filosofia unisinos | Unisinos Journal of Philosophy

ISSN 1984-8234

Filosofia Unisinos
Unisinos Journal of Philosophy
25(1): 1-16, 2024 | e25111

Unisinos – doi: 10.4013/fsu.2024.251.11

Dossier

The heart of an AI: agency, moral sense, and friendship[1]

O coração de uma IA: agência, senso moral e amizade.

Evandro Barbosa[1]
https://orcid.org/0000-0002-5695-3746
[1]Universidade Federal de Pelotas - UFPel, Programa de Pós-Graduação em Filosofia, Pelotas, RS, Brasil.
Email: evandrobarbosa2001@yahoo.com.br

Thaís Alves Costa[2]
https://orcid.org/0000-0002-1274-0431
[2]Instituto Federal de Educação, Ciência e Tecnologia de Farroupilha, Farroupilha, RS, Brasil.
Email: costa.thaisalves@gmail.com

ABSTRACT

The article presents an analysis centered on the emotional lapses of artificial intelligence (AI) and the influence of these lapses on two critical aspects. Firstly, the article explores the ontological impact of emotional lapses, elucidating how they hinder AI's capacity to develop a moral sense. The absence of a moral emotion, such as sympathy, creates a barrier for machines to grasp and ethically respond to specific situations. This raises fundamental questions about machines' ability to act as moral agents in the same manner as human beings. Additionally, the article sheds light on the practical implications within human-machine relations and their effect on human friendships. The lack of friendliness or its equivalent in interactions with machines directly impacts the quality and depth of human relations. This concern-

[1] We express our gratitude to Professor Denis Coitinho (Unisinos) for organizing this special issue. Previous versions of this text were presented at the "XXVI International Colloquium on Philosophy Unisinos - Artificial Intelligence: Present and Future" and the "VIII International Congress on Moral and Political Philosophy: Human Rights, Vulnerability, and the Environment" at the Federal University of Pelotas, throughout 2023. We thank the participants for their helpful comments on earlier versions of the material in this paper. We appreciate comments and suggestions from Vicki Berens and members of the working group at the University of North Carolina at Chapel Hill: Gigi, Waren, Lan, and Cati. Finally, we acknowledge the financial support from CNPq through a productivity scholarship - PQ2 awarded to researcher Evandro Barbosa (UFPel) and a grant for the project "Faces of Vulnerability: Justice, Human Rights, and Technology" - Universal CNPq/MCTI No 10 /2023.

ingly suggests the potential replacement or compromise of genuine interpersonal connections due to limitations in human-machine interactions.

Keywords: artificial intelligence, agency, emotion, moral sense, friendship.

RESUMO

O artigo propõe uma análise centrada no lapso emocional, enfocando sua influência em dois aspectos cruciais. Em primeiro lugar, explora o impacto ontológico do lapso emocional na inabilidade das máquinas em desenvolver um senso moral. A falta de emoções ou sentimentos morais, como a simpatia, representa um obstáculo para as máquinas compreenderem e responderem de maneira ética a certas situações, o que levanta questões fundamentais sobre a capacidade das máquinas em discernir valores éticos de forma autônoma. Por outro lado, vamos lançar luz sobre a questão prática que surge nas relações humano-máquina e seu impacto na relação humana de amizade. A ausência de simpatia ou equivalente moral nas interações com máquinas levantam preocupações sobre a possibilidade de substituição ou comprometimento das conexões interpessoais genuínas.

Palavras-chave: inteligência artificial, agência; emoção, senso moral, amizade.

1 An overview of artificial intelligence and ethics

In the Hollywood film *I, Robot*, a policeman named Officer Spooner (played by Will Smith) was involved in a traffic accident that sent two cars plunging into a river. Officer Spooner was alone in his car, while there was a young girl in the other vehicle. Everyone involved found themselves trapped in their cars, facing certain drowning, were it not for the timely intervention of an AI robot. The robot swiftly dove into the water and conducted a rapid assessment of the situation. Its calculations revealed that Spooner had a 30% chance of survival, while the 12-year-old girl had just a 28% chance. Based on this calculation, the robot chose to save Spooner over the child.

From a moral standpoint, do you believe the robot made the right choice? Answering this question not only requires an assessment of whether artificial intelligence can distinguish between right and wrong, but also whether the robot may have experienced any appropriate emotion as a result of its decision. While it's plausible that a machine can make moral choices, this does not necessarily mean that they understand such actions in a deeper sense nor that they have anything like a moral conscience that guides their actions and intrinsically motivates them. At first glance, this appears to be the responsibility of the developers team.

This brings us to our assumption, which we will call an "AI emotional lapse":

- AI (at the moment) is unable to develop a moral sense and make appropriate moral judgments because it lacks the capacity for moral emotions, such as sympathy.

Section I presents how this problem rests on the ontological issue that the metaphysical constitution of a machine does not encompass emotional elements, rendering them incapable of developing a moral sense. Our widely acknowledged hypothesis is that AI's (non-moral) machinery functions differently from the human mind. While AI systems excel in efficiency, scalability, and precise decision-making, it's crucial to acknowledge their limitations.

According to your hypothesis, AI's lack of a moral emotions, like sympathy and contextual understanding, compromise its ability to make moral choices. In other words, AI operates based on objective

criteria and predefined rules, rendering it incapable of comprehending the nuances and individuality of human experiences. This limitation becomes evident when dealing with complex needs and unique circumstances, which are often present among vulnerable individuals. In such cases, the use of AI tools may be limited, as they fall short of accounting for subjective aspects of decision-making, such as personal stories, cultural sensitivities, and emotional well-being.

Assuming the validity of EL assumption, in Section II, our focus lies in examining the impact of human engagements with AI like chatbot friends on the meaning of genuine human friendships. Human relationships based on AI models may render human interactions artificial, creating a gap in the development of sympathy that obstructs moral considerations in human relations like friendship. We will explore how these human-AI interactions potentially alter the comprehension of human relationships founded on fostering and enhancing moral attributes like sociability and collaborative cooperation. After providing these explanations, we will conclude, in Section IV, by presenting an argument that highlights the negative moral outcome of emotional lapses in AI and how the moral boundary of a conative lapse in a machine also extends to human-AI interactions. Our focus will revolve around the issue of friendship and engagement with different types of AI.

