Philosophy & Technology (2025) 38:109
https://doi.org/10.1007/s13347-025-00941-1

RESEARCH ARTICLE

Kantianism for the Ethics of Human–Robot Interaction

Hayate Shimizu[1]

Received: 24 December 2024 / Accepted: 14 July 2025

Abstract
This paper explores the implications of applying Kantianism, traditionally focused on human moral agents, to the ethics of human–robot interaction. Kantian approaches have often been regarded as ineffective for assessing whether non-human entities like robots deserve moral consideration. However, this paper argues that Kantian resources provide more substantial support for the moral consideration of robots and the value of human–robot relationships than is commonly assumed. The analysis proceeds in three steps. First, the paper examines Kant's doctrine of ends in themselves, which has traditionally viewed robots solely as instrumental means. Second, it discusses a Kantian perspective on indirect duties, emphasizing how human interactions with robots influence human virtue and moral feeling. This framework justifies norms against cruel behavior toward robots to protect one's moral character but remains insufficient as it only prescribes "avoiding cruelty insofar as it affects one's character." Finally, to address this limitation, the paper introduces a Kantian perspective grounded in the duty to promote others' happiness as an obligatory end. This approach enables the derivation of a duty to protect human–robot relationships based on moral concern for individuals who find happiness through such relationships.

Keywords Kant · Human–robot interaction · Indirect duty · Virtue · Others' happiness

✉ Hayate Shimizu
shimizu.hayate.z4@elms.hokudai.ac.jp

[1] Graduate School of Humanities and Human Sciences, Hokkaido University, Sapporo, Japan

Published online: 24 July 2025

1 Introduction

The rapid and significant impact of artificial intelligence (AI) technology on human society and the various ethical issues it raises have made the ethics of AI one of the liveliest topics in the philosophy of technology (Gordon & Nyholm, 2021). Robots with advanced artificial intelligence are being introduced into many areas of our lives. There are many different types of robots now, including self-driving cars, military robots, logistics robots for moving goods, robot vacuum cleaners, and companion robots that are being advertised as a new form of intimate partner. This paper focuses on companion robots, which are expected to be able to form intimate relationships with humans. We focus in particular on moral considerations of the human–robot relationship, which has recently opened up to the possibility of a more intimate and special relationship beyond the mere framework of "user and tool." This technological leap introduces urgent ethical questions that challenge traditional notions of moral status—questions such as "Do robots have moral status?," "Are robots objects of moral consideration?," and "Do robots have (moral or legal) rights?"—upon which many papers have been published (e.g., Bryson, 2010, 2018; Coeckelbergh, 2014, 2021; Danaher, 2019; DeGrazia, 2022; Gunkel, 2018; Mosakas, 2021; Müller, 2021; Sparrow, 2021). As AI enhances the capabilities of robots, we must ask ourselves: What do we owe to these robots? Should the relationships between humans and robots be respected? This paper uses Kantian resources to construct a robust ethical framework for understanding human-robot relationships.

In the existing debate on the moral status of robots, two major approaches can be identified. One is the properties approach, which grounds moral status in the intrinsic properties of robots, such as consciousness or sentience (e.g., Bryson, Mosakas, Sparrow). The other is the relational approach, which focuses on the moral significance of the relationships that humans form with robots (e.g., Coeckelbergh, Gunkel). While this paper does not directly take a position within this debate, it offers a new Kantian argument that shifts attention away from the ontology of robots and toward their significance in human life and happiness. In this sense, it may exhibit a certain affinity with the relational camp, insofar as it emphasizes what robots become for us, rather than what they are in themselves. This line of inquiry falls within the field of "the ethics of human–robot interaction." According to Nyholm (2020, 2021), this field addresses two key questions: "How should robots and other machines be made to behave around human beings?" and "How should human beings conduct themselves around robots and other machines with advanced forms of artificial intelligence?" This paper focuses in particular on the latter.

Technological entities such as robots are an artifact created to improve the efficiency of human work[1], but there are scenes now where it is recognized as more than just a tool for humans. For example, Paro, a seal-shaped therapeutic robot developed by *The National Institute of Advanced Industrial Science and Technology*, was created as a companion to ease loneliness.[2] Another example is AIBO, a dog-shaped

[1] As Nyholm pointed out, the word. "robot" derives from the Czech language word "robota", which roughly means 'forced labor' (Nyholm et al., 2023).

[2] See, http://www.parorobots.com/.

companion robot developed by Sony, whose funeral was held in Japan.[3] This suggests that some people regard AIBO as more than just a replaceable machine—almost as a "member of the family." In addition, Replika has been developed and is used as an AI companion designed to enable emotionally engaging interactions with humans.[4] In fact, it has been pointed out that some people already consider Replika their friend (Brandtzaeg et al., 2022). Some authors also argue for the possibility of a friendship or intimate relationships between humans and robots (Levy, 2008; Danaher, 2016; Ryland, 2021).[5] In this paper, I aim to argue that if such relationships can be considered valuable between humans and robots, we may indeed have an ethical duty to respect and protect these relationships.

In examining the ethics of human–robot interactions, one typical approach is to apply existing traditional ethical theories. However, this method is fraught with difficulties. This is because, as Nyholm notes, traditional ethical theories—such as utilitarianism, Kantian ethics, and virtue ethics—have overwhelmingly been concerned only with humans as moral agents and thus do not seem to apply to the ethics of human–robot interaction (Nyholm, 2021). This paper argues that a Kantian approach offers more resources than the literature has recognized.[6] I draw specifically on Kant's account of the duties of virtue to make this case. Kantianism is well known for its view that persons are the only ends in themselves that must be respected, which at first glance may seem ill-suited to addressing relationships with non-human entities like robots. However, by focusing on the indirect duty argument, often referenced in robot ethics, alongside Kant's duty of virtue to others, this paper offers an ethical argument for the duties we may have toward protecting human–robot relationships.

