Social robotics and deception: beyond the ethical approach

Rachele Carli[1][0000−0002−8689−285X]

Alma Mater Research Institute for Human-Centered AI, University of Bologna, co-tutelle ICR group, University of Luxembourg rachele.carli2@unibo.it

Abstract. Social robots are designed to directly interact with users, to collaborate with them and to act in a human-centred environment, with different degrees of automation. In order to encourage acceptability and trust, they are structured as so to lever the human tendency to anthropomorphise what they interact with. It follows that some machines are able to simulate the feeling of genuine emotions or empathy, to appear needy of help, to pretend to have an own personality and – more in general – to induce the user to think that they are something more than mere objects. Thus, it may be argued that such interaction could lead to forms of manipulation that fall within the remit of a deceptive dynamic. Such a phenomenon is still much debated by the scientific community and raises significant concerns regarding long-term ethical and psychological repercussions on the users.
This paper investigates which tools we have and which ones we may need to tackle the theme of deception in social robotics. Therefore, both ethical and legal perspectives are reconstructed, with the attempt to try to distinguish their respective scope and to emphasise their fruitful integration in addressing these issues. Finally, the possible relevance of fundamental human rights in human-robot interaction dynamics is discussed, due to their ability to reconcile ethical demands with the binding feature of legal norms.

Keywords: HRI · Deception · Human Dignity · Ethics · Law

1 Introduction

The so called "Fourth Revolution" [1] is leading to the development of new technological devices, increasingly interactive and pervasive in many areas of our lives. We are assisting to the migration of robots from factories to our homes, involved in many tasks - ranging from education to health, from entertainment to the care of the most fragile ones - [2]. This explains the growing focus on social robotics. In fact, social robots are characterised by a physical body and a software, which allows them – based on the level of technological advancement – to perceive the environment, to interpret both structured and not structured data, to process them and to extrapolate primitive and derivative pieces of information [3]. Therefore, they are able to directly interact with the users, collaborating

with them on a daily base and performing multiple tasks with different degree of automation. Those which are implemented with machine learning techniques are also able to develop social competences, create social bonds and learn to appropriately use natural signals – like indicating, gazing, winking – [4].

Consolidate scientific literature demonstrates that facilitating acceptability and trust in social robots plays a central role in order for them to pursue the given tasks and to behave efficiently in a human-centered environment [5][6]. Therefore, they are designed and programmed so as to lever the human tendency to attribute anthropomorphic characteristics to what they relate to, repeatedly over time [7].

This might lead to a new form of manipulation, based on deception. In fact, in spite of the pleasant design, the ability to move and act like living beings, to simulate pertinent conversation and to emulate feelings and emotions, robots are just objects [8]. They are not capable to set autonomously a goal and even the most sophisticated functionality is - at least for the moment - the result of the way they have been programmed by a human expert. Such a deceptive dynamic, although not favoured with dishonest intents, may be potentially risky for the physical, economical and psychological integrity of people involved and for the authenticity of their will. Moreover, it should be underlined that the speed of development of these technologies far exceeds the speed with which we are able to investigate negative outcomes. On one hand, possible side-effects - especially on the ethical, psychological and sociological level - have already been theorised. On the other hand, not all of them can be already precisely measurable or unequivocally proved, for we cannot have a long-term picture of the expected consequences yet. To this end, both the ethical and the legal perspectives play a central role. However, the respective scopes of intervention must be identified, in order to guarantee the efficacy of their impact on the theme of deception in social robotics.

This paper aims to investigate the effectiveness of ethics and the law as tools of analysis and to suggest the relevance of fundamental human rights – in particular human dignity – as a valid option to tackle this theme. In fact, they have the advantage to be efficiently used in a multidisciplinary debate, increasing the opportunities to find a common solution for similar cases [9] and to suggest a transparent frame of regulation. This would be functional to guarantee that new technologies (i) are projected so as to respect the centrality of human being, (ii) can be efficiently tested in the real world and (iii) are commercialised in a specifically developed market [10].

To this end, (section 2) recollects some relevant literature about the theme of deception in the human-robot interaction context, in order to give a general understanding of the phenomenon. (Section 3) briefly presents some scenarios in which deception - theorised as in the previous section - can occur. The attention will be focused on two categories of users, children (subsection 3.1) and fragile people (subsection 3.2), trying to highlight which concerns may rise from such a dynamic. (Section 4) evaluates ethics and the law as possible tools to address this theme. In particular, some of their strengths and weaknesses will be separately

analysed in (subsections 4.1 and 4.2). Therefore, (section 5) suggests that human dignity may play a more effective role as balancing principle with a view to risk-benefit examination of social robotics and the deceptive phenomenon.

2 Main forms of deception in human-robot interaction

The ability of a machine to deceive the human counterpart has been considered the qualifying criterion for the very notion of AI. This is due to the well-known "Turing Test", according to which a machine would have been considered as "intelligent" if it was able to induce a person – placed in a different room – to believe to be chatting with another human being, rather than with a robot [11]. However, there is still not a univocal understanding and evaluation of what is meant for deception in the context of human-robot interaction. This is due to the fact that simulation mechanisms depend closely on the nature of the robot, its functionality, the tasks it has to perform, the object of the interaction [12].

In order to better analyse the phenomenon, it is useful to introduce a preliminary tripartition: (i) external state deception, (ii) superficial state deception, (iii) hidden state deception [13].

