Artificial Emotions and the Evolving Moral Status of Social Robots

Arianna Sica
Computer Science and Communication
Østfold University College
Halden, Norway
ariannas@hiof.no

Henrik S. Sætra
Informatics
University of Oslo
Oslo, Norway
henrsae@ifi.uio.no

ABSTRACT

This article aims to explore the potential impact of artificial emotional intelligence (AEI) on the ethical standing of social robots. By examining how AEI interacts with and potentially reshapes the two dominant perspectives on robots' moral status, namely the property-oriented approach and the social-relational approach, we aim to offer fresh insights into this pressing dilemma. Our analysis reveals that although the incorporation of AEI does not conclusively confer moral status to current social robots, it might challenge the boundaries that separate robots from other entities customarily considered to have more status, thereby increasing the complexity of the debate.

CCS CONCEPTS

• Human-centered computing → HCI theory, concepts and models; • Computer systems organization → Robotics; • Computing methodologies → Philosophical/theoretical foundations of artificial intelligence.

KEYWORDS

Moral status; Artificial Emotional Intelligence; Social robots; Emotions; Consciousness; Sentience; Reason; Relational turn.

ACM Reference format:

Arianna Sica and Henrik Skaug Sætra. 2024. Artificial Emotions and the Evolving Moral Status of Social Robots. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3610977.3634934

HRI '24, March 11–14, 2024, Boulder, CO, USA.

ACM ISBN 979-8-4007-0322-5/24/03.
https://doi.org/10.1145/3610977.3634934

1 INTRODUCTION

With the rapidly shifting landscape of technology, an intensified debate about the rights and moral standing of robots has emerged [1-3]. These questions have added a fresh dimension to our understanding of our own nature and morality itself, driven in part by how the status of machines is largely examined through explorations of what gives humans, and animals, moral standing [4, 5]. With exponential strides in the fields of robotics and artificial intelligence, the debate surrounding robots' moral status has become part of the mainstream of ethical discussions. For instance, Blake Lemoine, a former engineer at Google, ignited discussions when it suggested that their AI chatbot generator was conscious[1].

One key tenet of this discourse is the exploration of social robots – autonomous, physically embodied machines, designed to interact socially with humans [6]. Various iterations of social robots have already permeated the society – from healthcare to education and companionship [7, 8] – and it is expected that their presence, utility, and capabilities will see substantial growth in the future [9].

The moral status of social robots shape individuals' interactions with, and societies' application and regulations of, them. When the technology changes, so could the arguments in favour of – or against – granting them some moral status. This article aims to delve into this pressing dilemma, by exploring whether *artificial emotional intelligence* (AEI) offers a fresh lens through which we can evaluate and possibly redefine the ethical standing of social robots. Specifically, we aim to understand how AEI interacts with and possibly reshapes two dominant perspectives that influence the contemporary debate of robots' moral status: the property-oriented approach and the relational approach [2].

The approach of considering AEI as a possible key determinant for the moral status of social robots has been partially covered [10] but still leaves much to be comprehensively understood and investigated. Through this analysis, we hope to shed light on the evolving discussion surrounding the moral standing of robots in our rapidly progressing tech-driven world. Furthermore, exploring the moral landscape in a world where technology and humanity increasingly converge serves as a mirror to our evolving understanding of morality itself, pushing us to question and redefine long-held beliefs about consciousness, the concept of person and human, and the essence of moral value and ethical

[1] https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

consideration [11].

The article is structured as follows. In section 2, we explore the concept of moral status and the kind of machines addressed, namely social robots equipped with AEI. In section 3, we describe two dominant theories for ascribing moral status – the property-based and the social-relational approach. The aim of this section is to evaluate the potential implications of integrating AEI into social robots on the prevailing theories. Specifically, it seeks to discern whether AEI challenges or consolidates the frameworks of these theories and assesses if it offers adequate grounds for granting moral consideration to robots. We conclude in section 4 with final reflections on the topics explored.

2 BACKGROUND

Before we ask how our understanding of the moral status of machines changes with AEI, we must establish what sorts of machines we focus on. Even if our analysis will provide insight into the more general question of machines' moral standing, we arrive at this through a specific focus on social robots with AEI.

2.1 Social robots with AEI

A social robot is a physically autonomous entity designed to interact with humans on a social and emotional level, emulating human or animal-like behaviours and mindsets, and learning from these interactions [6]. While humanoid social robots might seem particularly likely to evoke reactions resembling those evoked by other human beings, we have chosen to focus on social robots in general. This is because we start from the premise that a social robot's *form* will not be the key determinant of its moral consideration. Supporting this view, various ethical perspectives claim that moral status should be granted to a range of non-human animals [12-14], suggesting that human form is not central for the ascription of moral status.

The objective of social robotics is crafting machines that are perceived as more than mere instruments – to elevate these machines to genuine interaction partners [15]. Such a goal seems to become more likely with the incorporation of AEI. AEI, as defined by Schuller and Schuller [16], equips technology with the capacity to discern, generate, and utilise emotions for problem-solving and goal achievement, in order to create machines that are genuinely intelligent and able to interact with us seamlessly and authentically [17]. More specifically, AEI can detect and recognise others' emotions by analysing emotion-related data (*emotion recognition*). It can express emotion towards interlocutors through facial and vocal expressions, head position, gestures, and body language (*emotion generation*). Finally, it can determine behaviour and activity selection based on internal simulated states (*emotion augmentation*) [16].

We also stress that AEI technology is still in development. Despite advancements in various related fields [18] and notable progress in social robots such as Pepper [19], Buddy [20] and Ryan [21], fully equipping robots with human-like empathic abilities remains a challenge [22]. Given the limitations of its current applications, our exploration of AEI in this article is largely theoretical, assuming a future where AEI is fully realised as per the aforementioned definition.