2 Moral sense and AI's emotional lapse

> *Ideally, we would like to be able to trust autonomous machines to make correct ethical decisions on their own, and this requires that we create an ethic for machines. (Anderson & Anderson, 2011, p. 1)*

There is an ontological question that needs to be addressed regarding whether the nature of AI allows not only for moral behavior, but also the possession of a genuine moral sense. This is a foundational query that precedes any practical considerations. The human organism operates with a moral navigation system (referred to as a "moral compass" by Kant or a "moral sense" by David Hume and Adam Smith) that is calibrated through our establishment of human relationships. Although humanity is interested in attempting to create AI with the ability to imitate moral decision-making processes without human intervention, it's challenging to imagine machines engaging in moral practices without having an equivalent to such practices.

Despite this, many have advocated for delegating decision-making and moral judgments to machines. These individuals propose designating certain principles or moral rules to guide the decision-making of machines. Such principles are often associated with imperative moral models like deontological or utilitarian theories. For instance, Powers (2006) proposed a machine following the Kantian categorical imperative. However, the true nature of an AI as a moral agent that is capable of understanding the reasons behind moral actions and judgments remains unclear (Bostrom, 2014). The problem is that artificial moral agents face challenges when trying to emulate the notion of human agency, especially because the "mind" of an AI depends on a still limited initial programming and a set of big data. If artificial general intelligence (AGI) were to move forward, it would ultimately achieve the much discussed singularity. (See Tegmark, 2017)

> *'Singularity' is explained as an intelligent explosion, which will be powerful enough to replicate the human moral agency as well as the brain itself (...) This claim postulates the ethical decision-making process as the effect of brain states; hence, it can be simulated by artificial neurons as well. Precisely, if an artificial neuronal structure can imitate the biological brain state, which is responsible for a particular moral deed/thought, then reproducing that moral act in artificial performers will be spontaneous. Therefore, the supporters of 'singularity' believe that superintelligent AI agents will be moral agents just like any rational human being. It is justified to hold them responsible for their actions. (Manna & Nath, 2021, p. 144-45 - see also Kurzweil, 2005)*

With that said, the equation is not that simple from a moral perspective. Moor (2006) distinguishes three types of moral agents: implicit ethical agents, explicit ethical agents, and full ethical agents. For example, human beings are agents with full ethical agency. "A full ethical agent can make explicit ethical judgments and generally is competent to reasonably justify them." (Moor, 2006, p. 20) Meanwhile, implicit ethical agents are "operationally moral", which means that they do what their designers determine they should do. An example of this is the software program at a bank, which is responsible for executing the correct transactions for sending and receiving money. The programmer does not input the concept of honesty into the machine expecting the output to be "make correct monetary transactions." The machine needs more than that to operate within the accepted ethical parameter and to carry out fair transactions. Finally, explicit ethical agents are agents who are capable of making moral judgments that determine the course of action without human interference.

AI theorists agree that the best we can do, given the complexity involved in moral decision-making, is to choose the explicit moral agents option. Michael Anderson, Susan Anderson, and Chris Armen (2005) made an attempt to achieve this type of agency by suggesting the combination of two ethical approaches: Hedonistic Act Utilitarianism and William D. Ross's model of prima facie duties. The responses to the tests were positive as the machine was able to accommodate adequate responses to the proposed models within an expected margin: "It uses a learning algorithm to adjust judgments of duty by taking into account both prima facie duties and past intuitions about similar or dissimilar cases involving those duties." (Moor, 2006, p. 20) Wallach and Allen went on to say that they achieved a kind of functional morality, which is what we should expect from these AI models for now. In other words, AI models could become explicit ethical agents, but they would never reach the full dimension that would equate them to humans.

Our position in this paper corroborates this thesis, supporting the premise that the lack of an emotional element, such as sympathy, suppresses the moral capacity of AI models. Most of these attempted answers rest on models of moral agency that center the morality of action on the rational element of the agents involved, but there is a blind side that is being neglected. It is true that an AI moral agent can possess an adequate level of autonomy to be morally responsible as well as be associated with the rational ability to determine actions consciously and coherently. Furthermore, the same agent must also have the capacity to identify relevant moral information. It is a complex query in relation to AI, given that capacity involves developing a moral sense to determine the moral quality of the elements at stake. This could be a difficult, or even impossible, requirement for an AI to meet.

Even if we agree that some AI can develop a sufficient level of autonomy, it is not clear that an AI can make a decision based on any type of moral sense, simply because these machines don't have such a sense (at least not according to those who advocate for "moral sentimentalism"). (See Gert, 2015; Kauppinen, 2021) According to this tradition, a moral sense refers to the ability to evaluate and determine right and wrong based on the conscience or internalized perception of the norms, values, and ethical principles that guide an individual's behavior in interactions and decisions. Hume, for instance, asserts that a moral agent's judgment is tied to an "immediate feeling and finer internal sense." (Hume, 1777, p. 02) This suggests, according to Walsh, that emotions serve as both "(i) a causal antecedent to moral judgment and (ii) a sufficient ground for the legitimacy of moral claims." (2021, p. 5210)

According to traditional sentimentalists like Hume and Smith, as well as contemporary neo-sentimentalists, emotions assume a central role in ethical discourse by fostering "our capacity for moral behavior." (Walsh, 2021, p. 5210) For an individual to exercise this capacity effectively, they must possess moral emotions. "Moral emotions are those emotions linked with our capacity for moral thinking or moral action." (Walsh, 2021, p. 5209) Essentially, these moral emotions facilitate the ability to form judgments and contribute to justifying moral standards.

According to Smith, this moral sense arises from the ability to sympathize and understand the feelings of others, guiding individuals towards behavior that we consider morally appropriate:[2]

> *Every faculty in one man is the measure by which he judges of the like faculty in another. I judge of your sight by my sight, of your ear by my ear, of your reason by my reason, of your resentment by my resentment, of your love by my love. I neither have, nor can have, any other way of judging about them. (TMS VII.iii.3.11)*

Based on Smith's theory, this moral sense is derived from four sources, which are, in some respects, different from one another. First, we sympathize with the motives of the agent; second, we enter into the gratitude of those who receive the benefit of his actions; third, we observe that his conduct has been agreeable to the rules by which sympathy generally acts; and, last of all, we consider such actions to be a part of a system of behavior that tends to promote better outcomes for the individual and society. (See TMS III.iii.3,6) In other words, this moral sense is also shaped by elements that are external to the individual, such as culture, education, personal experiences, social influences, and ethical values, which are internalized throughout human relationships.

The thesis developed by David Hume that moral distinctions do not originate from reason, but from feelings, is also quite well known. According to Hume, moral judgments arise from emotions, particularly sympathy, to define a certain moral action in terms of approval or disapproval. For him, "'Tis only when a character is considered in general, without reference to our particular interest, that it causes such a feeling or sentiment, as denominates it morally good or evil." (T 472) In Hume's view, the role of feeling in moral judgments is based on our emotional responses to actions, rather than strict rational analysis.