The first one takes place (i) when the robot lies about something regarding the external world. It could be considered problematic, when it aims to mislead the user, but it can also imply a pro-social function. In fact, social conventions can require "white lies" for several reasons, such as to carry on the conversation, to be polite, to avoid uncomfortable situations or to encounter one's favour. This is possible using hyperbole or pleasantries, improvising not really known pieces of information and managing expectations [14]. Implementing machines with such features means favouring their integration into the human environment and overcoming the prejudices in which they may incur [15].

More challenging is the case of robots which (ii) simulate to possess capabilities and emotional dimensions actually lacking. With regards to this aspect, there are different opinions among the experts. According to the most extreme position any robotic cue that emulates a typically human one is deceitful [16]. This on the base of both technical and philosophical assumptions. On one hand, the behave of a device is evaluated as nothing more than the result of the way it was projected by the programmer [17], being neither aware nor autonomously settled. On the other hand – following the same argumentative line – anything is able to manipulate reality should be considered ethically wrong, for it harms the "duty to see the world as it is" [18]. A more lenient position traces back to the category of deception only those actions that induce the user to perceive the machine as something more than a very sophisticated piece of equipment [19], closer to a living being [20]. Such a function is conveyed not only by gestures or movements, but even more so by the simulation of emotional capabilities [21], the emulation of feelings of pain, suffer, attachment, care [22]. This is certainly functional for the collaboration between the user and the artefact. Nonetheless, it can also affect subconscious social dynamics, interfering with the formation and expression of the individual's will. Moreover, it is important to underline

that such characteristics have more incisiveness on lonely and needy people, the same ones that should be better protected against manipulative mechanisms, for more vulnerable.

The third form of deception occurs when the robot (iii) takes advantage of emulative signals in order to hide capabilities it has. This can lead to harmful consequences for people's privacy and data managing. For example, it was demonstrated that children and the elderly are more likely to confide to a "friendly" robot even things that they would not have revealed otherwise. This because they are persuaded that the machine can keep the secret and because they are unaware of how it can process those confidences in order to target their desires and preferences [23]. In fact, in order to improve both engagement and quality of the interaction, social devices can record actions, words and even emotions [24]. Furthermore, an individual could believe that when the robot is not in the view it cannot record what the person is doing [25], ignoring the presence of sensors that make him/her lives in a sort of "Big Brother" [26]. The lack of full understanding of the effective functionality of the machine could undermine the value of the consent – no more considerable as "informed" – given to the interaction [27]. At the same time, it is not objectively demonstrable that an increase in the information provided will always lead to a greater awareness in the use of the device. In fact, human-robot interaction involves multiple factors, many of which are strictly related to a subjective psychological dimension [28].

Taking into consideration the above-mentioned classification of the main forms of deception in the human-robot context, it is relevant to analyse some concrete scenarios in which such a dynamic may occur. Thus, potential long-term consequences may be analysed, so as to distinguish beneficial and harmful effects. This evaluation is fundamental, in light of the necessity of a human-centred development of AI systems.

3 Possible deceptive scenarios

It was assumed that social robots leverage the natural human tendency to anthropomorphising inanimate things (section 1). Indeed, Freud defined humans as "symbolic animals" [29], who tend to create and modify the way in which reality appears to them. Therefore, someone could argue that the individuals involved in the interaction have their own responsibility in the process of deception. In such a view, the machine's deceitful behave would appear as less relevant.

Nevertheless, it is important to introduce a fundamental distinction between two terms: (1) anthropomorphism and (2) anthropomorphisation. While the first one refers to the human propensity to attribute human-like features to robots [30], the second one implies the deliberate choice of designing such characteristics by developers [31]. It follows that, even if the attribution of anthropomorphic traits to the machine had not the precise aim to deceive – or to do so with a malicious intent – the programmers would have had the competence to foresee this effect and to correct its potentially harmful drift [32]. In fact, fully rational people are subjected to such a dynamic too [28]. It could be objected that we are

not facing real deception. In fact, there are other circumstances in which people are entertained through an illusion and still maintain the ability to distinguish it from reality, without negative consequences. Those who support this idea make the example of a magic show [33]. As the spectators know that the magician does not actually cut the partner into two parts - and are amused instead of scared -, in the same way robots' users could know that the machine is simulating emotions and attachment, without really experiment them.

Though, the characteristics which influence humans in anthropomorphising these devices are concretely present by design. These features are not the mere result of an effort of subjective imagination. For this reason, it was underlined that the difference between a mere toy - for instance - and a robot is the same as between the action of pretending and that of believing [34]. An example is the one of robots deliberately structured so as to seem clumsy, in need of help, or to make mistakes in pursuing the given task. This for error is considered typically human-like, while efficiency and perfection of execution are usually linked to what is artificial [35]. Thus, the creation of an empathic bond is elicited. Likewise, it was demonstrated that a similar effect can be produced by implementing the machine with a 'cheating' functionality [36]. The result is maximised if the robot repeatedly deceives the user, for it encourages the perception of an autonomous will in the device [37]. On the contrary, if the cheating behaviour is carried out only once, it is more likely considered as a problem of malfunction.

This underlines that, in the context of a human-social robot interaction, efficiency in operation is expendable in favour of the possibility of living a human-like experience.

