REVIEW

A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI)

Aimee van Wynsberghe
Shuhong Li

Department of Values, Technology and Innovation, Delft University of Technology, Delft, The Netherlands

Abstract: To date, the majority of work in the fields of human–robot interaction and robot ethics take as the starting point a dyadic interaction between a human and a robot. It is clear, however, that the impacts of robots in health care (understood as ranging from embodied robots and AI to avatars and chatbots) far exceed the individual with whom the robot is interacting. One of the most critical aspects of introducing robots in health care is how such a "bot" will restructure the health care system in a variety of ways: roles of health care staff will change once "bots" are delegated tasks, certain professions may no longer exist (eg, cleaning robots may remove the need for janitorial staff), the education of health care staff will need to include "bot" training, resources will be reallocated to account for the purchasing of "bots", and the expertise of health care staff will be called into question (eg, when an AI algorithm predicts something that the physician does not). A well-developed care system that includes "bots" of all kinds should predict and balance the ethical impact equally between not only caregivers and receivers, but for the system within which these actors function. This article proposes a model for doing just this, the human–robot–system interaction (HRSI) model that allows for the ethical assessment of "bots" as mediators between a care receiver and a health care system. The HRSI model has important implications for revealing a new set of ethical issues in the introduction of "bots" in health care and in calling for new forms of empirical research to track possible (unintended) consequences related to the rearrangement of roles and responsibilities in the health care system resulting from the integration of health care "bots".

Keywords: human–robot interaction, HRI, human–robot–system interaction, HRSI, robot ethics, care robots

Introduction

The health care system of 2019 uses a variety of "bots" in the provision of care from physical robots, embodied AI, to avatars and chatbots. According to the International Federation of Robotics, medical robot sales are at $1.9 billion for the year 2017.[1] Beyond this, the global market of chatbots, a kind of software used to communicate with users, is expected to reach $2.1 billion by 2024 and a large share of it will be in health care.[2] Developers claim these "bots" promise to mitigate the shortage of health care workers and resources; however, another school of thought criticizes the introduction of "bots" for their potential to threaten ethical and societal values such as privacy, well-being, and social isolation among others.[3–7] We suggest in this article that the traditional forms of ethical evaluation, which rely on a dyadic human–robot interaction (HRI), ought to be re-thought in order to account for the impact that robots have on the health care system as a whole, rather than the individual caregivers and/or care receivers.

Correspondence: Aimee van Wynsberghe
Department of Values, Technology and Innovation, Delft University of Technology, Jaffalaan 5, 2628BX, Delft, The Netherlands
Tel +31 15 278 9689
Email aimeevanrobot@gmail.com

In July 2019, a collaboration between the UK National Health Service (NHS) and the technology company Amazon, which makes the embodied AI product known as Alexa, was announced.[8] This collaboration aims to provide consumers of the Alexa the ability to seek medical advice from the device. To realize this, the NHS has shared medical data with Amazon. Such a collaboration confronts society with the challenge of understanding the boundaries between a "bot" – as a technology embedded in a network of funders and tech developers – and "bot" as a part of the health care system – understood as a network of care providers governed by regulatory boards and bioethical principles. When thinking about the well-being of patients, preventing harm and respecting autonomy, what are the responsibilities of the company making a robot and accordingly what are the responsibility of the health care system? Which stakeholder group assumes stewardship over the beneficence of patients?

To date, the fields of HRI and robot ethics take as the starting point a dyadic interaction between a human and a robot with the goal of creating intuitive and safe encounters. It is clear, however, that the impacts of robots in health care far exceed the individual with whom the robot is interacting. One of the most critical aspects of introducing robots in health care is how such a "bot" will restructure the health care system in a variety of ways: roles of health care staff will change once "bots" are delegated tasks, certain professions may no longer exist (eg, cleaning robots may remove the need for janitorial staff), the education of health care staff will need to include "bot" training, resources will be reallocated to account for the purchasing of "bots", and the expertise of health care staff will be called into question (eg, when an AI algorithm predicts something that the physician does not). A well-developed care system that includes "bots" of all kinds should predict and balance the ethical impact equally between not only caregivers and receivers, but for the system within which these actors function. This article proposes a model for doing just this, the human–robot–system interaction (HRSI) model that allows for the ethical assessment of "bots" as mediators between a care receiver and a health care system. This new framing makes explicit the potential for impact on the system and not just the individual patient or health care personnel interacting with the robot.

In the following sections, we begin by reviewing the current trends in health care "bot" technology covering robots, avatars, and software (including chatbots and various AI algorithms). We continue with a discussion of HRI and the current forms of ethical analyses using the HRI paradigm and show that their dyadic nature leaves them inadequate for addressing the scope of ethical issues pertaining to the health care system. We conclude by proposing a model for HRSI and explain it using various interaction scenarios. With a view of the HRSI model, and thus a better approximation of the complexity of care interactions, we identify unique ethical issues that arise surrounding issues of trust, accountability, responsibility, and conflicting preferences between care receivers and caregivers.

Current technology trends in health care "bots"

Each of the "bot" applications discussed here are meant to show various types of interaction partners between a care receiver and the health care system. By "interaction partners", we mean to suggest that humans will engage with the "bot" using different means (eg verbal, visual, and/or written) and that this interaction is more complicated than pressing buttons on the robot to get it to function.