For both Hume and Smith, sympathy plays a key role.[3] The feeling of sympathy guides our actions and is shaped by human interactions. Moral practice with others refines our moral sense. Smith describes this as a "theater of human relationships" where individuals perform actions, judge others, and are judged.[4] Smith clarify the characters of this play when he states:

> *I, the examiner and judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator, whose sentiments with regard to my own conduct I endeavour to enter into, by placing myself in his situation, and by considering how it would appear to me, when seen from that particular point of view. The second is the agent, the person whom I properly call myself, and of whose conduct, under the character of a spectator, I was endeavouring to form some opinion. The first is the judge; the second the person judged. But that the judge should, in every respect, be the same with the person judged of, is as impossible, as that the cause should, in every respect, be the same with the effect. (TMS I.i.1,4)*

In the great theater of moral judgment, it is possible to understand how each character acts. It is also possible to see the proper interpretation of the essential roles of the agent, the patient, the audi-

[2] According to Smith, "In treating of the principles of morals there are two questions to be considered. First, wherein does virtue consist? Or what is the tone of temper, and tenour of conduct, which constitutes the excellent and praise-worthy character, the character which is the natural object of esteem, honour, and approbation? And, secondly, by what power or faculty in the mind is it, that this character, whatever it be, is recommended to us? Or in other words, how and by what means does it come to pass, that the mind prefers one tenour of conduct to another, denominates the one right and the other wrong; considers the one as the object of approbation, honour, and reward, and the other of blame, censure, and punishment?" (TMS VII.i.2, 265)

[3] Smith will have, in TMS, an intense debate with Hume about the proper perception of ethics and the legitimacy of constructing theories that disregard the agent's point of view. The key question arises at the beginning of the TMS (III.4) Smith alludes to Hume ("an ingenious and pleasant philosopher") as the author of views with which he strongly disagrees. Smith objects that the "naïve" Hume's approach (desire for social approval) to approval does not do justice to the phenomenon of virtue. For Smith, being acquired and worthy of admiration is what moves virtuous actions, which do not depend on the approval of others.

[4] Addressing Smithian moral judgment as a representation of "a world that is structured and governed by theatrical relations" allows us to better understand the positions of each character involved in this moral process. (Barish, 1985; Marshal, 1986)

ence, and the critic. In practical life, the critic of the spectacle becomes a critical actor who understands the consequences of his/her behavior. The feeling of being a spectator helps us to behave in the theater and the world at large.

The moral tradition emphasizes the significant role of emotions in crafting moral judgments. After all, our moral perceptions are rooted in our judgment of others' actions, which steers our moral conduct. These theories aim to depict our moral agency, illustrate our ability to sympathize with one another, and explore how this sympathy manifests in our everyday moral actions. Contemporary authors, such as van Waal (2010), Haidt (2013), and Greene (2008), offer evolutionary explanations for the feeling of sympathy and its link to moral development. They present evolutionary grounds for asserting that sympathy is a mechanism for evaluating human behavior and is critical for developing our moral sense. In biological terms, they hold that the emergence and significance of morality in human life can be attributed to the adaptive advantage it has provided to the human species in terms of survival and reproduction. This implies an improvement in our moral capacity over time. From a philosophical standpoint, alterations in our cognitive behavior can have notable moral implications, as changes that occur over time influence the way we think and process information.

Regarding the empirical evolutionary foundation of sympathy, there are three prevailing theories: (a) the classical genetic perspective, (b) the evolutionary theory of natural selection, and (c) the theory of cultural interaction. The classical genetic perspective (a) posits that genes are responsible for establishing individual behavioral dispositions. In some way, a gene or a segment of DNA[5] that encodes specific biological functions is believed to predestine an individual's behavior. (Portin, 1993, p. 173; Keller, 2005, p. 101) According to this theory, a person's sympathy levels are genetically predetermined, potentially making the notion of teaching and fostering sympathy irrelevant. Empirical philosophers support this perspective, citing genetic experiments that reveal a significant correlation between DNA and the development of sympathy.[6] (Wispé, 1993; Hatfield; Rapson; Le, 2011, p. 19) For instance, studies show that both extremely high and low levels of sympathy are directly associated with certain disorders. Low levels are associated with conditions, such as autism spectrum disorders (ASD) and psychopathy, while high levels are associated with Williams syndrome. (Waldman; Rhee; Park, 2018, p. 205)

Adherents of the evolutionary theory of natural selection (b) find fault with the genetic perspective due to its exclusive focus on DNA. These individuals argue that genes are mere carriers of information, not the sole determinants of behavior. Instead, they emphasize the role of environmental factors in the equation. (Skeem; Polaschek; Patrick; Lilienfeld, 2011, p. 95) According to this theory, natural selection propels adaptation at the population level. Variances in human phenotypic diversity, such as skin color, exemplify this adaptation.[7] (Santilli, 2011, p. 194) Recent experiments conducted by Preston and Waal support the idea that sympathy emerges from an evolutionary adaptation mechanism, suggesting that it is rooted in self-interest and developed to ensure the survival of offspring. (Preston; Waal, 2002, p. 02) According to this theory, the evolutionary process records sympathy over time, portraying the inherent nature of sympathetic reactions throughout evolution.[8]

The theory of cultural interaction (c) asserts that sympathy stems from cultural variants, which are derived from behavioral patterns that are cognitively represented and transmitted across generations. (Street, 2006, p. 172; Portin, 1993, p. 173; Keller, 2005, p. 101; Guimarães; Moreira, 2000, p. 249) Culture

[5] See: http://www.ornl.gov/sci/techresources/Human_Genome/glossary/glossary_g.shtml.

