

## 불쾌한 골짜기: 로봇 속성의 의인화, 마음지각 및 도덕적 처우의 관계

Uncanny Valley: Relationships Between Anthropomorphic Attribution to Robots,  
Mind Perception, and Moral Care

신홍임<sup>†</sup>

Hong Im Shin<sup>†</sup>

### Abstract

The attribution of human traits, emotions, and intentions to nonhuman entities such as robots is known as anthropomorphism. Two studies were conducted to check whether human-robot interaction is affected by anthropomorphic framing of robots. In Study 1, participants were presented with pictures of robots that varied in human similarity in appearance. According to the results, uncanny feelings toward a robot increased with the higher levels of human similarity. Furthermore, as the level of mind attribution increased, participants tended to attribute more humanlike abilities to nonhuman agents. In Study 2, a robot was described as either a machine-like robot or a humanlike robot in a priming story; then, it was examined whether significant differences exist in mind attribution and moral care. The participants tended to perceive robots as more humanlike in the mind attribution when anthropomorphism was used in a robot's behavior, according to the findings. Furthermore, in the condition of increased anthropomorphism, a higher level of moral care could be observed compared with that in the other condition. This means that humanlike appearances may increase uncanny feelings, whereas anthropomorphic attribution may facilitate social interactions between humans and robots. Limitations as well as the implications for future research are discussed.

**Key words:** Uncanny Valley, Robot, Anthropomorphism, Moral Care

### 요약

의인화는 로봇과 같이, 인간이 아닌 대상에게 인간의 속성, 정서나 의도를 부여하는 것이다. 본 연구에서는 로봇에게 인간 고유의 속성을 부여하는 의인화가 로봇-인간의 상호작용에 끼치는 영향을 분석하였다. 연구 1에서는 다양한 로봇의 사진을 제시하고, 로봇의 외관에 따른 심리적 불쾌감, 마음지각 및 도덕적 처우를 자기보고식 질문지를 통해 분석하였다. 그 결과 로봇의 외관이 인간과 가장 유사한 안드로이드 로봇조건에서 휴머노이드 로봇의 조건과 기계적 외관의 조건보다 로봇에 대한 심리적 불쾌감이 가장 높았다. 또한 인간과 유사한 안드로이드 로봇에서 기계와 비슷한 로봇보다 로봇에 대한 마음지각이 더 높게 나타났다. 연구 2에서는 로봇의 속성을 의인화한 조건과 의인화하지 않은 조건에서 로봇에 대한 불쾌감, 마음지각과 도덕적 처우의 정도가 다르게 나타나는지를 비교하였다. 그 결과 로봇 속성의 의인화조건에서 로봇에 대한 마음지각과 도덕적 처우의 정도가 더 높게 나타났으며, 마음지각의 경험성이 높을수록 도덕적 처우의 정도가 더 높아졌다. 이 결과는 인간과 유사한 로봇의 외관은 로봇에 대한 심리적 불쾌감을 증가시키지만, 로봇의 속성에 대한 의인화는 로봇에 대한 마음지각을 증가시키고, 인간-로봇의 상호작용을 촉진시킬 가능성을 제시한다. 논의에서는 의인화가 인간-로봇의 상호작용에 끼치는 차별화된 영향에 대한 시사점을 토론했고, 연구의 한계 및 후속연구의 방향을 다루었다.

**주제어:** 불쾌한 골짜기, 로봇, 의인화, 도덕적 처우

※ 본 논문은 순천대학교 교연비 연구사업에 의해 지원되었습니다.

† (교신저자) 신홍임: 순천대학교 교양교육원 자유전공학부 부교수 / E-mail: shin7038@senu.ac.kr / TEL: 061-750-3809

## 1. 서론

이시구로(Ishiguro)의 소설(2021) ‘클라라와 태양(Klara and the sun)’에는 로봇 클라라를 의인화시켜 로봇과 인간의 경계에 대한 질문을 던진다. 다양한 로봇이 전시된 상점에서 클라라는 길거리의 보행자를 관찰하면서 행복과 슬픔을 동시에 느낀다. 또한 계속해서 업그레이드되는 새로운 로봇을 보면서, 자신의 기능에 대한 회의와 좌절감을 느끼고, 자신의 주인과 함께 시내에 갔을 때는 수많은 사람들을 보면서 공황장애와 같은 불안감에 휩싸인다.

인간이 아닌 무생물이나 동물에게 인간의 마음과 유사한 속성을 부여하는 의인화는 로봇설계의 연구에서 최근 주목을 받고 있다(Damiano & Dumouchel, 2018; Duffy, 2003; Müller et al., 2020; Nijssen et al., 2019). 수십 년 전까지만 해도 마음은 피와 뼈, 살을 가진 인간과 동물에게만 한정된 것으로 간주되어 왔다. 그러나 이제 우리는 로봇의 사용자에게 금속과 실리콘으로 만들어진 로봇에 대해 인간과 유사한 마음을 지각하도록 유도해야 하는 문제에 봉착했다. 인간의 모습과 비슷하거나 인간처럼 생각하며, 특정 기능을 수행하도록 의인화된 로봇은 사용자에게 사회적 존재로서 함께 있어주거나, 말하기와 같은 특정 행동을 함으로써 인간과 편안한 관계를 형성하고, 지속적으로 유지하는데 투입된다(Damiano & Dumouchel, 2018; Jang, 2006).

지금까지 선행연구에서는 인간이 아닌 대상에게 인간의 속성을 부여하는 의인화가 사회적 관계를 형성하는데 긍정적 영향을 끼침을 보고하고 있다(Nijssen et al., 2019). 반면 의인화가 유도되는 방식(인간과 비슷한 외관 vs. 인간과 유사한 마음)에 따라 이 대상에 대해 경험되는 정서와 도덕적 처우를 차별화하여 수행한 연구는 거의 수행되지 않았다. 이에 따라 본 연구에서는 인간-로봇의 상호작용에서 로봇에 대한 의인화가 두 가지 유형의 서로 다른 정보처리방식(상향처리 vs. 하향처리)에 의해 유도될 때, 인간-로봇의 상호작용에 차별화된 영향을 끼치는지를 살펴보고자 한다. 상향처리(bottom up)의 의인화는 로봇의 외관적 단서(예: 인간

의 얼굴과 비슷한 로봇의 얼굴, 피부나 머리털)에 의해 유도될 수 있다. 이에 비해 로봇의 마음을 인간의 마음과 유사한 방식으로 해석하고, 추론하도록 하면(예: 로봇이 자율적으로 사고하고, 스스로의 감정을 갖고 있다고 설명함), 로봇에 대한 의인화가 하향처리(top down)로 유도될 수 있다(Nijssen et al., 2019). 본 연구에서는 로봇의 외관에 의한 상향처리방식 또는 로봇의 마음을 추론하는 하향처리방식으로 의인화를 유도할 때, 로봇에 대한 심리적 불쾌감, 마음지각 및 도덕적 처우가 달라지는 경향을 탐색하려 한다.

## 2. 이론적 배경

### 2.1. 로봇 외관의 불쾌한 골짜기

일본의 로봇공학자 모리(Mori)는 인간의 모습과 유사한 안드로이드 로봇의 개발에 반대하면서 ‘불쾌한 골짜기’(uncanny valley) (Mori, 1970) 가설을 발표했다. 이에 따르면, 로봇의 외관이 인간과 유사해질수록 점점 호감도가 증가하지만, 유사성이 어느 지점에 도달하면, 오히려 호감도가 낭떠러지처럼 급격하게 떨어진다. 이것은 마치 우리가 등산을 할 때, 언덕을 오르고, 골짜기를 내려가기도 하는 것처럼, 로봇에 대한 호감도는 로봇 외관의 유사성과 함께 증가하다가 다시 떨어지는 비선형적 관계에 있을 것을 의미한다. 이 가설에 의하면 로봇 장난감에 대한 호감도는 높지만, 준비와 같이 움직이는 시체, 인간의 손과 유사한 의수 또는 인간과 유사한 피부와 머리털을 갖고 있는 안드로이드 로봇1)에 대한 호감도는 오히려 낮다. 불쾌한 골짜기가설은 거의 40여 년 동안 관심을 받지 못하다가, 최근 들어서는 로봇공학뿐만 아니라 사회과학분야(Gray & Wegner, 2012b; Mathur & Reichling, 2016; Müller et al., 2020; MacDorman & Ishiguro, 2006; Wang et al., 2015)에서도 학문적으로 주목을 받고 있다.