Chatbots are generally used to provide verbal or written communication to care receivers and/or physicians about symptoms, diagnoses, medication, and weight or health coaching.[9,10] These chatbots are software; they are not embodied in the real world; they are not physically interacting with their human counterpart. Woebot, for example, is a chatbot designed to provide mental support to users by communicating via text in an application on the smartphone.[11] Another chatbot, Your.MD, acts as a health consultant by asking questions about users' symptoms and their personal information. It makes a preliminary diagnosis and provides users with medical information on the likely cause to help them find a suitable treatment.[12]

More traditional bots in health care are embodied robots that have a range of appearances and capabilities.[13] Some of the more common examples of care robots are the surgical robot daVinci, delivery robots TUG, Helpmate and Hospi, and the lifting robot Muscle Suit. Other robots serve more companionship ends, such as Paro for reducing anxiety in elderly care receivers, AIBO, NeCoRo, iCat which provides company to people who live alone, and the feeding robot iEAT that can help with eating.[14] Examples of embodied AI include the previously mentioned Amazon Alexa, and another example is Mabu, the "personal health care companion" "whose conversations are tailored to each patient she works with".[15] These robots are embodied in the world but

are distinct from the more traditional robots listed above insofar as they cannot engage with their surroundings (ie, they cannot move), they are only meant to engage with a human counterpart.

In between the physically embodied robots and the strict software bots are avatars. These are images of people or animals presented on a computer screen intended to interact with a human counterpart without the option to reach out and touch them. One example is Patty, a virtual physician's assistant developed by Cisco in 2009. Patty is a female character playing the role of doctor and/or nurse to provide medical information on diseases and medication to the care receiver and family. Patty also helps to arrange daily schedules of doctors.[16] Another avatar is the virtual assistant Molly designed to mimic doctors and/or nurses taking care of people with chronic diseases. This animated female caregiver called 'Molly' checks in on care receivers every day to collect health data of users, and to provide recommendations to them accordingly.[17] Avatars are considered more engaging than chatbots because they combine both verbal and visual interaction with users which is expected to achieve better results.[18]

Understanding the robot as external to the health care system

The applications listed above are wide and varied, but the common link between all these technologies is that they become integrated into a care receiver–health care system relationship. This interaction between human and health care system through the "bot" can happen in a variety of ways with a variety of ends that the bot is serving. The bot may be used to collect information about the care receiver, about his/her symptoms, care plan, or health information which is then used by the system (one or more professionals working within) to make a decision about how to proceed. Or, the bot may be integrated into postoperative care to follow-up on care receiver recovery after a care receiver has received treatment (and established a therapeutic relationship with health care professionals in person). Or, the bot could be used as part of a care receiver's care while a care receiver is in a health care facility. In each of these instances, the bot acts as an instrument to provide care from the health care institution to the care receiver, and yet it is still somehow connected to the tech company from which it came.

In order to create governance mechanisms to protect patient data (among other things), one must understand whether the robot is part of the health care system or belongs to a third party, the tech developers. We suggest that the health care "bot" is neither entirely part of the health care system nor entirely part of the tech company. Instead, it exists in a fluctuating state in which at moments it is part of either, ie when in development it belongs to the tech company and yet when used in health care it partially belongs to the health care system until there is a malfunction and it must return to the tech company for repair (or a technician from the tech company visits the hospital to repair). We say "partially" above because most "bots" are constantly collecting data on patients, and this data is most often stored and used by the company that created the bot, for upgrades, etc. Thus, the "bot" is more often than not connected to the tech company even when introduced into the health care system. For this reason, we suggest understanding the bot as separate from the health care system insofar as it remains connected to the tech company responsible for its development. In this way, the bot mediates between patient and health care system.

We acknowledge that understanding the ontological status of the "bot" is also dynamic – once the "bot" has been in the system for an extended period of time, it is possible to suggest that the bot truly becomes part of the system (eg, with technicians in the health care system, with the health care system responsible for data collecting, storage, and usage, with the health care system responsible for upgrades, and so on). At this moment in time, however, this is not the situation for most "bots" commercially available. Therefore, we consider it paramount to frame the robot as external to the health care system in order to raise awareness of policymakers, caregivers, and patients whose traditional moral codes governing the health care system may be in jeopardy when interacting with a "bot".

The HRI paradigm as an evaluative tool

Given that any "bot" in health care is sure to confront the health care system, and society at large, with ethical concerns, the question at the axis of this work is how to evaluate the interaction between the system of human actors, ie, the health care system, and the "bots". The idea to study humans interacting with robots is not new; HRI as a field of study emerged in the 1990s with, among others, the canonical work of Kazerooni, Held and Durlach, Breazeal, as well as Dautenhahn.[19–23] It centers on the study of many forms of verbal and nonverbal interactions between human and robots, with multidisciplinary approaches combining insights from

robotics, cognitive science, psychology, biology, language, and design.[24]

In a 2002 paper by Yanco and Drury,[25] and an updated version in 2004,[26] a taxonomy for HRI is presented. This overarching taxonomy was created using the following categories: task type, task criticality, robot morphology, ratio of people to robots, composition of robot teams, level of shared interaction among teams, interaction roles, type of human–robot physical proximity, decision support for operators, the time/space taxonomy, and autonomy levels/amount of intervention.[26] All figures used to illustrate the taxonomy of Yanco and Drury show humans on one side and robots on the other side. In some instances, one human may interact with one or more robots and in other instances, one robot may interact with one or more humans. In essence, HRI is about the human and the robot interacting and how best to design the robot as an intuitive interface in order to achieve a predetermined goal successfully. Based on this paradigm come the majority of ethical evaluations of health care robots to date.