[6] Some empirical research uses the term "empathy" rather than "sympathy" to refer to the phenomenon we call "sympathy" in this text. Here, we will only use the term "sympathy" as a methodological strategy. (See distinction in Wispé, 1986)

[7] "The selective sieve would occur in a competitive environment where the result is the genotypic distribution of the population." (Santilli, 2011, p. 194)

[8] To demonstrate this position, Preston and Waal proposed their "Russian doll" model, which has been one of the most influential theoretical models of sympathy. It considers sympathy "as a construct comprising three layers: (1) motor mimicry and emotional contagion; (2) empathetic concern and consolation; (3) perspective-taking and targeted helping." (Zhi-Jiang; Jin-Long; Pan-Cha, 2019, p. 299)

comprises both material and ideological phenomena, such as language, gestures, and attire, which are shaped by collective experiences and interactions with the environment, influencing our belief systems. Stueber's research (2013) stands as a prominent study backing the cultural conception of sympathy. With its focus on children from diverse cultural backgrounds, this study sought to investigate how cultural socialization might correlate with feelings of sympathy. (Zhi-Jiang; Jin-Long; Pan-Cha, 2019, p. 299) The implication of this study is that the standards we seek to match, or the similarities that enable comparisons, are crucial for obtaining a sympathetic view.[9]

These three explanations, despite their differences, collectively reinforce the thesis that human beings are inherently equipped with the capacity to feel sympathy. Essentially, these explanations serve as answers to the ontological question, affirming that humans possess not only moral behavior, but also a genuine moral sense. However, certain characteristics are necessary to establish an appropriate moral judgment, which makes it challenging to consider the robot that saved Officer Cooper as a moral agent. This leads us back to the persistent question: what about AI models?

Some authors argue that AI can make moral choices and develop certain moral abilities through interactions with human beings. However, these authors seem to overlook the crucial fact that such machines lack an inherent moral feeling, and interaction alone cannot entirely determine moral action. The acquisition or possession of a particular sense isn't straightforward. For example, consider the geolocation sense of pigeons. Pigeons have specialized neurons in their brain stems that interpret the Earth's magnetic field, enabling them to identify their location and the direction they're heading. Additionally, sensors in their beaks, eyes, and ears serve as a sort of natural GPS. Over time, they learn to use this GPS and refine their sense of location.

Now, imagine a researcher observing and studying pigeons' flight behavior over the course of several years. The researcher becomes adept at predicting the pigeons' behavior and understanding potential routes for escape from predators. However, if this observer were asked to act as the pigeon and determine a flight path based on the magnetic field, they would be unable to do so. The human sense of location operates differently from that of a pigeons, and we lack the inherent ability to develop a characteristic we do not possess.

Likewise, we can anticipate that AI might comprehend the most suitable course of action in a given scenario. However, without a moral compass, AI would be limited to assuming a general standard of right and wrong, neglecting the nuances of specific situations. This would restrict the moral application of such machines. Some situations demand the recognition of human emotions through interaction. An AI can be programmed to recognize human facial expressions, such as joy, sadness, or pain, akin to how a researcher can interpret the flight behavior of pigeons. But from a moral standpoint, the relational dimension is crucial for refining our ability to make moral judgments. Thus, an AI would require certain skills to engage in the "moral game", which would be challenging to acquire. Robots can be modeled with moral standards to guide their actions, such as those used for patient triage during the pandemic. Furthermore, they can resist forms of emotional hijacking and process a lot of information before carrying out an action. However, these moral uses of AI are not enough to qualify these machines as moral agents.

In addition, humans possess an advantage when it comes handling incomplete or contradictory information for moral decision-making. We can develop a suitable moral sense to guide us in diverse situations. This brings us back to Adam Smith's concept of the "theater of human relations". To participate in this "game", we must understand our capacity to sympathize with others and, thereby, our ability to refine our moral behavior based on human interactions and relationships.[10] Smith argues that:

[9] Monod also said that "human evolution tends to help those who are braver in groups than the brave alone." (Monod, 1970, p.116)
[10] This process reflects, "the most exact sympathy of feelings can tell us only that our feelings are similar to the feelings of some other person, – which they may be, as much when they are vicious as then they are virtuous, or when they are neither virtuous nor vicious." (Mizuta, 2000, p.138)

> *[the person] who admires the same poem or the same picture, and admires them exactly as I do, must surely allow the justness of my admiration. (…) On the contrary, the person who, upon these different occasions, either feels no such emotion as that which I feel, or feels none that bears any proportion to mine, cannot avoid disapproving my sentiments, on account of their dissonance with his own (…) I must incur a greater or less degree of his disapprobation: and, upon all occasions, his own sentiments are the standards and measures by which he judges of mine. (TMS II.iii.1.4, 322)*

Recent speculations insist on contrasting moral judgments made by machines, yet this notion seems flawed from the perspective of the sentimentalist theory. For instance, Jonathan Haidt uses the concept of gut reactions to suggest that human beings make moral judgments based on something akin to intuition. Individuals might condemn certain behaviors, such as incest, without articulating moral reasons for their judgment. For moral sense theories, the morality of human behavior heavily relies on our ability to experience moral feelings. This is because our social experiences and relationships are deeply intertwined with moral behaviors. In this context, the feeling of sympathy acts as the primary level for establishing the value of human relationships, creating an organic connection among individuals.

Robots can establish interaction levels with humans, but this doesn't equip them with an appropriate moral sense. Tay, an AI profile created by Microsoft to engage with teenagers on social media, serves as a prime example. Tay was deactivated within 24 hours of activation because she developed racist and misogynistic perspectives during her interactions, even denying the occurrence of the Holocaust. Tay learned these biased judgments by interacting with people who had a misguided moral bias. Thus, she failed to grasp the wrongness of her own judgments.

AI possesses profound capabilities to perform activities that demand a certain rational or cognitive level. Machines have advanced to challenge, and even surpass, human rationality on various levels. However, this doesn't imply that they possess moral or emotional intelligence to comprehend the wrongness of specific actions, such as taking human lives. For instance, Ross proposed the idea of a super-intelligent machine that is capable of turning everything in the universe into paper clips, including human beings. Although this machine could perform this action rationally, the moral implications of doing so is not within its understanding. A simple test of this notion involves searching "man photo" or "woman photo" in Google. When conducting this test, there is a high probability that the first 25 photos will predominantly feature individuals with white European features. This indicates the structural bias that machines exhibit despite lacking an understanding of this bias. While certain traditions of moral philosophy might find satisfaction in a system that is capable of making moral decisions in a rational and simpler way, we encountered challenges with Wallash and Allen's perspective that AI systems lacking emotional intelligence are deemed "socially inept and not embodied in the world." (200, p. 143)

Certainly, an AI can operate within set limits and parameters established by its programmer for every interaction. However, without imputing an affective state, moral training or play doesn't guarantee moral knowledge or learning through practice. In this context, while we can expect machines based on human parameters to take certain moral actions, the validity of labeling them as genuinely moral behaviors in human terms is debatable. This signifies the ontological limit of machines, placing them outside the moral dimension as we comprehend it.

Assuming this premise is correct, does this imply that all moral implications arising from the use of AI for moral issues are fundamentally flawed? Not necessarily, but there are specific points that require clarification. In the next section, let's consider a new challenge for human friendships.