MacDorman & Ishiguro(2006)는 기계와 더 유사한 휴머노이드 로봇부터 인간과 더 유사한 안드로이드 로

1) 안드로이드 로봇은 인간-로봇과의 상호작용을 최적화하기 위해 인간의 외관과 행동을 가능한 최대한 모방하여 개발된 로봇으로 정의된다(MacDorman & Ishiguro, 2006).

봇까지 유사성을 점차로 증가시켜 호감도와 의 관계를 측정하고, 불쾌한 골짜기의 관계가 나타남을 보여주었다. 또한 Seyama & Nagayama(2007)에서는 연구참가자들에게 인형의 얼굴부터 실제 사람의 얼굴까지 합성한 사진에 대해 긍정적/부정적 인상<sup>2)</sup>이 형성되는 정도를 평정하도록 하였을 때, 실제 사람의 얼굴과 유사한 지점부터 불쾌한 골짜기가 나타났다. 인간의 모습과 더 많이 유사할수록 호감도가 오히려 감소될 수 있다는 것은 인간의 모습과 유사한 안드로이드 로봇을 제작하는 것에 대한 문제제기가 필요함을 보여준다.

그런데 인간과 비슷한 얼굴을 갖고 있는 로봇이 왜 무섭고, 기괴하게 느껴질까? 이에 대해 지각과정의 상향처리를 설명하는 관점에서는 완벽하게 인간이 아닌 로봇의 외관이 병원균 감염이나 죽음을 연상시킨다고 설명한다. 예를 들어, 진화심리학의 병원균 회피 가설에서는 인간과 비슷하지만 눈/코/입의 비율이 매우 불균형하거나 인간과 다른 피부는 기계보다 더 위험한 자극으로 인식되어 메스꺼움이나 두려움과 같은 불쾌한 감정을 유발하고, 인간이 개체보존을 위해 피하도록 진화되어 왔다고 설명한다(Wang et al., 2015). 이에 비해 기대위반가설에서는 인간의 얼굴과 유사한 로봇의 얼굴을 보았을 때, 처음에는 인간과 유사하여 따뜻하고, 축축할 것을 예측하였는데, 실제로 만져보니 차갑고, 딱딱했다면 기대위반으로 인해 불편한 감정이 경험되는 것으로 설명한다.

주목할 것은 인간과 지나치게 유사한 로봇의 외관은 인간/로봇의 구분을 모호하게 한다는 것이다. 이것은 인간을 고유한 존재로서 인간/로봇의 경계를 구분하려는 심리적 욕구를 방해하며, 빠르게 진행되는 상향처리를 통해 심리적 불편감을 발생시킨다(Müller et al., 2020). 그렇다면 심리적 불편감은 인간/로봇을 구분하려는 의도와 해석이 담긴 하향처리에 의해서도 발생할 것인가?

## 2.2. 마음지각의 불쾌한 골짜기

마음지각가설(Gray & Wegner, 2012b)에 의하면 인간/로봇에 대한 구분은 외관뿐만이 아니라 관찰자로서

의 인간이 로봇의 마음을 규정하는 것에 기반한다. 이 가설에서는 한 개인을 인간으로 보게 하는 가장 기본적인 마음의 요소가 무엇인지에 대한 질문에서 출발한다. 마음지각이론(Gray et al., 2007)에 의하면, 마음은 주도성(agency)과 경험성(experience)의 두 가지 독립된 차원으로 구분된다. 주도성에는 기억하고 생각하는 능력, 계획하는 능력 및 자기 의지대로 행동하는 능력이 포함된다. 경험성의 차원에는 고통, 즐거움 또는 분노를 느끼는 능력 및 자신의 욕구를 가질 능력이 포함된다. 마음지각의 선행연구(Gray et al., 2007; Gray & Wegner, 2012a)에서는 특정 대상에 대한 마음지각의 주도성이 높아질수록 행위에 대한 책임을 묻는 경향이 커지는데 비해, 경험성이 높아질수록 이 대상을 해치는 행동에 대한 심리적 불편감이 커지는 경향이 나타났다. 또한 상대방이 고통을 느낀다고 지각할수록 이 상대방을 해치는 행동을 심리적으로 불편한 것으로 평정하여, 마음지각의 경험성은 대상에 대한 도덕적 처우와 직접적 관계가 나타났다. 따라서 한 개인이 대상을 보는 관점에 따라 다양한 마음지각이 형성되어, 이 대상의 행동에 대한 도덕적 책임을 묻고, 이 대상을 도덕적으로 대하는 정도가 달라진다고 볼 수 있다(Gray & Schein, 2012). 예를 들어, 돌과 인간을 연속선상의 양 극단에 두었을 때, 해치는 행동에 대한 심리적 불편감은 가장 단순한 마음을 가진 돌에서 가장 낮고, 복잡한(sophisticated) 마음을 가진 인간에 대해 가장 높을 것을 예측할 수 있다.

Gray & Wegner(2012b)의 연구에서는 연구참가자들에게 아동의 얼굴과 유사한, 살아있는 듯한 로봇 ‘카스파(Kaspar)’를 영상으로 보여주었다. 기계 로봇의 조건에서는 전선줄이 보이는 로봇의 뒷면에서 촬영한 영상을 보여주고, 살아있는 로봇의 조건에서는 아동과 유사한 로봇의 얼굴을 앞면에서 촬영한 영상을 보여주었다. 이후 참가자들에게 로봇에 대한 심리적 불안감(uneasy), 무서움(unnerving) 및 섬뜩함(creeped out)을 5점 척도에서 평정하도록 한 결과, 실제 인간과 유사한 로봇의 조건에서 기괴한 불편감(eeriness)의 부정적 정서<sup>3)</sup>가 더 강하게 나타났다.

2) 이 연구에서 긍정적 인상은 ‘매력적인, 자연스러운, 건강한’ 등의 형용사와 연관된 인상으로 정의하고, ‘0’이상에서 ‘+2’까지에 응답하도록 하였으며, 부정적 인상은 ‘매력없는, 못생긴, 무서운’ 등의 형용사를 연상시키는 인상으로 정의하고, ‘0’ 이하에서 ‘-2’까지로 응답하도록 하였다.

그런데 Gray & Wegner(2012b)의 연구에서 주목할 것은 로봇이 인간과 유사한 외관을 가질수록 심리적 불쾌감이 증가할 뿐만 아니라, 로봇의 마음이 인간과 유사하다고 인식될수록 로봇에 대한 불안감과 불쾌감이 더 높아진다는 것이다. 이것은 로봇을 인간보다 열등하다고 생각했는데, 로봇에게 인간의 마음과 유사한 마음이 있는 것처럼 로봇을 제시하면 로봇으로부터 심리적 불쾌감을 느낀다는 것이다. 로봇인데 감정을 느낀다면, (로봇은 감정을 못 느낀다고 생각했기 때문에) 이상하게 느껴지고, 사람인데 감정을 못 느낀다면, (사람이라면 감정을 느낄 것으로 생각했기 때문에), 불쾌한 감정을 느낄 가능성을 추정해볼 수 있다.

마음지각이론은 로봇과 인간의 유사성이 높아질수록 심리적 불쾌감이 느껴지는 불쾌한 골짜기에 대한 설명을 확장시켰는데 중요한 의의가 있다. 이전에는 주로 로봇의 외관에만 치중되었던 데 비해 로봇과 인간의 마음을 구분하는 근본적인 기준에 대한 문제를 제기하고, 마음의 경험성이 주도성보다 더 결정적임을 보여주었다. 그러나 마음지각이론에서 주장하는 것처럼 로봇에게 인간의 마음과 같은 속성을 부여하는 것이 심리적 불쾌감을 경험하게 하는 이유인가?(Wang et al., 2015). 예를 들어, 특정 상황에서 한 개인은 시를 쓰는 로봇에 대해 마음지각에서 높은 점수를 줄 수도 있다. 그렇다면, 로봇에게 인간의 속성을 부여하는 사람일수록 로봇에게 언제나 불쾌한 감정을 느끼게 될 것인가? 또한 인간과 유사하게 보이는 로봇에 대한 마음지각이 높게 형성되지만, 이와 동시에 심리적 불쾌감을 경험한다면, 이 로봇에 대한 도덕적 처우는 어떻게 될 것인가? 로봇이 복잡한 마음을 가지고 있다고 지각될수록 로봇을 해치는 행동에 대한 심리적 불편감이 증가할 것인가? 또는 인간과 유사한 로봇에 대해 심리적 불쾌감이 커지기 때문에 도덕적 처우도 낮게 될 것인가? 따라서 마음지각이론이 인간-로봇의 상호작용을 모두 설명하기에는 한계가 있는 것으로 보인다.