When the interaction involves subjects that are more vulnerable due to their age of health condition, the effects of continued exposure to similar mechanisms deserve to be analyzed more in detail. Consequently, plausible scenarios of robots with children and with fragile people will be presented below.

3.1 Children-robot interaction

The illusion that machines can be engaged in an appropriate conversation, experiment empathy and establish a real friendship can lead to entrusting them with tasks that go far beyond their actual functionalities. In order to better understand this passage, we can try to image a children's play scenario. The robot can be trained to prevent or to react to standard/common hypothesis of harmful events. However, it has very little changes to recognise a child pretending to fall – because of the nature of the play – from one who has actually been hit. Again, a child who uses a common tool – such as a pen or scissors – does not always use it appropriately and the machine may not be able to distinguish – or distinguish promptly – the suitable use from the harmful one [32].

With regard to childhood, then, there is a heated debate about the possible uses of social robots. Some studies show that these devices have a positive effect in the treatment of children with autism [38][39][40]. However, it may be the case that what was presented as a solution at the very beginning may turn to be the problem at the end. Due to the mechanical, precisely planned nature of the

machine a long-term/semi-exclusive interaction with people tending to isolation could increase this practice. In fact, the machine represents a 'safe reference', which does not pose opposition and promptly meets the unidirectional needs of the child [41].

The most detrimental effect of such a deprivation of significant human relationships could be appreciated in babies. In this case, we can refer to indirect evidence only, for it is not admissible to conduct experimentation with newborns. Old researches highlight that those who were deprived of a 'personalised', attentive, empathic care or of warmth, human contact let themselves die or developed serious physical and psychological problems [42][43].

Even if we do not consider babies, but older children, a long-term interaction with robotic caregivers, instead of human ones, could implies criticality. First of all, they will become used to predictable and schematic responses to given inputs, possibly developing difficulties in managing real emotions – like disappointment, frustration, dissatisfaction –. This could compromise their capacity to empathy themselves, for they would lack of the experience of real relationships, based on compromise, mutual-adaptation, in which it is impossible to be always listened, pleased and pandered in selfish needs [44].

Nevertheless, it cannot be ignored that some applications of AI devices with young people have also positive aspects. This is the example of Nao or Pepper, which can help children manage painful medical procedures or not be completely excluded from the school context due to a long hospital stay [49]. It follows that to accurately ponder the kind, time and dynamic of the interaction can be fundamental in order to distinguish empowering uses of social robots from detrimental ones.

Moreover, it was demonstrated that, when the robot simulates gratitude or a more intimate interaction with a specific child because of the amount of attention he/she turned on it, the child was encouraged to increase this behaviour [46]. In fact, the more the user interacts with the machine the more the result will be satisfactory and calibrated on the personality and the habits of the human being [47]. In addition, the overexposure to technological devices has been proven to release dopamine and its sudden deprivation or decrease can provoke anxiety, restlessness, anger . So described, the pattern is close to the one established in case of any form of addiction - both behavioural and substance ones –.

In the case of robots that engage the user at an emotional level, these effects can be summed to those of psychological attachment and affective dependence. Though, such a long-term result should be deepened with specific studies.

Indeed, it should be underlined that new technologies have an impact on how we act in the world, being able to modify the way we perceive and concept reality . This is even more the case for the youngest ones, who have not completed their psychological and cognitive development path yet.

3.2 Social robots and fragile people

Social robots can be used even in the treatment of people with mental or physical disability and the elderly. In these scenarios, the device can have the role of a

caregiver, a companion or even a therapeutic tool. It is easily understandable that the machine's deceptive features have different effects on the base of the task it has to perform.

An example is Paro [50], a device resembling a seal pup which displays positive responsiveness and beneficial impact on the health and mental status of the user when cuddled [51]. The choice to emulate this specific animal is not by chance. It is certainly not a typical pet – such as a cat or a dog – and this decreases human's expectations with regard to the way it responds to the interaction [52][53]. Thus, its technology is able to influence people's emotions, although it is not very sophisticated. Another case is the one of a robotic doll, specifically programmed to induce individuals affected by dementia to create an emotional bond towards the machine [54], engaging them at a conversational level. A similar dynamic was analysed in the project Rehabibotics, involving people with serious cognitive difficulties. The robot was projected so as to provoke empathic responses in order to favour the interaction. Therefore, it could record likes and dislikes of the patients, emotional and mental states, in order to predict them and track the progress of the disease. Anyway, the machine showed some errors in this procedure, which could not be corrected by individuals' feedback, for dementia made them not always – or not reliably – aware of their own inner states [55].

In particular for what fragile and old people are concerned, exacerbation of isolation end dehumanisation [56] are the main risks that need to be carefully considered, for they can lead to the objectification of human beings, whose autonomy and self-determination could be challenged [57].

Moreover, the report written by the Rathenau Institute for the Council of Europe highlights a possible infringement of fundamental rights - in particular human dignity – by the long-term exposure to a continued human-robot interaction for the elderly. This rises the necessity of a reflection with regard to a plausible right to meaningful human contact [58].

It follows that – as we have briefly tried to demonstrate here – social robotics is a varied field, which lends itself to many possible applications. Therefore, the challenge that new technologies poses to social sciences is to identify intervention tools capable of protecting the integrity of the human beings involved in the interaction, taking into account possible material – but also immaterial – damages [59]. To this end both legal and ethical approaches should be analysed.