The design of social robots equipped with AEI with the express intention of evoking various feelings in human users, and to simulate social cues, differentiates them from other robots that may unintentionally elicit human emotional reactions [23]. One example of the latter would be how some individuals grow attached to robotic devices such as the vacuum cleaner Roomba [24]. Given the machine's functions, the human or animal likeness of their interactions, and the potential attachment and feelings they aim to evoke in users, it has been argued that social robots inhabit a grey area between machines and sentient beings [25-27], providing reason to question whether and how we should attribute moral standing to them, and distinguish them from other technological entities.

2.2 Moral status

This subsection delves into the concept of moral status, a necessary and yet far to be complete exploration given the scope and space constraints of this article. To begin with, we refer to Warren's [28] comprehensive and intuitive definition of moral status, described as a delineation of entities to which we have moral duties. Entities with moral status, also called moral patients, are seen as deserving of moral consideration based on their needs, interests, and wellbeing, regardless of benefits or disadvantages we might derive from them. Some authors equate moral status with having moral rights [29]. While there is debate of this equivalence, both perspectives agree that having moral status, or being morally considerable, imposes obligations on moral agents. Moral agents are those from whom we expect adherence to morality and the respect of others' rights.

Moral status, unlike legal status, is and will remain something individuals determine based on a wide variety of considerations. It consequently remains subjective. Legal status, on the other hand, can be granted by policymakers and would apply to all in a given legal context, regardless of their opinions and perception of the machines. While a philosophical treatise about the moral status of social robots with AEI could persuade some to change their opinion of these machines, we do not presume to adjudicate these questions for all. Changes of opinion could change social norms, however. Furthermore, dominant subjective perceptions of moral status tend to inform and change our legislation, which means that debates about robot's moral status can, but need not, change their legal status. In some instances, legal status is given purely for practical reasons, such as a delivery robots' right of way as a necessity for performing a function. A totally different case would be to grant robots some legal protection against, for example, abuse based not on protecting someone's property but because they become perceived as entities with rights to privacy and not to be subject to suffering. The latter case illustrates how debates about robots' moral standing overlaps with environmental ethics and animal rights discourse [4, 5]. We focus exclusively on moral standing in this paper.

3 TWO PATHS TO ROBOTS' MORAL STATUS

The question of what moral standing machines have is often explored through two dominant perspectives: the property-based and the relational approach to morality [2, 4]. The *property-based approach* posits that moral standing is inherently linked to the possession of certain properties, such as consciousness or the ability to experience pain. Within this framework, the question becomes whether social robots with AEI can genuinely possess these properties or merely imitate them.

The *relational approach* suggests that moral standing is not a matter of inherent properties but is rather intricately tied to the relationships that entities forge with others. In other words, it is not about what an entity "is" or "has" but about the quality, depth, and nature of its interactions with surrounding entities. In the context of social robots, this viewpoint underscores the significance of the interactions and relationships these robots cultivate with human users, irrespective of the authenticity of their emotions.

Our primary objective is to assess whether AEI could have points of connection or relevance that either support or challenge these established theories, and consequently provide new perspectives on the moral standing of machines. In doing so, we do not aim to defend or attack either theory. However, our analysis highlights certain strengths and weaknesses of each approach through how AEI changes – or does not change – how machines should be conceived. There is also significant debate about both what properties "count" in property-based approach [4, 5], and what the relational approach really entails [30, 31]. This means that both approaches can be used to grant – or deny – robots' moral status, and our intent is to highlight how, and not to determine which approach is more or less "correct".

3.1 Property-based morality

Some theoretical accounts of moral status define it as a concept which indicates a being's particular attribute which confers it moral standing [2, 4, 28]. In other words, these perspectives suggest that an entity's moral value is determined by a specific characteristic or quality inherent to it. The crucial question concerns, therefore, the specific normative criterion on which moral status is based.

Different theories confer moral status based on factors such as being a human, a person, a living individual, a member of a community of forms of life, a sentient being [28]. There is no definite consensus on which specific properties are morally relevant, and a comprehensive analysis of the most suitable theory of moral status is beyond the scope of this article. Hence, for the sake of our discussion, we lean towards the two most discussed criteria when it comes to AI, namely reason and sentience [5, 32-36]. We assume that aligning our discussion with previous works not only validate our approach but also enable us to engage with existing debates and contribute meaningfully to the discourse surrounding robots' moral consideration when introducing the novel dimension of AEI.

In the next subsections, we define the concepts of reason and sentience and, following previous authors, explore their interconnection with emotions and broader moral considerations. We then pose the question whether and how artificial emotions could impact these theories, whether they could be equated to human emotions, and what are the consequences for the debate of the moral status of social robots possessing AEI. We finally raise some objections and concerns.

3.1.1 Reason. Reason is the capability to govern one's action by thinking, understanding, and making logical judgements about the world [37]. In the pursuit to comprehend the dynamics of human cognition, modern research has shown that reason and emotions are integrally entwined, contradicting the classical dichotomy that posits them as separate entities [38, 39]. Emotions play a significant role in human behaviour, communication, and interaction and are essential for the effective functioning of human intellect and thinking processes [40-42]. An example of the complex interactions between cognition and emotion in cognitive processes is the mood-congruent effect on memory: when we are in a particular mood, it affects not only what we remember but also how we take in new information and how we feel when we recall these memories [43, 44]. Also working memory performance is affected by emotional states, as Levens and Phelps' study [45] has shown. Furthermore, in the Dual Competition Model proposed by Pessoa [46], emotions impact how information is processed in two specific ways: one that is driven by external stimuli and another that depends on the individual's internal state. In both cases, a competitive interaction occurs at both the sensory perception level and the executive function level, indicating a complex interplay in how information is processed under emotional influence.