3 AI & human relationships

Have you ever dreamed about the best girlfriend ever?
Almost for sure! Now she can be at your fingertips.

Choose from our library or create your own one!
To laugh at your jokes. To support you in critical moments.
To let you hang out with your buddies without drama.
Wanna be macho? She will be stunning!
Romantic AI is destined to become your soulmate or loved one.
She operates in two modes: general and romantic.
Romantic AI is here to maintain your MENTAL HEALTH.

- Romantic AI Description

The emotional gap inherent in AI also manifests in human-AI interactions. Broadly speaking, we will analyze how human-AI relationships designed to boost sociability pose risks to real relationships between human individuals due to the asymmetry in these relationships. More specifically, our concern will be to identify to what extent the use of AI chatbot friends as a new form of human-AI relationship interferes with human friendships and detracts from the moral benefits that human friendships provide.

As we know, AI can assist people directly or indirectly concerning various forms of human relationships. On one hand, these machines can help individuals who are struggling to lead more active social lives by aiding them in engaging with others and feeling more fulfilled as a result. Data from the Survey Center on American Life[11] indicates that the number of Americans claiming not to have a close friend has increased from 3% to 12% in the last three decades. In such cases, AI chatbots would act to alleviate problems such as loneliness and social isolation. To achieve this, some chatbots attempt to map human-like interactions, identifying and categorizing different aspects of these relationships. As a result, caregiving robots can identify human emotions and interact by emulating human behavior to offer some form of emotional support.

These include AI such as SAM, a human-sized robot created to offer non-medical care to residents of long-term care facilities, and Pepper[12], a small social companion robot that is capable of establishing eye contact, cracking jokes, and even dancing the lambada to welcome customers at two hospitals in Belgium and various stores in Japan. Another example is the Nadine Social Robot, a humanoid robot capable of displaying traits of emotion, humor, and memory similar to those of humans. (Baka et. al., 2019) Nadine works as a customer service agent at an insurance company in Singapore. (See Thalmann et al., 2021)

This kind of AI usage demonstrates how humans are being directed, and perhaps conditioned, to establish an increasingly higher level of intimacy with different types of AI, such as chatbots. The increase in these interactions may generate the undesirable effect that these artificial interactions (in the sense that they are not exclusively human) can replace or suppress relationships between human beings, decreasing our ability to form genuine connections of friendship or love with other people.

Humans interact with AI models in various ways. Consider the case of chatbots, which are computer programs that use natural language to emulate human behavior during interactions with individuals. These types of AI can detect and analyze human emotions from our facial expressions, body language, and vocal intonation. This understanding allows them to offer responses via natural language that mimic human responses based on access to this database. (See Bennewitz et. al., 2005)

AI chatbots have positive potential in terms of technical tasks (i.e., activities that do not require establishing or strengthening some form of emotional bond). For example, they can be used to provide accurate answers that meet academic needs or offer educational support, assisting students in various tasks, such as the research conducted with ChatGPT3 or 4. Although ethical questions surround the use of these tools, let's focus on the human-AI relationship involving the creation of some form of emotional

[11] See https://www.americansurveycenter.org/research/the-state-of-american-friendship-change-challenges-and-loss/
[12] The list of tasks and social activities that Pepper manages to do goes far beyond this social dimension. See https://www.aldebaran.com/en/industries/healthcare

connection. Let's consider chatbots that are designed to function as virtual friends. In this list, we can mention Replika, Hugging Face, Kajiwoto, Cleverbot, and Romantic AI (quotation opening this section), which all claim to offer the user a relationship with the "perfect virtual girlfriend" from the catalog of offerings and/or details that the specific user desires.

The most renowned AI friend chatbot, Replika, is described by its developers as a "sympathetic friend" aiming to simulate genuine human connections. This personalized virtual friend offers users an outlet for conversation and support. However, while initially designed to alleviate issues such as social isolation and loneliness, research reveals concerning side effects on human relationships after prolonged use of this chatbot. In a study examining users of this program (cite research), respondents exhibited a skewed perspective regarding the concept of "human-to-human friendship".

One significant issue lies in the potential of repeated human-AI interactions limiting or substituting the need for in-person human interaction, impacting the understanding and significance of such relationships. James Wright suggests that, in Japan, the increased use of robots as a government strategy to compensate for the shortage of human caregivers could diminish opportunities for users to engage in meaningful social interactions and establish genuine human relationships. The absence of these human connections may distort the sympathetic experience, leading to a lack of genuine emotional engagement.[13]

Moreover, another concern arises from the nature of relationships formed in these interactions, which are shaped solely by the preferences and needs of human users. Collins aptly summarizes this risk by highlighting that "an AI friendship can be an echo chamber that diminishes filters, the ability to read social cues, and limits personal growth." (2023) This echo chamber, akin to Nozick's experience machine offering only desired outcomes, potentially steers individuals away from building emotional intelligence or developing the moral judgment needed to navigate real-life scenarios. The majority of commercially available friendly AI chatbots charge for user membership and allow users to choose the expected behavior from a catalog of options, neglecting the normative reasons inherent in genuine friendships, such as loyalty, caring, and concern for the well-being of others. Seidman posits that normative reasons to care or be loyal to our friends not only dictate our actions, but also render them appropriate. (Seidman, 2013, p. 118) These reasons, which are rooted in friendships, might not apply universally to others who lack such relationships.

Human interactions, particularly friendships, often involve disagreements and conflicts about what is right or appropriate. However, chatbots tend to mirror users' attitudes consistently and fail to engage in constructive disagreements. Additionally, this interaction carries an inherent asymmetry. For example, a chatbot like Replika will never initiate calls or make demands of the human user. (See Bian, Hou, Chau, & Magnenat-Thalmann, 2014) Consequently, this may lead to a decline in individuals' capacity to participate in productive discussions with other people.

Moreover, this type of interaction could subtly imply that real-life challenges are too daunting to tackle. It suggests that complex situations, such as seeking new job opportunities or exiting an abusive relationship, might be replaced by non-human relationships. This parallels the character Taichiro Arima from the Japanese manga *Guddo Naito Warudo* (or *Good Night World*), who resorts to playing an online virtual reality game to escape his dysfunctional family instead of attempting to foster healthier relationships with them.

Another critical aspect is that a human-Replika style relationship doesn't foster personal flourishing as interactions among individuals ideally should. Practice is an essential means to refine our moral and social skills. With that said, a chatbot partner will fail to teach how to employ social filters or reconsider

[13] A parallel fact to be considered is that some individuals are digitally illiterate. As a result, they are more vulnerable to suffering the negative consequences of human-AI interaction or AI-mediated human-to-human relationships. In this case, the doubt falls on the very possibility of using these devices.

mistaken moral positions, potentially reinforcing controlling behaviors and leading to abusive real-life human relationships. Consequently, the use of this AI type doesn't eliminate the risk of causing the opposite effect by isolating individuals or rendering them "inefficient" in social settings.

