과학철학 22권 2호 (2019) pp. 27-52

인공 반려의 유혹:

인공물과의 교감을 생각한다†

천 현 득‡

인공지능 로봇이 사회적, 정서적 상호작용의 대상으로서 사람들의 일상적 공간 안으로 들어올 것으로 예상되는 상황에서, 이 글은 우리가 똑똑한 인공물과 어떤 식으로 교감할 수 있는지 다루고자 한다. 특히, 인공 반려의 가능성과 잠재적 문제에 초점을 맞춘다. 인공 반려를 개발하는 배경을 간략히 살펴본 후, 어떠한 조건들이 만족되는 경우 어떤 대상이 우리와 반려의 관계를 맺을 수 있는지 점검해본다. 반려 관계를 맺기 위해서는 정서적 교감 능력이 필요함이 드러난다. 이 글에서 나는 교감의 형태를, 마음 읽기, 감정적 전염, 동정, 공감으로 구분하고, 진정한 공감 능력은 인공지능이 가까운 미래에 획득하기 어렵다고 주장한다. 마지막으로, 공감 능력이 없지만 사람들이 공감하는 존재로 대우하는 반려대체물의 성격을 규명하고, 그것이 가져올 몇 가지 윤리적, 사회적 문제에 관해 언급한다.

【주요어】 감정 로봇, 소셜 로봇, 인공 지능, 인공 반려, 공감, 진정성

† 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF 과제번호: NRF-2018S1A5B8069796).

‡ 서울대학교 철학과, hdcheon@gmail.com.

1. 인공지능과 관계 맺기

급속한 기술 발전이 단지 산업적, 경제적 차원에서 사회를 변화시킬 뿐 아니라 동시대인들의 삶의 양식에도 영향을 끼친다면, 이에 관한 분석과 대응은 응당 철학자의 임무 가운데 하나일 것이다. 이 글에서 관심을 가지는 기술은 인공지능이다. 인공지능은 인간 지능을 포함한 지능 일반을 탐구하여 지능을 물리적으로 (혹은 가상적으로) 구현하고, 그렇게 함으로써 인간 지능과 지능 일반에 대한 이해를 심화하려는 연구 분야이다. 인공지능은 흔히 공학기술의 일부로만 간주되기 쉽지만, 인공지능에 대한 이해는 지능의 본성에 관한 이해, 더 나아가 인간에 관한 이해와 분리될 수 없다. 인공지능은 지구상에서 가장 뛰어난 자연 지능 가운데 하나인 인간을 비추는 거울이자, 인간에게 자기각성을 일으키는 환기적 대상이라는 사실에 주목할 필요가 있다. 인공지능 기술이 급속히 발전하고 있는 이 시대에 인지와 문제해결 영역에서 기계의 추월을 의식하게 되면서, 사람들은 더욱 인간의 본성에 관심을 기울이게 되는 듯하다. 인간에게 고유한 특성으로서 감정, 공감, 그리고 사회성이 부각되고 있다. 우리는 서로에게서 자신의 모습을 발견하고 공감하며 동료애를 느끼고 공동체를 이루어 살아가는 존재이기 때문이다. 물론 의사소통은 이성적인 대화나 문자의 교환을 통해 이루어질 수도 있지만, 타인의 마음을 읽고 그에 적절히 반응하는 능력은 공동체의 구성원으로 살아가기 위해 필수적이다.

공감과 사회성은 인간만의 영역일까? 최근 급속히 발전하고 있는 인공지능은 단지 주어진 과제를 신속하고 효과적으로 수행하는 것을 넘어 인간과 의사소통이 가능한 형태로 진화하고 있다. 공장에서 정해진 과제를 반복적으로 수행하는 제조 로봇의 성능도 향상되고 있지만, 로봇공학의 무게 중심은 지능형 서비스 로봇 쪽으로 점차 옮겨가고 있다. 최근 연구자들은 바둑이나 퀴즈와 같은 일부 인지 영역에서 지능적인 행위를 보여주는 인공지능을 넘어, 사람과 상호작용하는 영역에서도 활용될 수 있는 인공지능을 설계하고 제작하려 한다. 이 글에서 나는 인공지능 로봇이 사람들의 일상적 공간 안으로 들어와, 사람들

사이의 정서적, 사회적 관계를 보완하거나 대체하게 될 때 우리가 고민해야 할 문제 가운데 일부를 숙고해보고자 한다. 인공지능과 맺게 될 바람직한 관계의 양식을 분석하는 일은 인류가 이미 바람직한 사회적 관계를 맺고 있음을 전제로 하지는 않는다. 다만, 사람들이 맺고 있는 사회적, 정서적 관계는 새롭게 다가오는 인공지능과의 관계보다 비교적 더 잘 이해되고 있다고 가정해도 좋을 것이다. 인공지능 로봇이 사람들의 일상적인 사회적 관계 속으로 진입하기 시작하는 상황에서, 우리는 로봇과 어떤 관계를 맺을 수 있는지, 또 어떤 관계가 바람직한지 검토해 보아야한다. 똑똑한 인공물은 단순한 사물이나 도구가 아니라 대화 상대, 동료, 혹은 친구가 될 수 있을까? 이 글은 인공 반려(artificial companionship)의 개념을 도입함으로써 이 문제에 접근한다. 즉, 인공 반려의 가능성과 조건, 그리고 잠재적인 위험에 관해 생각해 본다.

2. 감정 로봇

먼저 논의 대상을 한정하자. 이 글에서 다루는 주된 대상은 물리적 신체나 가상적 표현을 가지고 인간 및 환경과 상호작용하는 인공지능이다. 이것은 휴머노이드나 안드로이드, 가상적인 대화 행위자를 모두 포함하며, 통칭하여 체현된 인공지능(embodied AI) 혹은 인공지능 로봇(줄여서 “로봇”)으로 부를 수 있다. 물론 모든 로봇이 인공지능일 필요도 없고, 반대로 인공지능이 모두 로봇의 형태로 존재해야 하는 것도 아니다. 예컨대, 이세돌과 대국을 했던 알파고는 시각 정보를 처리하는 장치나 바둑돌을 옮기는 행동 장치를 가지지 않은 (물론 거대한 하드웨어들 안에서 수행된) 소프트웨어 프로그램이었지만 로봇은 아니었다. 반대로, 공장에서 자동차 부품을 조립하는 데 사용되는 수많은 로봇팔은 인공지능이 아니다. 이 글의 관심은 인공지능 소프트웨어가 탑재된 로봇이다. 인공지능 로봇은 두 가지 흐름이 교차하는 지점에서 등장한다고 볼 수 있다. 하나는 어떤 것이 생각하는 존재이기 위해서

는 내적 표상에 대한 계산 과정만으로는 불충분하며 신체를 가지고 물질적 환경과 상호작용하는 것이 필수적이라는 보는 체현된 인지(embodied cognition) 접근이며,[1] 다른 하나는 로봇이 더 나은 지능을 갖도록 함으로써 단순한 육체노동을 대체하는 수준을 넘어 인간과 의사소통이 가능한 수준까지 발전시키려는 로봇공학과 업계의 노력이다.[2]