These examples illustrate how cognitive and emotional processes extend beyond mere interaction; their neural mechanisms are intertwined both influence behaviour [47]. Given the indispensability of emotions for human reason, some have suggested to reconceptualise humans not as primarily rational but as fundamentally emotional animals [48]. In light of this, we seem to further accentuate the dichotomy between AI and humans, particularly when considering the conventional definition of AI as associated with systems with the computational capacity for goal-oriented tasks and problem-solving, notably excluding emotional considerations from its scope [49, 50]. Thus, if we accept the proposition that reason lays the foundation for moral status, then we might conclude that AI, despite its advanced rational capabilities, does not qualify for moral valuation due to its inherent lack of an emotional component [5].

The advance of AEI, however, introduces some complexities to the discourse, implying a possible recalibration of the moral assessment of entities (in our case, robots) possessing artificial emotions. The aim of the research in AEI is to develop computers "to be genuinely intelligent" [17], an intelligence analogous to human cognition which, as shown, consists of both emotional and rational capabilities. Should advancements in AEI eventually make it possible for artificial emotions to guide rational thought processes analogous to human cognitive functioning, a critical question arises: could such developments suffice for the conferral of moral status to AEI social robots?

A few objections could be raised. First, the choice of reason as the criterion for moral consideration is quite controversial and often objected, as it marginalises entities that many argue warrant

inclusion within our moral community, such as individuals with mental impairments [51]. However, assuming we adhere to the assertion that reason is the fundamental moral criterion, it could still be objected that AEI inherently lacks the capability to genuinely experience emotions [48] and, therefore, AI can never fully align with the constructs of human reason. Whether artificial and organic emotions could be equated, however, it is a discussion that we leave after introducing the property of sentience.

3.1.1 Sentience, emotions, interests. Sentience refers to the capacity to have pleasant or unpleasant experiences [14]. Contrary to a narrow interpretation that might limit sentience to basic sensory experiences, such as physical pleasure and pain, we follow DeGrazia [35] by arguing that sentience encompasses a range of emotional states (e.g., satisfaction and frustration), as well as instances of showing care or concern. Furthermore, Singer's [52, 53] work posits a foundational connection between the capacity for emotional experiences and the recognition of interests, as he suggests that the very ability to undergo suffering or relish enjoyment serves as a prerequisite for the existence of interests. In Singer's view, an entity that remains incapable of experiencing these emotional nuances, and by extension, possessing interests, falls outside the realm of moral consideration.

Along similar lines, Rodogno [34] argues that emotions are deeply intertwined with interests, identified as those moral values which hold profound significance for the subjects deserving moral considerations. He claims that, when an interest is satisfied or thwarted, emotional responses emerge, acting as indicators or reflections of these underlying values. This suggests that emotions, in many cases, serve both as motivators for action and as signals to what matters most to the individual [54]. Hence, the presence of such emotional responses underscores the possession of certain interests and moral values and, consequently, the candidacy for moral consideration.

This comprehensive understanding of sentience, emotions, and interests sets the stage for the dilemma at hand, namely the potential for AEI robots to acquire these attributes. This leads us back to the issue previously raised: is it feasible for a machine to experience emotions, or are these programmed responses mimicking genuine sentiment something else? If a robot does have the capability to exhibit emotions, as informed by earlier discourse linking emotions to moral worth, it naturally follows that they too might be deserving of moral status.

3.1.2 When it gets artificial. Hereafter, we first delve into theoretical frameworks of emotions, to determine how artificial emotions might affect ascription of moral status based on the selected properties. As we explore some of these theories, it becomes evident that there is no straightforward answer to the potential equivalency of organic and artificial emotions. This ambiguity stems from the very fact that even within human psychology, there is no universal consensus of the nature, origins, and purposes of emotions. In this section, rather than delving deep into each theory or advocating for one over another, our aim is to highlight that the debate around the equivalency of organic and artificial emotions is at least partially tied to the broader, unresolved discussions about the essence and role of emotions themselves. As we will see, some theories appear to facilitate comparisons between the emotional capacities of humans and AEI robots. We conclude this section by advancing objections to these views.

Any interpretation of a robot's emotional capacity is largely influenced by the theoretical paradigm chosen. For instance, views such as the James-Lange theory consider emotions as processes related to physiological factors. According to such views, emotions are "sensory feelings constituted by perceptions of changes in physiological conditions relating to the autonomic and motor functions" [55]. Accordingly, we do not tremble because we are afraid; rather, we feel fear because we tremble. Such views seem to confine emotions to organic entities capable of undergoing such physiological changes and likely challenge the possibility for AEI robot's emotions to equate humans' ones, however advanced their programming is[2].

However, motivational theories of emotions present a broader perspective, by considering emotions as "distinctive motivational states" and "internal causes of behaviours aimed at satisfying a goal" [56]. This definition transcends the organic boundary and resonates more with artificial emotions. If emotions fundamentally drive behaviour, aiming to fulfil certain objectives, then robots with algorithms designed to mimic this process could, theoretically, be said to "experience" emotions. Advances in AEI have shown that machines can be constructed with emotion-based algorithms that do not just mimic emotions, but use them as central components in decision-making, responsiveness, and interactions [57-59], much in line with Damasio's [60] theories of the role of emotions in human reasoning and decision-making. Here the emphasis is on the role of emotion augmentation in AEI social robots. As mentioned before, this function pertains to the deliberate integration of emotional capabilities in machines, with the objective of either enhancing their functionality or making their interactions more relatable and human-like.

