82.

Russell, Stuart J., and Peter Norvig. 2010. *Artificial Intelligence: a Modern Approach*. 3rd ed, Upper Saddle River, N.J.: Prentice Hall.

Scheutz, Matthias, Paul Schermerhorn, James Kramer, and David Anderson. 2007. “First steps toward natural human-like HRI.” *Autonomous Robots* 22 (4):411-423.

Searle, John R. 1992. *The rediscovery of the mind, Representation and mind*. Cambridge, Mass.: MIT Press.

Turkle, Sherry. 2011. *Alone Together: Why we expect more from technology and less from each other*. New York: Basic Books.

Waytz, Adam, and Michael Norton. 2014. “How to Make Robots Seem Less Creepy”, *The Wall Street Journal*, June 1, 2014.

Abstract

Artificial intelligence and artificial emotions

- Is an emotion robot realizable? -

Hyundeuk Cheon

As artificial intelligence outperforms humans in some cognitive tasks which have been regarded as properly belonging to the human being, there is growing anxiety that people experience. Now, many people attempt to place the uniqueness of the humans in emotion rather than reason or intelligence. Since artificial emotions and emotional robots are emerging as new topics in artificial intelligence and robotics, however, philosophical treatment of them is required. I begin with discussing what has motivated to develop emotional robots and why emotions matter in robotics. After reviewing the current status of emotional robots, I examine whether the emotion-possessing robots are possible. The possibility of artificial emotions depends on what emotions really are. Instead of stipulating the definition of emotions, I suggest criteria on which we can assign emotions to other people, animal, or alien objects including artificial agents by considering the key roles played by emotions in intelligent beings. On these criteria, I argue, it is unlikely for genuinely emotional robots to be realized in the near future. Still, even if emotional robots would not appear shortly, unidirectional emotional bonds with social robots equipped with some degree of autonomy are potentially detrimental, which ought to be seriously considered in an ethical discussion of robots.

Subject Areas: Philosophy of Technology, Philosophy of Mind

Keywords: Artificial Intelligence, Emotion, Artificial Emotions, Dis-quotational Phenomenon

투고일: 2017년04월04일 **심사일:** 2017년04월15일~05월19일 **게재확정일:** 2017년05월22일