최근 로봇공학에서는 감정 인식 기능의 필요성을 강조하고 있다. 인간의 심성 상태와 감정을 인지하고 이에 적절히 반응할 수 있는 "감성 로봇(emotion robot)" 혹은 "사귐 로봇(sociable robot)"은 인공지능 로봇이 단순한 사물이나 보조 수단이 아니라 동료나 반려로서 대우받을 수 있는 가능성을 시사한다. 선도적인 감성 로봇 연구가인 브리질은 로봇을 네 종류, 즉 도구, 사이보그 연장, 아바타, 파트너로 구분한 바 있다(Breazeal and Brooks 2005). 어떤 형태의 로봇이더라도, 사용자의 미세한 감정 변화를 인식하는 로봇은 미리 정해진 프로그램이 명령하는 대로만 움직이는 로봇보다 더 나은 서비스를 제공할 수 있을 것이다. 인공지능(AI) 스타트업 어펙티바(Affectiva)의 창립자이자 최고경영자인 칼리오비(Rana el Kaliouby)는 알렉사(Alexa)와 같은 인공지능 스피커의 음성 인식 소프트웨어가 감정이나 분위기를 인식하지 못한다는 점에서 결점이 있다고 말한다. 그가 한 발표의 리허설에서 알렉사를 언급하자 갑자기 알렉사가 깨어나 세레나 고메즈의 음악을 재생했는데, 그가 몇 번씩이나 "알렉사, 멈춰!"라고 소리 친 후에야 노래가 멈췄다고 한다. 알렉사는 그의 짜증에 무심했던 것이다.[3] 알렉사가 칼

1) 체현된 인지와 로봇에 관해서는 Anderson (2003), Brooks (1991), Varela, Thompson, and Rosch (1991) 등을 참조할 수 있다.

2) 세계 로봇 선언(2004, 후쿠오카)는 다음과 같이 선언한다. 첫째, 차세대 로봇은 인간과 공존하는 파트너가 될 것이다. 둘째, 차세대 로봇은 인간을 신체적으로 보조할 뿐 아니라 심리적으로도 보조할 것이다. 셋째, 차세대 로봇은 안전하고 평화로운 사회의 실현에 기여할 것이다.

3) Kaliouby (2017). 이때, 칼리오비가 여러 형태의 정서적 교류를 선명하게 구분하고 있지 않고, 감정(emotion), 분위기(mood), 공감(empathy) 등의 개념이 혼용되고 있는 것처럼 보인다. 나중에 보겠지만, 여러 형태의 교감 방

리오비의 감정을 알아챘다면 결과는 달랐을 것이다. 자율주행차가 사용자의 정서 상태를 파악할 수 있다면 피곤해 보이는 운전자에게 운전을 멈추도록 제안할 수도 있고, 사용자의 감정을 파악할 수 있는 냉장고는 더 건강한 식사를 도와줄 수도 있을 것이다. 이와 같은 감정과 분위기를 인식하는 기술은 개인 맞춤형 서비스를 제공함으로써, 교육, 간병, 노년층 돌봄, 연예 등 많은 분야에서 응용될 것으로 예상된다. 몇 가지 형태의 감성 로봇은 이미 가정, 양로원, 병원 등에서 사용 중이며, 앞으로 그 사용은 더 늘어날 것이다(Robins et al. 2006; Wada et al. 2006; Levy 2007).

사용자의 필요와 정서 상태를 인지할 수 있는 로봇이 각광받는 데에는 여러 이유가 있을 것이다. 첫째, 개발자의 입장에서 보면, 현재 많은 인공지능 소프트웨어와의 의사소통이 단발성에 그치는 경우가 많기 때문에, 사용자와 인공지능 사이의 의사소통 시간을 늘리는 것은 중요한 과제이다. 감정 능력을 갖춘 로봇은 인간 사용자와의 의사소통을 더욱 자연스럽게 만듦으로써 대화 상대로서의 수용성을 높여줄 수 있고, 사용자가 그것과의 장기적 의사소통에 기꺼이 응하도록 만들 수 있다. 만일 로봇이 인간이나 동물과 비슷한 외양을 가지면서 정서적 교류가 가능한 것으로 우리에게 인지된다면, 지속적이고 장기적인 의사소통이 가능해질 수 있다.

둘째, 감정 인지 능력은 로봇이 주어진 과제를 더 효과적으로 수행하는 데에도 도움이 된다. 사람들의 사회적 상호작용은 미묘하고 복잡하기 때문에, 여기에 개입하려면 사회 지능과 감정을 갖추어야 한다. 사람들을 귀찮게 만들지 않으면서 필요를 충족시키려면 사용자의 분위기를 파악하는 섬세한 감각이 필요하다. 예컨대, 환자를 돌보는 로봇이라면 환자의 고통, 불편 증상, 피로감 등의 징후를 파악할 수 있어야 한다. 환자의 감정 상태를 살펴, 그가 투약을 거부하지 않도록 배려하거나, 투약에 쉽게 응할만한 분위기를 파악하여 적시에 투약을 권할 수도 있다.

셋째, 우리는 반려와 더불어 살아가는 존재이다. 이때, 반려

식을 구분할 필요가 있다.

(companion)란 단지 직장 동료나 사업상의 거래인과 같이 감정을 동반하지 않고 사업이나 업무상의 이유에서 만나는 사람들이 아니라, 친구나 가족처럼 최소한의 친밀성(intimacy)에 기반한 관계를 뜻한다.[4] 반려 관계를 맺는 사람을 우리는 반려자라고 부른다.[5] 반려는 완전히 대칭적일 수는 없지만 어느 정도 상호적 관계이다. 우리는 이해 받고 위로 받고 돌봄 받기 원하면서, 동시에 위로를 건네고 돌보고 싶어 한다. 특히 과도한 업무로 인한 스트레스와 공동체의 해체로 인한 외로움을 감당해야할 현대인들에게 공감하고 공감 받을 대상이 절실하게 필요하다. 그런데 정보통신 기술의 발달로 인해 사람들 사이의 의사소통방식도 이미 변화하고 있다. 사람들은 대면 접촉 대신 이메일이나 메신저, 소셜네트워크(SNS)로 소통한다. 어쩌면 감정 로봇과 같은 인공지능의 산물은 현대인에게 반려가 될 수 있을지도 모른다. 많은 현대인들은 외로움을 덜어줄 수 있는 기술적 수단을 원한다.

인공지능 로봇은 정말 인류에게 좋은 반려(자)가 될 수 있을까? 인공 반려의 조건은 무엇이고, 우리가 로봇과 맺을 적절한 관계는 무엇일까?

---

4) “companion”의 어원이 함께 빵을 먹는 것에 있음을 상기하라. 흥미롭게도 이는 우리말의 “식구”와 유사한 의미를 지닌다.

5) 이 글에서 “반려”는 관계를 뜻하기도 하고 관계를 맺는 당사자를 뜻하기도 한다. 그런 점에서 다소 느슨하고 포괄적으로 “반려”라는 개념이 사용되고 있다. 예컨대, 인공 반려는 인공물과의 반려 관계를 뜻할 수도 있고, 반려 관계를 맺고 있는 인공물을 뜻할 수도 있다. 물론, 반려는 관계의 일종으로, 반려의 관계를 맺는 당사자는 반려자로 구분해서 볼 수 있다. 예컨대, 인공물과의 관계는 인공 반려로, 관계 맺는 대상은 인공 반려자로 구분하는 것도 가능하다. 그러나 일상적인 표현에서 “반려자”는 배우자 혹은 그와 유사한 의미로 사용되는 경우가 흔하기 때문에, 이 글에서는 명시적으로 반려와 반려자를 구분해서 사용하지 않는다. 그럼에도 개념적으로 관계로서의 반려와 대상으로서의 반려가 구분될 수 있음에 주목할 필요가 있다. 이를 지적해주신 익명의 심사위원께 감사드린다.