Strömfelt, Zhang and Schuller [59] provide a comprehensive overview of the different ways artificial emotion has been applied in AI at an infrastructure level, emphasising cases where emotion is *intrinsic* to the machine's architecture. They mention, among others, a model of AEI that provides a unique perspective on the role of emotion in decision-making processes, particularly in balancing exploration and exploitation. Its emotion generator comprises various elements like sensations, feelings, emotions, and a hormone system which interacts and influences the feeling component. Some of the emotions this system includes are: anxiety, activated when rewards decrease; confidence, which works inversely to anxiety; fear, which increases with anxiety and affects the choice of strategies; and warmth, which indicates when to stop the algorithm based on levels of fear and the number of iterations. Based on these sensations, a dominant emotion emerges, which serves an executive role as it adjusts specific parameters of the algorithm and potentially alters its strategy or actions.

If we accept that AEI social robots can "feel" in this manner, does it not suggest that there is a potential for such robots to possess genuine emotions, although in a format distinct from human

[2] It could be counterargued, however, that bio-inspired and synthetic nervous system could achieve similar processes in generating emotions.

experience? Furthermore, it is worth noting that such AEI processes seem to significantly bridge the gap for robots to align more closely with the above-described reason criterion. If the line between human emotions and artificial emotions becomes ambiguous, it beckons us to revisit and possibly redefine our understanding of emotions and experiences and, consequently, for the arguments stated above, social robots' position in our moral community. In the next subsection, we present counterarguments to such a view.

3.1.3 Objections. As we have seen, motivational theories of emotions could offer support to the idea that artificial emotions might be functionally equated to human emotions. However, this perspective is not without its challenges. A significant objection is rooted in the rich subjective experience of emotions. Arguably, emotions extend beyond mere behavioural responses. They manifest as sensations that individuals can introspect upon, recall, and be overwhelmed by.

This intrinsic nature of emotions leads to a crucial question: can an entity truly experience emotion without possessing the capability for subjective experience, or in essence, consciousness? A comprehensive exploration and conceptualisation of consciousness would far exceed the bounds of this article. However, a modest and general understanding of consciousness as the capacity to have private mental experiences [35] seems enough to argue that consciousness becomes an indispensable component for an entity to experience any form of emotion, sensory perception, or care.

When considering social robots with AEI, the distinction between authentic emotional experience and programmed response becomes crucial. Causing distress in a system devoid of the capability to genuinely perceive it raises the question whether any real suffering occurs. Let's refer again to the example of the AEI model described in Strömfelt, Zhang and Schuller's [59] work and consider an AEI model implemented in a robot that exhibits, e.g., the emotion of fear. Some might argue that such fear is, in essence, a programming design to avoid hazardous situations. Without consciousness acting as a backdrop, the genuineness of such emotion remains in question.

After all, as Mosakas [33] suggests, it appears absurd to claim that pain carries meaningful implications for an entity that cannot genuinely perceive it. It follows that entities that lack consciousness naturally lack the capability to truly experience pleasure, pain, or any emotional nuance in between; namely, the capacity of sentience. For non-sentient beings, the external world's actions upon them are met with indifference, due to their inherent incapability to perceive and process these experiences meaningfully [53]. For the argument elaborated above, it follows that, without consciousness, an entity's claim to moral rights and consideration becomes tenuous. This places us on the threshold of reevaluating the weight we assign to artificial emotions and subsequently reshaping our moral obligations towards such entities.

However, a few objections can be advanced. First, it might be argued that, in the future, more sophisticated robots could reach a certain degree of consciousness [61, 62]. Furthermore, the very definition and understanding of consciousness are not set in stone and, depending on which lens we view consciousness, we could posit that robots, even in their current state, possess a form of it [63, 64].

Secondly, there is an ongoing debate which asks whether sentience can be attributed to living things without consciousness. For instance, Warren [28] claims that non-self-aware beings, however unconscious, such as embryos and foetuses, might possess interests in, e.g., being alive. Embryos and foetuses might still have interests in continuing to live until birth, however being unaware of it.

Thirdly, when it comes to reason as the criterion for moral consideration, it might be argued that the enhancements that artificial emotions impart to machine capabilities may *sufficiently* mirror the contribution of emotions to human reasoning. It could be said that the presence of consciousness and authentic emotions may not be indispensable, as the approximation of artificial emotions to organic emotions may well enable robots to function in ways comparable to human cognition.

And, finally, as observed by some in the context of robots' moral agency, adhering to a property-based approach could be inherently flawed when it comes to robots [65]. According to Kahn's New Ontological Category Hypothesis [27], robots represent a unique ontological category, consequently challenging the validity of applying the same criteria for moral patiency to robots as we do to living beings. As such, it may be necessary to adopt a different approach when considering robots in terms of moral status.

Such arguments are complex and warrant a more thorough examination than we cannot offer in this article. Nonetheless, by presenting such objections, our intention is to bring attention to the prevailing absence of unanimity on whether consciousness is essential for sentience and reason. This current lack of consensus renders the assertion - that artificial emotions are not comparable to human emotions and that AEI does not sufficiently justify attributing moral status to robots – significantly contentious.

In conclusion, this section has analysed the influence of social robots' artificial emotions on discussions about their moral status, based on a property-related approach. Using the property-based approaches anchored in the concept of reason and sentience, we investigated the parallels between artificial and human reasoning and emotions. Our findings indicate that AEI in social robots clearly influences the degree to which robots can aspire to the properties chosen. This implies that, depending on how one defines and approaches these concepts, AEI could indeed change people's perception of their moral status and blur the boundaries between humans and machines.