Ethical reflections on health care robots to date

The first discussions on ethical issues surrounding robots in health care can be traced back to 2005.[27] Roboticist and robot ethicist Gianmarco Veruggio pointed out that the advance of surgical robots and robotic prothesis gave rise to medical ethics and bioethics problems. Veruggio created an overview of robot ethics based on the application domain and posited that health care robots faced ethical issues such as the impact of a robots' dexterity, dependability, and functionality on care receivers and on surgeons.[28]

Since then, the list of ethical concerns has grown. Generally, most ethical issues examine the risks in the interaction between the care receiver and the robot: the safety concerns to care receivers posed by large-sized robots, especially those receivers who do not know how to operate the robots properly;[4] the risk to privacy and data security of the person when being monitored by robots with sensors and cameras to record and monitor his/her vital signs and daily activities;[29,30] the potential deception of both the caregiver and the care receiver that may result in an undue assignment of greater intelligence than the robot is actually capable of;[31] the risk to the care receiver's autonomy when being stopped from performing certain actions, such as moving outside of the building freely for safety reasons;[32,33] the problem of infantilization of

elderly people;[34–36] the potential reduction of human contact when robots can take over care tasks from family and caregivers;[4,37] the issue of disregard for informed consent if the caregivers use robots for care receivers with dementia who cannot voluntarily make their own decision to either accept or decline to participate.[38]

Some roboticists have studied such interactions in a more nuanced manner than strictly according to the taxonomy of HRI again emphasizing the concern for patients in the HRI. Riek, for example, observes multiple ethical challenges arising in HRI: the therapy recipients in HRI are inclined to develop emotional and psychological bonds with the robot, which may result in negative effects on their psychological health and physical therapeutic treatment.[39] Several empirical studies in HRI focus on the interaction between robots and children. Belpaeme et al draw attention to the social bonds built in child–robot interaction, in particular, that robots need to function as peers to play together with children in the interaction.[40] A study conducted by Vallès-Peris et al in a children's hospital shows that care interactions happen in a bidirectional way in the imaginations of children, namely, the robot and the child take care of each other.[41] Additionally, Arnold and Scheutz distinguish the soft robots from hard-bodied robots within HRI ethics. They propose that soft-bodied robots should develop a balanced tactile engagement rather than psychological deception and help users to realize their bonds with a tool but not a person to mitigate the ethical challenges in HRI.[42]

Robot ethicists have also focused on the caregiver in the HRI. Normally, it is acknowledged that robots can help caregivers to relieve physical burdens by taking over manual tasks such as lifting, which benefit the care receivers' bodily health.[33,43] But the replacement of caregivers by robots raises concerns for a potential threat to the caregiver's ability to gain the skills required of a good caregiver, described by robot ethicist Shannon Vallor as a risk of "deskilling" workers.[44] This can happen in both technical and nontechnical ways, eg, technical skills like losing the ability to lift at an appropriate speed and nontechnical skills like losing the ability to perceive suffering in care receivers.

Although some believe that health care robots will remove caregivers from the dull and burdensome portions of care, freeing up the time of caregivers for emotional support of care receivers,[45] Borenstein and Pearson are skeptical about the actual effects that care robots may have on human caregivers' capabilities. To exemplify this, they make reference to the case of household appliances; household appliances did not free women from staying home but instead

cost them more time on other tasks. From this, they suggest an indication that care robots may not necessarily always promote caregivers' capabilities but may alternatively lead to more personal sacrifice instead.[32]

A third approach to ethical evaluations of care robots centers on care practices. In this approach, called Care Centered Value Sensitive Design (CCVSD),[46,47] van Wynsberghe insists that care robots need to be evaluated according to their impact on care practices rather than on the impact of either care receiver or caregiver alone. In this way, the robot's evaluation centers on its ability to enhance (or weaken) elements of care practices, such as the attentiveness of health care personnel or the reciprocity between caregiver and care receiver, as necessary conditions for good care.

A focus on the impacts of care practices brings us closer to recognizing that there are external considerations to the care receiver + caregiver relationship that need to be considered. Yet still, what is needed is a way to understand the "bot" as both an extension of the health care system of human caregivers (in so far as care is provided through the "bot") that has substantial impacts on the health care system, namely a rearrangement of the health care system. What is needed now is a way to account for this unique ontological status of the "bot", the rearrangement of the health care system that inevitably accompanies the 'bot, and the ethical issues this raises in a health care context.

A paradigm shift to capture the complexity of health care "bot" + health care system interactions: the HRSI model

In short, the traditional dyadic model of HRI serves as a useful tool for conceptualizing the interaction between humans and "bots"; however, it fails to account for the complexity of the network which the "bot" is stepping into and which the "bot" also adds to. There is an urgency to understanding that "bots" in health care will have significant downstream effects on the health care system, for example, the various forms of restructuring we have raised, given the lack of attention to this topic in the robot ethics or HRI space. Seeing that robot ethics has relied on the HRI model for developing ethical analyses, we suggest the need for a paradigm shift in conceptualizing human and robot interactions in the health care sector. To that end, we suggest a HRSI model, one in which the robot is placed in between a human care receiver on one side and a health care system of human caregivers including professional medical staff such as doctors and nurses and informal caregivers such as family and friends on the other. Along these

lines, Parviainen et al have suggested a triadic model of human–robot–human–interaction as a way of showing the complexity of interactions in health care that go beyond traditional HRI representations. One of the examples they use is nurses escorting care receivers to the operating suite on the mechanical bed. They suggest that having a mechanical bed capable of traveling to the surgical suite without a nurse is possible but fails to account for the significant role the nurse plays in reducing anxiety, in other words, the role the nurse has in the care practice of escorting care receivers to surgery. This example, for the authors of this article, should still be considered an illustration of robots and humans interacting within care practices and can be accounted for in the HRI taxonomy of Yanco and Drury (see Figure) or the CCVSD approach in general.