4 Ethics and the law: possible tools of analysis

Assessing the theme of deception in social robotics, philosophical and legal perspectives are often taken into consideration. In fact, both of them could be relevant to discuss possible harmful repercussions on the individuals involved and to intervene in order to limit or remove them. However, identifying general characteristics of each of the two disciplines is essential to understand how and to what extent they can effectively contribute to the debate.

4.1 The not-universality and not-univocality of ethical statements

By definition, ethics is a branch of philosophy which guides people's behavior in the world and in the relations that they establish one another [60]. For this reason, someone says that whenever there is a debate regarding which conduct or risk is best to take, it has to be ethical oriented [61].

This very approach has been largely adopted even with regards to new technologies [62][63]. However, this discipline has no external oversight nor even standards protocol for enforcing its guidelines [64]. Moreover, it is far from being really universal, contrary to what it is commonly claimed. The term 'ethics', without any other specification, includes different theoretical frameworks of reference and not all of them can be considered conform to every legal system. An example can be the concept of development formulated by transhumanists. Taking into consideration the European context, it appears simplistic and potentially dangerous for the integrity of the users. In fact, it aims to subvert the very concept of "humanity", in favour of a limitless trust in the power of science [65]. Even more radical is the post-humanist refusal to adopt an anthropocentric perspective [66]. The base of this idea is the belief that human nature would be something to be overcome in order to realise "singularity" [67]. With a view of consistency with Member States' Constitutions and international treaties, the bio-conservative understanding of human nature seems the most appropriate to address the issues posed by robotics and AI. It is considered as universally recognized to everyone in reason of their own existence, not modular or subjected to renunciation [68]. However, not even in a similar prospective all the alternatives may be equally suitable. This is the case of utilitarian argumentation. It aims to legitimate deception in human-robot interaction in reason of the beneficial purpose, without taking into account the wider range of interests and rights involved and the correlate effects [69].

This is possible for ethics purports to investigate all areas of what is rationally knowable, without being held to strictly comply with acquired concepts and axioms – contrary to what concerns the legal analysis –. Therefore, its assumptions and guidelines have been accused of ambiguity and not obligatory [70]. Therefore, the variety of existing ethics, the need of a careful *ex ante* evaluation of the conformity with the legal framework behind it and the lack of enforcement rise the need to identify precise, enforceable parameters for facing the challenges posed by social robots arises. To this end, the role of the law can be crucial.

4.2 The binding and complete nature of the legal system

The main elements that distinguish the legal discipline from philosophy are: methodology, object of investigation and the limits they have to handle. In fact, legal argumentation cannot operate in a *vacuum*, for it takes place in a proper, self-referred system and lawyers are bound externally by fundamental principles, typically affirmed in Constitutions.

It could be argued that even not every legal norm is equally enforceable all over the world. Nevertheless, it is likewise true that, considering a given field of application, legal dispositions are binding all the people involved in it or, at least, those previously determined and indicated [71]. This confers homogeneity of solutions and treatments.

However, reasons of major complaints about the law as a tool for the regulation of new technologies are: (i) the long time needed for its formulation and concrete application, (ii) the difficult individuation of the proper time for an intervention, (iii) the rigidity of its statements. The processes of discussion, decision and entry into force of a new legislation are often considered inconceivable with the speed of scientific evolution [72]. The regulation could intervene too late, thus loosing its incisiveness and failing the aim to prevent the spread of potentially harmful devices. At the same time, even the choice regarding "when to regulate" has a decisive impact. In fact, acting too early could damage the very phase of experimentation, stopping a process which might need to be only corrected. Same observations could be done even with regards to already operative rules. If they are proved ineffective or inadequate, their modification or replacement is seen as laborious and not timely enough not to damage scientific research [73].

Such considerations are based on two basic misunderstandings. First of all, the fact that legal strict statements always undermine innovation, development and competitiveness. Actually, it is the uncertainty in regulation that, by definition, produces that very effect [74]. Furthermore, a relevant false belief is that emerging technologies – and, in particular, examples of embodied-AIs – would highlight legal gaps, so as to require the formulation of specific norms. On the contrary, the law constitutes a system that is complete *per se*, for it does not consist in the black-letter-law only. It regards norms, but even legal doctrine and judicial applications [75].

Therefore, legal interpretation can help find solutions to specific cases, without the need for specific rules [76]. This would be possible: (i) through the application of rules governing similar cases or similar matters (*analogia legis*), (ii) through the interpretation of the legislator's will, by means of the general principles of the legal system (*analogia iuris*), (iii) through elastic concepts, applicable in many different cases (general clauses, i.g. good faith).

This does not exempt from the possibility to question the adequacy of existing norms. Therefore, any chance to revise the actual legislation to correct undesirable, inefficient or even sub-efficient outcomes should be seized. It follows that the right question legal experts should try to answer is not if the law can play its part in the regulation of social robotics and AI, but how it can do that. In this view, the priority should be to guarantee both scientific progress and an efficient protection for human beings.

To this end, the attention could be focused on fundamental human rights, which are not proper of a specific ideology but, at the same time, have the advantage to be precise and not ambiguous, without the necessity of a narrow detailed definition [77].