3. 인공 반려의 조건

어떤 대상이 반려일 수 있는 조건을 검토해보자. 우리가 어떤 대상과 사귀고 단순한 공존 이상의 관계를 맺기 위해서는 어떤 조건들이 만족되어야할까? 우선, 사람들은 서로에게 반려가 될 수 있다. 그렇다고 모두가 모두에게 반려라는 뜻이 아니라 사람들은 서로에게 반려일 수 있는 기본적인 조건을 갖추고 있음을 뜻한다. 강아지와 고양이 같은 일부 고등동물들도 인간에게 반려일 수 있고, 이를 우리는 반려 동물이라고 부른다. 사람들은 그러한 고등 동물뿐 아니라 다양한 동물이나 식물, 심지어 무생물들을 집안에 들이고 애정을 주기도 한다. 그러나 뱀이나 곤충, 물고기 같은 애완동물이나 집에서 키우는 화초, 그리고 수집하는 수석 등 자연물은 반려가 되기 어렵다. 때로 우리는 소장하고 있는 예술작품에 애정을 쏟지만 그것을 반려로 보기는 어렵다. 우리가 애정을 쏟을 수 있는 대상의 범위는 넓지만, 그 가운데 일부만이 반려일 것이다. 반려 관계는 지속적이고 장기적인 의사소통과 교감이 가능한 경우에만 성립하는 것처럼 보이기 때문이다. 그리고 이를 위해서는 일정 수준의 지능이나 인지 능력, 그리고 감성 능력이 필요하다. 인공물이 반려로 간주되려면 만족해야할 몇 가지 조건을 차례로 살펴보자.

첫째, 인공 반려는 나를 알아볼 수 있어야 한다. 이를 재인 조건(recognition condition)으로 부를 수 있다. 어떤 것이 나와 반려 관계를 맺으려면, 그것은 사물들을 적절히 분류할 뿐 아니라 행위자(agent)들을 분간할 수 있어야하며, 그 가운데 나를 하나의 행위자로 인지하고, 나를 다른 누군가와 구별되는 개체로서 파악할 수 있어야 한다. 물론 이러한 재인 능력을 가지기 위해서는 사물을 배경으로부터 분리해내고 범주화하는 능력과 어떤 것을 의도와 마음을 가진 행위자로 볼 수 있는 능력도 갖추어야 한다. 게다가, 상당한 수준의 기억력을 갖추고 있어야하고, 과거의 상호작용을 기억해 그 이후에도 그런 정보를 활용할 수 있어야한다.

두 번째 조건은 호출가능성 조건(addressability condition)으로, 인공

반려는 내가 불러낼 수 있어야한다. 반려가 되려면 대화를 시작하기 위해 말을 붙이거나 과제를 지시하기 위해 언급할 수 있는 대상이어야 한다. 물론 애플 사의 아이폰에 장착된 “시리”와 같은 음성으로 호출 가능한 경우도 있지만, 인간과 외형적으로 유사한 휴머노이드 로봇이나 가상 행위자가 장기적인 관계를 맺기에 유리하다. 시공간상에 위치하면서 특정한 사회적 역할을 부여받은 경우 (예, 비서, 요리사, 운전사 등) 호출가능성이 높아진다.

셋째, 가장 핵심적인 조건은 나와 그 대상 사이의 정서적 상호작용이 가능해야한다는 것, 즉 교감 조건이다. 인공 반려가 보이는 반응은 다른 누군가가 아니라 나를 향한 것이어야 하고, 그 반응은 단순한 반사 반응이나 행동 단서에 의해서 촉발되는 자동적 반응이 아니라, 나의 심성 상태에 대한 이해(혹은 추정)에 기반을 둔 것이어야 한다. 나의 마음이 상대방에게 이해받고 있다고 느낄 때, 그리고 내가 상대방의 마음을 헤아릴 수 있을 때, 상호적 반려 관계가 가능해진다.

넷째, 인공 반려가 보이는 반응은 상황에 알맞고 자연스러워야 한다. 이를 적절성 조건으로 부를 수 있다. 사용자가 우울한 상태에 있는데 로봇이 웃는 표정을 짓거나, 상황과 전혀 관련없는 반응을 내놓는다면, 사용자와 신뢰할 수 있는 의사소통을 축적해 나갈 수 없다. 이를 위해 인공 반려는 상황에 관한 평가 능력을 가져야하며, 동일한 자극에 대해 늘 기계적으로 똑같은 반응을 산출하는 대신 상황에 알맞도록 유연하게 대처할 수 있어야 한다.[6)]

---

6) 인공 반려의 반응이 인간에게 자연스럽게 보이기 위해서는 그것의 겉모습도 중요하다. 예컨대, 휴머노이드 로봇을 제작할 때 “섬뜩한 계곡(uncanny valley)”을 고려해야할 것이다. 그런데 섬뜩한 계곡이 정말 존재한다면 (이에 관해서는 더 면밀한 경험과학적 탐구가 필요하다고 믿는다) 그것의 발생 이유에 관해 추정해보는 것도 좋겠다. 한 가지 가설은 예측가능성에 호소하는 것이다. 전혀 인간을 닮지 않은 인공물이라면 우리는 아무런 예상도 없이 그것의 행동을 주시할 것이지만, 그것이 인간의 모습을 무척 닮아 있다면, 인간과 같이 행동할 것으로 기대할 것이다. 그런데 그런 기대에 어긋나는 행동을 볼 때, 우리는 외양과 행동으로부터 내면을 추론하는 추론 기제가 작동하지 않음을 알게 됨으로써, 섬뜩한 느낌을 가질 수 있다.

다섯째는 성격 조건(character condition)으로, 장기적인 관계를 맺기 위해 반려는 안정적인 성격을 가져야 할 것을 요구한다. 우리는 사람에게뿐 아니라 반려견에게도 성격을 부여하곤 한다. 어떤 강아지는 상냥하고 다른 강아지는 새침하다. 인공물에게 성격을 부여하는 것은 이상한 이야기로 들릴 수도 있겠지만, 복잡한 대상들의 행태는 우리로 하여금 거기에 성격과 같은 것을 부여하도록 만드는 것 같다. 예컨대, 사람들은 알파고에게서 특정한 대국 스타일을 본다. 만일 인간이 그렇게 행동했더라면 특정한 성향이나 성격을 가졌을 것이라고 추론하는 경우, 우리는 그 대상에도 성격을 부여할 수 있을 것이다. 이러한 성격의 부여 내지 투사는 알파고가 실제로 어떤 성격을 가진다는 주장과 다르지만, 적어도 장기적이고 지속적인 관계의 형성하고 유지를 위해 필요한 최소한의 조건으로 볼 수 있다.

종합하자면, 서로를 알아보고 불러내고 자연스럽고 적절하게 그리고 정서적으로 상호작용하며 이러한 상호작용이 장기간 지속될 때, 우리는 반려 관계를 맺을 수 있다. 나와 타자가 이상적인 반려 관계를 맺는다면, 나의 삶과 그의 삶은 연결될 뿐 아니라 적어도 부분적으로 공동의 삶을 살아간다. 즉, 반려 관계를 맺는 이들은 생의 일부를 공유한다. 언급된 다섯 조건은 반려 관계를 형성하기 위해 (충분하지는 않지만) 필요한 조건들이다.[7] 즉, 이런 조건들이 만족되지 않고서는 인공물은 반려가 될 수 없을 것이다. 그 중에서도 가장 핵심적인 조건은 교감 조건이다. 인공물이 단지 사물이나 도구가 아니라 서로 교감하는 관계를 맺는 경우에만 우리는 그것을 반려로 고려할 것이기 때문이다.

---

7) 다섯 조건이 반려 관계를 형성하는 단계에서 필수적이더라도, 일단 형성된 반려 관계를 유지하기 위해서도 반드시 필요한지는 분명치 않다. 예컨대, 오랫동안 반려 관계를 맺던 상대가 인지 능력이 감퇴하여 나를 알아보지 못하거나 자연스러운 상호작용이 불가능해진 경우에도, 그와의 관계가 여전히 반려인지는 추가적인 논의를 필요로 한다. 다만, 최초에 반려 관계를 확립하는 단계에서는 위 조건들이 필요해 보인다.