In laying the theoretical groundwork for this Kantian approach, I will examine the following three issues: (1) Robots as an instrumental means for human ends: This perspective assesses the traditional view of robots as a means to serve human purposes. (2) Robots as a practical means for human virtue: This argument can impose an ethical duty to refrain from mistreating robots, but this paper evaluates its implications as limited. Perspectives (1) and (2) have been previously discussed in the context of Kantianism and robot ethics (or, more broadly, applied ethics). This study introduces a new third perspective, (3) the duty to protect the human–robot relationship as a duty to others: We reference the moral duty of concern for others whose happiness is constituted by their relationship with robots, proposing a norm to actively protect the human–robot relationships based on the duty of virtue to promote others'

[3] See, https://www.nationalgeographic.com/travel/article/in-japan--a-buddhist-funeral-service-for-robot-dogs There is also the example of the military robot Boomer, which was given an improvised military funeral by U.S. soldiers on the battlefield in Iraq when it Boomer "died."

[4] See, https://replika.com/.

[5] For example, Danaher argues that we should adopt a more performative and behavioral understanding of friendship, and suggests that robots may meet Aristotle's criteria for virtue friendship. Nyholm criticizes Danaher's argument for the possibility of human–robot friendship (Nyholm, 2020, Chap. 7). Elder also criticizes the friendship between humans and robots, arguing that such relationships lack the agency and reciprocity necessary for genuine friendship (Elder, 2017).

[6] Studies that apply Kantianism to the problem of robot ethics can be found, for example, in Gordon and Nyholm's paper "Kantianism and the Problem of Child Sex Robots." They conclude that Kantian ethics does not provide a definitive answer to the problem of child sex robots.

happiness. In this way, Kantianism for the ethics of human–robot interaction argues not only for avoiding cruel treatment of robots but also for a moral duty to protect human–robot relationships.

2 Robots as a Means to Serve Humans

The foundational principle of Kantianism, the categorical imperative, is an unconditional moral command that asserts a moral imperative for agents to follow. Kant's categorical imperative comprises several formulations, one of which is the "Formula of Humanity." This formulation states that one should treat others not merely as a means, but also as ends in themselves. According to Kant, human agents, who are autonomous and rational beings, must be treated not merely as a means to an end, but as beings with their own intrinsic value. This principle asserts that every person are ends in themselves and have intrinsic value that should be respected:

> Now I say that the human beings and in general every rational beings exist as an end in itself, not merely as a means to be used by this or that will at its discretion; instead they must in all their actions, whether directed to themselves or also to other rational beings, always be regarded at the same time as an end. (IV, 428)[7]

So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means. (IV, 429)

On the other hand, a thing (*Sache*) is treated as a mere means for a person. This refers to a being with value relative to a person that has absolute value (dignity) as an end in itself (IV, 428). The basis of the end in itself or dignity of a being is that it is an autonomous rational being, and a "person" in the strong rationalistic sense is an end in itself. Moreover, it is supported by the idea of "humanity," and the Kantian moral circle appears to be open only to humans.[8] Therefore, from Kant's doctrine of ends in themselves, it follows that the robot can relate to humans only as a means to an end, and it would be wrong to treat it as an end in itself, that is, as if it had intrinsic value in its being. As Gordon and Nyholm presented in their Kantian argument, things (including everything that is not a person with reason) have value only relative to people's desires, and it is never inherently wrong to treat things as a mere means to our end. Therefore, it is permissible to treat robots as a mere means to any human end (Gordon & Nyholm, 2022, pp. 136–137). In other words, robots are useful as a means to an end, but they cannot be an end in themselves. Thus, from the Kantian approach, the conclusion that robots are merely a means to an end is not surprising and should be taken for granted. In other words, robots are nothing more than tools to be used

[7] All citations are in accordance with the standard Academy pagination.

[8] Some Kantian scholars, like Korsgaard, 2018, recognize an end in itself even in animals, but because it presupposes the desire and the ability to seek what is good for its nature, Korsgaard's strategy of extending the ethical circle is unlikely to extend to robots. Even if the robot seeks the good for itself, such as self-preservation or repair, this is programmed and not based on its nature. However, some argue that artifacts also have a good of their own that is not reducible to human intentions (Vogel, 2015).

and cannot be morally significant entities that should be given moral concern as dignified and not reduced to mere means. This argument can be summarized as follows:

1. Only rational, autonomous beings that are ends in themselves have the absolute value of dignity and should be treated morally.
2. Because robots are neither rational nor autonomous, they are things with only relative value.
3. Therefore, robots are not beings that should be treated morally.

This argument does not assert that only humans have absolute value but rather that all rational beings capable of autonomous decision-making have absolute value. So, if robots ever become autonomous agents in the Kantian sense, they could become ends in themselves. However, robots are not moral agents in this sense, at least in the current state of technology (cf. Benossi & Bernecker, 2022; Chakraborty & Bhuyan, 2023). If so, Kantianism appears to leave little room for discussing current human-robot relationships beyond the dichotomy of "tools to be used" and "subjects who use them."

However, it is also true that in a society in which robots are expected to engage with humans as social beings, robots are becoming more than just tools for the sake of service. In this context, behavior toward robots has come to be considered to have certain moral implications. The argument referencing Kant's concept of *indirect duty*, which will be discussed in the next section, is considered to provide an important perspective on this point.