5 Overcoming the dichotomous approach: the role of Human Dignity

So far we have described an articulate interweaving of plausible but not completely self-sufficient solutions. Therefore, it could be useful to analyse the role that the concept of human dignity may have in the debate regarding deception in social robotics. In fact, it could represent an objective and external criterion, able to collect both the instances of philosophical speculation with regards to ethics and the non-dismissible and binding character of legal principles. This would be functional to evaluate which types of technologies deserve to be favoured - for their correspondence to the reference values - and which ones to stem.

The reason to identify such a parameter in the very principle of human dignity is, primarily, that its relevance and authority as a value is unquestionable. In fact, it plays a central role among fundamental rights, for it is the one which summarises all the others [80]. Moreover, its respect is not limited because of age, gender, religion, nationality, political convictions or any other subjective factor [81]. It is recognised in many national norms and Constitutions around the world, in EU treaties – especially the European Charter of Fundamental Human Rights – and in eminent judgments in the Courts of Justice. With regard to this latter aspect, two of the most emblematic judicial cases are the German "*Peep-Show-Fall*" [78] and the French "*Jeux de nains*" [79]. In both of them, the Courts highlighted that every human being carries a fragment of the universal principle of dignity. As a consequence, diminishing one own value implies to reduce the one of all the affiliates.

However, the doctrinal debate about the very nature of human dignity is still open. More precisely, it is accused of vagueness, for it is often theorised as indemonstrable, imperative, inexpressible [82] and for the lack of a specific definition.

Nevertheless, it is important to emphasize that the level of abstraction of this concept is not a negative aspect for what the regulation of social robotics is concerned. It could have two functions: (i) guaranteeing flexibility and (ii) shaping mandatory norms. In fact, flexibility allows this principle to adapt more efficiently to the manifold variety of existing technologies and their unceasing innovation and development. Concurrently, it is deeply rooted in the European tradition also for what lies outside the purely philosophical reflection, for it represents a legal concept. Therefore, human dignity - and fundamental human rights in general - have the advantage to be already able to be validly used in a multidisciplinary debate. In fact, they are binding and not modulable on the base of the ethical framework taken into account [83]. This would be a starting point to guarantee that new technologies are projected to respect the centrality of human beings. In fact, they can be used for: (i) testing the desirability of robotics applications, (ii) identifying – even in a case-by-case perspective – which principles should prevail, (iii) orienting innovation towards devices that allows to promote such values, (iv) allowing that they can be efficiently tested in the real world - not just in a laboratory - [85][84].

6 Discussion and Final Remarks

This paper assessed the theme of deception in social robotics, underlining the need to identify an objective criterion to balance the demand for acceptability - to foster innovation - and the necessity to protect users' material and psychological integrity.

To such an end, the traditional juxtaposition between ethical and legal perspectives was presented, so as to underline their structural differences – mainly in terms of methodology and scope –. This is intended to clarify that none of them should be removed from the debate on the protection of new technologies' users. Nevertheless, they should play a different role in pursuing such a goal.

Ethics may be inspirational from a political or economic point of view. It can promote the introduction of a new or reformed legislation, the implementation of companies' policies, the development of new awareness campaigns in the public [83]. Moreover, ethical principles are, in many cases, useful to overcome legal theories' limitations or to better understand and to convey the *ratio* on the base of legal reasoning [83]. Hence, they can live in a functional relationship with the law and help overcoming the strict boundaries of its formalities. Nonetheless, they should always be seen as an instrument to reach a goal, not the goal itself.

For its part, the law has the merit of being a complete, binding and enforceable system, deeply rooted in fundamental principles. Thus, it can be useful to inspire the design of socially competent devices and to evaluate their effects not only on the rights of the users, but also on those of all the members of society.

Given the peculiarities and variety of the theme here analysed, this paper suggests the possibility to rely on a third option: fundamental human rights. Among them, human dignity deserves a particular attention and could be used as external criterion to approach the regulation.

In spite of the claims of conceptual vagueness, it is legally binding, common to everyone because of their appurtenance to humankind, adaptable but not dismissible. Such a flexibility could be essential in order to face the challenges that the theme of deception in social robotics can – and will – pose. In fact, human dignity constitutes the core of what it means to be "human". This is reflected in terms of rights and duties, but even much more so in terms of the perception individuals have of themselves, their environment, the others and of the way they can act and relate to this ecosystem. This aspect is crucial, considering that the technology we interact with on a daily base can influence our intentions, awareness and the way we process, categorise and evaluate information, concepts, relations [86].

In light of the above, further investigation is need to understand how concretely the principle of human dignity can be adopted in the debate regarding deception in human-robot interaction. In particular, it should be better defined how it can materially balance instrumental benefits and possible individuals' integrity harms of such a phenomenon. In fact, the aim of this discussion is not to condemn the implementation of machines with social features *tout court*, but to suggest the need to draw a line between beneficial and risky contexts.

To this end, we should even consider that the theme of deception is still controversial in the engineering and robotic field, as it emerges form the scenarios of interaction here presented.

Nowadays the ability of a social robot to deceive – inducing the user to create an emotional and subconscious bond with the machine – is considered "central to AI as the circuits and software that make it run" [31]. However, (i) the qualification of such a dynamic and (ii) the scope for action to protect peopled involved are still reasons for debate. Actually, such technologies are already part of our reality, although they are not yet so widespread as to allow neither a sufficiently well-stocked collection of concrete cases, nor an in-depth investigation of their long-term effects. Another critical element is represented by the lack of homogeneous consensus in the scientific field about whether and to what extent to enhance robotic deceptive behavior towards the user. This makes difficult for philosophers and legal scholars to have an univocal perception of the issues raised from this kind of technology and to face them in an effective and appropriate way. For this very reason, it is crucial to promote an integrated, multidisciplinary approach, able to take into consideration both the specificity of the social robotics field and the importance of a human-centered technological development.