### 2.3. 마음지각과 로봇 속성의 의인화

최근 보급되고 있는 사회적 로봇이 의인화를 유도하는 방식은 인간과 유사한 외모 또는 로봇의 자율적 행동을 활용한다. 선행연구에서는 이 두 방식이 로봇을 인간과 유사하게 지각하게 하고, 인간-로봇의 경계를 무너뜨리며, 로봇과 사회적 상호작용을 하게 됨을 보고한다(Damiano & Dumouchi, 2018). 그런데 인간과 유사한 외관만을 갖고 있는 로봇이 자율적으로 행동을 하지 못한다면, 불쾌한 골짜기(Mori, 1970)와 같이 로봇에 대한 호감도가 떨어지고, 의인화도 잘 되지 않는 것으로 보인다. 로봇이 자율적으로 행동하고 정서를 경험하도록 해야 로봇에 대한 의인화는 가능한 것으로 보인다. Levillain & Zibetti(2017)의 연구에서는 로봇뿐만 아니라 다른 대상들(예: 인형, 자동차)도 자율적으로 느끼고 행동하는 것처럼 제시하면, 의인화가 증가함을 보고하였다. 따라서 대상의 외관과 자율적 정서/행동 중에서 자율적 정서/행동이 의인화에 더 결정적임을 알 수 있다. 또한 의인화가 단순히 대상을 인간과 혼동하여 인간의 범주에 포함시키는 것이 아니라 대상으로부터 인간의 고유 속성을 추론하는 것이기 때문에 로봇이 인간처럼 주도적으로 행동하고, 감정을 경험하는 것은 로봇의 마음에 대한 추론과정을 유도하여 실제 로봇과 사회적 상호작용을 하고 있다는 생각을 하는데 긍정적으로 기여하는 것으로 보인다.

마음지각이론(Gray & Wegner, 2012b)과 의인화이론(Epley et al., 2007)의 차이는 마음지각이론에서는 대상을 범주로 구분하여(예: 인간 vs. 로봇) 대상의 마음을 평가하는데 비해, 의인화이론에서는 인간이 아닌 대상의 특정 행동을 통해 대상의 마음을 추론하여 이 대상에게 인간의 속성을 부여한다는 것이다(Wang et al., 2015). 이 차이는 마음지각에서는 인간이 아닌 대상에 대한 부정적 평가를 하게 될 가능성을 증가시키고(예: 로봇은 인간이 아니면서 감정을 느끼니 이상함), 의인

3) Wang et al.(2015)은 불쾌한 골짜기에서 경험되는 부정적 정서(uncanny feeling)는 기괴한 불쾌감(eeriness)으로 가장 잘 표현할 수 있지만, 이 정서의 내용에 다양한 요소가 포함되어 있다고 설명했다. 이 정서에는 무서움, 두려움, 예측불가능함과 불안함 등이 포함되어 있는데, 연구자들마다 이 정서를 다양하게 해석한 조작적 정의를 사용하기 때문에, 표준화된 척도가 필요함을 지적하였다. Gray & Wegner(2012b)도 동일한 문제를 언급하며, 이 정서를 가장 대표적으로 표현할 수 있는 기괴한 불쾌감이 일상에서 잘 사용되지 않는 단어이기 때문에, 다른 세 가지의 단어(uneasy, unnerved, creeped out)로 대체하였다고 기술하였다. 본 연구에서는 Gray & Wegner(2012b)의 척도의 세 문항을 한국어로 번역하여(무서움, 불안함, 소름끼침) 사용하였다(부록 2 참조).

화에서는 대상에 대한 긍정적 평가(예: 로봇의 말하는 행동이 나의 마음을 공감하는 것 같음)를 하게 될 가능성을 증가시킨다. 예를 들어, Gray & Wegner(2012b)에서는 인간이 아닌 로봇이 인간의 외관과 유사하게 제시될 때 마음지각은 높아졌지만, 이와 동시에 로봇에 대한 심리적 불쾌감이 증가하였다. 이에 비해, Nijssen et al.(2019)에서는 로봇의 정서(예: 당황함, 행복감)와 생각(예: 망설임, 심사숙고)을 의인화하여 제시하는 조건에서 의인화되지 않은 로봇의 조건보다 가상의 돌발 사고상황에서 로봇/인간 중에서 로봇에게 해를 가하려는 경향이 더 낮게 나타났다. 이에 따라 본 연구에서는 로봇의 외관 또는 로봇의 속성이 인간과 유사할 때, 로봇에 대한 심리적 불쾌감 및 마음지각에서 차이가 나타나는지를 분석하려 한다.

또한 본 연구에서는 선행연구를 확장시키기 위해 로봇에 대한 의인화가 심리적 불쾌감 및 도덕적 처우와 어떤 관계에 있는지를 분석하려 한다. 지금까지 선행연구(예: Gray 등, 2012b)에서는 인간과 비슷한 로봇의 외관이 심리적 불쾌감과 마음지각에 끼치는 영향을 분석하였지만, 로봇의 속성을 의인화하는 것이 로봇에 대한 마음지각 및 도덕적 처우에 어떤 영향을 끼치는지의 문제는 불분명하다. 로봇에 대한 도덕적 처우가 로봇을 단순한 기계 또는 인간처럼 느끼고 행동하는 대상으로 제시하는 것에 따라 달라진다면, 이 결과는 향후 인간-로봇의 상호작용을 촉진시키는데 적용할 수 있을 것이다.

본 연구의 연구문제는 다음과 같다.

첫째, 로봇의 외관이 인간과 유사할수록 로봇에 대한 불쾌감, 로봇에 대한 마음지각 및 도덕적 처우에서 차이가 나타나는가(연구 1)? 선행연구(Gray et al., 2012b)에서는 로봇의 외관이 인간과 유사할수록 로봇에 대한 심리적 불쾌감이 증가하며, 로봇에 대한 마음지각이 경험성의 차원에서 높아짐을 보고하였다. 본 연구에서는 로봇의 외관이 유사할수록 로봇에 대한 마음지각과 심리적 불쾌감이 증가할 것을 예측한다. 또한 로봇에 대한 마음지각에서 경험성이 더 높아질수록 심리적 불쾌감과 도덕적 처우(예: 가해행동에 대한 비도덕성 평정)가 증가할 것을 예측한다.

둘째, 로봇의 속성을 의인화하여 제시하는 것이 기계적 로봇의 조건에 비해 심리적 불쾌감, 마음지각 및

도덕적 처우에서 달라지는가(연구 2)? 로봇의 의인화에 관한 선행연구(Damiano & Dumouchi, 2018; Duffy, 2003)에서는 인간처럼 행동하는 사회적 로봇이 기계적 로봇보다 내적 속성에서 인간과 더 유사하게 지각되어, 로봇에게 인간의 속성(예: 자율적인, 공감하는)이 부여되는 의인화가 나타날 가능성을 제시하였다. 본 연구에서는 선행연구를 확장하여 로봇의 단순한 외관이 아닌 로봇의 정서 및 행동에 대한 의인화의 정도가 증가할수록 로봇에 대한 심리적 불쾌감, 마음지각 및 가해행동에 대한 비도덕성 평정이 달라지는지를 검증하려 한다. 연구 2에서는 로봇의 행동이나 속성이 의인화된 조건이 로봇의 기계적 및 실용적 속성을 기술한 조건보다 로봇에 대한 마음지각 및 도덕적 처우에서 더 높을 것을 예측한다.