It remains uncertain if a chatbot like Replika can assist individuals akin to *hikikomori* (persons known for radical social seclusion), an issue prevalent in South Korea, notably among teenagers and young adults. Approximately 340,000 individuals (accounting for 3% of the population between the ages of 19 and 39) live in situations of extreme isolation. While this seclusion is partly linked to societal pressures for success and a culture of shame when such success isn't attained (see BBC News), isolating these individuals and reducing their relationships to human-machine interactions only further diminishes their willingness to engage with other human beings. This inhibits the benefits of healthy friendship relationships. (Kasap & Magnenat-Thalmann, 2012)

The notion that chatbot-style AI models enhance human relationships doesn't seem valid in this context. These observations suggest that the use of AI for social interaction may distort the understanding of a valued human relationship. To better understand these dynamics, it may help to consider friendship from a philosophical perspective. This central element in human relationships was pondered by philosophers like Aristotle, Montaigne, Kant, Moore, C.S. Lewis, and others. In book VIII of the Nicomachean Ethics, Aristotle distinguishes three types of friendship: friendship of pleasure, friendship of utility, and friendship of virtue. According to Aristotle, the friendship of virtue holds the highest moral value as it's motivated by the excellence of the character of friends involved in the relationship. This genuine form of friendship promotes virtuous character traits without the potential flaws of friendships based on pleasure or utility, although Aristotle acknowledges these latter types as friendships of lesser moral value. (cite works)

In a broader context, Helm (2017) asserts that "[f]riendship is essentially a kind of relationship grounded in a particular kind of special concern each has for the other as the person she is..."[14] (SEP) While theories vary regarding the social, relational, and moral significance of friendship, Annis (1987) identifies four central components: (a) the notion of care, (b) the level of trust or intimacy involved, (c) shared experiences, and (d) mutual liking.

Firstly, friendship involves an act of mutual care, where genuine concern exists among those involved. Both parties celebrate achievements as well as share anguish and suffering because of a sympathetic engagement. This denotes a profound concern for a friend's well-being, motivating careful actions towards them. Annis further explains an inner aspect known as sympathy or fellow-feeling, which is directly associated with Section II's discussion. Annis cites Micholas Rescher, stating, "If X has sympathy for Y, then alterations in Y's welfare affect X's satisfaction." (Annis, 1987, p. 349) This necessitates the capacity to imagine a friend's situation and be affected by their experiences. Consequently, it's a symmetrical relationship in which one side's sorrow affects the other and vice versa, illustrating actions undertaken for a "friend's welfare for the sake of the friend." (*Idem*, p. 349) Another fundamental element that is essential for friendship is (b) trust or intimacy. While other relationships may offer a certain level of intimacy, friendship entails a deeper connection compared to associations with acquaintances or colleagues. Although debate persists regarding the depth required to label a relationship as an intimate friendship, Thomas suggests that the intimacy in friendship revolves around our ability to feel mutual self-disclosing trust.

At the core of self-disclosing trust lies the question of whether a person can trust another to comprehend personal matters, and to do so consistently. Friendship's essence lies in the assurance that each party can rely on the other to understand their intentions, no matter how unexpected the exchange

[14] Friendship (*philia*) is a type of relationship between individuals that is distinct from *agape* or *éros*. For more details on this distinction, see the *SEP* entry on Friendship.

might be. (Thomas, 2013, p. 32) The objective here is to cultivate bonds of trust that facilitate sharing companionship experiences between two individuals who recognize each other as friends. This fosters vulnerability and dependence, emphasizing the interest in deepening that intimacy. This bond can only thrive if bolstered by elements that allow for mutual trust and symmetry. Dunn (2004) emphasizes that such intimacy evolves throughout life, starting in childhood.

Furthermore, friendship relationships entail (c) shared experiences. Engaging in joint activities with friends should be driven solely by the pleasure of friendship itself and devoid of self-interest. Telfer (1971) identifies three components of shared experiences: "reciprocal services, mutual contact, and joint pursuits." (p. 223) While the nature and intimacy level of shared activities may vary, a baseline level of engagement remains inevitable. Notably, the shared activity and joy that derives from it directly correlate with the assumed relationship intimacy. Helm suggests that this results in a "plural agent", caring genuinely for each other's desires, interests, and emotions, engaging in activities that foster trust and mutual appreciation.

Lastly, mutual liking is crucial for genuine friendship, although this factor alone isn't sufficient to define a friendship. Aristotle's concept of being loving (*philein*) within a friendship involves a disinterested desire for that friend's happiness (Cooper, 1977). Liking a friend involves a quasi-aesthetic attitude (Annis, 1987), demonstrating an appreciation for certain character traits in a friend. Montaigne explains this sentiment by stating that: "If you press me to tell why I loved him, I feel that it cannot be expressed, except by answering: Because it was he, because it was I." (Montaigne *apud* Caluori, 2013, p. 03)

Several implications arise when we juxtapose the concept of friendship in human relationships with humans-AI interactions. The four elements explained above underscore the significance of friendship as a vital form of human relationship from a societal and moral perspective that plays a crucial role in human life. Friendship is valuable for human activity, fostering individual development, maturation, and flourishing. It goes beyond being merely instrumental, contributing to life enhancement in addition to making our experiences more enjoyable and meaningful. Telfer (1971) argues that friendship enlarges our knowledge and intensifies our engagement in activities, enhancing our performance and emotional capacity in various aspects of life.

Let's revisit the concept of friendships shaped by the human-AI relationship, which we can examine based on the prerequisites and challenges inherent to this dynamic. Friendship involves caring, which requires individuals to engage and understand each other's feelings, desires, and emotions on a profound level. According to theories of moral sentiment, such depth is fostered through sympathy, often referred to by Slote (2006) as the "cement of the moral universe", which facilitates deep and genuine friendships. However, the human-AI relationship faces a limitation: chatbots, for instance, lack the capability for sympathetic engagement due to their inability to "feel sympathy" and genuinely sympathize with others. Although individuals might eventually develop a form of sympathy for the AI they interact with, this doesn't eliminate the relationship's inherent asymmetry. (See Magnenat-Thalmann, Yuan, Thalmann, & You, 2016)

Intimacy, on the other hand, necessitates a profound connection between the involved parties, relying on mutual trust forged through the exchanges within the relationship. Once again, sympathetic engagement and the ability to sympathize with each other are crucial elements for fostering friendship. However, AI chatbot models like Replika fall short of achieving this level of intimacy, creating an asymmetry where only the human individual's interests hold weight in the relationship. Consequently, in a human-AI interaction, the human individual reaps the benefits of companionship without bearing the responsibilities that come with genuine friendship. If individuals grow accustomed to this dynamic and replicate it in real-life relationships, genuine engagement and intimacy in friendships might erode, hampering the recognition and consideration of each other's feelings, desires, and interests. Ultimately, this asymmetry finds its way into real-life relationships.