3.2 The social-relational approach

In this section, we present the social-relational approach as an alternative method for the conditions of moral status ascription. We will see that the approach itself carries inherent challenges that may not be resolved by the technology of AEI. Nonetheless, this perspective stands to be considerably strengthened by the incorporation of AEI into social robots.

In traditional viewpoints, as the one previously explored, the entity's intrinsic properties determine its moral value; in other words, something has value because of what it inherently is.

However, the social-relational approach, advanced by authors such as Coeckelbergh and Gunkel [2, 3, 66-68] suggests that, instead of looking within the entity for its moral value, we should look at the relationships and social context surrounding it. Here, moral value is not intrinsic but rather is an outcome of the social interplays and dynamics; i.e., how the entity is situated in our social world and how we relate to it.

Interestingly, this framework holds that moral consideration of social robots does not necessarily hinge on authenticity. According to the property-based view, this authenticity is equated to a robot's ability to have genuine emotions. Instead, the social-relational view focuses on our interactions with, perceptions of, and experiences with the robot. It is the way an entity *appears* to us that determines the quality and depth of the relational experiences. The entity's properties still play a role as the foundation on which we base our moral consideration, but their importance is reframed [30]. Instead of them being valuable for what they inherently are, they matter because of how we experience and interpret them [1].

After all, Coeckelbergh [69] argues, when interacting with another individual, we do not have direct access to their consciousness or emotions. Instead, we rely on external cues – behaviour, expressions, words – for how they appear to us, and interpret these as indicators of their internal states. Our moral considerations are based on such interpretations, not on direct knowledge or authenticity. This marks a crucial distinction from property-based frameworks and opens new possibilities for considering social robots as morally worthy.

Empirical work in HRI provides valuable insights, particularly concerning the impact of people's perception of robots on moral and social considerations. Thellman, Thunberg and Ziemke [70] investigated the impact of people's emotional state on their perceptions of human-like qualities and mental capacities of robots, along with their attitudes towards these machines. Malle et al. [71] demonstrated that robot appearance affects people's moral judgements about them. Carpinella et al. [72] have developed a scale (RoSAS) to measure social perception of robots. Research conducted by De Graaf, F. Hindriks and K. Hindriks [73] involving an online survey exploring layman's attitudes towards granting particular moral and legal rights to robots, reveals that public opinions on a robot's emotional and cognitive capabilities significantly influence the decision to grant robots rights. Also, Weiss and Hannibal [20], has proposed a study which explores how the relationship between social robots and users evolves over time.

According to the social-relational approach, a robot's moral standing is not reliant on its intrinsic properties but rather on the dynamics of its interaction within human society. It is about how the robot makes us feel, the kind of relationships we can form with it, and the experiences it can foster. Sex robots, falling under the broader range of social robots and sometimes referred to as love robots [74], stand as a prominent example in this context. These robots have been developed beyond their initial purpose of physical gratification [61], and are increasingly seen as capable of engaging in a broader spectrum of intimate relationships [75]. The experiential and relational aspects of interacting with sex robots have served as platforms for exploring how robotic entities can fulfil emotional, social, and psychological needs in ways previously not considered, such as in context of care [76]. A notable example in this regard is Davecat's relationship with his sex doll Sidore, which he has been married to for over 20 years [77]. This case suggests a foundational aspect of the social-relational approach, where the authenticity or the lifelike qualities of the doll are secondary to the emotional fulfilment, companionship, and psychological comfort Davecat derives from the relationship.

Nonetheless, Davecat's doll lacks any form of AI. Their relationship can be seen as an important, yet preliminary exploration in the discourse on the moral status of AI-enhanced robots from a social-relational perspective. As Coeckelbergh argues [1], the robot needs to be sufficiently advanced to be granted moral consideration. Which is to say, it has to simulate the ability to feel or experience so convincingly that, to an observer, they seem almost indistinguishable from genuine emotions. Within this framework, the capacity to simulate and generate emotions of AEI (what we have previously called emotion generation) in social robots becomes crucial. The simulation, if sophisticated enough, can potentially foster deeper and more genuine feeling relationships between humans and robots, consequently challenging the traditional boundaries we have established between the two entities.

The emotional interactivity AEI could bring about between robots and humans, might amplify the illusion of genuine subjectivity of the robot, making this latter more relatable in their interactions. As we do not demand direct evidence of the presence of subjectivity or genuine emotions in another human, we should not necessarily demand the same of AEI social robots [69]. If we interpret artificial emotions as sufficiently genuine, allowing us to foster meaningful interactions with the robot, this alone becomes the basis for the robot's moral status. Within the social-relational framework, therefore, AEI might significantly increase the likelihood for granting moral consideration to social robots equipped with such technology.

However, the social-relational approach raises criticisms which, as we will argue, might persist even in the face of AEI's advancements. Let us delve into some of these objections in the following.

3.2.1 Objections. A primary concern implicit in the social-relational approach is the adoption of a relativistic viewpoint regarding moral status. As discussed in the previous section, the moral significance of an entity is determined by the relationship and attachment individuals or communities form with that entity. In simpler terms, if people care about something or someone due to the connection they have established, then that entity is granted moral consideration. The problem is that the threshold for what qualifies as having moral status becomes highly variable and subjective. Since moral significance is influenced by feelings, personal biases, and cultural values, it can be argued that virtually anything can be given moral importance depending on individual or collective sentiments.

To highlight this problem, Muller [36] makes the hypothetical example of a community that holds deep regard for pencils. They might expect outsiders to respect their sentiment towards pencils. However, this subjective reference should not be mistaken for an objective assertion that pencils have universal moral value. Similarly, more controversial, just because a group might devalue a

particular set of individuals (red-haired women, in the given example) does not mean that those individuals lack moral value.