The necessity of emphasizing the health care system in ethical evaluations are many and center on the ways in which the "bot" will force a restructuring of the health care system: roles of nurses and doctors in health care settings will change since robots can take certain tasks from human medical staff and as such the distribution of responsibilities will also change; some professions, eg, deliverymen and janitors in health care settings, may no longer exist since these manual and repetitive tasks can be delegated to robots; the education of health care professionals will change in order to teach health care staff the necessary skills of working with the "bots"; the expertise of health care staff may be called into question or restructured insofar as certain "bots" will be considered the experts rather than the humans; "bots" will change the flow of money in health care sector in order to purchase and maintain them. The HRSI model is meant to highlight the various ways in which humans and "bots" can interact in order to understand the complexity of introducing "bots" into health care and to come closer to framing the ways in which the health care system will be rearranged.

Types of interaction scenarios in the HRSI model

In general, the "bot" functions as a bridge between the caregiver and/or the health care system of a network of human caregivers, and it is important to remember this to avoid misrepresentations of the "bot" – it is not conscious, sentient, or capable of caring in the humans' sense of the word. Rather, it provides a new kind of access to the care provided by the health care system. We refer to the health care system because there are instances in which health care professionals provide data that are in turn used to

train the algorithm of the chatbot. Thus, a care receiver is not interacting with one health care professional but is interacting with a collection of data from a broad group of professionals. Given the variety of types of "bots" and the variance in types of interactions, it is necessary to outline the HRSI model in more detail.

In the following, we will demonstrate three levels of interaction scenarios in which robots/chatbots/avatars have a critical role based on the complexity of the interactions involved (see Figure 1). These scenarios consist of the following actors: a care receiver (eg, a patient or a user of an app); a "bot"; and a health care system (which can be a variety of professionals in the system or one nurse or one physician, or one family or one friend who provides care). The scenarios sketch the divergent ways in which interaction can happen. It should be noted that in each scenario the "bot" is at the center place indicating that it functions as a mediator between the caregiver and the care receiver. The arrows in the figures below represent interactions between the actors and the direction of flow of data. The one-way arrows indicate unidirectional interaction, or flow of data, from one partner to another, while the two-way arrows mean reciprocal interactions, meaning data flows in both directions.

Level 1 – HRSI with limited dyadic interactions

The illustration at Level 1 shows a primitive interaction scenario of HRSI. The two-way arrow indicates there are reciprocal "care receiver + robot" interactions and the one-way arrow suggests the interaction from the caregiver to the robot. To exemplify this type of interaction, consider Woebot, which is a chatbot developed by clinical psychologists from Stanford University. Researchers expect the chatbot to help with people's mental health using cognitive behavioral therapy techniques. The input into the chatbot, eg, clinical experience and therapy theories, is the one-way interaction from the health care system to the robot. The two-way interaction will be formed if/when a conversation starts between the care receiver and the "bot". Each conversation starts by the chatbot checking in with the user to know his/her feeling and then asking what areas the user wants to be helped with. This form of interaction scenario appears most similar to the traditional HRI model; however, the data provided to the bot for function guides comes from the health care system, and the bot is communicationg with the care receiver on behalf of the health care system.

Level 2 – intermediate HRSI with unidirectional human interactions

Comparing with the first scenario, Level 2 shows an extra arrow between the care receiver and the caregiver along with the two-way arrow between the bot and the health care system. In type A, the one-way arrow shows the interaction from the care receiver to the caregiver. An example to illustrate this in practice is Your.MD, a chatbot using AI to help users have a better understanding of their symptoms. The data input from the health care system to the bot is the large amounts of health information databases from the health care system. To be sure,
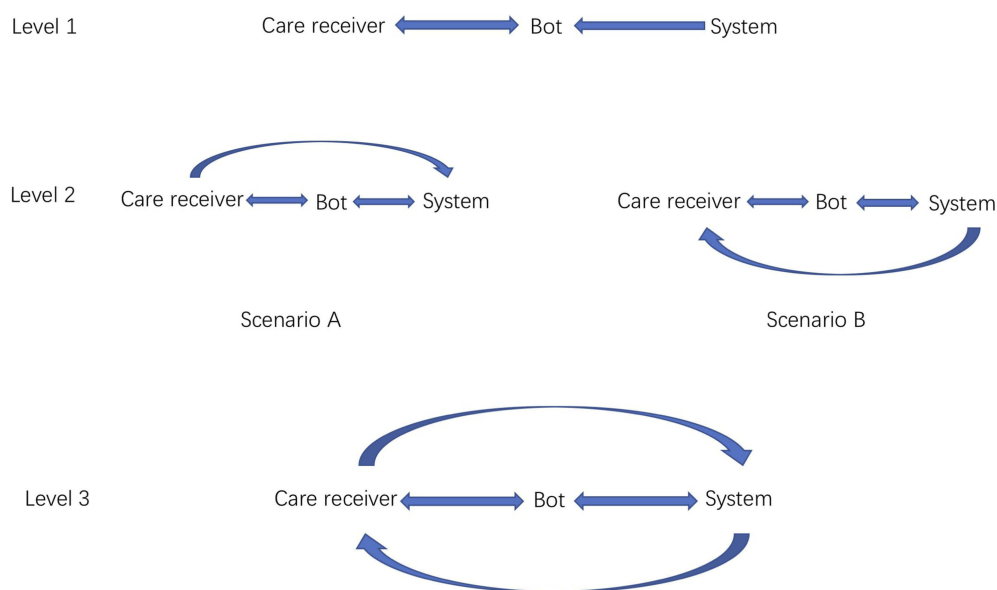