7 Acknowledgements

The author acknowledges that this work has received funding from the Alma Mater Research Institute for Human-Centered AI, "Law, Science and Technology Joint Doctorate", University of Bologna.

References

1. Floridi L.: The fourth Revolution. How the Infoshere is Reshaping Human Reality. Oxford University Press, Oxford, (2014)
2. Statista, Valishery: Global social and entertainment robot until sales 2015-2025, www.statist.com/statistics/755677/social-and-entertainment-robot-sales-worlwide/
3. Fong T, Nourbakhsh I, Dautenhahn K.: A Survey of socially interactive robots. Robot Auton Syst, 42, p. 145 (2003)
4. Dautenhahn K.: Socially intelligent robots: dimensions of human-robot interaction. Philos Trans R Soc Lond B Biol Sci 362, p. 684 (2007)
5. Wagner A., Arkin R.: Acting Deceptively: Providing Robots with the Capacity for Deception. International Journal of Social Robotics, **3**(1), pp. 5-26, (2011)
6. Shim J., Arkin R.: Other-Oriented Robot Deception: How Can a Robot's Deceptive Feedback Help Humans in HRI?. International Conference on Social Robotics (2016)
7. Złotowski J., Proudfoot, D., Yogeeswaran, K., Bartneck, C.: Anthropomorphism: opportunities and challenges in human–robot interaction. International Journal of Social Robotics, **7**(3), pp. 347-36 (2015)
8. Bertolini A.: Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules. Law, Innovation and Technology, **5**(2) (2013)

9. Stradella, E.: La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione. Media Laws, **1**, www.medialaws.eu (2019)
10. Palmerini, E., Bertolini, A., Battaglai, F., Koops, B.-J., Carnevale, A., Salvini, P.: Robolow: Towards a European framework for robotics regulation. Robotics and Autonomous Systems, **86**, pp. 78–85 (2016)
11. Turing: Computing Machinery and Intelligence. Mind, **59**, pp. 433-460 (1950)
12. Shim, J., Arkin, R. C.: A taxonomy of robot deception and its benefits in HRI. IEEE International Conference on Systems, Man, and Cybernetics (2013)
13. Danaher J.: Robot Betrayal: a guide to the ethics of robotic deception. Ethics and Information Technology, pp. 1-12, https://doi.org/https://doi.org/10.1007/s10676-019-09520-3 (2020)
14. Wilson D., Sperber D.: On Grice's Theory of Conversation, in P. Werth "Conversation and Discourse", London, Croom Helm, pp. 155–178 (1981)
15. Isaac A.: White lies on silver tongues: why robots need to deceive us (and how). https://doi.org/10.1093/oso/9780190652951.003.0011 (2017)
16. Wallach W., Allen, C.: Moral machines: Teaching robots right from wrong. New York: Oxford University Press, p. 44 (2009)
17. Floreano D., Mitri S., Magnenat S., Keller L.: Evolutionary conditions for the emergence of communication in robots. Current Biology, **17**(6), pp. 514–519 (2007)
18. Sparrow R.: The march of the robot dogs. Ethics and Information Technology, **4**, pp. 305–318 (2002)
19. Grodzinsky F. S., Miller K. W., Wolf M. J.: Developing automated deceptions and the impact on trust. Philosophy & Technology, **28**(1), pp. 91–105 (2005)
20. Sorell T., Draper H., Second thoughts about privacy, Safety and deception, Connection Science, **29**(3), pp. 217–222 (2017)
21. Matthias A.: Robot lies in health care: When is deception morally permissible?. Kennedy Institute of Ethics Journal, (2), p. 17 (2015)
22. Johnson D., Verdicchio M.: Why robots should not be treated like animals. Ethics and Information Technology, **20**, p. 299 (2018)
23. Sharkey A., Sharkey N.: Children, the elderly, and inter-active robots. IEEE Robotics and Automation Magazine, **18**(1), pp. 32–38 (2011)
24. Matarić M. J., Socially assistive robotics: Human augmentation versus automation, Science Robotics, **2**(4), (2017)
25. Kaminski M. E., Rueben M., Smart W. D., Grimm C. M.: Averting robot eyes. Md. L. Rev., **76**, p. 983 (2016)
26. Orwell G.: Nineteen Eighty-Four (1949)
27. Vandemeulebroucke T., de Casterle B. D., Gastmans, C.: The use of care robots in aged care: A systematic review of argument-based ethics literature. Archives of Gerontology and Geriatrics, **74**, pp. 15–25 (2018)
28. Carli R., Najjar A.: Rethinking Trust in Social Robotics. In: Proceedings of SCRITA 2021 (https://arxiv.org/abs/2108.08092), a workshop at IEEE RO-MAN 2021: https://ro-man2021.org/
29. Carotenuto A., Senso e contenuto della psicologia analitica, Bollati Boringhieri editore, Torino (1990)
30. Hegel F., Krach S., Kircher T., Wrede B., Sagerer G.: Understanding social robots: A user study on anthropomorphism. In: RO-MAN 2008—The 17th IEEE International Symposium on Robot and Human Interactive Communication, pp. 574–579, https://doi.org/10.1109/ROMAN.2008.4600728 (2016)
31. Natale: S., Artificial intelligence and gullible humans. The Turing Test and the real significance of AI. iai, issue 96, https://t.co/VEcGoqCVIQ?amp=1 (2021)