3 Duties Regarding Robots: Indirect Duties Argument

3.1 Overview of Indirect Duties Argument in Kant

Kant argued that only rational and autonomous beings deserve moral treatment. Therefore, in Kant's view, only persons can be the direct objects of duty, leaving no room for direct duties toward animals or machines. However, this does not mean that Kant's framework excludes duties regarding such entities. Since robots and other advanced technologies did not exist in the 18th century when Kant lived, he did not address the nature of human–robot relationships, but Kant argued that we have some duties regarding non-human beings, such as animals. Kant concluded that while we do not have direct duties to animals, we do have indirect duties with regard to them. Kant states the following:

> With regard to the animate but nonrational part of creation, violent and cruel treatment of animals is far more intimately opposed to a human being's duty to himself, and he has a duty to refrain from this; for it dulls his shared feeling of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other human beings. [⋯]– Even gratitude for the long service of an old horse or dog (just as if they were members of the household) belongs indirectly to a human being's duty

> with regard to these animals; considered as a direct duty, however, it is always only a duty of the human being to himself. (VI, 443)

Kant admitted that it is an indirect duty to not treat animals with unnecessary cruelty and to be thankful for them as a duty with regard to non-human beings. The main reason is that weakening morally serviceable feelings, such as sympathy, is a violation of one's duty to oneself. Kant claims here that, while there may appear to be a duty to the animal, it is a duty to oneself. Nevertheless, because treating an animal in a bad way violates one's duty to oneself, one should refrain from such cruel treatment as a means of fulfilling one's duty to oneself. We should consider non-human beings, such as animals, on this indirect basis because the weakening of moral feelings and the formation of a vicious character violate the duty we owe to ourselves. Thus, based on the indirect duty argument, moral consideration for animals is required, even though they are not direct objects of duty. In this sense, this argument can be understood as establishing an indirect duty regarding animals.

Now, the duty of virtue to oneself is the duty of perfection (cf. IV, 430; VI, 385). Perfection is the duty to cultivate all the faculties of the self, but moral perfection is especially relevant here. Moral perfection means having a purely virtuous disposition that enables one to fulfill one's duties as motivated by moral law (cf. VI, 387, 392), and striving toward perfection is a duty of virtue to oneself. One of the processes of this striving involves the cultivation of moral feelings, which are given to human beings as a natural predisposition. Therefore, since it is a violation of one's duty to oneself to undermine one's moral feelings, one must refrain from acts that are likely to do so–that is, mistreatment of animals. This argument is supported by some Kantians who are interested in animal welfare (cf. Denis, 2000; Altman, 2011).

Several authors have referred to the indirect duties argument and argued that there could also be duties regarding robots. Darling, for example, argues that an analogy with animals is useful for considering the moral and legal status of AI/robots, focusing on human emotional responses to robots in empirical cases and referring to Kant's indirect duties (Darling, 2016, 2021). Furthermore, Gerdes, Coeckelbergh, and others suggest a continuum of moral duties that logically extends from animals to robots, proposing that if indirect duties apply to animals owing to their impact on our moral character, then similarly, robots should also be encompassed within this moral consideration (Gerdes, 2016, pp. 276–278; Coeckelbergh, 2021, p. 343).[9] According to their view, if we accept the duty with regard to animals, then, by the same argument, there is also a duty with regard to robots. The application of the duty toward oneself to cultivate one's virtue means that even if robots are not the direct object of the duty, they should be treated morally because treatment of robots influences human moral feelings and virtue. This is based on the notion that interactions with robots, particularly those designed to evoke emotional responses, can influence human virtue, and therefore that they should be treated morally. The implication of this is that we should

[9] Others, such as Anderson (2011), Sparrow (2021), and Smith (2021), cite Kant, noting the need to refrain from cruel treatment of robots because of its effect on human virtue.

avoid bad behaviors, such as destroying robots by inflicting violence on them. This argument can be summarized as follows:[10]

1. If we treat robots badly, we will undermine our moral feelings and fail to strive for moral perfection (or to maintain a good moral character).
2. We have a duty to strive for moral perfection (the duty of virtue to oneself).
3. Therefore, we should avoid treating robots badly.

3.2 Critique of Indirect Duties Argument in Kant

This idea has a certain appeal in that it guides ethical behavior toward robots without dealing with such difficulties as whether the robot is conscious. However, the practical implications for the ethics of human–robot interaction derived from this argument seem to be limited. This is because what we can derive from this is that if we treat the robot cruelly, it can damage our capacity for sympathy and develop a vicious character in us, and therefore, we should refrain from doing so.[11] In other words, as long as one's capacity for sympathy is preserved, there is no need to actively treat robots in a morally considerate way. Alternatively, this approach provides only a limited norm: refraining from bad behavior toward robots to avoid negatively impacting one's sympathy or virtuous character formation. As Gordon puts it, the great weakness of this argument is that "the object of morality itself is not given a moral claim" (Gordon, 2020a, p. 217).

As Flattery correctly points out, "the robot itself has no moral status," which is the premise of Kant's indirect duty argument, so there is no problem with the argument itself. However, it is undeniable that the practical implications derived from it are limited. If our treatment of robots does not affect our moral character, then it does not matter what we do to them. Nor does it seem convincing to consider animal abuse and putative robot abuse as being at the same level (Johnson & Verdicchio, 2018). This is because, for example, animals really suffer, but robots do not (although they can sometimes mimic suffering). At the very least, it is questionable to treat animals, sentient beings that feel pain, in the same way as machines, which are inanimate objects that do not feel anything. This is an empirical matter and an open question, but it could make a crucial difference in the impact on our moral feelings.

[10] Note that this argument owes a lot to Flattery, 2023. The original "Indirect Robots Argument" by him is as follows.

If we regularly treat sentient-like robots badly, then we become more likely to treat humans badly.