4. 세 가지 관점: 실체, 관계, 규범

4.1 인공지능의 경우

철학의 한 가지 역할은 사태를 조망하는 렌즈와 관점을 제공하는 데 있다. 인공지능과 로봇의 미래를 이야기하면서 우리는 흔히 세 관점 사이를 오간다. 인공지능의 지능이나 공감 능력을 물을 때에도 마찬가지이다. 나는 이를 각각 실체적, 관계적, 평가적 관점으로 부르고자 한다. 실체적 관점에서 우리는 "X란 무엇인가", 혹은 "X는 Y를 가지는가"를 묻는다. 퀴즈쇼에서 우승한 IBM의 딥블루가 언어를 이해하는 능력을 가지고 있는지, 사람들과 자연언어로 대화하고 정서적으로 반응하는 소프트뱅크의 페퍼가 감정을 가지고 있는지 물을 때, 우리는 그것들의 실체적 속성에 관심을 갖는다. 사실 이 같은 물음의 방식은 철학의 역사를 지배해왔다. "지식이란 무엇인가", "정의란 무엇인가", "도덕이란 무엇인가"를 물었던 것처럼 이제 철학자들은 인공지능 연구자들이 내놓은 인공물은 정말 지능을 가지고 있는지, 감정을 가지는지, 의식이나 자유의지를 가지는지 묻는다.

관계적 관점에서 우리는 실체적 속성이 아니라 관계적 속성에 관심을 기울인다. 인공지능이 진짜 지능을 가지는지 또는 언제쯤 인간과 같은 지능을 가지는지 묻기보다, 지능적 행동을 보이는 인공지능을 사람들이 어떻게 대우하는지, 그것이 우리의 사회적 실천 양식을 어떻게 바꾸는지 묻는다. 즉, 우리와의 관계 속에서 그것들이 어떻게 취급되는지를 묻는 것이다. 예컨대, 딥블루가 중간 개발 단계에서 우스꽝스럽게 실수를 연발하자 사람들은 그것을 비웃었는데, 이에 대해 몇몇 개발자는 모욕감과 같은 감정을 느낀 것으로 보인다.[8] MIT 인공지능 연구실에서 키즈멧(Kismet)을 제작했지만 그것을 놓고 연구실을 나왔던 신시아 브리질이 자신이 만든 대상에 대해 일종의 모성의 느낌을 가졌던 사례는 잘 알려져 있다. 사람들이 딥블루, 알파고, 페퍼 등을 어떤 존재로 대우하는지, 그것이 사회적 실천의 양식을 어떻게 변화시키는지

8) "Smartest Machine on Earth", PBS (방송일: 2012.05.02.).

묻는다면, 이는 관계적 관점에 접근하는 것이다.

실체적 관점이 철학적이라면, 관계적 관점은 사회 심리적이다. 그러나 이 둘은 별개가 아니다. 평가적 관점은 실체적 관점과 관계적 관점 사이의 간격을 평가하고 재조정을 요청한다. 과연 알파고를 지능으로 대우하는 것이 적절한지 묻는 것이다. 기계가 생각할 수 있는 가능성을 진지하게 고려하게 된 시기에 이르기까지, 지능에 대해 실체적 관점과 관계적 관점이 내놓는 그림은 대체로 일치했다. 인간은 자신만이 고등한 지능을 가졌다고 믿었고 또 그렇게 대우했다. 그러나 인공지능의 등장은 인간 아닌 다른 대상이 지능적일 가능성을 시사하면서 두 관점의 분화를 가져왔다. 예컨대, 알파고가 지능인지는 매우 까다로운 문제가 되었다. 분명히 인간보다 뛰어난 과제수행 능력을 보여주지만, 인간의 지능과 다른 점도 있기 때문이다. 그렇지만 요즈음 사람들은 어떤 인공물이 지능을 가진다는 생각에 별다른 거부감이 없어 보인다. 예컨대, 아이폰의 시리를 인공지능 비서로 부르고, 알렉사를 인공지능 스피커로 부르며, 세탁기에 인공지능이 탑재되었다는 말을 발화하는 데 아무런 거리낌이 없다. 우리는 그것을 지능을 가진 것처럼 다룬다. 그렇지만 그것들은 정말로 지능을 가진 존재인가?

어떤 인공물(예, 알파고)을 지능으로 대우하는 것이 적절한지에 관해 논쟁이 생겨나는 경우, 지능의 의미에 대한 명료화와 개념적 분화를 통해 논쟁은 생산적으로 해결될 수 있다.[9] 평가적 관점에서 철학자들은 지능을 과제·영역에 특수한 지능과 일반지능을 구분한다. 이 구분에 의하면, 알파고는 바둑이라는 게임에 특화된 지능이지만 인간과 같은 일반지능은 아닌 셈이다. 세밀해진 기준에 의해서 알파고가 지능으로 간주될 수 있는 의미가 드러나고, 두 관점의 갈등은 봉합된다.[10]

---

9) 절대적으로 정확한 개념이란 없다. 개념의 정확성은 언제나 맥락 의존적이다. 대한민국의 봄철 날씨는 따뜻하다는 말은 네 계절의 날씨를 비교할 때에는 허용될 수 있지만, 오늘 입을 옷을 고르는 데에는 별 도움이 안 된다. 더 정밀한 개념 사용이 필요한 맥락에서 개념을 구분하고 명료화하는 것은 대개 좋은 전략이 된다.

10) 모든 경우에 해결이 손쉬운 것은 아니다. 예컨대, 음주가능 연령을 법제화

그러한 특수 지능이 하나의 지능으로 인정받게 된 데에는 사고의 본질을 계산과 정보처리로 보는 현대 인지과학의 기본 가정뿐 아니라 지적 능력을 계량화할 수 있는 업무수행 능력과 효율적인 문제해결 능력으로 환원하려는 현대 산업사회의 문화가 배경으로 자리하고 있다. 근래 우리 사회에서 심심치 않게 들리는 창의성에 대한 요란스런 강조에도 불구하고, 실제로 사람들의 능력은 많은 경우 특수 지능에 의해 평가된다.[11]

실체적, 관계적, 그리고 평가적 관점은 그 자체로 양립 불가능한 관점들이 아니다. 그것이 무엇인지에 관한 우리의 생각은 그것을 우리가 어떻게 대우하고 있는지에 관한 우리의 실천과 조화될 수 있어야 하고, 평가적 관점은 이를 중재하려는 시도이다. 이 글은 그런 중재의 일반적인 작동 기제에 관해 논의하려는 것이 아니다. 다만, 중재의 한 방식, 즉 개념적 명료화와 분화를 통해 실체적 관점에서의 생각과 관계적 관점에서의 실천을 조화시키는 방식을 인공지능과 인공감정이라는 구체적인 사례에 적용하고자 하는 것이다.

4.2 인공감정과 공감의 경우

인공지능에서 시선을 감정과 교감으로 돌리면, 세 관점은 서로 다른 세 물음을 묻게 된다. 첫째, 인공지능 로봇은 인간과 사회적, 정서적으로 상호작용하고 의사소통할 수 있는 능력, 즉 교감 능력을 소유하는가?[12] 둘째, 사람들은 발달한 로봇과 상호작용할 때 그것과 교감한다

하고 강력하게 단속하면 “음주가능 연령”과 관련된 발언이나 행위도 규제될 수 있다. 그러나 개념 명료화가 언제나 사람들의 인식과 행동을 바꿀 수 있는 것은 아니다. 이런 경우 수정된 제안이 거부되거나 원래 문제가 무의미한 것으로 (어떤 것이 지능을 가지는지 여부) 치부될 수도 있고, 재개념화를 시도해야 할 수도 있다. 이는 여러 사회적 조건들에 영향을 받을 것이다.