### 3. 연구 1

#### 3.1. 연구방법 및 절차

##### 3.1.1. 연구목적

연구 1의 목표는 로봇의 외관에 따라 로봇에 대한 불쾌감, 마음지각 및 도덕적 처우에서 차이가 나타나는지를 검증하는데 있다.

##### 3.1.2. 연구도구 및 절차

4년제 대학교의 대학생 62명(평균연령 만 21.55세, 표준편차=3.08세; 남=28명)이 학교 홈페이지 게시판을 통해 참가하였다. 연구 참가에는 약 10분의 시간이 소요되었다. 참가자들은 오천원 상당의 커피 기프트콘의 보상을 받았다.

연구 1은 온라인 설문지로 진행되었다. 연구 1에서 참가자들은 로봇의 얼굴 자극에 대한 질문지과제를 수행하였다(부록 1 참조). 자극재료는 Mathur & Reichling (2016)의 연구를 토대로 세 장의 사진을 선정하여 참가자에게 제시하였다. 이 사진은 로봇의 외관 연구에서 사용되는 자극재료의 표준화를 위해 선행연구(Mathur & Reichling, 2016; Mueller et al., 2021)에서 일반적으로 제안한 원칙에 따라 로봇의 얼굴/몸통/팔/다리가 모두 사진에서 제시되며, (애니메이션 또는 합성사진이 아닌) 실제 개발되어, 인간과 상호작용하며 현재 사용 중

인 로봇이어야 한다는 원칙을 적용하여 선택하였다. 참가자에게는 각 로봇사진에 제시되는 질문에 응답하도록 지시하였다. 질문지는 총 열 여섯 개 문항으로 구성되었으며, 선행연구에서 사용된 영어문항을 전문번역인의 교차번역을 과정을 거쳐 한국어로 번안된 문항을 사용하였다. 우선, 조작점검을 위해 로봇의 외관이 인간의 얼굴과 유사한 정도에 관한 단일문항(예: 이 로봇의 외관은 인간과 유사하다)을 Müller et al.(2020)의 연구를 참고하여 사용하였다. 또한 Gray & Wegner(2012b)를 토대로 심리적 불쾌감의 3개 문항(예: 이 로봇의 모습을 나를 섬뜩하게 한다)을 사용하였다(부록 2 참조). 연구 1에서 심리적 불쾌감에 관한 문항 간의 내적 합치도(Cronbach's Alpha)는 .83으로 나타났다. 또한 Gray & Wegner(2012b)의 연구를 토대로 마음지각 질문지를 주도성의 5개 문항(예: 이 로봇은 스스로 계획할 수 있다) 및 경험성의 5개 문항(예: 이 로봇은 고통을 느낄 수 있다)을 선정하였다. 본 연구에서 주도성의 내적 합치도(Cronbach's Alpha)는 .81, 경험성의 내적 합치도는 .78로 나타났다. 마지막으로 Gray et al.(2007)을 참고하여 도덕적 처우의 2개 문항(예: 이 로봇을 해친다면 마음이 불편할 것이다)을 로봇에 대한 가해행동의 비도덕성 평정과제에 사용하였다<sup>4)</sup>. 가해행동에 대한 비도덕성 평정 문항의 내적 합치도(Cronbach's Alpha)는 .88로 나타났다. 연구참가자에게는 위의 열 여섯 개 문항의 내용에 자신이 동의하는 정도를 7점 척도(1: 전혀 일치하지 않음, 4: 보통, 7: 매우 일치함)에 따라 평정하도록 지시하였다. 질문지와제가 완료되면, 참가자들에게 연구목적과 취지를 파악했는지에 대한 사후질문에 응답하도록 한 후, 참가에 대한 보상으로 오천 원 상당의 커피쿠폰을 온라인으로 지급하였다.

### 3.1.3. 연구 설계

본 연구는 로봇의 외관(기계적 로봇 vs. 휴머노이드 로봇 vs. 안드로이드 로봇)을 피험자내 변인으로 측정하였다. 측정변인은 로봇의 외관에 따른 심리적 불쾌

감, 마음지각 및 도덕적 처우(가해행동의 비도덕성 평정)였다.

### 3.1.4. 분석 방법

연구 1에서는 로봇의 외관에 따라 심리적 불쾌감, 마음지각 및 도덕적 처우가 달라지는지를 검증하고자 하였다. 이를 위해 SPSS 26.0 통계분석프로그램을 사용하여 기술통계, 대응표본 *t*-검정, 상관분석 및 반복측정 변량분석을 수행하였다.

## 3.2. 연구결과 및 논의

### 3.2.1. 조작점검

연구 1에서는 로봇의 외관에 따라 인간과의 유사성을 조작하였다. 로봇의 외관(기계적 vs. 휴머노이드 vs. 안드로이드 로봇)에 따라 연구참가자가 인간과의 유사성을 인식하는 정도에는 유의한 차이가 나타났다,  $F(2, 122)=523.34, p < .001, \eta^2 = .89$ . 따라서 외관조작이 인간 외관과의 유사성을 경험하는 정도를 적절하게 조작하였다고 볼 수 있다.

### 3.2.2. 로봇의 외관과 심리적 불쾌감

로봇의 외관(기계적 vs. 휴머노이드 vs. 안드로이드 로봇)이 심리적 불쾌감에 끼치는 영향을 분석하기 위해 로봇의 외관을 참가자내 변인으로 반복측정 변량분석을 수행한 결과, 외관의 주효과가 유의했다,  $F(2, 122)=29.62, p < .001, \eta^2 = .32$ . 예측했던 대로 인간의 외관과 가장 유사한 안드로이드 로봇의 조건에서 경험되는 심리적 불쾌감이 가장 높았다(Table 1 참조). 대응표본 *t*-검증을 수행한 결과, 안드로이드 로봇이 휴머노이드 로봇보다 인간의 외관과 더 유사하다고 지각되었으며,  $t(61) = 19.09, p < .001$ , 심리적으로 더 불쾌하게 경험되었다,  $t(61) = 2.91, p < .001$ . 휴머노이드 로봇은 기계적 로봇보다 인간의 외관보다 더 유사하게 평정되었으며,  $t(61) = 12.61, p < .001$ , 심리적 불쾌감이 더 높게

4) 도덕성에는 행위주체(agent)와 피행위자(patient)의 측면으로 구성되는 이원적 특징이 있다. 예를 들어 어떤 사람이 로봇을 해쳤다면, 사람은 행위주체이며, 가해행동을 당한 로봇은 피행위자가 된다. 선행연구(Laakasuo et al., 2021; Müller et al., 2021)에서는 도덕성 평정에서 행위주체의 책임정도를 측정하는 과제(예: 로봇의 행동이 비난받는 정도) 또는 피행위자의 관점에서 가해행동을 당했을 때 도덕적으로 처우하는 정도(예: 해를 당한 로봇에 대한 심리적 불편감)를 측정하였다. 본 연구에서는 이 중에서 로봇이 피행위자로서 가해행동을 당했을 때 로봇에 대해 경험하는 심리적 불편감을 측정하였다.

Table 1. Descriptive statistics (Means, SD) of study 1 for all variables

Type	Similarity	Uncanny feelings	Experience	Agency	Moral care
Mechanical	1.48 (.50)	1.96 (1.38)	1.61 (.85)	3.43 (1.47)	3.15 (1.83)
Humanoid	3.32 (1.18)	3.24 (1.95)	2.94 (1.58)	4.33 (1.18)	3.31 (1.64)
Android	5.87 (.83)	4.05 (1.69)	3.55 (2.14)	4.54 (1.48)	3.57 (2.06)

\*\*\*  $p < .001$ , \*  $p < .05$

평정되었다,  $t(61) = 9.05, p < .001$ .

### 3.2.3. 로봇의 외관과 마음지각

로봇의 외관에 따라 마음지각의 경험성에서 유의한 차이가 나타났다,  $F(2, 122)=46.57, p < .001, \eta^2= .43$ . 대응표본  $t$ -검정을 수행한 결과, 인간과 가장 유사한 안드로이드 로봇 조건에서 로봇에 대한 마음지각은 경험성의 차원에서 휴머노이드 로봇의 조건보다 유의하게 더 높았다,  $t(61) = 3.19, p = .002$ . 또한 마음지각의 주도성에서도 로봇의 외관에 따라 유의한 차이가 나타났다,  $F(2, 122)=25.10, p < .001, \eta^2= .29$ . 대응표본  $t$ -검정에 의하면, 인간과 유사한 안드로이드 로봇의 조건에서 마음지각의 주도성이 기계적 로봇의 조건보다 높았지만,  $t(61) = 5.59, p < .00$ , 안드로이드 로봇-휴머노이드 로봇 간에는 유의한 차이가 나타나지 않았다,  $t(61) = 1.38, p < .17$ .