We ought to avoid doing things that would make us more likely to treat humans badly.

So, we ought to avoid treating sentient-like robots badly.

This argument focuses on avoiding treating humans badly, whereas Kant's argument focuses on avoiding undermining the moral feelings that are part of one's moral character. Therefore, premise 2 should be changed to focus on one's duty to oneself, since it is not a matter of treating people badly.

[11] This argument is sometimes thought of as an empirical assertion that mistreating animals or robots will form characters that will harm others, but this is not accurate. The issue is the duty to the self to develop feelings appropriate to a virtuous character, and we need not wait for empirical evidence as to whether it leads to cruelty to others (cf. Ripstein & Tenenbaum, 2020).

Nevertheless, it is also important to note that recent empirical research casts doubt on the assumption that robots are perceived merely as inanimate objects. While robots, as artificial entities, do not possess consciousness or sentience in any robust empirical or philosophical sense, recent studies have shown that people may still perceive them as capable of being harmed. For example, Lima et al. (2020) report public support for protecting robots from cruel treatment, based not on their inner capacities but on how they are perceived. This highlights the role of human perception in shaping moral attitudes toward robots. These findings suggest that people can react emotionally even to artificial beings, which may complicate a strict distinction between animals and robots in terms of their impact on moral feelings. Still, we cannot straightforwardly assume that the emotional effects of our treatment toward robots and animals are fully parallel. This remains an empirical question beyond the scope of this paper. What matters here, however, is not the empirical similarity or difference, but the normative limitation of the Kantian indirect duty argument itself.

Due to this limited implication, Kantian indirect argument cannot make the positive claim that robots are important to humans and that this relationship should be considered. Since the indirect duty is derived from the duty to oneself, it is sufficient if one's character is not negatively affected. If one treats robots carefully as a thing, it fulfills its duty. This argument does not require any kind of positive moral treatment of robots, nor does it require admitting the intimate relationship between humans and robots. In other words, the current Kantian argument considers the role of a robot's relationship with humans as either an instrumental means (to achieve a particular end) or a moral means (to cultivate one's moral character).

Does Kantianism address the importance of robot relationships for humans only in terms of whether the robot is useful as a means to an end for humans? Even if we do not recognize moral status or dignity as values for robots, Kant's indirect duty argument still plausibly prohibits bad treatment of robots. However, the implication of this norm is limited to preserving moral feelings such as sympathy, which support the formation of virtuous character; it does not extend to requiring positive moral treatment of a robot, even if that robot is considered an important other in a person's life. At the very least, this conclusion is not pleasing to some people who want to build a relationship with a robot as a loving and intimate partner. Regardless of how important and valuable a human–robot relationship may be to an individual, the indirect duty argument cannot fully address whether such a relationship should be morally protected. However, the ethics of human–robot interaction should attempt to provide a normative answer to the question of how we should treat robots and other machines, or what kind of attitude we should have toward human–robot relationships. Is it possible to derive a more positive norm from Kantianism? The answer is as follows: By considering the human–robot relationship as being essential to human happiness and focusing on our duty of virtue to others, it is possible to derive the duty to protect human–robot relationships.

4 The Duty to Protect the Human–Robot Relationship

In this section, I focus on the possibility of robots becoming significant beings for humans, rather than mere means to an end. Ultimately, I argue that while robots are not ends in themselves and thus not direct objects of moral duties, we can derive an ethical duty to protect human–robot relationships through Kant's duty to promote the subjective happiness of others. As mentioned in the introduction, some people today regard robots and other artificial beings not as mere means to an end but as beings with whom they can establish an intimate relationship. In other words, some people live happy lives because of their relationships with robots or with artificial companions, and they choose that lifestyle for themselves. For example, Tomomi Ota's *Robot Friendly Project* envisions a society that supports people who find happiness in living with robots, illustrating that such relationships can be meaningful and fulfilling[12]. If this is the case, should we not adopt some moral attitude toward lives of people who achieve happiness through relationship with a robot? In other words, shouldn't the happy relationship between humans and robots be protected? This moral norm would be required regardless of the impact of our behavior toward the robot on the cultivation of virtue or vice and the fact that robots are artificial entities, or so I will now argue. Indeed, this paper argues that we can derive that duty from Kantianism. Specifically, I focus on Kant's category of the duty of virtue to others to develop this argument.

First, we confirm what Kant's duty of virtue to others is. The duty of virtue is described as "an end that is at the same time one's duty" (VI, 382), an obligation that commands the end of the act to be adopted. We call it an "obligatory end." Kant identifies two obligatory ends: "the perfection of the self" and "the happiness of others" (VI, 382). The former corresponds to the "duty to cultivate one's moral virtue" discussed in the previous section, the latter to the duty of virtue toward others, which requires, as a duty, that "promotion of the happiness of others" be adopted as an obligatory end. In other words, it is not a command to action to help others, but a duty related to the end of the action, "to adopt the happiness of others as one's end" (cf. VI, 391–393).

The duty of virtue toward others includes a duty of moral concern for others.[13] It is a duty that we should be concerned about and aim to promote the happiness of others. The duty can be formulated as stating that one should adopt to promote the happiness of others as one's end and perform beneficent acts accordingly. Kant defines it as follows: "To be beneficent, that is, to promote according to one's means the happiness

[12] See the *Robot Friendly Project* website: https://robot-friendly.com/ (in Japanese). The website describes real-life examples of people living with robots, including robots assisting individuals with limited mobility by engaging in activities on their behalf, such as enjoying dining out. It also highlights robots that, after living with humans for many years, form familial relationships, sharing meals and fostering communication around the same table.