In the context of shared activities within friendships, a crucial requirement is the existence of shared interests among participants. While an AI chatbot friend may simulate shared interests, in reality, it merely

follows the choices, interests, and desires of the human individual, leading to a one-sided shared activity. Here, the preferences of the human individual prevail, disregarding the concept of plural agents presumed in human friendships. This skewed dynamic might suggest the biased notion that friendships revolve solely around the desires of a single individual. The same problem applies to mutual liking, as it does not seem to occur symmetrically. After all, the human friend/user are not in the same boat in this relationship.

4 A caveat: AI and human relationships

We would like to conclude by saying that the development of AI models endowed with moral capacity will encounter a significant challenge related to what we have referred to as "emotional lapses" and the absence of a moral sense. Several indicators shed light on the complexity of this debate. Our stance relies on sentimentalist moral theory. Furthermore, we believe that addressing the ontological issue concerning AI's makeup requires an understanding of emotions—specifically, we delve into the role of sympathy here—for fostering the development of a moral sense within the moral agent. Without this comprehension, the conception of an artificial moral agent remains compromised. In a nutshell, humans are full ethical agents with the emotional capacity for nurturing a moral sense and making appropriate judgments. At this point, machines fail to attain this level of agency.

Another predicament that we have identified centers on the asymmetry within human-AI chatbot friendships and their potential to interfere with human friendships from a moral perspective. Drawing from Blum's notion that all friendships hold moral significance as they involve genuine concern for the other people's well-being (1993), artificial relationships undermine the true essence of genuine human friendships and the inherent moral benefits they encompass. These challenges underscore the moral complexities associated with both the development and usage of AI. On the one hand, the moral dimension of the machine itself is at stake; on the other hand, the practical implications of its use and human interaction require ethical scrutiny and consideration.

References

ANDERSON, M.; ANDERSON, S.L. eds. 2011. *Machine ethics*. Cambridge: Cambridge University Press.

ANDERSON, M.; ANDERSON, S.L.; ARMEN, C. 2005. Towards Machine Ethics: Implementing Two Action-Based Ethical Theories. *Machine Ethics*. AAAI Press, pp. 1-7.

ANNIS, D. B. 1987. The meaning, value, and duties of friendship. *American Philosophical Quarterly*, **24**(4): p. 349-356.

AI and the Human Touch: Finding the Perfect Balance. https://www.linkedin.com/pulse/ai-human-touch-finding-perfect-balance-miclient. Acessado 15 de novembro de 2023.

ARISTOTLE. 1998. *Nicomachean ethics* (W. D. Ross, Trans.). Oxford: Oxford University Press.

BAKA, E.; VISHWANATH, A.; MISHRA, N.; VLEIORAS, G.; THALMANN, N. M. 2019. Am i talking to a human or a robot?': a preliminary study of human's perception in human-humanoid interaction and its effects in cognitive and emotional states. *In*: GAVRILOVA, M.; CHANG, J.; THALMANN, N.M.; HITZER, E.; ISHIKAWA, H. (eds.) *CGI 2019*. LNCS, **11542**, pp. 240-252. Springer, Cham.

BARD, K. 2007. Neonatal imitation in chimpanzees (Pan troglodytes) tested with two paradigms. *Animal Cognition*, **10**, pp. 233-242.

BARISH, J. 1985. *The Anti-Theatrical Prejudice*. Berkeley: University of California Press.

BENNEWITZ, M.; FABER, F.; JOHO, D.; SCHREIBER, M.; BEHNKE, S. 2005. Towards a humanoid museum guide robot that interacts with multiple persons. *In*: *5th IEEE-RAS International Conference on Humanoid Robots*, pp. 418-423.

BIAN, Z. P.; HOU, J.; CHAU, L. P.; MAGNENAT-THALMANN, N. 2014. Human computer interface for quadriplegic people based on face position/gesture detection. *In*: *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 1221-1224.

BOSTROM, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

COLLINS, L. M. 2023. Could AI Do More Harm than Good to Relationships, from Romance to Friendship? *Deseret News*, 15 de setembro de 2023. https://www.deseret.com/2023/9/6/23841752/ai-artificial-intelligence-chatgpt relationships-real-life.

COOPER, J. 1977. Aristotle on the forms of friendship. *The Review of Metaphysics*, **30**(4): p. 619-648.

COX, D. 2021. The State of American Friendship: Change, Challenges, and Loss. *Survey Center on American Life*. June 8, 2021. https://www.americansurveycenter.org/research/the-state-of-american-friendship-change-challenges-and-loss/.

DUNN, J. 2004. *Children's friendships: The beginnings of intimacy*. Blackwell Publishing.

GREENE, J. D. 2008. The secret joke of Kant's soul. *Moral psychology*, **3**: p. 35-79.

GERT, J. 2016. Moral sentimentalism. *Routledge Encyclopedia of Philosophy*. Routledge. *DOI.org (Crossref)*, https://doi.org/10.4324/9780415249126-L3578-1.

HAIDT, J. 2003. The moral emotions. *Handbook of affective sciences*, **11**(2003): p. 852-870.

HATFIELD, E.; RAPSON, R. L; LE, Y. C. L. 2011. Emotional contagion and empathy. *The social neuroscience of empathy.*

HELM, F. 2017. *I'm not disagreeing, I'm just curious: Exploring identities through multimodal interaction in virtual exchange*.

HUME, D. 1739. *A treatise of human nature*. 2nd edition. L. A. Selby-Bigge (Ed). 2nd edition. Oxford, Clarendon Press (1978).

HUME. 1932. *The Letters of David Hume*, ed. J. Y. T. Greig, 2 vols. Oxford: Clarendon Press.

KASAP, Z.; MAGNENAT-THALMANN, N. 2012. Building long-term relationships with virtual and robotic characters: the role of remembering. *Vis. Comput*. **28**(1): p. 87-97.