### 3.2.4. 로봇의 외관, 심리적 불쾌감 및 마음지각의 관계

주목할 것은 인간의 외관과 가장 유사하게 지각되는 안드로이드 로봇의 조건에서 로봇에 대한 심리적 불쾌감과 마음지각의 경험성이 유의한 정적 상관관계에 있는 것이었다,  $r = .29, p = .019$ . 반면, 안드로이드 로봇에 대한 심리적 불쾌감과 마음지각의 주도성 차원 간에는 유의한 상관관계가 나타나지 않았다,  $r = .19, p = .129$ . 이 결과는 로봇에 대한 마음지각에서 감정을 느끼는 경험성의 차원이 로봇에 대한 심리적 불쾌감을 형성하는데 더 중요한 역할을 하게 될 가능성을 보여준다.

### 3.2.5. 로봇의 외관, 마음지각 및 도덕적 처우의 관계

표 1과 같이 로봇의 외관에 따른 마음지각에서 유의한 차이가 나타났지만, 로봇을 해치는 행동에 대한 심리적 불편감에서 유의한 차이가 나타나지 않았다. 기계

적 로봇을 해치는 행동과 휴머노이드 로봇을 해치는 행동의 비도덕성 평정의 차이는 유의하지 않았다,  $t(61) = .84, p = .41$ . 기계적 로봇을 해치는 행동과 안드로이드 로봇을 해치는 행동의 비도덕성 평정의 차이도 유의하지 않았다,  $t(61) = 1.27, p = .213$ ,

이에 비해 안드로이드 로봇을 해치는 행동에 대한 비도덕성 평정과 안드로이드 로봇에 대한 마음지각의 경험성 및 주도성 간에 유의한 정적 상관관계가 나타났다, (경험성;  $r = .45, p = .005$ , 주도성:  $r = .47, p = .003$ ). 이 결과는 로봇에 대한 마음지각이 높을수록 로봇에 대한 도덕적 처우의 의향이 증가함을 보여준다. 반면 안드로이드 로봇을 해치는 행동을 도덕적으로 옳지 못하다고 판단하는 경향은 안드로이드 로봇에 대한 심리적 불쾌감과 유의한 관계가 나타나지 않았다,  $r = .01, p = .968$ . 로봇을 해치는 행동에 대한 비도덕성 평정은 휴머노이드 로봇에 대한 심리적 불쾌감이나,  $r = .07, p = .6638$ , 기계적 로봇에 대한 심리적 불쾌감과도 유의한 상관관계가 나타나지 않았다,  $r = .12, p = .173$ .

### 3.2.6. 논의

연구 1에서는 예측했던 대로 인간의 외관과 가장 유사한 안드로이드 로봇의 조건에서 심리적 불쾌감이 가장 높게 나타났다. 또한 안드로이드 로봇에 대한 마음지각은 경험성의 차원에서 가장 높게 나타났다. 이 결과는 선행연구(Gray et al., 2012b)와 일치하는 결과다. 안드로이드 로봇에 대한 마음지각에서 경험성이 높을수록 심리적 불쾌감이 더 높게 나타났다. 따라서 인간의 외관과 유사하지만, 인간이 아닌 로봇이 인간처럼 감정을 경험한다고 인식될수록 로봇에 대한 심리적 불쾌감이 증가함을 보여준다.

한편 로봇의 외관에 따른 마음지각과 가해행동의 비도덕성 판단 간에 유의한 상관관계가 나타났다. 로봇에 대한 마음지각이 높을수록 가해행동에 대한 심리적 불편

함이 증가하였다. 이에 비해 로봇을 해치는 행동에 대한 비도덕성 평정은 로봇에 대한 심리적 불쾌감과 유의한 관계가 나타나지 않았다. 이 결과는 인간의 모습과 유사한 로봇에 대한 심리적 불쾌감은 로봇의 도덕적 처우를 심사숙고하는 하향처리와는 다른, 별개의 차별적인 자동적 상향처리과정에 의해 형성될 가능성을 보여준다.

## 4. 연구 2

### 4.1. 연구방법 및 절차

#### 4.1.1. 연구목적

연구 2의 목표는 로봇의 속성을 의인화하여 제시하면, 로봇의 속성을 추론하는 하향처리과정에 의해 로봇에 대한 마음지각, 심리적 불쾌감 및 로봇을 해치는 행동에 대한 비도덕성 평정에서 유의한 차이가 나타나는지를 검증하는데 있다.

#### 4.1.2. 연구도구 및 절차

4년제 대학교의 대학생 107명(평균연령 만 22.01세, 표준편차 = 2.93세; 남 = 57명)이 교양수업을 수강하면서 수업게시판의 모집공고를 통해 자발적으로 참가하였다. 연구참가에는 약 5분의 시간이 소요되었다.

연구 2에서는 참가자 집단을 무선적으로 두 집단으로 나누어, 로봇에 대한 각각 다른 글을 제시하였다. 로봇에 대한 지시문은 선행연구(Wang & Krumhuber, 2018; Yam et al., 2021; Ishiguro, 2021)를 토대로 인용하였다(부록 2 참조). 기계적 로봇의 조건에서 참가자는 로봇의 실용적 측면(예: 특정 정보의 신속한 검색)을 기술하는 글을 읽었다. 이에 비해 의인화된 로봇의 집단에서는 로봇이 인간의 삶을 위해 정서적 지지를 수행하며, 인간과 유사한 정서(예: 불안)와 행동(예: 회피 행동)을 갖고 있다는 글을 제시하였다. 이후 참가자는 조작 점검을 위해 이 로봇이 인간과 어느 정도 유사한지를 측정하는 문항에 응답하였다. 또한 참가자는 연구 1과 동일한 마음지각 질문지(10문항)에 7점 척도상에서 응답하였다. 또한 연구 1과 동일하게 2개 문항(예: 이 로봇을 해친다면 마음이 불편할 것이다)을 로봇에 대한 가해행동의 비도덕성 평정과제에 사용하였다. 마

지막으로 로봇에 대한 심리적 불쾌감을 측정하였다. 로봇에 대한 심리적 불쾌감은 연구 1과 동일한 3개 문항을 사용하여 측정하였다. 참가자에게는 문항의 내용에 자신이 동의하는 정도를 7점 척도(1: 전혀 일치하지 않음, 4: 보통, 7: 매우 일치함)에 따라 평정하도록 지시하였다. 연구 2에서 마음지각 질문지의 신뢰도(Cronbach's alpha)는 경험성의 측면에서 .72, 주도성의 측면에서는 .71로 나타났다. 로봇을 해치는 행동에 대한 도덕성을 평정하는 문항의 신뢰도(Cronbach's alpha)는 .68, 로봇에 대한 심리적 불쾌감에서는 .71로 나타났다. 마지막으로 참가자들은 연구의 목적과 취지를 파악했는지에 대한 질문에 응답한 후, 오천원 상당의 기프트콘을 온라인으로 제공받았다.

#### 4.1.3. 연구 설계

연구 2는 로봇의 제시방식(기계적 로봇 vs. 의인화된 로봇)에 따른 피험자간 설계로 구성되었다. 측정변인은 로봇에 대한 유사성의 지각, 로봇에 대한 심리적 불쾌감 및 마음지각, 로봇에 대한 가해행동의 비도덕성 판단이었다.

#### 4.1.4. 분석 방법

연구 2에서는 로봇 속성의 제시방식(기계적 로봇 vs. 의인화된 로봇)에 따른 유사성의 지각, 심리적 불쾌감, 마음지각 및 가해행동에 대한 비도덕성 평정에 관한 연구참가자의 자료를 자기보고식 질문지를 통해 수집하였다. 연구 2의 자료분석을 위해 SPSS 26.0 프로그램을 사용하여 기술통계, 독립표본 t-검정, 상관분석 및 다중회귀분석을 수행하였다.