The underlying challenge is that the social-relational runs the risk of diluting the concept of moral status to mere personal inclinations. Without a more concrete and shared foundation for determining moral status, this approach could lead us into moral ambiguity, where each individual's or group's feelings become their own moral compass.

Furthermore, applying the social-relational approach to social robots possessing artificial emotions might introduce additional concerns, such as wrongly prioritising them over entities that truly deserve moral patiency, such as humans and animals [78]. Limited resources and attention could be diverted towards ensuring the rights of robots, neglecting moral obligations towards other beings. Implications of granting moral status to social robots should be critically evaluated.

Finally, authors such as Sætra [30] have argued that the social relational approach may exhibit anthropocentric tendencies, as it places importance on how humans decide to respond to entities and what they decide to afford to them based on their performance, or even simply based on the human's projection of characteristics or performance of inanimate objects, such as a simple doll as in the case of Davecat. Therefore, this perspective ignores the intrinsic value of other beings or entities when not belonging to human relationships and establishes humans as the sole arbiters of moral status. This reconnects to the problem of prioritising entities designed to foster human emotions such as AEI social robots, over natural entities that might not immediately elicit strong human emotions but have inherent worth, such as animals like fishes in the sea.

To conclude this section, the challenges embedded within the social-relational approach are not easily surmounted, and the advent of artificial emotions does not seem to present resolutions to these issues. Nevertheless, the role of AEI becomes pivotal within this approach, opening up novel avenues for its proponents and possessing the potential to alter perspectives regarding the moral status of social robots. While the predominance of the social-relational view in conferring moral status remains uncertain, it is plausible to suggest that the advancements in AEI render this perspective increasingly viable and worthy of consideration.

4 CONCLUSION

In this paper, we have asked whether the introduction of artificial emotions in social robots might influence the ongoing discourse regarding their moral status. We delved into the two predominant frameworks on this subject: the property-based and social- relational approaches, re-evaluating them specifically in the context of social robots enhanced with AEI. From our analysis, we deduce that there is not a definitive resolution, as the implementation of AEI brings robots closer to humans in more intrinsic, fundamental ways, introducing additional complexity to the debate.

In this last section, we aim to address some final considerations on the matter of robots' moral status. Although AEI does not seem to definitively bestow moral value upon existing social robots, these entities could still hold instrumental value [79]. This aligns with the concept of derived moral status [36], indicating a kind of moral status based on the entity's instrumental, extrinsic or emotional value [33, 36].

Robots' derived moral status suggests that, while social robots might not possess intrinsic value akin to sentient or rational beings, their treatment and role in human societies warrant certain moral considerations. Indeed, it is imperative to consider the broader social implications of our interactions with these robots. Treating robots with cruelty or insensitivity, simply because they lack consciousness or genuine emotions, could foster an environment where such behaviour becomes normalised [80, 81]. While this is a potential concern with any object, social robots with artificial emotions merit special attention due to their appearance and their specific function in society.

A pertinent example relates to sex robots. While they lack consciousness, and thus one could argue that they cannot be directly harmed, harassed, or abused [61], they still mimic human features and behaviours. The ethical concern is not about causingharm to the robot itself, but about what such interactions might represent: a potential undermining of consent norms in human sexual relationships [82]. Any act that objectifies or dehumanises these robots can have a ripple effect, potentially reinforcing or normalising demeaning behaviours in real-life human relationships [7, 83]. As such, advocating for respectful treatment of sex robots is less about safeguarding their wellbeing and more about preserving the dignity and rights of actual humans.

Finally, we want to highlight a few potential counterarguments to the propositions we have presented, indicating areas for further exploration in future research. For instance, one could posit that, as technological advancements persist, the future could bring forth artificial entities endowed with morally relevant attributes – such as emotions, interests, and a susceptibility to suffering – which exist solely in digital environments [84]. Such entities would be entirely devoid of a tangible, physical presence, whether it mirrors human form or any other recognisable structure.

Our focus on social robots might be seen therefore as a limitation in the scope of our study, challenging the necessity to restrict moral considerations to entities with physical embodiments and urging for research on purely digital entities with the potential to manifest genuine emotions.

However, in response to these objections, our focus on robots is informed by theories which maintain that both reason and emotions are necessarily embodied experiences [41, 85, 86]. These theories assert that emotions are deeply intertwined with our bodily experiences and interactions with the world, suggesting that the presence of a physical body is crucial for the manifestation of morally relevant emotions. Thus, in this view, social robots offer a more authentic and meaningful parallel to human experiences and moral considerations than entities that are purely digital.

In closing, this article prompts discussion regarding the evolving moral landscape we find ourselves in as social robots are equipped with AEI. It is crucial to reflect on AEI's transformative role in eroding the demarcations traditionally drawn between humans and machines. As we have seen, AEI, by integrating emotional processing within the computational frameworks of robots,

endangers a profound re-evaluation of our perceptions and interactions with these entities. Regardless of the specific theoretical perspective one adopts concerning the moral status of robots, the implications of AEI remain significant and far-reaching. This convergence of cognition and emotion within machines provokes deeper reflections on the essence of consciousness and the nature of emotional experience, and demands a meticulous reconsideration of our ethical obligations, moral values and societal norms. It is essential, as technology advances, to continue to explore and re-evaluate our moral frameworks and ethical standpoints in relation to artificial entities, to foster a harmonious coexistence and to uphold the values and dignities that define our humanity.