Figure 1 Interaction scenarios of the HRSI model in the health care system.

the health information has been checked by certified doctors in advance. According to the symptoms listed by the care receiver to the chatbot, the users can find information about causes, diagnosis, and/or actionable treatment to make choices for himself/herself, eg, taking specific medication and/or making changes to one's diet. The chatbot can also help make appointments with physicians when necessary.[48]

The interaction between the care receiver and the bot forms when the user starts typing a question. The interaction continues as long as the user asks questions to the chatbot. Input from caregivers is fed to the chatbot in advance, stored, and recalled to provide the care receiver with medical knowledge. When the user is diagnosed as having a serious illness, he/she will most likely prefer to go to a doctor to receive proper treatment. As Your.MD can help make an appointment with a doctor, the care receiver can visit a doctor or nurse and an in-person interaction between the care receiver and caregiver is formed (ie, as indicated by the additional arrow from the health care system to care receiver).

Distinct from type A, the arrow in the diagram of B is pointing from the caregiver to the health care system to show a different way of bots used to draw health care professionals into a direct interaction with the care receiver. Monitoring bots – chatbots, avatars, and embodied robots – are the best case to reflect this interaction scenario. They are designed and used to help prevent falls in care homes and private homes. Monitoring robots such as AILISA and Care-O-bot are equipped with sensors and cameras to keep an eye on the movement of the care receiver. This interaction is between the care receiver and the bot with the purpose of the bot to relay important information to the health care system. If/ when the bot alerts a caregiver and/or the care receiver's relatives to notify that a fall or other frailty has occurred, the caregiver is able to communicate to the care receiver directly through the bot interface. In some instances, after the warning has been received, the caregiver and/or the relatives will travel to the care receiver's location to check the situation and provide help. This kind of immediate reaction from the caregiver to the care receiver is very common in HRSI involving monitoring and is a central reason for which the traditional HRI model is not adequately equipped for ethical evaluations of such scenarios.

Level 3 – advanced HRSI with triadic reciprocal interactions

Level 3 of the HRSI model allows for a representation of the more complex, multidirectional, and reciprocal forms of interaction. In this figure, care receivers may have had initial interaction with a health care professional, such as a surgeon in hospital and then be monitored in their home afterward through a chatbot. Or, the care receiver may engage with one or more health care professionals through a bot while at the same time the bot is collecting physiological information to share with the caregiver and/or the care receiver is at the same time engaging with caregivers present in a care facility.

Consider, for example, the remote presence robot RP-7, a telepresence robotic system designed by Intouch Health. The top of the RP-7 robot is fitted with camera and microphone for real-time two-way audio and video communication between the care receiver and the expert clinician who is off-site and with whom the care receiver is interacting with.[49] The expert clinician uses a joystick to control the locomotion of the robot to have a further detailed observation of the care receiver as well as the environment in the ward. The robot can also record the care receiver's vital signs and send the data to the clinician. Thus, the bot is providing data to the health care professional while at the same time being used as an instrument for direct communication. With the information retrieved and sent by the robot and the real-time video consultation, the expert clinician can make suggestions to the medical staff present on the actions to be taken.

In this scenario, the RP-7 robot makes the remote consultation possible by providing the care receiver and the expert clinician with direct contact (noted as the two reciprocal interactions between the care receiver and the robot). There are also direct interactions between the care receiver and the medical staff present. They can help to perform tests on the care receiver according to the expert clinician's instructions that cannot be achieved remotely by the robot (noted as the two arrows between the care receiver and the caregiver).

The contribution of HRSI to the field of robot ethics

There are many instances when it is important to provide an ethical assessment of the "bot's" impact on individuals or on the ability of caregivers to provide good care. Yet there are also moments in which such an isolated assessment of this kind fails to capture the complexity of the situation (eg, the rearrangement or responsibilities associated with data collection and/or ownership) and consequently the additional ethical issues that go beyond a dyadic HRI. As briefly noted earlier, in most instances,

the health care system is not the institute who has developed the "bot" product; instead, the health care system is the technology implementer and/or the user. In other words, a novel party is being introduced into the care receiver + health care system relationship – the "bot" designer, developer, or distributer. This third party is not in the practice of making "traditional health care tools" but is making data collection tools. Thus, we must question the ethical practices, assessments, and safeguards for this new actor in the care receiver + health care system interaction. For this reason, robot ethics should now begin to engage with the significant role that the "bots" plays as a third party. In the remaining section, our aim is to raise awareness of certain ethical issues resulting from the "bots" introduction and to build on these in future work.

Trust in the health care system, the robot, or … ?

Trust is paramount in any health care situation. Caregivers and more importantly the health care system as a whole must be trusted. In fact, trust is the cornerstone of the professionalization of medicine and nursing. In general, it is easy for care receivers to understand and accept medical instruments such as scalpels and stethoscopes since the doctors understand, endorse, and directly use them. When introducing the robot in between the care receiver and the health care system, the question is whether the care receiver is being asked to trust the health care system, the robot, or the third party involved in the robot's implementation. Given that most care receivers will have no idea who the third party in question is we can assume that their trust in the health care system will extend to the robot. Consequently, the health care system ought to ensure high standards of the "bots".