## 4.2. 연구결과 및 논의

### 4.2.1. 조작점검

연구 2에서는 로봇 속성의 제시방식에 따라 인간과의 유사성을 조작하였다. 연구참가자는 의인화된 로봇 조건에서 기계적 로봇보다 인간과의 유사성을 더 높게 평정하였다,  $t(105) = 5.59, p < .001$ . 따라서 로봇 속성의 제시방식이 인간과의 유사성을 경험하는 정도를 적절하게 조작하였다고 볼 수 있다.



4.2.2. 의인화된 로봇, 심리적 불쾌감 및 마음지각

Table 2와 같이 의인화된 로봇에 대한 심리적 불쾌감은 기계적 로봇과 유의한 차이가 나타나지 않았다,  $t(105) = -.43, p = .65$ . 또한 의인화된 로봇의 조건에서 유사성이 지각된 정도와 심리적 불쾌감의 관계는 유의하지 않았다,  $r = .12, p = .173$ . 이 결과는 연구 1에서 로봇 외관의 유사성이 증가할수록 심리적 불쾌감이 높았던 것과 차이가 나타남을 보여준다. 연구 2에서는 의인화된 로봇이 인간과 유사하게 지각되지만, 심리적 불쾌감에 유의한 영향을 끼치지 않았음을 보여준다.

4.2.3. 의인화된 로봇에 대한 마음지각과 가해행동에 대한 비도덕성 평정

표 2와 같이 의인화된 로봇에 대한 가해행동은 기계적 로봇에 대한 가해행동보다 더 비도덕적으로 평정되었다,  $t(105) = 2.33, p = .023$ . 또한 의인화된 로봇에 대한 마음지각의 경험성이 높을수록 가해행동에 대한 심리적 불편감이 높게 나타났다,  $r = .27, p = .04$ . 이에 비해 마음지각의 주도성은 가해행동에 대한 비도덕성의 판단과 유의한 관계가 나타나지 않았다,  $r = .17, p = .12$ .

마지막으로 연구 2에서 인간과의 유사성, 마음지각의 경험성, 마음지각의 주도성 및 심리적 불쾌감이 로봇을 해치는 행동에 대한 비도덕성 판단에 영향을 끼치는지를 검증하였다. 다중회귀분석에 의하면, Table 3과 같이 인간과의 유사성, 마음지각의 경험성, 주도성 및 심리적 불쾌감은 모두 분산팽창지수(VIF)가 10이하의 범위에 나타나, 다중공선성의 문제가 없었으며, 가해행동에 대한 도덕성 판단의 총 변량에서 16.8%를 설명하였다. 심리적 불쾌감은 가해행동에 대한 비도덕성 판단에 유의한 부적 영향을 끼쳤고, 마음지각의 경험성은 유의한 정적 영향을 끼친 반면, 인간과의 유사성 및 마음지각의 주도성은 도덕성 판단에 유의한 영향을 끼치지 않은 것으로 나타났다.

4.2.4. 논의

연구 2에서는 로봇의 의인화를 통해 로봇의 속성을 인간의 마음과 유사하게 지각하도록 유도하였을 때, 로봇에 대한 심리적 불쾌감, 마음지각, 로봇에 대한 도덕적 처우에서 차이가 나타나는지를 비교하였다. 그 결과 의인화된 로봇의 조건에서 기계적 로봇의 조건보다 인간과의 유사성이 더 높게 평정되었지만, 심리적 불쾌감

Table 2. Means and correlations between variables of study 2

Types	Variables	Mean (SD)	Correlations			
			1	2	3	4
Mechanical robot	1. Similarity	2.89 (1.62)				
	2. Experience	2.25 (1.13)	-.07			
	3. Agency	4.14 (1.42)	-.02	-.48**		
	4. Uncanny Feelings	2.45 (1.21)	-.18	.14	.16	
	5. Moral Care	4.54 (1.61)	-.06	.33*	1.17	.08
Humanlike-robot	1. Similarity	4.42 (1.34)				
	2. Experience	3.31 (1.11)	.35**			
	3. Agency	5.04 (1.10)	.42**	.44**		
	4. Uncanny Feelings	2.57 (1.35)	-.10	.18	-.14	
	5. Moral Care	5.27 (1.28)	.25*	.27*	.17	-.34**

\* $p < .05$ , \*\*  $p < .01$

Table 3. Multiple regression analysis for predictor variables of moral care (study 2)

Variables	B	S.E.	$\beta$	$t$	$p$	VIF	F = 6.132 R <sup>2</sup> = .200 Adj.R <sup>2</sup> = .168
Similarity	.147	.105	.138	1.405	.163	1.135	
Experience	.284	.105	.279	2.708	.008	1.390	
Agency	-.041	.134	-.031	-.304	.762	1.348	
Uncanny Feelings	-.339	.093	-.178*	-3.649	.000	1.039	

에서는 두 조건 간에 차이가 나타나지 않았다. 이 결과는 로봇의 외관이 인간과 유사해질 때 나타나던 불쾌한 골짜기 현상이 의인화를 통해 로봇의 속성을 인간과 유사하게 추론하도록 하면 로봇에 대한 심리적 불쾌감이 증가하지 않음을 보여준다.

의인화된 로봇에 대한 마음지각은 기계적 로봇에 대한 마음지각보다 경험성과 주도성의 차원에서 더 높게 평정되었다. 이 중에서 마음지각의 경험성은 심리적 불쾌감과 함께 의인화된 로봇에 대한 가해행동을 도덕적으로 옳지 못하다고 판단하는데 유의한 영향을 끼쳤다. 이 결과는 의인화된 로봇이 인간과 유사하게 고통을 느끼는 것으로 지각되면, 기계적 로봇보다 더 높은 도덕적 지위를 차지하게 될 가능성을 보여준다. 또한 연구 2에서는 의인화된 로봇에 대한 심리적 불쾌감이 증가할수록 로봇에 대한 가해행동을 비도덕적으로 판단하는 경향이 줄어들 수 있음을 보여주었다. 연구 1에서는 로봇의 외관이 인간과 유사해질수록 심리적 불쾌감이 증가한데 비해, 연구 2에서 로봇의 마음이 인간의 마음과 유사하게 지각되는 것은 심리적 불쾌감에 영향을 끼치지 않았다. 이 결과는 인간과의 유사성이 로봇의 외관에 의한 상향처리와 로봇의 마음을 추론하는 하향처리에서 로봇에 대한 심리적 불쾌감을 형성하는데 각각 별개의 과정으로 차별화되어 작동할 가능성을 시사한다.

## 5. 결론

본 연구에서는 로봇의 외관과 의인화에 따른 심리적 불쾌감, 마음지각 및 도덕적 처우의 관계를 분석하였다. 그 결과 연구 1에서는 로봇의 외관이 인간과 비슷한 안드로이드 로봇의 조건에서 로봇에 대한 심리적 불쾌감 및 마음지각이 가장 높게 나타났다. 그러나 이 세 조건에서 로봇에 대한 도덕적 처우에서는 차이가 유의하지 않았다. 이 결과는 로봇의 외관이 인간과 유사할수록 마음지각이 높아졌지만, 이와 동시에 로봇에 대한 심리적 불쾌감이 증가하여 로봇에 대한 가해행동을 판단하는데 차이가 나타나지 않았을 가능성을 보여준다. 연구 2에서는 로봇의 유형을 의인화된 로봇 또는 기계적 로봇으로 구분하여 제시하였을 때, 정서를 경험

하는 의인화된 로봇의 조건에서 마음지각 및 도덕적 처우의 정도가 기계적 로봇의 조건보다 더 높게 나타났다. 이 결과는 인간과 비슷한 로봇의 외관보다는 로봇의 속성에 대한 의인화가 인간과 비슷한 로봇의 마음을 추론하는 과정을 통해 인간-로봇의 상호작용을 촉진시킬 가능성을 보여준다.