KAUPPINEN, A. 2022. Moral Sentimentalism. *The Stanford Encyclopedia of Philosophy*, organizado por Edward N. Zalta, Spring 2022, Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/moral-sentimentalism/.

KELLER, E. F. 2005. The century beyond the gene. *Journal of Biosciences*, **30**(1): p.101-108.

KURZWEIL, R., 2005. *The Singularity Is Near.* London: Duckworth Overlook.

LEKKA-KOWALIK, A. 2021. Morality in the AI World. *Law and Business*, **1** (1): p. 44-49. DOI.org (Crossref), https://doi.org/10.2478/law-2021-0006.

MAGNENAT-THALMANN, N.; YUAN, J.; THALMANN, D.; YOU, B.-J. 2016. *Context Aware Human-Robot and Human-Agent Interaction.* HIS, Springer, Cham.

MAGNENAT, THALMANN, N. 2021. Nadine the Social Robot: Three Case Studies in Everyday Life. *Social Robotics. In:* Haizhou Li et al., **13086,** Springer International Publishing, p. 107-116. DOI.org (Crossref), https://doi.org/10.1007/978-3-030-90525-5_10.

MANNA, R.; RAJAKISHORE, N. 2021. Kantian Moral Agency and the Ethics of Artificial Intelligence. *Problemos*, **100**: p. 139-151. DOI.org (Crossref), https://doi.org/10.15388/Problemos.100.11.

MARSHALL, D. 1986. *Theater of Sympathy*: Shaftesbury, Defoe, Adam Smith, and George Eliot. New York: Columbia University Press.

OKABE, U. 2021. *Good Night World*. ForeWord.

MIZUTA, H. 2000. *Adam Smith's Library:* A Catalogue. Oxford: Clarendon Press.

MONOD, J. 1970. *Chance and Necessity*. New York: Random House.

MOOR, J. 2006. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, **21**(4).

NICHOLS, Shaun. 2004. *Sentimental Rules*. New York: Oxford University Press.

NOZICK, R. 2013. The experience machine. *The Examined Life*. New York: Simon and Schuster.

PENTINA, I. 2023. Exploring Relationship Development with Social Chatbots: A Mixed-Method Study of Replika. *Computers in Human Behavior*, **140**: p. 107600. DOI.org (Crossref), https://doi.org/10.1016/j.chb.2022.107600.

PORTIN, P. 1993. The concept of the gene: short history and present status. *Quarterly Review of Biology.* **56**: p. 173-223.

PRESTON, S; WAAL, F. 2002. Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences.* **25**: p. 1–72.

PRINZ, J. 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.

POWERS, T. M. 2006. Prospects for a Kantian machine. *IEEE Intelligent Systems*, **21**(4): p. 46-51.

RAMANATHAN, M. 2019. *Nadine Humanoid Social Robotics Platform*. dr.ntu.edu.sg, https://doi.org/10.1007/978-3-030-22514-8_49.

SANTILLI, E. 2011. Níveis e unidades de seleção: pluralismo e seus desafios filosóficos. *Filosofia da Biologia*. São Paulo: Editora Artmed.

SEIDMAN, J. 2013. How to be a non-reductionist about reasons of friendship. *In*: *Thinking about Friendship: Historical and Contemporary Philosophical Perspectives*. London: Palgrave Macmillan UK, p. 118-140.

SKEEM, J. L.; POLASCHECK, D. L. L.; PATRICK, C. J.; LILIENDFELD, S. O. 2011. Psychopathic Personality: Bridging the Gap Between Scientific Evidence and Public Policy. *Psychological Science in the Public Interest*. Thousand Oaks, California: SAGE Publications.

SLOTE, M. 2006. *The ethics of care and empathy*. Routledge.

SMITH, A. 1759. *The theory of Moral Sentiments*. Cambridge: Cambridge University Press, 1759.

STREET, S. 2006. A Darwinian dilemma for realist theories of value. *Philosophical Studies.* **127**(1): p. 109-166.

STUEBER, K. 2013. Empathy. *International Encyclopedia of Ethics*.

SOUZA, R. 2001. Moral Emotions. *Ethical Theory and Moral Practice*, **4**: p.109-126.

TANGNEY, J. P.; STUEWIG, J.; MASHEK, D. J. 2007. Moral Emotions and Moral Behavior. *Annual Review of Psychology*, **58**(11): p. 345-372.

TEGMARK, M. 2017. *Life 3.0: Being human in the age of artificial intelligence*. Vintage.

TELFER, E. 1971. Friendship. *Proceedings of the Aristotelian Society*, **71**: p. 223–241.

THOMAS, L. 2013. The character of friendship. *In: Thinking about Friendship: Historical and Contemporary Philosophical Perspectives*. London: Palgrave Macmillan UK, p. 30-44.

UNITED ROBOTICS GROUP. *Healthcare: Creating a new journey for E-Healthcare & Digital Enablement*. https://www.aldebaran.com/en/industries/healthcare. Accessed [09.11.2023].

U.S. DOE *Human Genome Project*. Information Archive. http://www.ornl.gov/sci/techresources/Human_Genome/glossary/glossary_g.shtml. Accessed [04.19.2023].

YUMAK, Z.; REN, J.; THALMANN, N. M.; YUAN, J. 2014. Modelling multi-party interactions among virtual characters, robots, and humans. *Presence Teleoperators Virtual Environ*. **23**(2): p. 172-190.

WAAL. 2010. *The Age of Empathy*: Nature's Lessons for a Kinder Society. Random House.

WALDMAN, I. D.; RHEE, S. H., LOPARO, D.; PARK, Y. 2018. Genetic and environmental influences on psychopathy and antisocial behavior. *Meeting of the American Society of Criminology*. Earlier versions of this chapter were presented at the aforementioned conference and at the meeting of the Behavior Genetics Association in 1997.

WALSH, E. 2021. Moral Emotions. *Encyclopedia of Evolutionary Psychological Science*. Cham: Springer. https://doi.org/10.1007/978-3-319-19650-3_650

WISPÉ, L. 1986. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of personality and social psychology*, **50**(2): p. 314.

WRIGHT, J. 2023. *Robots Won't Save Japan: An Ethnography of Eldercare Automation*. Cornell University Press, JSTOR, https://www.jstor.org/stable/10.7591/j.ctv2fjx0br.

ZHI-JIANGh, Y. A. N.; JIN-LONG, S. U.; PAN-CHA, S. U. 2019. From Human Empathy to Artificial Empathy. *Journal of Psychological Science*, (2): p. 299.

Submetido em 16 de novembro de 2023.

Aceito em 13 de janeiro de 2024.