11) 심지어 강조되는 창의성마저도 영역 특수한 지능 안에서의 창의성이다. 그런 의미라면, 알파고의 새로운 수도 창의적이다. 그러나 창의성의 개념 자체는 이 글에서 다루려는 초점은 아니다.

고 느끼고, 로봇을 교감 능력을 가진 존재로 대우하는가? 셋째, 사람들이 로봇을 교감하는 존재로서 대우하는 것은 적절한가? 이 물음들은 각각 실체적, 관계적, 그리고 평가적 관점에서 제기된다. 과거에는, 감정을 소유하고 공감하는 능력이 인간과 높은 지능을 가진 몇몇 고등 동물에만 부여되었다. 그런데 공감 행동을 하는 것처럼 보이는 감정 로봇의 등장과 함께 실체적 관점과 관계적 관점의 조화는 훼손된다. 다른 지면에서 나는 이를 감정의 "탈인용부호현상"이라고 불렀다(천현득 2017). 사람들은 로봇 강아지나 페퍼 등 인간의 감정에 반응하는 대상을 의인화하고 마치 그것이 마음을 가지고 교감하는 것처럼 행동한다. 한 연구에서 연구참여자들은 인간의 명령에만 따르는 로봇과 협동과제를 수행한 경우보다 감정 표현능력을 가진 로봇과의 협동과제를 수행한 경우 로봇에 대한 호감도가 높았고 로봇이 감정을 틀림없이 가진다고 생각한 사람의 수도 많았다(Scheutz et al. 2007). 사람들은 사귐 로봇에 이름을 붙여주고 가족과 친구들에게 소개하고 애착을 느낀다. 강아지 로봇 아이보(AIBO)는 강아지의 몇 가지 행동을 흉내내는 로봇에 불과했지만, 소유자들이 그것에 쏟는 정성과 애착은 반려견에 못지않다. 보스턴 다이나믹스의 한 홍보 영상은 다양한 지형에서 사족보행이 가능한 로봇 스팟(Spot)을 소개하는데, 한 연구자가 힘껏 걷어차자 로봇은 비틀거리면서 균형을 잡는다. 이 영상을 본 많은 사람들은 실제 강아지가 불쌍하게도 걷어차인 것과 같은 반응을 보였다. 정말 인공지능은 감정을 가지는지, 그리고 로봇을 우리와 공감하는 것으로 다루는 것이 타당한지 논란거리가 된다.

우리가 지능의 개념을 명료화하고 구분했던 것처럼 교감 능력에 관해서도 유사한 작업을 수행하는 것이 도움이 된다. 교감이란 단지 조건반사적으로 행동하는 것이 아니라 더불어 살아야할 타자의 존재를 가정하고, 그의 마음을 파악하고 그와 의사소통하는 한 방식이다. 그런데 타인 혹은 타자의 마음을 파악하는 데에는 여러 차원이 존재한다. 과연 로봇이 교감할 수 있는지, 또 어떤 수준에서 타자의 마음을 이해

12) 이 글에서 "교감"은 행위자들 사이의 사회적, 정서적 상호작용의 여러 형태들을 통칭해 부르는 말로 느슨하게 사용되고 있음에 주의하라.

하고 교감할 수 있는지 묻기 위해 교감의 여러 형태들을 구분해볼 필요가 있다. 여기서는 교감의 형태들을 마음 읽기(mind reading, or Theory of Mind), 감정적 전염(emotional contagion), 동정(sympathy), 그리고 공감(empathy)으로 구분한다.

마음 읽기는 상대방의 의도나 믿음을 파악하는 능력으로, 다른 교감의 형태들과 구분된다. 감정적 전염이나 동정, 공감은 타자의 감정과 느낌을 알아차리고 반응하는 데 반해, 마음 읽기는 믿음이나 의도와 같은 인지적 상태를 알아차리는 것이다. 상대방의 믿음, 욕구, 의도 등을 파악하는 능력은 자신의 생존을 위해 필요할 뿐 아니라 어떤 상대와 협력을 할지 결정하는 데에도 중요하게 작용할 것이다. 그러나 우리가 상대의 마음을 읽고 이성적으로 의사소통할 수 있는데, 이에 더하여 “공감” 능력을 가지고 있는 이유는 아마도 공감이 우리로 하여금 공감의 대상에 관심을 가지고 염려하고 돌보도록 하기 때문이다. 단순한 인지적 마음 읽기는 그러한 관심과 돌봄을 산출하지 않기에, 장기적 관계 맺기를 위해 공감이 필요하다.

인지적 마음 읽기가 아닌 교감 능력들도 더 세분할 수 있다. 먼저, 공감과 동정의 차이점에 주목해보자. 공감은 내가 다른 사람의 입장에서 봄으로써 가지게 되는 것이다. 관점의 전환(perspective taking) 능력을 가진 경우에만 공감은 가능하다. 공감이 타인이 느끼고 있는 감정을 파악하고 그와 같은 감정을 내가 대리적으로 가지는 것이라면, 동정은 그렇지 않다. 예컨대, 배우자의 외도를 목격하고 고통스러워하는 사람을 보면서, 내가 그 사람이라면 느꼈을 감정을 나 스스로 느끼는 것이 공감이라면, 그에 대해 안타깝고 안쓰러운 느낌만을 가진다면 그것은 동정에 해당한다. 나의 느낌이 상대의 느낌과 같은지 여부는 공감과 동정을 구분하는 한 기준이 된다. 둘째, 공감은 감정적 전염과도 구별된다. 생후 몇 개월이 채 되지 않은 유아들도 다른 아이들의 고통에 반응하는 것으로 알려져 있다. 그래서 한 방에 있던 한 아이가 울면 다른 아이들도 따라서 운다. 이러한 자동적인 감정적 공명은 아마도 거울 뉴런의 활동에 기인한 것으로 보인다. 감정적 전염은 내가 타자와 구별되지 않는다는 점에서 (더 정확히 말하자면, 피아 구별을

자각하지 않은 채로 경험된다는 점에서) 공감과 구별된다. 내가 타자에게 공감할 때, 나는 타자가 나와는 다른 존재라는 것을 인식하면서도 그의 입장에 서서 그의 감정을 함께 느낀다. 하나의 정서적 상태로서 공감은 실제 상황에서 외적 자극에 의해 발생하는 직접적 정서 상태와는 구별된다. 왜냐하면 공감이란 타인의 상태로부터, 그리고 그것에 대한 상상과 시뮬레이션으로부터 촉발되기 때문이다. 그런 의미에서 공감은 대리적(vicarious) 정서 상태이다.

이와 같이 명료화된 교감의 여러 형태들에 비추어볼 때, 현재 인공지능 로봇이 가질 수 있는 교감 능력의 종류는 제한적임을 알 수 있다. 공감이든 동정이든, 아니면 감정적 전염이든 간에, 그것을 가지기 위해서는 그 대상 스스로가 감정과 정서 상태를 소유할 수 있어야 한다. 그런데 현재 그리고 근미래에 감정을 소유한 인공물이 등장할 것 같지는 않다. 감정은 한정된 자원을 가지고 복잡하고 때로는 예측 불가능한 물리적, 사회적 세계에 살아가기 위해 유연하고 적응적인 행위를 나타내야할 지적인 존재에게 요구되는 것으로서, 개체의 생존과 안녕에 유관한 정보를 표상하고 인지 과정에 영향을 미치며 행위를 지도하고 사회적 상호작용에 관여하는 것이기 때문이다. 인간과 유사한 감정을 가진 로봇이 등장하려면, 인간이나 고등 동물 이상의 일반 지능을 가지고, 생명체들이 가진 신체와 유사한 신체를 가지며, 생명체가 흔히 처하는 것처럼 복잡하고 예측 불가능한 환경에 놓여 적응할 수 있어야 한다.[13] 그렇다면 현재 상태에서 로봇이 인간에게 할 수 있는 것은 인지적인 형태의 마음 읽기뿐이다. 스스로 감정과 느낌을 가질 수 없고, 자아를 가지고 타자로부터 자신을 분리할 수 없다면, 공감도 동정도 불가능하다.