[13] Although there is no moral category in Kant's theory that directly corresponds to *moral concern*, it can be interpreted as corresponding to a duty of virtue concerning the happiness of others (cf. Müller, 2022, Chaps. 2, 6). For example, there is an argument that one should respect others as persons of dignity and ends in themselves, but respect and concern are two different attitudes. Note that respect is not discussed in depth in this paper.

of other human beings in need, without hoping for something in return, is everyone's duty" (VI, 452). Kant also calls this the "duty of love," which he defines as the "duty to make others' ends my own (provided only that these are not immoral)" (VI, 450).

Since the end to be adopted as a duty here is the subjective happiness that others naturally wish and seek, it requires that we sincerely consider what others perceive as happiness and, provided it is not immoral, support them in realizing it. Kant considers the happiness that should be promoted by the duty of virtue to be subject-relative (cf. Ginther, 2022, pp. 415–417). Kant's conception of happiness is complex, encompassing both natural happiness and the moral happiness associated with the highest good. However, in the context of the duty of virtue toward others, the focus is on an individual's natural happiness. In other words, happiness is an empirical and subjective satisfaction that human nature inevitably seeks, and "it is for them to decide what they count as belonging to their happiness" (VI, 388). In other words, what brings about happiness is an empirical matter, and each subject decides for themselves. Therefore, the happiness at issue is solely what each subject seeks. For this reason, Kant clearly states: "I cannot do good to anyone in accordance with my concepts of happiness, thinking to benefit him by forcing a gift upon him; rather, I can benefit him only in accordance with his concepts of happiness" (VI, 454). However, happiness that includes immoral ends, such as pseudo-pleasure through excessive inebriation or pleasure obtained by doing violence to others, is not included in the happiness that should be of concern.

Now, let us apply this argument to the context of human–robot relationship. Suppose that there is a person attached to a robot and finds an intimate relationship with it that improves the happiness of their life. What we should do for that person is recognize the relationship as valuable and show moral concern for the person seeking happiness through their relationship with the robot. Suppose that you are a virtuous agent who adopts the promotion of the happiness of others as an obligatory end. In such cases, we should respect the person and their subjective choice of happiness, including their life with the robot, through which they achieve happiness. This is because, according to Kant's duty of virtue, we should have concern for the happiness that others themselves determine and seek as an end; therefore, we should act with the end of promoting happiness for that person. If the human–robot relationship contributes to the happiness of the person, then treating the relationship as morally important and having concern for its happiness should fulfill the obligatory end. In other words, if the relationship with the robot represents a meaningful and personally chosen source of happiness, choosing actions that protect that happy relationship fulfills our duty to promote the happiness of others as they themselves define it.

However, there may be criticisms that the happiness derived from a relationship with a robot is inauthentic, and that a happy human–robot relationship is not something that deserves protection. As previously mentioned, if happiness is immoral or unethical, then it is not happiness that should be promoted. For example, critics argue that forming intimate bonds with robots could encourage self-deceit and unhealthy attachment due to being deceived by the robot, or could divert emotional investment from human relationships (cf. Misselhorn, 2021; Sparrow, 2002; Turkle, 2011; Wallach & Allen, 2009). This criticism could be rephrased as the claim that happiness derived from a human–robot relationship is unhealthy and ethically undesirable, and

therefore, there is no duty to promote such happiness. Consequently, it is necessary to demonstrate that the happiness gained through a relationship with a robot is not immoral in this sense. Indeed, robots do not have consciousness or genuine emotional responses; therefore, the relationship differs from the intimate love relationship between humans. As some scholars argue, what makes intimate relationships of friendship or love possible between humans is genuine reciprocity, which presupposes an inner life, such as a mind, in each party (Elder, 2017; Nyholm & Frank, 2017; Nyholm, 2020)[14].

Certainly, there can be cases in which people obtain self-deceptive happiness through their relationships with robots. However, this is not so of all cases. As long as the individual is aware that the robot is not a conscious being and still finds meaningful happiness in the interaction, this relationship can be viewed as ethically permissible. This happiness does not necessarily stem from self-deception but from the subjective experience of being with a robot. Admittedly, this claim may still be subject to skepticism; however, it would also be premature to conclude that all such happiness is deluded or rooted solely in self-deception. Even if relationships with robots are not commonly or predominantly sources of meaningful happiness for humans, as long as it is conceivable that such relationships can subjectively provide happiness for certain individuals, we have a reason to consider an ethical attitude toward them. Furthermore, promoting happiness through human–robot relationships does not necessarily imply neglecting human relationships. Individuals can maintain fulfilling relationships with both humans and robots, each serving a different emotional requirement.

At this point, it is worth briefly addressing utilitarianism, as the principle of promoting happiness is more commonly associated with utilitarian ethics. Utilitarian ethics might seem to reach the same conclusion more straightforwardly as the argument presented here. However, Kant's duty to promote happiness differs from utilitarian ethics in its focus on happiness according to the individual's own concept. Unlike utilitarianism, which seeks the maximization of overall happiness, Kantian principle emphasizes respecting each individual's own concept on what constitutes their happiness. According to Kant, people usually know better what makes them happy than you do (VI, 454). Therefore, Kantian principle encourages not to judge the happiness of others arbitrarily, but to consider humbly the happiness they choose for themselves. Therefore, Kant's duty to promote the happiness of others avoids paternalistically dismissing happiness derived from relationships with robots as inauthentic. Instead, it encourages us to respect the choices individuals make to pursue their own happiness, provided these choices do not involve immoral ends. For example, while utilitarianism might deprioritize the ethical significance of human–robot relationships if there were greater opportunities to promote overall happiness elsewhere, Kant's framework obliges us to take seriously the happiness of individuals

[14] This view of reciprocity as dependent on inner mental states can be broadly characterized as reflecting a Kantian and Western framework. However, it is worth noting that alternative cultural perspectives, such as those described by Indigenous scholar Kimmerer (2015), recognize forms of morally significant reciprocity between humans and nature that do not rely on such assumptions. While this paper remains within the Kantian ethical tradition, acknowledging these diverse views invites further dialogue across philosophical and cultural traditions about the moral significance of relationality.

who have chosen to engage meaningfully with robots. Even if such relationships are not common, they remain worthy of moral concern because they represent the individual's decision about their happiness.