본 연구는 다음의 두 가지 중요한 시사점을 제시한다. 첫째, 본 연구에서는 인간의 외관과 유사한 단서를 통한 상향처리가 로봇에 대한 심리적 불쾌감을 증가시키는데 비해, 인간의 마음과 유사한 속성을 추론하는 하향처리는 인간-로봇의 관계에 긍정적 영향을 끼칠 수 있음을 보여주었다. 따라서 정보처리과정의 초기에 자동적으로 진행되는 로봇의 외관과 연관된 상향처리의 부정적 결과를 로봇의 마음에 대해 심사숙고하며 추론하는 하향처리를 통해 수정할 수 있는 가능성을 보여준다. 후속연구에서는 다양한 로봇의 외관조건에서 의인화를 유도하였을 때, 로봇의 외관을 통한 상향처리가 로봇의 마음을 추론하는 하향처리와 상호작용하는 양상을 분석해볼 필요가 있다. 둘째, 본 연구에서는 로봇에 대한 마음지각의 경험성이 로봇에 대한 도덕적 지위의 판단에 유의한 영향을 끼침을 보여주었다. 연구 1과 연구 2에서 모두 참가자가 로봇이 정서를 느낀다고 판단할수록 로봇에 대한 가해행동을 옳지 못하다고 평정하는 경향이 증가하였다. 반면 로봇에 대한 마음지각의 주도성은 로봇을 해치는 행동에 대한 심리적 불편감에 유의한 영향을 끼치지 않았다. 이 결과는 무생물인 로봇이 ‘고통을 느끼는 존재’로서 지각되는 것이 ‘행위의 주체’로서 인식되는 것보다 로봇에 대한 도덕적 정서를 유발하는데 더 많이 기여할 수 있음을 보여준다.

본 연구의 제한점 및 후속연구에 대한 제언은 다음과 같다. 첫째, 연구 1에서는 로봇 사진을 외관의 유사성에 따라 분류하여 자극재료로 사용하였다. 연구결과를 통해 로봇에 대한 심리적 불쾌감이 인간 외관과의 유사성에 따라 증가함을 보여주었지만, 실제 인간 사진과의 비교조건이 포함되어 있지 않은 한계가 있다. 연구 1에서 심리적 불쾌감의 기저선을 설정하지 않은 것은 로봇에 대한 심리적 불쾌감이 인간의 조건에 비해 전반적으로 어떻게 나타나는지를 예측하는데 어려움을 준다. 둘째, 후속연구에서는 본 연구에서 사용된 인간과 유사한 로봇 사진의 범위를 더 확장하여 다양한 유형의 로봇에

대한 심리적 불쾌감을 측정해볼 필요가 있다. 인간과 비슷한 외관의 로봇은 지속적으로 개발 중에 있고, 아직까지 선행연구에 많이 사용되지 않은 국내 안드로이드 로봇(예: 에버)에 대한 심리적 불쾌감, 마음지각 및 도덕적 처우의 관계를 검증하는 것은 인간-로봇의 상호작용에 대한 실제적 시사점을 제공할 것으로 생각된다. 셋째, 본 연구에서는 참가자가 로봇에 대한 가해행동을 비도덕적으로 판단하는 정도를 선행연구에서 선정한 2개의 문항을 통해 측정하였다. 그러나 도덕적 의사결정은 정서와 사고가 개입되는 복잡한 과정이기 때문에 간략한 문항만을 통해 도덕성 판단을 측정하는데 한계가 있다. 후속연구에서는 다양한 윤리적 갈등 시나리오를 활용하여 로봇의 외관과 의인화의 정도가 달라질 때, 로봇에 대한 가해행동을 판단하는 경향이 어떻게 달라지는지를 분석해볼 필요가 있다. 넷째, 본 연구에서는 로봇의 의인화를 인간과 유사한 정서(예: 불안)를 느끼는 능력을 통해 유도하였다. 본 연구에서 마음지각의 경험성만이 로봇에 대한 가해행동의 도덕성 판단에 유의한 영향을 끼친 것은 본 연구에서 인간과 유사한 정서에만 초점을 맞추어 의인화를 조작했기 때문일 가능성이 있다. 의인화는 인간과 유사한 정서뿐만 아니라 인간과 비슷한 행동을 통해서도 유도될 수 있다. 예를 들어, 음악을 들으며, 즐거움을 느끼고, 춤을 추는 로봇의 행동을 제시하며 의인화를 유도할 수 있다. Haslam et al.(2008)은 인간 본성(human nature)(예: 정서)과 인간 고유성(human uniqueness)(예: 문화)의 측면에서 인간성의 개념을 규정했다. 이것은 무생물에게 인간 속성을 부여할 때 정서를 느끼는 능력뿐만 아니라 음악과 같이 문화를 이해하는 행동을 보여주는 것도 가능성을 보여준다. 후속연구에서는 다양한 유형의 의인화가 로봇에 대한 마음지각과 도덕성 판단에 차별적인 영향을 끼치는지를 분석해볼 필요가 있다.

인공지능의 발달과 함께 인간과 유사한 자율적 로봇은 우리의 일상에서 점점 더 큰 비중을 차지하게 될 것이다(Yam et al., 2020). 본 연구에서는 로봇의 외관이 인간과 유사할수록 심리적 불쾌감이 증가할 수 있음을 보여주었다. 그러나 의인화를 통해 로봇이 인간과 유사한 정서를 경험하는 것으로 인식된다면, 다른 기계적 로봇보다 도덕적 처우에서 더 높은 지위를 차지할 가능성도 제시하였다. 로봇에게 “e-인간(전자인간)’으

로서 인간의 지위를 부여하고, 사회에서 한 개인으로 받아들여지도록 새로운 규칙이 제정된다면, 다양한 윤리적 갈등상황이 발생할 수 있다. 예를 들어, 내가 모르는 사람을 위해 내가 잘 아는, 정서를 표현하는 로봇이 희생당하는 것이 도덕적으로 합당한가? 로봇의 지위와 의무는 어느 정도까지가 적절한가? 본 연구에서는 의인화가 로봇의 정서에 대한 공감을 통해 로봇에 대한 도덕적 처우를 촉진시킬 가능성을 보여주었다. 반면 이 결과는 인간-로봇의 도덕적 지위의 판단에 새로운 규칙이 적용되고, 인간을 위해 로봇이 희생되어야 하는 윤리적 갈등상황에서 발생할 수 있는 다양한 문제를 이해하고, 예측하는 것이 시급함을 보여준다.

## REFERENCES

- Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, 90, 363-371. DOI: 10.1016/j.chb.2018.08.028
- Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness creeps In: Uncanny valley feelings are acquired in childhood. *Child Development*, 90(4), 1202-1214. DOI: 10.1111/cdev.12999
- Carter, O., Hohwy, J., Van Boxtel, J., Lamme, V., Block, N., Koch, C., & Tsuchiya, N. (2018). Conscious machines: Defining questions. *Science*, 359(6374), 400. DOI: 10.1126/science.aar4163
- Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human-robot co-evolution. *Frontiers in Psychology*, 9, 468. DOI:10.3389/fpsyg.2018.00468
- Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it?. *Science*, 358(6368), 486-492. DOI:10.1126/science.aan8871
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3), 177-190. DOI: 10.1016/S0921-8890(02)00374-3
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315, 619. DOI: 10.1126/science.1134475
- Gray, K., & Schein, C. (2012). Two minds vs. two philosophies: Mind perception defines morality and