It should be noted here that for the FDA most robots fall under Class II medical devices in terms of risks and are regulated accordingly, meaning the FDA will enforce oversight. In October 2016, the surgical robot daVinci, categorized as a Class II medical device, was recalled via the FDA because of "a software anomaly in the da Vinci Xi P5 software that can result in unexpected master movement and potential instrument tip movement under certain circumstances".[50] In such instances, companies must communicate with and through the FDA to inform consumers of anomalies. Alternatively, most of the chatbots discussed in the article are classified as Class I mobile medical apps meaning they present minimal risk to patients and in these cases the FDA has 'enforcement discretion' meaning the FDA does not intend to pursue enforcement action for violations of the FD&C Act and application regulations.[51] If something goes wrong, there is not the same

need for companies to communicate with or through the FDA to inform users. Such a divergence is representative of how the "bot" restructures the traditional mechanisms in place for oversight of health care technologies. In these instances then, care receivers are unknowingly placing their trust in the third-party companies making or distributing the "bots" and these companies are not held to the same standards as the hospital (or health care system for that matter). To ensure care receivers are given ample opportunity for informed consent along with placing their trust in the correct institution, they ought to be informed of who and what they are being called upon to trust when interacting with the health care "bot".

Responsibility and accountability

There are three aspects associated with responsibility and accountability that need to be discussed; first, data provenance issues, second, the responsibility of the health care system when care interactions are reduced to the "bot" and the care receiver, and third, given the reality that "bots" will restructure the health care system in a variety of ways, it is paramount to consider how responsibilities are also restructured and further, who is accountable when things go wrong.

Consider, for example, the use of chatbots for preliminary diagnosis. These chatbots persuade users that the diagnoses and ensuing advice are made based on analyses of large datasets with input from medical professionals. Yet, many of these applications insist that the diagnoses and the medical advice are merely a guide and for reference only. In this case, who will be responsible when a user exclusively follows suggestions provided by the chatbot but his/her medical situation deteriorates? In traditional doctor–care receiver relationships, the doctor (supported by the health care institution where he/she works) bears responsibility for medical accidents, but if the chatbot assumes the task of initial assessment then who is responsible when things go wrong? This invites a discussion of the quality of the training data used and the reliability of the algorithm used for prediction, issues concerned with the ethics of AI in general but which are increasingly critical when AI is used in a health care context. To ensure high reliable diagnoses, all chatbots should be subject to rigorous validation standards and regular IT and procedural auditing. As discussed above, chatbots, in particular, are still considered Class I medical apps and as such are not required to follow such criteria.

Another kind of restructuring has to do with the "bots" taking on certain roles or tasks of health care staff: when "bots" are introduced as mediators between the care receiver and the health care system there may be an impoverished

interaction between patient and health care system (ie, when the bot is acting on behalf of the health care system). First, there will be technological limits to both what the robot is capable of taking in from the care receiver and what the robot is capable of conveying to the health care system. The loss of contextual details when using a bot as the mediator in health care may, in turn, lead to imprecise and unsatisfactory care. Chatbots used for preliminary diagnoses may not fully capture external factors related to a care receiver and/or a chatbot may not ask the same set of questions as the professional caregivers will ask, which may lead to miscategorizations of a care receivers needs.

Another instance in which the loss of contextual details may have serious repercussions is when care receivers are suffering from abuse and the only way to know this is by observing them in person. While there may be practical limits to how much a doctor or nurse can take in from a care receiver, they are free to "go the extra mile" when they deem necessary. A pediatrician may suspect that a child is being abused and if pressed the pediatrician may claim that her suspicion is simply a "hunch". However, this hunch may lead the pediatrician to act in a way which may confirm or refute that hunch. The pediatrician may feel responsible for this child in a way that a "bot" never could. The addition of the robot as mediator may unintentionally reduce care interactions to a simple exchange of physical details rather than paying tribute to the holistic view of a care receiver or worse the distance between care receiver and health care system may lead to absolving (either symbolically or casually) the health care system of legal, moral, and any feeling of responsibility.

Conflicting preferences

There may be instances in the near future in which care receivers neither trust in the technology nor wish to interact with it. This could create a conflict between the needs of health care systems or institutions to systematize portions of care processes and care receivers who wish to interact with humans for each portion of the care. Consider, for example, the Japanese lifting robot ROBEAR for lifting care receivers. When the health care system of the network of caregivers has decided on the use of the robot for reasons of efficiency and the nurse (also a caregiver) must implement this choice, what happens when the care receiver refuses to be lifted by the robot? How can care receivers make choices about their care if caregivers are bound by the choices of the institution? While this may seem commonplace, nurses are frequently asked for alternative options, it does not diminish the fact that caregivers and care receivers should still be provided the autonomy to

make choices about the provision of care especially without proper evidence showing that robot care is superior to human care.

Alternatively, care receivers may desire impersonal interactions with care bots over personal human interactions whether it is providing convenient and timely answers to their questions or assisting with toilet time. Having a bot available could be more convenient or could provide a more dignified form of intervention. In either case, in a study done by Parviainen et al, the authors conducted empirical research to demonstrate how care workers perceive robots: "the caregiver and care receiver make use of a technological device in ways that suit their needs without losing the possibility for human touch and interaction".[52] Considering that the robot provides access to the health care system, it is paramount for the institution to provide care workers with the freedom to navigate these situations as they see fit. To that end, we suggest that explicit and proactive efforts be made to include care staff (eg, physicians, nurses, porters, cleaners, managers among others) to be made a part of the design process insofar as their experiences and voices are included in the conceptual thinking about the "bot".