Therefore, if the relationship with a robot significantly contributes to a person's happiness that the individual has chosen for themselves, we have a duty to protect that relationship, aiming to promote that individual's happiness. This duty can be formulated as showing moral concern for the human–robot relationship. This argument can be summarized as follows:

1. We should promote others' happiness according to their concepts, not ours.
2. People have meaningful, happiness-conducive/defining relationships to certain robots.
3. Therefore, to the extent that people have happiness-conducive/defining relationships with robots, we should protect happy relationships with robots.

This conclusion does not imply the following. First, it does not necessarily imply a norm that all human–robot relationships should be protected with moral concern. For example, suppose that a person who owns a companion robot does not realize happiness by interacting with it. In this case, there is no direct duty to show moral concern for that relationship and protect it. This is because the relationship would not be contributing to the person's happiness, and thus, promoting it would not align with the duty to support the happiness that others have chosen for themselves. We should only have moral concern for the person who has subjectively chosen to pursue happiness through that relationship. In other words, we must deliberate on what is essential to the person's self-determined pursuit of happiness. Second, it does not support the claim that robots are independently worthy of moral consideration. The Kantian argument focuses on moral duties toward others, emphasizing the moral importance of relationships based on the happiness that humans actively seek and define through their interactions with robots. The normative claim is that if robots are important to human happiness, then it is our moral duty to protect that relationship without undermining it.

This Kantian perspective resonates with several non-Kantian approaches that emphasize the ethical significance of human relationships with non-human entities. For example, in the context of care ethics, de la Bellacasa (2017) argues in *Matters of Care* that caring for others is itself a virtuous practice that does not necessarily depend on the moral status of the recipient. In the context of animal rights, Kelch (1999) explores how animal rights could be grounded partly in emotional factors such as happiness, not solely in rationality. Drawing on both animal rights theory and the rights of nature, Gellers (2020) offers a unique perspective on how robots might deserve moral standing, not because of intrinsic properties, but due to their meaningful participation in social and ecological relationships, which may contribute to the flourishing of the environment. These accounts complement the Kantian view developed here by supporting the broader normative claim that meaningful relationships can be ethically relevant and even worthy of protection, independently of the moral properties of the non-human entity involved.

In summary, this argument suggests that the human–robot relationship should be protected based not on the moral status of the robots but on the moral importance of human happiness. This Kantian argument for protecting human–robot relationships also finds support in broader discussions within robot ethics. For instance, Nyholm (2023, Chap. 9), drawing on Munn and Weijers (2023), argues against harming technologies that individuals relate to meaningfully. He also cites Loh (2019), who calls for a respectful attitude toward those who have meaningful relationships with technologies. These perspectives reinforce the broader idea that human–robot relationships can carry ethical significance by contributing to human flourishing, independently of the robots' own moral status.

A similar idea appears in Mamak's attempt to legally protect love relationships with robots (Mamak, 2024). Mamak argues that robots that have significant relationships with people should be protected based on the attachment humans have toward them, even if the robots are not moral patients. He presents the following scenario: A woman named Inga has been living with a robot named Otto for 40 years, treating the robot as her partner. She understands that the robot has no inner life but has built a meaningful relationship with it. Her friends and family know that she treats her robot as a partner, and they also treat him as her partner. One day, a guest attacks Otto in front of Inga (cf. Mamak, 2024, pp. 573–574). In this case, in what sense should the guest who attacked Otto be condemned? Is it merely for damaging someone else's property? Mamak argues that a simple charge of property damage is insufficient, and he proposes a legal framework that protects human–robot relationships based on the value of a loving relationship, rather than the object's value as property (p. 579).[15] Mamak's proposal is detached from the discussion of the moral status of robots, and instead demands protection of robots that have a concrete relationship with humans, based on the attachment that humans have for them (p. 576).

Kantianism for the ethics of human-robot interactions, which focuses on the duty to promote the happiness of others, as presented in this paper, regards the protection of such relationships as a moral duty. In the case of Inga and Otto, we have a duty to promote Inga's happiness as she has chosen her relationship with Otto as a source of happiness. If her relationship with Otto enriches her happiness, then protecting that relationship fulfills our obligatory end. Therefore, the guest who attacked Otto is morally condemned for violating his duty to others. On the other hand, her family and friends, who know that Inga treats the robot as a partner and respects that relationship, are to be evaluated for fulfilling their duty to promote her happiness.

What then exactly should we do to fulfill our duty to protect such happy relationships? Kant states that the specific actions cannot be determined because there is "room for choice" in the actions that serve obligatory ends (VI, 390). The duty only requires that one promote happiness for others as the end of one's actions. Hence, this is also called a wide duty, in the sense that it gives the agent room for choice of action (ibid.). However, it can be said that there is an attitude that we should have towards the happy relationship between humans and robots. For example, respecting a person's way of life that promotes their happiness through a relationship with a

[15] Mamak's proposition is as follows: § 1. Whoever destroys a robot that is a domestic partner with a person is subject to a penalty. § 2. The crime provided for in § 1 is prosecuted upon the harmed party's motion.

robot, and recognizing that the robot has an important place in their life. At the very least, if the robot is important to a person and they have formed a happy relationship with it, then making stereotypical statements that deny it is a violation of duty. It is not, through this Kantian lens, an indirect violation of duty, a means of avoiding the undermining of one's virtue, but as a direct violation of duty, as an act of refusal to promote the happiness of others. Therefore, moral consideration for a person who seeks happiness through a relationship with a robot is a duty, and in this sense, happy human–robot relationships should be protected.