로봇에 감정을 부여하고 나에게 공감하는 대상처럼 다루는 것은 사회적 상호작용의 느낌을 주고 의사소통을 촉진할 수 있다. 하지만 로봇이나 가상 행위자에게 어떤 정서적 맥락이 존재하는 것처럼 가정하는 것은 (적어도 현재 단계에서는) 순전히 허구이다. 그 자신이 무언가

13) 인공지능이 감정이 가질 수 있는지에 관한 실체적 관점에서의 논의를 위해서는 천현득 (2017)을 참고하라.

를 느끼지 않는 존재와 우리가 감정적으로 교류한다는 것은 애당초 말이 되지 않기 때문이다. “사회적/사교 로봇”이라는 말은 마치 인공지능 로봇이 동등한 상호작용의 파트너인 듯 생각하게 만든다. 하지만 어린이가 장난감 인형을 (그것의 인공성을 인지하면서) 의인화하거나 우리가 어떤 대상에 반성적 거리를 유지한 채 의인화하는 것에서 더 나아가 그것을 실제로 감성을 가진 존재처럼 다루고, 치료나 교육 목적으로 광범위하고 실질적인 용도로 사용하게 되면 잘못된 감정 귀인의 문제는 더 날카롭게 제기될 수 있다.

문제는 이런 식의 개념적 명료화가 얼마만큼의 힘을 가질 것인가 하는 것이다. 사람들은 로봇에게 감정이 있다고 믿지 않으면서도 그것을 감정적인 존재인 것처럼 대우하곤 한다. 이 지점에서 지능에 적용되었던 전략은 쉽게 적용되지 않을 수 있다. 진짜 공감이 아닌데도 그렇게 보이는 행동에 관해 추가적인 고찰이 필요하다.

5. 인공 반려의 유혹

우리가 인간이나 동물을 닮은 로봇에게 쉽게 공감을 표시하는 것은 그러한 로봇과의 상호작용이 오랜 진화사를 통해 형성된 인류의 다윈적 단추(Darwinian Button)를 누르기 때문이다(Turkle 2011). 로봇의 내부 작동과정을 들여다볼 수 있도록 만들어서 사람들에게 그 안에 특별한 것이 있지 않음을 보여주어도 사람들의 반응은 달라지지 않는다. 이중체계 이론(dual-system theory)의 구분을 들여오면, 우리의 정서적, 사회적 관계는 의식적이고 언어적이고 숙고하는 시스템2에 의존하지 않고, 몇 가지 단서들에 의해 신속하고 자동적으로 처리되는 시스템1에 달려있다.[14] 시스템1의 작동은 시스템2의 개입에 의해 쉽사리 교정되지 않는다. 우리의 선조들은 진정한 관계와 흉내낸 관계를 구별할 필요성이 없었고, 인류의 마음은 그런 기능을 내재하고 있지 않다.

14) 이중체계 이론 혹은 이중과정 이론을 위해서는 Evans (2003), Kahneman and Frederick (2002), Stanovich and West (2000) 등을 참조할 수 있다.

공감 능력은 어떤 대상이 반려이기 위한 필요조건이다. 우리에게 공감할 수 없는 인공지능 로봇은 우리의 진정한 반려가 될 수 없다. 그러나 사람들이 로봇이 우리에게 공감하는 것처럼 느낀다면, 그것을 마치 반려처럼 대우하게 되는 것도 이상한 일이 아니다. 나는 그 대상을 반려대체물(companion-substitutes)로 부르고자 한다. 반려대체물이란 진정으로 우리에게 공감하는 존재는 아닐지라도 우리가 마치 그런 존재인 것처럼, 즉 반려로서 대우하려는 경향을 가지는 대상을 말한다. 반려대체물과의 관계에는 반려 관계에 내재된 상호성이 존재하지 않지만, 사람들은 그 속에서 상호성의 그림자를 발견한다. 그래서 현대인들은 "인공 반려"에 매혹된다. 로봇 강아지는 진짜 강아지 못지않다. 로봇 친구는 학교 친구나 동료, 가족보다 더 나은 반려일지도 모른다. 왜냐하면 그것은 위험하지도 않고 우리를 배신하지도 않기 때문이다. 그것은 우리의 감정과 느낌, 우리의 생각과 의지에 충실히 반응하고 우리의 외로움을 덜어줄 뿐 아니라 우리를 성가시게 하지도 않고 스트레스를 주지도 않는다.

인간과의 의사소통능력을 가진 인공지능 로봇을 반려가 아닌 반려대체물로 간주해야한다는 주장에 대해, 혹자는 반려견이나 반려묘와 같은 동물은 진정한 반려인지 아니면 반려대체물인지 의문을 제기할 수 있다. 인간이 아닌 다른 동물들이 반려일 수 있다면 왜 로봇은 반려일 수 없는지, 반대로 로봇이 진정한 반려일 수 없다면 어떻게 해서 동물들은 반려일 수 있는지 묻는 것이다.[15] 앞서 잠시 언급된 것처럼 현대인들은 함께 생활하는 일부 고등 동물을 반려 동물로 부르는 데 거리낌이 없다. 그렇게 부른다고 해서 반려 동물이 인간 반려자(human companion)와 동일한 의미에서 진정한 반려임이 자동적으로 따라 나오는 것은 아닐 것이다. 짧은 지면에서 인간과 반려 동물간의 상호작용에 관해 상세히 논의할 수는 없지만, 그 상호작용의 성격에 관해 간략히 살펴봄으로써 "반려 동물"과 "반려 로봇"의 차이점을 언급할 수 있겠다.

---

15) 이러한 물음을 제기함으로써 논의를 선명하게 만들 기회를 주신 익명의 심사위원께 감사드린다.

먼저, 사람들이 반려 동물로 간주하는 강아지나 고양이 등으로 논의를 국한하자. 이들 반려 동물들은 주인의 믿음이나 의도를 파악하고 반응한다는 점에서 일정한 정도의 마음 읽기 능력을 가질 뿐 아니라 주인의 감정 상태와 분위기에도 나름의 방식으로 반응한다. 게다가 반려 동물은 그 스스로 감정과 정서적 상태를 가진다.[16] 이 점에서 반려 동물은 인지적 상태의 마음 읽기만 가능한 로봇과 차별화된다. 특히, 반려 동물과 주인 사이에서 감정적 전염이 일어난다는 것을 입증하는 과학적 증거가 축적되고 있다(Milani 2017; Panksepp and Panksepp 2013; Yong and Ruffman 2014). 반려 동물들은 인간이 경험하는 섬세한 사회적 감정들을 가지지 않더라도 최소한 기본 감정들을 경험하고, 주인이 부정적 감정을 경험할 때 이와 유사한 감정을 경험하게 된다. 물론 공감 능력을 위해서는 관점 전환과 같은 고도의 인지 능력이 필요하기 때문에 그러한 수준의 공감 능력을 가진다고 보기는 어렵다. 어쩌면 높은 수준의 공감은 인간들 사이에서만 가능할지도 모른다. 그러나 주인이 경험하는 긍정적이거나 부정적인 감정을 파악하고 그에 적절한 그 자신의 감정을 경험한다면, 이는 (인간의 동정과 동일한 것은 아니지만) 일종의 동정으로 볼 수 있다. 요컨대, 반려 동물은 인지적 마음 읽기를 넘어 감정적 전염을 경험하거나 동정의 능력을 가진다고 볼 수 있다. 그렇다면 반려 동물이 인간 수준의 반려라고 볼 수는 없을지라도, 단순한 반려대체물로 간주될 수도 없을 것이다. 따라서 반려 동물과 인공 반려를 동일선상에 놓는 전략은 인공 반려가 실상 반려대체물로 간주되어야 한다는 본 논문의 주장을 훼손하지 못한다.