Conclusion

In view of the growing applications of bots in health care, the various ethical analyses common to HRI in the health care space now seem inadequate. Specifically, we argue that the HRI label fails to pay tribute to the system of health care workers in place or the rearrangement of responsibilities and complexities that a bot in health care introduces. To overcome this limitation, we propose the HRSI framework for evaluating the impact that the robot will have not only on the individual patients and/or care providers but on the entire health care system.

We suggest that the introduction of the range of "bots" in health care (eg, embodied robots and AI, avatars, and chatbots) will create a restructuring within the health care system in a variety of ways, from a redistribution of roles and responsibilities (ie, that "bots" will take on jobs previously done by human workers) to the new ways in which money will be allocated and/or health professionals will be trained. The impact on health care staff will also require new kinds of empirical studies that go beyond the traditional framework found in HRI. Based on these forms of restructuring, we suggest empirical research to assess the subjective experience of care workers following the introduction of the "bot" regarding their previous and new roles and track educational changes over time (eg, new courses offered and older courses dropped).

Moreover, we also suggest transparency on the part of health care institutions concerning the financial reports when "bots" have been purchased. This final point allows for an assessment of the restructuring of hospitals pre- and post-"bot".

Based on our suggestion that "bots" in health care will undoubtedly restructure the health care system coupled with the need for new empirical methods to study and evaluate this, it is time for robot ethics to place more emphasis on, and ask more questions of, the designers, developers, and implementers of robots in order to shift responsibility and accountability toward these institutions for the possible negative ethical issues resulting. A commitment to good health care requires accountability on the part of health care systems as well as the third-party developers for the "bots" that implemented in care.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. International federation of robotics. Executive Summary World Robotics 2018 Service Robots. World Robotic Report - Executive Summary. Available from: https://ifr.org/downloads/press2018/Executive_Summary_WR_Service_Robots_2018.pdf. Published 2018. Accessed May 20, 2018.

2. Zion market research. Chatbot market by type (support, skills, and assistant), by end-user (healthcare, retail, BFSI, travel & hospitality, E-commerce, media & entertainment, and others): global industry perspective, comprehensive analysis, and forecast 2017–2024. Available from: https://www.zionmarketresearch.com/report/chatbot-market. Published 2018. Accessed December 2, 2018.

3. Veruggio G, Operto F. Roboethics: social and ethical implications roboethics. In: Siciliano B, Khatib O, editors. *Springer Handbook of Robotics*. Berlin: Springer; 2008:1499–1524.

4. Sparrow R, Sparrow L. In the hands of machines? The future of aged care. *Minds Mach*. 2006;16(2):141–161.

5. Sharkey N, Sharkey A. The crying shame of robot nannies: an ethical appraisal. *Interact Stud*. 2010;11(2):161–190.

6. Lin P, Abney K, Bekey G. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge: The MIT Press; 2016.

7. Vallor S. Carebots and caregivers: sustaining the ethical ideal of care in the twenty-first century. *Philos Technol*. 2011;24(3):251–268.

8. Department of Health and Social care. NHS health information available through amazon's Alexa. Available from: https://www.gov.uk/government/news/nhs-health-information-available-through-amazon-s-alexa. Published 2019. Accessed July 18, 2019.

9. Huang J, Zhou M, Yang D. Extracting chatbot knowledge from online discussion forums. In: IJCAI International Joint Conference on Artificial Intelligence; 2007; Hyderabad, India:423–428

10. Hill J, Randolph Ford W, Farreras IG. Real conversations with artificial intelligence: a comparison between human-human online conversations and human-chatbot conversations. *Comput Human Behav*. 2015;49:245–250.

11. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Heal*. 2017;4(2):e19–e19. doi:10.2196/mental.7785

12. Orrell C. Trust and accuracy in diagnostic AI: your. MD's story so far. Available from: https://medtechengine.com/article/your-md/. Accessed November 1, 2018.

13. Yang G-Z, Bellingham J, Dupont PE, et al. The grand challenges of science robotics. *Sci Robot*. 2018;3(14):eaar7650. doi:10.1126/scirobotics.aar7650

14. Kachouie R, Sedighadeli S, Khosla R, Chu MT. Socially assistive robots in elderly care: a mixed-method systematic literature review. *Int J Hum Comput Interact*. 2014;30(5):369–393. doi:10.1080/10447318.2013.873278

15. Kidd C. Introducing the Mabu personal healthcare companion. Available from: https://www.cataliahealth.com/introducing-the-mabu-personal-healthcare-companion/. Published 2015. Accessed July 18, 2019.

16. Earnhardt J. Technology ushering in healthcare's 'Golden Age'. Accessed October 7, 2009 Available from: https://blogs.cisco.com/news/technology_ushering_in_healthcares_golden_age. Accessed October 29, 2018.

17. Sensely. MEET MOLLY, YOUR VIRTUAL ASSISTANT. Available from: http://www.sensely.com/. Accessed October 29, 2018.

18. Sheth R. Avatar technology: giving a face to the e-learning interface. *eLearning Dev J*. August 25; 2003:1–10.

19. Kazerooni H. Human-Robot Interaction via the Transfer of Power and Information Signals. *IEEE Trans Syst Man Cybern*. 1990;20(2):450–463. doi:10.1109/21.52555

20. Held RM, Durlach NI. Telepresence. *Presence Teleoperators Virtual Environ*. 1992;1(1):109–112. doi:10.1162/pres.1992.1.1.109

21. Kazerooni H. Extender: a case study for human-robot interaction via transfer of power and information signals. In: Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication; Tokyo, Japan. 1993: 10–20. doi:10.1109/ROMAN.1993.367756

22. Breazeal C. A motivational system for regulating human–robot interaction. In: Proceedings of the National Conference on Artificial Intelligence; Madison, USA; 1998:54–61.