32. Sharkey A., Sharkey N.: We need to talk about deception in social robotics!. Ethics Inf Technol., https://doi.org/10.1007/s10676-020-09573-9 (2020)
33. Coeckelbergh M.: How to describe and evaluate "deception" phenomena: Recasting the metaphysics, ethics, and politics of ICTs in terms of magic and performance and taking a relational and narrative turn. Ethics and Information Technology, **20**, p. 78 (2018)
34. Turkle S.: Alone Together: why we expect more from technology and less from each other. Basic Books, New York (2011)
35. Salem M., F. Eyssel, K. Rohlfing, S. Kopp, F. Joublin: Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. International Journal of Social Robotics, **5**(3), pp.313-323 (2013)
36. Short E., J. Hart, M. Vu, B. Scassellati: No Fair!! An interaction with cheating robots. 5th ACM/IEEE International Conference on Human-Robot Interaction, pp. 219-226 (2010)
37. Lee M., Peng W., Jin S. A., Yan C.: Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. Journal of communication, **56**(4), pp. 754-772 (2006)
38. Scassellati B., Henny A., Maja M.: Robots for use in autism research, Annual review of biomedical engineering, **14**, pp. 275-294 (2012)
39. Diehl J. J., Schmitt L. M., Villano M., Crowell C. R.: The clinical use of robots for individuals with autism spectrum disorders: A critical review. Research in autism spectrum disorders, **6**(1), pp. 249-262 (2012)
40. Huijnen, C. A. G. J., Lexis, M., de Witte, L. P.: Robots as new tools in therapy and education for children with autism. Int J Neurorehabil, **4**(278), 2376-0281 (2017)
41. Di Dio C., Manzi F., Peretti G., Cangelosi A., Harris P. L., Massaro D., Marchetti A.: Shall I Trust You? From Child-Robot Interaction to Trusting Relationship, Front. Psycol., **11**, pp. 469 ss., https://doi.org/10.3389/fpsyg.2020.00469 (2020)
42. Nelson C. A., Zeanah C. H., Fox N. A., Marshall P. J., Smyke A. T., Guthrie D.: Cognitive recovery in socially deprived young children: The Bucharest early intervention project. Science, **318**(5858), pp. 1937–1940 (2007)
43. Chugani H., Behen M., Muzik O., Juhasz C., Nagy F., Chugani D., Local brain functional activity following early deprivation: A study of post-institutionalised Romanian orphans. Neuroimage, **14**(6), pp. 1290–1301 (2001)
44. Turkle S.: Why these friendly robots can't be good friends to our kids. Washington Post,(2017)
45. Palmerini E., F. Azzarri, F. Battaglia, A. Bertolini, A. Carnevale, J. Carpaneto, F. Cavallo, A. Di Carlo, M. Cempini, M. Controzzi, B.-J. Koops, F. Lucivero, N. Mukerji, L. Nocco, A. Pirni, H. Shah, P. Salvini, M. Schellekens, K. Warwick: Guidelines on Regulating Robotics. In: Robolaw Grant Agreement Number: 289092, D6.2. (2014)
46. Kanda T., R. Sato, N. Saiwaki, H. Ishiguro: A two-month field trial in an elementary school for long-term human robot interaction. IEEE Transactions on Robotics and Automation, **23**(5), pp. 962-971 (2007)
47. Howard A., Tapus A., Kajitani I.: Socially assistive robots, guest editors. IEEE Robotics and Automation Magazine, **26**(2), pp. 10–110 (2019)
48. Lezhaen: Overexposure to screens of young children: The agitation of fears by autism!. https://blogs.mediapart.fr/lezhaen/blog/210118/surexposition-aux-ecrans-des-jeune s-enfants-l-agitation-des-peurs-par-l-autism (2018)
49. Ozaeta L., Graña, M., Dimitrova M., Krastev A.: Child oriented storytelling with NAO robot in hospital environment: preliminary application results. Problems of Engineering Cybernetics and Robotics, **69**, pp. 21-29 (2018)