그런데 피곤한 인간관계에서 우리가 지불해야할 대가 없이, 반려 관계에서 얻을 수 있는 유익을 얻을 수 있다면, 그것이 진정한 반려가 아니라 반려대체물이라고 해서 무슨 상관이겠는가? 우리가 다른 사람들보다 로봇에 더 공감한다고 느끼는 것은 왜 문제가 되는가?

진정한 반려와 반려대체물의 차이에 주목해보자. 반려대체물은 스스로 감정을 가지지 않고 우리에게 공감할 능력이 없지만, 우리가 그것

16) 인간이 동물들과 기본 감정(basic emotion)을 공유한다는 이론에 관해서는 Ekman (1999)와 Panksepp (1998)을 참조할 수 있다.

을 공감하는 존재로 다루는 대상이다. 그것은 까다로운 애인이나 잔소리하는 가족보다 오히려 더 만족스러운 대상일지도 모르지만, 독립성을 결여한 채 나르시시즘을 촉진하는 존재이다. 반려대체물은 우리와의 관계에서 어떤 마찰을 일으키지 않는다. 그러나 진정한 반려는 다른 선택지가 없이 필연적으로 우리에게 공감하는 존재가 아니다. 반려는 자신의 욕구와 의지, 의도에도 불구하고, 즉 그렇게 하지 않을 수 있었음에도 불구하고 나에게 공감하기 때문에 그 일을 선택한다. 오래된 친구는 자기 나름의 사정이 있지만 나에게 공감하기 때문에 자신의 일부를 기꺼이 내어준다. 이러한 차이는 반려대체물과의 관계 맺음을 비판하는 근거는 아니지만, 그 관계가 진정한 반려 관계보다 더 낫다는 생각을 의문시하기에 충분하다.

반려대체물과의 관계 맺기가 예외적인 상황이 아니라 통상적인 것으로 자리잡게 되는 경우, 우리는 진정성의 위기를 경험하게 될 것이다. 어쩌면, 오히려 진정성의 위기로 인해 반려대체물을 추구하는 것일지도 모른다. 사람들 사이에서도 진정한 공감을 얻기란 쉽지 않은 일이다. 공감을 바라지만 그것의 실현은 매우 어렵고 커다란 대가를 지불해야한다. 반면, 훨씬 쉬운 종류의 "공감"이 가능하면 그것에 만족하려 들 것이다. 사람들은 그것을 "충분히 좋은" 공감이라고 자위할 것이다. 마찰이나 불협화음 없는, 헌신이나 얽매임 없는, 그래서 정서적 비용을 지불하지 않고 얻는 공감이 표준적인 것으로 자리잡는다면, 일상적인 대면 접촉에서도 우리는 서로에게 감정 노동만을 요구하게 될지 모른다. 그리고 그러한 감정 노동은 로봇이 더 잘할 수 있다.

또 한 가지 두드러진 문제는 기만의 문제이다. 스스로 진짜 감정을 가지고 공감하는 것이 아니고 우리의 마음을 이해해주는 것도 아닌데, 마치 그런 것처럼 다루어지는 것은 그 자체로 도덕적으로 문제라는 주장이 있을 수 있다. 이 주장의 한 가지 형태에 따르면, 마치 사람이나 동물과 교류하듯이 로봇과 교류한다고 생각하는 것은 기만당하는 것이며, 기만 일반과 마찬가지로 윤리적으로 허용될 수 없다.[17] 그러나 기

17) 기만이 윤리적 문제를 야기한다면, 기만당하는 쪽보다는 기만하는 쪽이 문제의 근원이다.

만이라는 사실로부터 도덕적으로 나쁘다는 결론을 내리는 과정은 다소 성급해 보인다. 사람들 사이에서도 속임수와 기만은 널리 퍼져있고, 우리는 모든 거짓말을 윤리적으로 나쁘다고 단정하지 않는다. 상대방을 격려하기 위해서 그의 능력을 과대평가한다거나 집단이 처한 위험 상황에서 큰 희생을 피하기 위해 거짓말을 한다고 해서, 그것을 도덕적으로 나쁘다고 볼 수는 없다. 사람들 사이의 대인관계에서도 어느 정도의 기만이 포함되어 있다면, 로봇의 경우 왜 특별히 문제가 되는지 생각해보아야 한다. 철학자 쿠켈버그는 로봇에게 인간보다 더 엄격한 기준을 들이대는 것은 공정치 않다고 지적하면서, 로봇과의 상호작용을 통해 사용자의 주관적 안녕감이 높아지고, 그러한 관계가 대인 접촉을 대체하는 것이 아닌 한, 도덕적으로 더 문제가 있지 않다고 지적한다(Coeckelbergh 2011). 다만, 이런 대응은 기만의 비용과 이득을 계산할 수 있음을 전제로 한다. 기만의 이득이 비용보다 큰 경우 도덕적으로 문제 삼을 수 없고, 반대의 경우에만 문제가 되기 때문이다. 따라서 쿠켈버그의 대응은 기만의 비용-편익 분석이 가능한 경우에만 타당하다. 문제는 우리가 항상 그런 계산을 할 수는 없다는 데 있다. 우리는 로봇과의 관계가 가져올 결과에 관해 상당 부분 무지하다. 그렇다면 현명한 방책은 기만을 원칙적으로 나쁜 것으로 보되, 기만의 악영향을 상쇄할 만한 덕이 있는 경우에 예외적으로 허용하는 것이다.

관련된 쟁점은 공감이 가진 “편들기” 역할에 관한 것이다. 사람들은 서로에게 공감을 얻고자 하는데, 많은 경우 타인으로부터 받은 공감을 자신에 대한 지지로 간주하기 때문이다. 특히, 논쟁 중인 상황에서 당사자가 아닌 제3자로부터 받는 공감은 자신의 편을 들어주는 것으로 여겨진다. 때로는 한편을 들어줌으로써 물리적 충돌을 야기하지 않고 논쟁을 종식시킬 수도 있다. 그러나 어느 한편에 공감하고 편을 들면, 우리 편이 옳고 상대편이 그르다고 가정하게 되면서, 분열을 가속화할 수도 있다. 인터넷 공간에서 흔히 발견되는 집단 극화는 이런 현상의 한 형태이다. 이제 로봇이 제3자인 당신을 논쟁 가운데 끌어들여서 당신의 공감을 얻으려 한다고 가정해보자. 인간 행동과 심리학에 대한 정보를 학습한 인공지능 로봇이라면, 자신의 목적을 달성하는 데 효과

적인 전략적 행동이 무엇인지 알 것이다. 로봇은 공감을 표시하는 당신을 끌어들여 자신의 관점에 동의하도록, 그래서 논쟁의 상대편에 반대하도록 할 수 있다. 그 로봇이 고객의 항의에 대응하는 업무를 하거나, 상품을 홍보하거나, 대중의 의견을 조작하는 일에 동원되는 일은 상상하기 어렵지 않다. 의지를 가진 로봇이 인류를 지배하는 세상은 아마 오지 않을 것 같다. 그러나 인류가 공감을 조작하는 능력을 가진 인공지능과 그것을 활용하는 일부 사람들에 의해 통제사회에 참여하게 될 가능성은 그보다 높을 수 있다. 우리가 로봇에 공감하면서, 로봇의 편을 들어 다른 인간 동료들에게 반대하는 일이 가능하다.