REFERENCES

[1] Coeckelbergh, M. Robot rights? Towards a social- relational justification of moral consideration. *Ethics and Information Technology*, 12, 3 (2010), 209-221.
[2] Gunkel, D. J. *Robot rights*. MIT Press, 2018.
[3] Gunkel, D. J. *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. MIT Press, 2023.
[4] Gellers, J. C. *Rights for Robots: Artificial Intelligence, Animal and Environmental Law*. Routledge, 2020.
[5] Sætra, H. S. *Man and his fellow machines: An exploration of the elusive boundary between man and other beings*. Waxman, 2019.
[6] Darling, K. Extending legal protection to social robots. *IEEE Spectrum* (2012).
[7] Sica, A. The Robot will Feel You Now: The Ethics of Artificial Emotional Intelligence in Sex Robots. In *Proceedings of the IEEE Ro-Man 2023* (Busan, 2023).
[8] Broadbent, E. Interactions with robots: The truths we reveal about ourselves. *Annual review of psychology*, 68 (2017), 627-652.
[9] Grace, K., Salvatier, J., Dafoe, A., Zhang, B. and Evans, O. When will AI exceed human performance? evidence from AI experts. *Journal of Artificial Intelligence Research*, 2 (2018), 729-754.
[10] Corti, L., Stefano, N. D. and Bertolaso, M. Artificial emotions: toward a human-centric ethics. *International Journal of Social Robotics* (2022), 1-15.
[11] Sætra, H. S. Robotomorphy: Becoming our creations. *AI and Ethics*, 2, 1 (2022), 5-13.
[12] Regan, T. *'The Case for Animal Rights'*. Routledge, 2013.
[13] Singer, P. *Animal liberation*. Routledge, 2004.
[14] Gruen, L. The moral status of animals (2003).
[15] Damiano, L. and Dumouchel, P. Anthropomorphism in Human–Robot Co-evolution. *Frontiers in Psychology*, 9 (2018-March-26 2018).
[16] Schuller, D. and Schuller, B. W. The age of artificial emotional intelligence. *Computer*, 51, 9 (2018), 38-46.
[17] Picard, R. W. *Affective Computing*. MIT Press, Cambridge, 1997.
[18] Kaushik, R. and Simmons, R. *Perception of emotion in torso and arm movements on humanoid robot quori.*, 2021.
[19] Fiorini, L., Loizzo, F. G., D'Onofrio, G., Sorrentino, A., Ciccone, F., Russo, S., Giuliani, F., Sancarlo, D. and Cavallo, F. *Can I Feel You? Recognizing Human's Emotions During Human-Robot Interaction*. Springer, 2022.
[20] Weiss, A. and Hannibal, G. *What makes people accept or reject companion robots? A research agenda.*, 2018.
[21] Abdollahi, H., Mahoor, M., Zandie, R., Sewierski, J. and Qualls, S. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. *IEEE Transactions on Affective Computing* (2022).
[22] Marcos-Pablos, S. and García-Peñalvo, F. J. *Emotional intelligence in robotics: a scoping review*. Springer, 2022.
[23] Reeves, B. and Nass, C. *The media equation: How people treat computers, television, and new media like real people.*, Cambridge, UK, 1996.
[24] Scheutz, M. *The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots*. MIT Press, 2009.
[25] Sica, A. and Sætra, H. S. *In Technology We Trust! But Should We?* Springer, 2023.
[26] Damiano, L. and Dumouchel, P. Anthropomorphism in human–robot co-evolution. *Frontiers in psychology*, 9 (2018), 468.
[27] Kahn Jr, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., Ruckert, J. H. and Gill, B. *The new ontological category hypothesis in human-robot interaction.*, 2011.
[28] Warren, M. A. *Moral Status: Obligations to Persons and Other Living Things*. Oxford University Press, Oxford, 2000.
[29] Reichlin, M. *Moral Status*. Springer, 2014.
[30] Sætra, H. S. Challenging the Neo-Anthropocentric Relational Approach to Robot Rights. *Frontiers in Robotics and AI*, 8 (2021-September-14 2021).
[31] Mosakas, K. On the moral status of social robots: considering the consciousness criterion. *AI & SOCIETY*, 36 (2020), 429-443.
[32] Sytsma J, M. E. T. s. o. m. s. R. P. and 3:303–324., P. Two sources of moral standing. *Review of Philosophy and Psychology*, 3 (2012), 303-324.
[33] Mosakas, K. On the moral status of social robots: considering the consciousness criterion. *AI & SOCIETY*, 36 (2021), 429-443.
[34] Rodogno, R. *Robots and the Limits of Morality*. Ashgate, 2016.
[35] DeGrazia, D. Robots with moral status? *Perspectives in Biology and Medicine*, 65 (2022), 73-88.
[36] Muller, V. C. Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology*, 23 (2021), 579-587.
[37] Honderich, T. *The Oxford companion to philosophy*. OUP Oxford, 2005.
[38] Keltner, D. T. and Lerner, J. S. *Emotion*. Wiley, 2010.
[39] Cacioppo, J. T. and Gardner, W. L. Emotion. *Annual review of psychology*, 50, 1 (1999), 191-214.
[40] Damasio, A. Descartes' error: Emotion, rationality and the human brain. *New York: Putnam*, 352 (1994).
[41] Damasio, A. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Harcourt inc., Orlando, 2003.
[42] Lerner, J. S., Li, Y., Valdesolo, P. and Kassam, K. S. Emotion and decision making. *Annual review of psychology*, 66 (2015), 799-823.
[43] Buchanan, T. W. Retrieval of emotional memories. *Psychological bulletin*, 133, 5 (2007), 761.
[44] Bower, G. H. and Forgas, J. P. Affect, memory, and social cognition. *Cognition and emotion* (2000), 87-168.
[45] Levens, S. M. and Phelps, E. A. Emotion processing effects on interference resolution in working memory. *Emotion*, 8, 2 (2008), 267.
[46] Pessoa, L. How do emotion and motivation direct executive control? *Trends in cognitive sciences*, 13, 4 (2009), 160-166.
[47] Liu, Y., Fu, Q. and Fu, X. The interaction between cognition and emotion. *Chinese Science Bulletin*, 54, 22 (2009), 4102-4116.
[48] Turkle, S. *Alone together: Why we expect more from technology and less from each other*. Basic Books, New York, NY, US, 2011.
[49] McCarthy, J. What is artificial intelligence? (2004).
[50] Copeland, B. J. Artificial intelligence. Encyclopedia Britannica. *Inc.: Chicago, IL, USA* (2020).
[51] Agnieszka, J. and Tannenbaum, J. *The Grounds of Moral Status.*, 2023.
[52] Singer, P. *Animal liberation*. HarperCollins, New York, 1975.
[53] Singer, P. *The expanding circle: ethics, evolution and moral progress*. Princeton University Press, New Jersey, 2011.
[54] Stark, L. and Hoey, J. The ethics of emotion in artificial intelligence systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2021). ACM, 2021.
[55] James, W. What is an emotion? *Mind*, 9 (1884), 188-205.
[56] Scarantino, A. and De Sousa, R. *Emotion.*, 2018.
[57] Moerland, T. M., Broekens, J. and Jonker, C. M. Emotion in reinforcement learning agents and robots: a survey. *Machine Learning*, 107 (2018), 443-480.
[58] Rosenbloom, P. S., Gratch, J. and Ustun, V. *Towards emotion in sigma: from appraisal to attention*. Springer, 2015.
[59] Strömfelt, H., Zhang, Y. and Schuller, B. W. Emotion- Augmented Machine Learning: Overview of an Emerging Domain. In *Proceedings of the Affective Computing and Intelligent Interaction (ACII 17)*, 2017.
[60] Damasio, A. *Descartes' Error: Emotion, Reason, and the Human Brain. Quill*. Penguin, 2006.
[61] Frank, L. and Nyholm, S. Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable? *Artificial intelligence and law*, 25 (2017), 305-323.
[62] Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C. and Ji, X. Consciousness in Artificial Intelligence: Insights from the science of consciousness. *arXiv preprint arXiv:2308.08708* (2023).
[63] Bryson, J. J. A role for consciousness in action selection. *International Journal of Machine Consciousness*, 4, 02 (2012), 471-482.
[64] Dennett, D. C. *Elbow room, new edition: The varieties of free will worth wanting*. mit Press, 2015.
[65] Kim, B., Phillips, E., Zhu, Q. and Williams, T. Perspectives on Moral Agency for HRI: Cognitive Construct or Ontological State? (
[66] Gunkel, D. J. *Mark Coeckelbergh: Growing moral relations: critique of moral status ascription: Palgrave Macmillan, New York, 2012, 239 pp, ISBN: 978-1-137-02595-1. Springer, 2013.*
[67] Gunkel, D. J. The other question: can and should robots have rights? *Ethics and Information Technology*, 20 (2018), 87-99.
[68] Gunkel, D. J. *The right (s) question: Can and should robots have rights?* Brill mentis, 2020.
[69] Coeckelbergh, M. Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, 12 (2010), 235-241.
[70] Thellman, S., Thunberg, S. and Ziemke, T. *Does Emotional State Affect How People Perceive Robots?* , 2021.
[71] Malle, B. F., Scheutz, M., Forlizzi, J. and Voiklis, J. *Which robot am I thinking about? The impact of action and appearance on people's evaluations of a moral robot*. IEEE, 2016.
[72] Carpinella, C. M., Wyman, A. B., Perez, M. A. and Stroessner, S. J. *The robotic*