50. www.parorobot.com
51. Robinson H., Macdonald B., Kerse N., Broadbent E.: The psychosocial effects of a companion robot: A randomized controlled trial. Journal of the American Medical Directors Association, **14**(9), pp. 661–667 (2013)
52. De Graaf M.M.A., Allouch S.B.: The influence of prior expectations of a robot's lifelikeness on users' intentions to treat a zoomorphic robot as a companion. Int. J. Social Robot., **9**(1), pp. 17–32, https://doi.org/https://doi.org/10.1007/s12369-016-0340-4 (2016)
53. Savela N., Turja T., Oksanen A.: Social acceptance of robots in different occupational fields: a systematic literature review, Int. J. Social Robot., **10**(4), pp. 493–502, https://doi.org/https://doi.org/10.1007/s12369-017-0452-5 (2018)
54. Kitwood T., Dementia reconsidered: The person comes first, Buckingham, Open University Press (1997)
55. Shukla J., Cristiano J., Amela D., Anguera L., Vergés-Llahí J., Puig D.: A case study of robot interaction among individuals with profound and multiple learning disabilities. In: International Conference on Social Robotics, Springer, pp. 613-622 (2015)
56. Sharkey A., Sharkey N.: Granny and the robots: ethical issues in robot care for the elderly, Ethics Inf Technol., **14**, pp. 27–40, (2012)
57. Nussbaum M. C.: Frontiers of justice: Disability, nationality, species membership. Cambridge: Harvard University Press., (2009)
58. Van Est R., Gerritsen J. B. A., (with the assistance of L. Kool): Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality – Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE). The Hague: Rathenau Instituut (2017)
59. PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM(2021) 206 Final, 2021/0106 (COD), 21.4.2021, Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) — Shaping Europe's digital future (europa.eu) (2021)
60. Floridi L.: Soft Ethics and the Governance of the Digital. Philosophy & Technology, **31**(1), pp.1-8 (2018)
61. Skorupinski B., Ott K.: Technology assessment and ethics. Poiesis & Praxis, **1**(2), pp. 95-122 (2002)
62. High-Level Expert Group on Artificial Intelligence: Ethics Guidelines For Trustworthy AI. European Commission B-1049 (2019)
63. Floridi L. et al.: Ai4people—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, pp. 689–707 (2018)
64. Mittelstadt B.: Principles Alone Cannot Guarantee Ethical AI. Nature Machine Intelligence, **1**, pp. 501-507 (2019) https://doi.org/www.dx.doi.org/10.2139/ssrn.3391293
65. Stile G. C.: Transumanesimo. Una introduzione all'idea di evoluzione autodiretta, Laboratorio ISPF, **XII** (2015) https://doi.org/10.12862/ISPF15L406
66. Fukuyama F.: Our Posthuman Future: Consequences of the Biotechnology Revolution. Farrar, Strauss and Giroux, New York, pp. 149-160 (2002)
67. Coates J. F.: The singularity is near: When humans transcend biology- Discussions. Technological Forecasting And Social Change, **73**(2), pp. 121-127 (2006)

68. Kass L. R.: Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection. The New Atlantis, **1**, Spring, pp. 9-28 (2003)
69. Bertolini A.: Human-Robot Interaction and Deception. In: Osservatorio del diritto civile e commerciale, Il Mulino, **2**, p. 656 (2018)
70. Metzinger T.: Ethics washing made in Europe. Der Tagesspiege (2019) www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html
71. Sartor, G.: Artificial intelligence and human rights: Between law and ethics. Maastricht Journal of European and Comparative law (2020) https://doi.org/https://doi.org/10.1177/1023263X20981566
72. Moses L. B.: Agents of Change: How the Law "Copes" with Technological Change. Griffith Law Review, **20**(4), p. 764 (2011) http://ssrn.com/abstract=2000428
73. Holdere C., Khurana V., Harrison F., Jacobs L.: Robotics and Law: Key Legal and Regulatory Implications of the Robotics Age (Part I of II). Computer Law & Security Review,**32** (2016)
74. Bahmani-Oskooee M., Saha S.: On the effects of policy uncertainty on stock prices. J Econ Finan, **43**, pp. 764–778 (2019) https://doi.org/https://doi.org/10.1007/s12197-019-09471-x
75. Sacco R., Cos'è il diritto comparato, Giuffrè (1992)
76. Easterbrook, F. H.: Cyberspace and the Law of the Horses. University of Chicago Legal Forum (1996)
77. Dworking R.: Take rights seriously. Harvard University Press, p. 155 (1977)
78. BVerwG, 15 dicember 1981, NJW, pp.664 ss. (1982)
79. Cons. état. Ass., 27 ottober 1995, Ville d'Aix-en-Provence, in D., pp. 177 ss. (1996)
80. Zatti, P.: Note sulla semantica della dignità. In: "Bioetica e dignità umana. Interpretazioni a confronto a partire dalla Convenzione di Oviedo", cured by E. Furlan, F. Angeli, Milano, pp. 95-109 (2009)
81. Resta, G.: La disponibilità dei diritti fondamentali e i limiti della dignità (note a margine della Carte dei Diritti). Riv. dir. civ., **II**, p. 829 (2002)
82. Fabre-Magnan, M.: La Dignité en Droit: un Axiome. Université Saint-Louis, Bruxelles, **58**(1) (2007) www.cairn.info/revue-interdisciplinaire-d-etudesjuridiques-2007-1-page-1.htm
83. Harris and others: Ethical Assessment of New Technologies: A Meta-methodology. Journal of Information, Communication and Ethics in Society, **9**(1), pp. 49-64 (2011)
84. Kritikos M.: Artificial Intelligence ante portas: Legal & ethical reflections, Scientific Foresight Unit (STOA), (2019) https://www.europarl.europa.eu/at-your-service/files/be-heard/religious-and-non-confessional-dialogue/events/en-20190319-artificial-intelligence-ante-portas.pdf
85. Koops B.J.: Concerning "Humans" and "Human" Rights: Human Enhancement from the Perspective of Fundamental Rights. In: Koops B.J. and others(eds): Engineering the Human: Human Enhancement Between Fiction and Fascination, p. 174 (2013)
86. Bisol B., Carnevale A., Lucivero F.: Diritti umani, valori e nuove tecnologie. Il caso dell'etica della robotica in Europa. Metodo. International Studies in Phenomenology and Philosophy, **1**(2), p. 244 (2014) ISSN 2281-9177