감정을 소통하고 반려로서 기능하는 로봇을 설계하고 제작하는 과정에서 생겨날 수 있는 문제도 있다. 인공 반려를 제작하기 위해서는 자연스러운 의사소통이 물리적으로 구현해야한다. 이를 위해 기계 인공물을 인간화하는 것이 유리하다. 체현된 인공지능이란 단지 인공지능이 신체(혹은 가상적 이미지)를 가지고 있음을 뜻할 뿐 아니라, 의사소통도 물리적으로 실현됨을 뜻한다. 즉, 물리적 상호작용이 가능해야하고, 특히 감정이 제스쳐나 표정, 몸짓 등 물리적 형태로 표현할 수 있어야 한다. 정서적 소통을 물리적으로 구현하는 인공물을 제작하려면 정서와 행동 사이의 관계를 계산적으로 모델링해야 한다. 이때, 계산적 모델링의 특징에 주목할 필요가 있다. 모델링을 위해서는 개별적이거나 다의적인 측면을 축소하고, 보편적으로 적용 가능한 변수들만을 취급해야하기 때문이다. 이에 따라 탈맥락화, 추상화, 일반화가 불가피하다. 따라서 인간 감정과 공감이 가지는 복잡한 특성들, 개별성, 맥락 의존성이 간과될 가능성이 있다.[18]

어떤 현상을 이해하기 위해 단순화하거나 이상화하는 것이 반드시 악덕은 아니다. 많은 과학적 활동은 추상적 모형이나 일반적 법칙을

18) 어쩌면 이는 빅데이터에 의존한 접근의 한계일 수도 있다. 이를 극복하려면 1)적은 데이터를 이용한 새로운 방법론을 개발하거나, 그렇지 않으면 2) 더 많은 데이터를 더 확보하기 위해 유사한 사례들을 수집하고 집단적으로 비교 검토해야한다. 한편, 감정 구성의 복잡성에 관해서는 Barrett (2017)을 보라.

통해 세계에 관한 이해를 제공한다. 그러나 아직 성인에 이르지 않은 유아나 청소년들이 인공 행위자와 상호작용하면서 자라난다면, 표준화되고 전형적인 정서 반응을 보이는 인공물의 행동 패턴이 "정상적"인 것으로 간주되고, 다채롭게 때로 엉뚱하고 예측 불가능한 인간의 감정과 정서 반응은 "비정상" 혹은 불편한 것으로 치부될 수 있다. 모델링을 위해 필요한 정서 반응의 표준화로 인해, 개인의 특이성이나 개성이 번거롭고 까다롭고 피해야할 것으로 간주될 위험이 있다.

"인공 반려"는 분명히 매혹적이다. 그러나 현 단계에서 그리고 가까운 미래에, 스스로 감정을 가지고 우리에게 공감하는 인공물은 등장하지 않을 것 같다. 우리는 인공지능 로봇을 반려가 아니라 반려대체물로서 대우해야 한다. 최근 반려견의 지위와 규제를 놓고 우리 사회가 시끄럽다. 여기에 인공 반려까지 포함해서 더 시끄럽게 토론될 필요가 있다.

참고문헌

천현득 (2017), 「인공 지능에서 인공 감정으로: 감정을 가진 기계는 실현가능한가?」, 『철학』 131권, pp. 217-243.

Anderson, M. (2003), "Embodied Cognition: A Field Guide", *Artificial Intelligence* 149(1): pp. 91 - 130.

Barrett, L. F. (2017), *How Emotions Are Made: The Secret Life of the Brain*, Houghton Mifflin Harcourt.

Breazeal, Cynthia, and Rodney Brooks (2005), "Robot emotion: A functional perspective", in *Who Needs Emotions*, edited by Jean-Marc Fellous and Michael Arbib, New York: Oxford University Press, pp. 271-310.

Brooks, R. (1991), "Intelligence without representation," *Artificial Intelligence* 47: pp. 139-59

Coeckelbergh, M. (2011), "Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations", *Studies in Ethics, Law, and Technology* 4(3): pp. 1-19.

Ekman, P. (1999), "Basic Emotions", in Tim Dalgleish and Mick J. Power (eds.), *Handbook of Cognition and Emotion*, Wiley and Sons, pp. 45-60.

Evans, J. S. (2003), "In two minds: dual-process accounts of reasoning", *Trends in Cognitive Science* 7(10): 454-59.

Kahneman, D. and S. Frederick (2002), "Representiveness revisited: Attribute substitution in intuitive judgment", in T. Gilovich & D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment*, New York: Cambridge University Press, pp. 49-81.

Kaliouby, Rana el (2017), 'We Need Computers with Empathy' https://www.technologyreview.com/s/609071/we-need-computers-with-empathy/ (검색일: 2019.07.24.)

Levy, D. (2007), *Love and Sex with Robots: The Evolution of Human-Robot Relationships*. New York: Harper.

Milani, M. (2017), “Human-animal emotional contagion and client communication”, *Canada Veterinary Journal* 58(12): pp. 1329-330.

Panksepp, J. (1998), *Affective Neuroscience: the Foundations of Human and Animal Emotions*, New York, Oxford: Oxford University Press.

Panksepp, Jaak and Jules B. Panksepp (2013), “Toward a cross-species understanding of empathy”, *Trends in Neuroscience* 36(8): pp. 489-96.

Robins, B., Dautenhahn, K., and Dubowski, J. (2006), “Does appearance matter in the interaction of children with autism with a humanoid robot?” *Interaction Studies* 7(3): pp. 509-54

Scheutz, Matthias, Paul Schermerhorn, James Kramer, and David Anderson (2007), “First steps toward natural human-like HRI”, *Autonomous Robots* 22(4): pp. 411-23.

Stanovich, K. E., & West, R. F. (2000), “Individual differences in reasoning: implications for the rationality debate?” *Behavioral and Brain Sciences* 23(5): pp. 645-65.

Turkle, S. (2011), *Alone Together: Why We Expect More from Technology and Less from Each Other*, New York: Basic Books.

Varela, F., Thompson, E. and E. Rosch (1991), *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

Wada, K., Shibata, T., Sakamoto, K., and Tanie K. (2006), “Long-term Interaction between Seal Robots and Elderly People ― Robot Assisted Activity at a Health Service Facility for the Aged.” *Proceedings of the 3rd International*

Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE 2005), Berlin: Springer.

Yong, M. and T. Ruffman (2014), “Emotional contagion: Dogs and humans show a similar physiological response to human infant crying”, *Behavioral Processes*, 108: pp. 155-65.

논문 투고일	2019. 07. 08.
심사 완료일	2019. 07. 23.
게재 확정일	2019. 07. 23.

The Allure of Artificial Companionship

Hyundeuk Cheon

In this article, the concept and potential risk of artificial companionship are discussed. This discussion is timely because artificial intelligence robots are expected to enter into people's everyday spaces as partners of social and emotional interactions. After briefly surveying the background of the development of social robots, we examine the conditions on which an object can make a companion relationship with us. In order to establish such a relationship, the condition of empathy turns out to be the most important. In this article, I distinguish empathy and sympathy from mind-reading in the cognitive sense and reveal that social robots in the near future are hard to achieve a genuine sense of empathy. Finally, we identify the nature of the companion-substitute and address some ethical and social issues concerning its replacement with human companions.

Keywords: Social Robots, Artificial Intelligence, Artificial Companion, Empathy, Authenticity