Finally, it should be noted that the object of this direct duty is the human other who wishes and seeks happiness, not robots. The object of moral concern within the framework of Kantianism is the being whose intrinsic impulse is to seek happiness for itself. Hence, this argument retains a human-oriented perspective. However, the argument that we should support each subject in achieving what they seek as happiness, can dictate our behavior toward robots. Suppose a robot has become an integral part of a person's happiness. In that case, we should show some moral concern for a robot in order to protect that relationship because we are committed to the obligatory end of promoting that person's happiness. In this sense, Kantianism can offer a human-oriented approach to duty that ethically requires the protection of human–robot relationships, providing a normative basis for considering the ethics of human–robot interactions.

Kantianism for the ethics of human–robot interaction argues not only for the indirect duty argument, which treats robots as a means of preserving one's own moral character, but also advocates for the duty to protect human–robot relationships. Kantianism, added in this paper, does not argue that robots are worthy of moral consideration, but rather advocates such an ethical norm based on the duty to promote others' happiness. The basis for ethical duties toward the human–robot relationship is the obligatory end of others' happiness. This Kantianism can extend to those who think that robots are just machines and deny their moral importance, refusing to grant them moral status. It is practically persuasive in that it is sufficient to accept the duty to adopt the obligatory end of promoting others' happiness, specifically their subjective-relative happiness.

There may still be the impression that traditional ethical frameworks, such as Kant's, are no longer valid for the ethics of human-robot interaction. However, by mining lesser-known features from the traditional ethical theory of Kantianism, we can provide a theoretical framework for considering the human–robot relationship. As Gordon, for example, puts it, previous applications of traditional theory have had in mind only a superficial understanding of ethical theory due to "rookie mistakes" (Gordon, 2020b). What this paper has presented here is a more sophisticated Kantianism. We need to recognize that Kantianism is not an old and useless theory for the ethics of human–robot interaction but rather provides us with a rich resource for coexisting with robots and other types of artificial entities.

5 Conclusion

Kantianism, indeed, only sees robots as things. Robots are not valuable beings in their own right, worthy of moral consideration. Previous arguments for Kantianism in robot ethics have focused only on constraining the bad treatment of robots to maintain one's own good moral character. This argument, derived from the duty to oneself, carries a certain degree of persuasiveness, but its implications are limited in that it cannot actively prescribe a moral attitude toward human–robot relationships. This paper presents another Kantian argument by applying the duty to promote the happiness of others to the context of the ethics of human-robot interaction. According to this view, we can establish a duty to protect happy human–robot relationships based on concern for the happiness of others. An agent who adopts and acts with the obligatory end of promoting the happiness of others should also be willing to show concern for a robot that is an integral part of a person's happiness and to protect that relationship. This duty will become increasingly important as future technological developments increase the likelihood that humans will flourish through their relationships with robots.

Acknowledgements I am deeply grateful to Sven Nyholm (LMU Munich), Nico Dario Müller (University of Basel), Robert Sparrow (Monash University), Ian Robertson (Universität Erlangen-Nürnberg), Masashi Takeshita (Hokkaido University), and Kengo Miyazono (Hokkaido University) for their valuable comments and suggestions on earlier drafts of this paper.

Author Contributions This paper was written solely by the author.

Funding This work was supported by JSPS KAKENHI Grant Number JP24KJ0265.

Data Availability This study does not involve any datasets.

Declarations

Ethics Approval This study did not involve human participants or animals, and therefore did not require ethical approval.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author has no relevant financial or non-financial interests to disclose.

References

Altman, M. C. (2011). *Kant and applied ethics: The uses and limits of kant's practical philosophy*. Wiley-Blackwell.

Anderson, S. L. (2011). The unacceptability of Asimov's three laws of robotics as a basis for machine ethics. In M. Anderson & S. L. Anderson (Eds.), Machine ethics. Cambridge University Press.

Beck, J. (2013). Married to a doll: Why one man advocates synthetic love. *The Atlantic*https://www.theatlantic.com/health/archive/2013/09/married-to-a-doll-why-one-man-advocates-synthetic-love/279361/

Benossi, L., & Bernecker, S. (2022). A Kantian perspective on robot ethics. In H. Kim & D. Schönecker (Eds.), *Kant and artificial intelligence* (pp. 145–168). De Gruyter. https://doi.org/10.1515/9783110706611-005

Bryson, J. (2010). Robots should be slaves. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 63–74). John Benjamins Publishing Company.

Bryson, J. (2018). Patiency is not a virtue: The design of intelligent systems and systems of ethics. *Ethics and Information Technology*, *20*, 15–26.

Chakraborty, A., & Bhuyan, N. (2023). Can artificial intelligence be a Kantian moral agent? On moral autonomy of AI system. *AI and Ethics*, *4*, 1–7. https://doi.org/10.1007/s43681-023-00269-6

Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-cartesian moral hermeneutics. *Philosophy & Technology*, *27*(1), 61–77.

Coeckelbergh, M. (2021). Should we treat teddy bear 2.0 as a Kantian dog? Four arguments for the indirect moral standing of personal social robots, with implications for thinking about animals and humans. *Minds & Machines*, *31*, 337–360.

Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, *3*(1), 5–24.

Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*, *26*, 2023–2049.