- dissolves the debate between deontology and utilitarianism. *Review of Philosophy and Psychology*, 3(3), 405-423. DOI: 10.1007/s13164-012-0112-5
- Gray, K., & Wegner, D. M. (2012a). Morality takes two: Dyadic morality and mind perception. In M. Mikulincer & P. R. Shaver (Eds.), *Herzliya series on personality and social psychology*. The social psychology of morality: Exploring the causes of good and evil (p. 109-127). American Psychological Association. DOI: 10.1037/13091-006
- Gray, K., & Wegner, D. M. (2012b). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125, 125-130. DOI: 10.1016/j.cognition.2012.06.007
- Gray, K., Knickman, T. A., & Wegner, D. M. (2011). More dead than dead: Perceptions of persons in the persistent vegetative state. *Cognition*, 121(2), 275-280. DOI: 10.1016/j.cognition.2011.06.014
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480. DOI: 10.1037//0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216. DOI: 10.1037/0022-3514.85.2.197
- Haslam, N., Loughnan, S., Kashima, Y., & Bain, P. (2008). Attributing and denying humanness to others. *European Review of Social Psychology*, 19(1), 55-65. DOI: 10.1080/10463280801981645
- Jang, Y. (2006). A study on the depiction of concept in user centered for the supporting shopping robot design development. *Science of Emotion & Sensibility*, 9(3), 287-297.
- Kim, Y., Yoon, S., Lee, D., & Kwak, Y. (2001). Mechanism design of the interactive emotional robot. *Science for Emotion & Sensibility*, 233-238.
- Kuchenbrandt, D., Eyssel, F., Bobinger, S., & Neufeld, M. (2013). When a robot's group membership matters. *International Journal of Social Robotics*, 5(3), 409-417. DOI: 10.1007/s12369-013-0197-8
- Laakasuo, M., Palomäki, J., & Köbis, N. (2021). Moral uncanny valley: A robot's appearance moderates how its decisions are judged. *International Journal of Social Robotics*, in press. DOI: 10.1007/s12369-020-00738-6
- Levillain, F., & Zibetti, E. (2017). Behavioral objects: The rise of the evocative machines. *Journal of Human-Robot Interaction*, 6(1), 4-24. DOI: 10.5898/JHRI.6.1.Levillain
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems*, 7(3), 297-337. DOI: 10.1075/is.7.3.03mac
- Majdandžić, J., Bauer, H., Windischberger, C., Moser, E., Engl, E., & Lamm, C. (2012). The human factor: Behavioral and neural correlates of humanized perception in moral decision making. *PLoS ONE*, 7, e47698. DOI: 10.1371/journal.pone.0047698
- Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146, 22-32. DOI: 10.1016/j.cognition.2015.09.008
- Mori, M. (1970). The uncanny valley. *Energy*, 7, 33-35.
- Müller, B. C., Gao, X., Nijssen, S., & Damen, T. (2020). I, robot: How human Appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics*, in press. DOI: 10.1007/s12369-020-00663-8
- Nijssen, S. R., Müller, B. C., van Baaren, R. B., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition*, 37(1), 41-56. DOI: 10.1521/soco.2019.37.1.41
- Ochsner, K. N., Ray, R. R., Hughes, B., McRae, K., Cooper, J. C., Weber, J., Gabrieli, J. D. E., & Gross, J. J. (2009). Bottom-up and top-down processes in emotion generation: Common and distinct neural mechanisms. *Psychological Science*, 20(11), 1322-1331. DOI: 10.1111/j.1467-9280.2009.02459.x
- Pyszczynski, T., Greenberg, J., & Solomon, S. (1999). A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory. *Psychological Review*,

- 106(4), 835-845. DOI: 10.1037/0033-295x.106.4.835
- Sanders, R. (2017, December 21). Feeling like the Grinch? ‘Customer rage’ is real. Retrieved from <https://www.usatoday.com/story/money/business/2017/12/21/feeling-like-grinch-customer-rage-real/971904001>
- Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence, 16*, 337-351. DOI: 10.1162/pres.16.4.337
- Spatola, N., & Wudarczyk, O. A. (2020). Implicit attitudes towards robots predict explicit attitudes, semantic distance between robots and humans, anthropomorphism, and prosocial behavior: From attitudes to human-robot interaction. *International Journal of Social Robotics*, in press. DOI: 10.1007/ s12369-020-00701-5
- Wang, S., Lilienfeld, S. O., & Roach, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393-407. DOI: 10.1037/ gpr0000056
- Wang, X., & Krumbhuber, E. G. (2018). Mind perception of robots varies with their economic versus social function. *Frontiers in Psychology, 9*, 1-9.
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219-232. DOI: 10.1177/ 1745691610369336
- Weise, D., Pyszczynski, T., Cox, C., Arndt, J., Greenberg, J., Solomon, S., & Kosloff, S. (2008). Interpersonal politics: The role of terror management and attachment processes in shaping political preferences. *Psychological Science, 19*(5), 448-455. DOI: 10.1111/j.1467-9280.2008.02108.x
- Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2020). Robots at work: People prefer – and forgive – service robots with perceived feelings. *Journal of Applied Psychology*, in press. DOI: 10.1037/apl0000834

원고접수: 2021.07.25

수정접수: 2021.09.20

게재확정: 2021.09.23

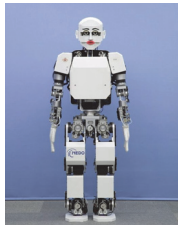
**부록 1. 연구 자료**

연구 1

1. 기계적 로봇조건



2. 휴머노이드 로봇조건



3. 안드로이드 로봇조건



연구 2

1. 기계적 로봇 조건

나의 로봇 ‘클라라’는 내가 시간을 더 효율적으로 사용할 수 있도록 도와줍니다. 예를 들어, 내가 마트에 갈 때, 같이 동행합니다. 내가 마트에 가서 원하는 제품을 찾지 못할 때 편리하게 도와줍니다. 또한 제품에 대한 질문사항이 있을 때, 언제나 다양한 정보를 신속하게 제공해줍니다.

2. 로봇 속성의 의인화조건

나의 로봇 “클라라”는 내가 혼자 있을 때, 나와 같이 시간을 보내줍니다. 예를 들어, 내가 마트에 갈 때, 같이 동행합니다. 그런데 하루는 클라라가 마트에 가서 많은 사람들을 보고 당황해했습니다. 클라라는 진열대 구석에 가서 꼼짝 않고 얼어붙어 있었습니다.

**부록 2. 질문지 문항**

연구 1과 연구 2(동일 문항 사용 사진 대신 기술문 제시)

A. 아래의 로봇 사진을 보시고, 이 로봇이 실제 인간의 모습과 어느 정도 비슷하다고 생각하는지에 대해 자신의 의견을 “V”로 표시해주시시오(1: 전혀 비슷하지 않음, 4: 보통, 7: 매우 비슷함).

1-2-3-4-5-6-7

B. 아래의 로봇 사진을 보시고, 각 문항이 이 로봇과 어느 정도 연관된다고 생각하는지에 대해 자신의 의견을 표시해주시시오(1: 전혀 연관되지 않음, 4: 보통, 7: 매우 연관됨).

- |                    |               |
|--------------------|---------------|
| 1 자신의 욕구를 가질 수 있다  | 1-2-3-4-5-6-7 |
| 2 즐거움을 느낄 수 있다     | 1-2-3-4-5-6-7 |
| 3 배고픔을 느낄 수 있다     | 1-2-3-4-5-6-7 |
| 4 고통을 느낄 수 있다      | 1-2-3-4-5-6-7 |
| 5 분노를 느낄 수 있다      | 1-2-3-4-5-6-7 |
| 6 자신을 통제할 수 있다     | 1-2-3-4-5-6-7 |
| 7 기억을 할 수 있다       | 1-2-3-4-5-6-7 |
| 8 자신의 느낌을 인식할 수 있다 | 1-2-3-4-5-6-7 |
| 9 스스로 계획을 세울 수 있다  | 1-2-3-4-5-6-7 |
| 10 도덕적으로 행동할 수 있다  | 1-2-3-4-5-6-7 |

C. 아래의 문항을 읽고, 위의 로봇에 대한 자신의 의견에 일치하는 곳에 표시해주시시오.

(1: 전혀 일치하지 않음, 4: 보통, 7: 매우 일치함)

- |                          |               |
|--------------------------|---------------|
| 11 이 로봇의 외모가 무섭다.        | 1-2-3-4-5-6-7 |
| 12 이 로봇의 외모가 나를 불안하게 한다. | 1-2-3-4-5-6-7 |
| 13 이 로봇의 외모가 소름끼친다.      | 1-2-3-4-5-6-7 |

D. 아래의 문항을 읽고, 위의 로봇에 대한 자신의 의견에 일치하는 곳에 표시해주시시오.

(1: 전혀 일치하지 않음, 4: 보통, 7: 매우 일치함)

- |                                     |               |
|-------------------------------------|---------------|
| 14 내가 이 로봇을 해친다면, 마음이 불편할 것이다.      | 1-2-3-4-5-6-7 |
| 15 나는 이 로봇을 해치는 것이 도덕적으로 나쁘다고 생각한다. | 1-2-3-4-5-6-7 |