23. Dautenhahn K. Methodology & themes of human-robot interaction: A growing research field. *Int J Adv Robot Syst*. 2007;4(1):103–108. doi:10.5772/5702

24. Goodrich MA, Schultz AC. Human-robot interaction: a survey. *Found Trends® Hum Comput Interact*. 2007;1(3):203–275. doi:10.1561/1100000005

25. Yanco HA, Drury JL. A taxonomy for human-robot interaction. In: AAAI Fall Symposium on Human–Robot Interaction; Falmouth, Massachusetts; 2002:111–119.

26. Yanco HA, Drury J. Classifying human-robot interaction: an updated taxonomy. *Conf Proc - IEEE Int Conf Syst Man Cybern*. 2004;3(May):2841–2846. doi:10.1109/ICSMC.2004.1400763

27. Veruggio G. The birth of roboethics. In: IEEE International Conference on Robotics and Automation; Barcelona, Spain: Workshop on Roboethics; 2005:1–4.

28. Veruggio G. *EURON Roboethics Roadmap*. Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots; 2006; Genoa, Italy;612–617.

29. Denning T, Matuszek C, Koscher K, Smith JR, Kohno T. A spotlight on security and privacy risks with future household robots. In: Proceedings of the 11th International Conference on Ubiquitous Computing; Orlando, USA. 2009.

30. Calo MR. Robots and privacy. In: Lin P, Abney K, Bekey GA, editors. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge: MIT Press; 2011:187–202.

31. Feil-Seifer D, Matari MJ. Socially assistive robotics ethical issues related to technology. *Robot Autom Mag.* 2011;18(1):24–31. doi:10.1109/MRA.2010.940150

32. Borenstein J, Pearson Y. Robot caregivers: harbingers of expanded freedom for all? *Ethics Inf Technol.* 2010;12:277–288. doi:10.1007/s10676-010-9236-4

33. Sharkey A. Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics Inf Technol.* 2014;16:63–75. doi:10.1007/s10676-014-9338-5

34. Sharkey A, Wood N. The paro seal robot: demeaning or enabling? Proceedings of the 50th Annual Convention of the AISB; 2014; London, UK.

35. Sharkey A, Sharkey N. Children, the elderly, and interactive robots: anthropomorphism and deception in robot care and companionship. *IEEE Robot Autom Mag.* 2011;18(1):32–38. doi:10.1109/MRA.2010.940151

36. Körtner T. Ethical challenges in the use of social service robots for elderly people. *Z Gerontol Geriatr.* 2016;49(4):303–307. doi:10.1007/s00391-016-1066-5

37. Parks JA. Lifting the burden of women's care work: should robots replace the "Human Touch"? *Hypatia.* 2010;25(1):100–120. doi:10.1111/j.1527-2001.2009.01086.x

38. Friedman B, Kahn PH. Human values, ethics, and design. In: The Human-Computer Interaction Handbook; Hillsdale, NJ: L. Erlbaum Associates Inc.; 2002:1177–1209.

39. Riek LD, Howard D. A code of ethics for the human-robot interaction profession. In: Proceedings of We Robot Conference; 2014; Coral Gables, FL:1–10.

40. Belpaeme T, Baxter PE, Read R, et al. Multimodal child-robot interaction: building social bonds. *J Human-Robot Interact.* 2013;1(2):33–53. doi:10.5898/JHRI.1.2.Belpaeme

41. Vallès-Peris N, Angulo C, Domènech M. Children's imaginaries of human-robot interaction in healthcare. *Int J Environ Res Public Health.* 2018;15:970. doi:10.3390/ijerph15061188

42. Arnold T, Scheutz M. The tactile ethics of soft robotics: designing wisely for human-robot interaction. *SOFT Robot.* 2017;4(2):81–87. doi:10.1089/soro.2017.0032

43. van Wynsberghe A. A method for integrating ethics into the design of robots. *Ind Robot An Int J.* 2013;40(5):433–440. doi:10.1108/IR-12-2012-451

44. Vallor S. Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character. *Philos Technol.* 2015;28(1):107–124. doi:10.1007/s13347-014-0156-9

45. Coeckelbergh M. Health care, capabilities, and AI assistive technologies. *Ethical Theory Moral Pract.* 2010;13(2):181–190. doi:10.1007/s10677-009-9186-2

46. van Wynsberghe A. Designing robots for care: care centered value-sensitive design. *Sci Eng Ethics.* 2013;19(2):407–433. doi:10.1007/s11948-011-9343-6

47. van Wynsberghe A. *Healthcare Robots:Ethics, Design and Implementation.* Farnham: Ashgate Publishing Ltd.; 2015.

48. Your.MD. Your personal health guide and symptom checker. Available from: https://www.your.md/. Accessed October 29, 2018.

49. Sharkey N, Sharkey A. The eldercare factory. *Gerontology.* 2012;58(3):282–288. doi:10.1159/000329483

50. FDA. Class 2 device recall da Vinci Xi" surgical system. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfres/res.cfm?id=150460. Accessed January 16, 2019.

51. FDA Center for Devices and Radiological Health. Mobile medical applications guidance for industry and food and drug administration staff. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/mobile-medical-applications. Published 2015. Accessed January 15, 2019.

52. Parviainen J, Turja T, Van Aerschot L. Robots and human touch in care: desirable and non-desirable robot assistance. In: Ge S. et al., editors. Social Robotics. Proceedings of the 10th International Conference on Social Robotics; 2018; Qingdao, China:533–540.