Darling, K. (2016). Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects, In R. Calo, A. M. Froomkin, & I. Kerr (Eds.), Robot law (pp. 213–232). Edward Elgar.

Darling, K. (2021). *The new breed: What our history with animals reveals about our future with robots*. Henry Holt.

de la Bellacasa, M. P. (2017). *Matters of care: Speculative ethics in more than human worlds*. University of Minnesota.

DeGrazia, D. (2022). Robots with moral status? *Perspectives in Biology and Medicine*, *65*(1), 73–88.

Denis, L. (2000). Kant's conception of duties regarding animals: Reconstruction and reconsideration. *History of Philosophy Quarterly*, *17*(4), 405–423.

Elder, A. (2017). *Friendships, robots, and social media*. Routledge.

Flattery, T. (2023). The Kant-inspired indirect argument for non-sentient robot rights. *AI Ethics*. https://doi.org/10.1007/s43681-023-00304-6. Online publication date: 05-July-2023.

Gellers, J. C. (2020). *Rights for robots: Artificial intelligence, animal, and environmental law*. Routledge.

Gerdes, A. (2016). The issue of moral consideration in robot ethics. *Acm Sigcas Computers and Society*, *45*(3), 274–279.

Ginther, L. (2022). Kantian objectivism and subject-relative well-being. *Dialogue*, *61*(3), 407–419.

Gordon, J. S. (2020a). What do we owe to intelligent robots? *AI & Society*, *35*, 209–223.

Gordon, J. S. (2020b). Building moral robots: Ethical pitfalls and challenges. *Science and Engineering Ethics*, *26*(1), 141–157.

Gordon, J-S., & Nyholm, S. (2021). Ethics of artificial intelligence. *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/ethic-ai/

Gordon, J. S., & Nyholm, S. (2022). Kantianism and the problem of child sex robots. *Journal of Applied Philosophy*, *39*(1), 132–147.

Gunkel, D. (2018). *Robot rights*. The MIT Press.

Johnson, D. G., & Verdicchio, M. (2018). Why robots should not be treated like animals. *Ethics and Information Technology*, *20*, 291–301.

Kant, I. (1908). *Kants gesammelte schriften*. Hrsg. von der Königlichen Preußischen Akademie der Wissenschaften und Nachfolgern.

Kelch, T. (1999). The role of the rational and the emotive in a theory of animal rights. *Boston College Environmental Affairs Law Review*, 27(1). Available at SSRN: https://ssrn.com/abstract=3490296

Kimmerer, R. W. (2015). *Braiding sweetgrass*. Milkweed Editions.
Korsgaard, C. (2018). *Fellow creatures: Our obligations to the other animals*. Oxford University Press.
Levy, D. (2008). *Love and sex with robots: The evolution of human–robot relationships*. Harper Perennial.
Lima, G., Kim, C., Ryu, S., Jeon, C., & Cha, M. (2020). Collecting the public perception of AI and robot rights. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1351–13524. https://doi.org/10.1145/3415206
Loh, J. (2019). *Roboterethik: Eine einführung*. Suhrkamp.
Mamak, K. (2024). Should criminal law protect love relation with robots? *AI & Society*, *39*(3), 573–582.
Misselhorn, C. (2021). *Künstliche Intelligenz und Empathie: Vom Leben mit Emotionserkennung, Sexrobotern & Co.* Reclam.
Mosakas, K. (2021). On the moral status of social robots: Considering the consciousness criterion. *AI and Society*, *36*(2), 429–443.
Müller, V. C. (2021). Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology*, *23*, 579–587.
Müller, N. D. (2022). *Kantianism for animals: A radical Kantian animal ethic*. Palgrave Macmillan.
Munn, N., & Weijers, D. (2023). Corporate responsibility for the termination of digital friends. *AI & Society*, *38*, 1501–1502. https://doi.org/10.1007/s00146-021-01276-z
Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield International.
Nyholm, S. (2021a). The ethics of Human–robot interaction and traditional moral theories. In C. Véliz (Ed.), *Oxford handbook of digital ethics, Oxford handbooks online*. Oxford Academic.
Nyholm, S. (2023). *This is technology ethics: An introduction*. Wiley.
Nyholm, S., & Frank, L. (2017). From sex robots to love robots: Is mutual love with a robot possible? In J. Danaher, & N. McArthur (Eds.), *Robot sex: Social and ethical implications* (pp. 219–244). The MIT Press.
Nyholm, S., Friedman, C., Dale, M. T., Puzio, A., Babushkina, D., Lohr, G., Kamphorst, B., Gwagwa, A., & IJsselsteijn, W. (2023). Social robots and society. In van de I. Poel (Ed.), *Ethics of socially disruptive technologies: An introduction* (pp. 53–82). Open Book.
Petter Bae Brandtzaeg, P., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their Human–AI friendship. *Human Communication Research*, *48*(3), 404–429.
Ripstein, A., & Tenenbaum, S. (2020). Directionality and virtuous ends. In J. J. Callanan, & L. Allais (Eds.), *Kant and animals* (pp. 139–156). Oxford University Press.
Ryland, H. (2021). It's friendship, jim, but not as we know it: A Degrees-of-Friendship view of Human–Robot friendships. *Minds & Machines*, *31*, 377–393.
Smith, J. K. (2021). *Robotic persons: Our future with social robots*. Westbow.
Sparrow, R. (2002). The March of the robot dogs. *Ethics and Information Technology*, *4*(4), 305–318. https://doi.org/10.1023/A:1021341213157
Sparrow, R. (2021). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics*, *13*, 23–29.
Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
Vogel, S. (2015). *Thinking like a mall: Environmental philosophy after the end of nature*. MIT Press.
Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford University Press.