social attributes scale (RoSAS) development and validation, 2017.
[73] De Graaf, M. M., Hindriks, F. A. and Hindriks, K. V. Who wants to grant robots rights? , 2021.
[74] Whitby, B. 15 Do You Want a Robot Lover? The Ethics of Caring Technologies. *Robot ethics: The ethical and social implications of robotics* (2011), 233.
[75] Scheutz, M. and Arnold, T. *Are we ready for sex robots?* IEEE, 2016.
[76] Fosch-Villaronga, E. and Poulsen, A. *Sex robots in care: Setting the stage for a discussion on the potential use of sexual robot technologies for persons with disabilities*. 2021.
[77] Nyholm, S. *Robotic animism: the Ethics of attributing minds and personality to Robots with Artificial Intelligence*. Springer, 2023.
[78] Birhane, A. and van Dijk, J. Robot Rights? Let's Talk about Human Welfare Instead. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2020).
[79] Nyholm, S. *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield Publishers, 2020.
[80] Bryson, J. J. Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology*, 20, 1 (2018), 15-26.
[81] Parthemore, J. and Whitby, B. Moral agency, moral responsibility, and artifacts: What existing artifacts fail to achieve (and why), and why they, nevertheless, can (and do!) make moral claims upon us. *International Journal of Machine Consciousness*, 6, 02 (2014), 141-161.
[82] Danaher, J. *The Symbolic-Consequences Argument in the Sex Robot Debate*. MIT Press, 2017.
[83] Nascimento, E. C. C., Silva, E. d. and Siquiera-Batista, R. The 'use' of sex robots: A bioethical issue. *Asian Bioethics Review*, 10, 3 (2018), 231-240.
[84] Harris, J. and Anthis, J. R. The moral consideration of artificial entities: a literature review. *Science and engineering ethics*, 27, 4 (2021), 53.
[85] Müller, V. C., & Hoffmann, M. What is morphological computation? On how the body contributes to cognition and control. *Artificial life*, 23, 1 (2017), 1-24.
[86] Shapiro, L. and Spaulding, S. *Embodied Cognition*. Metaphysics Research Lab, Stanford University, 2021.