Ethics and Accountability of Care Robots

Jyri Rajamäki and Jaakko Helin
Laurea University of Applied Sciences, Espoo, Finland
jyri.rajamaki@laurea.fi
jaakko.anton.helin@gmail.com

Abstract: The subject of this paper is ethical and responsibility issues relating to the development and acquisition of robotics in healthcare. The purpose of the paper is to study previous scientific publications and research related to the topic and to clarify which questions, aspects, and concerns are most relevant when considering ethics and responsibility issues related to care robots. In the second phase, ideas from different stakeholders regarding the viewpoints are studied, and those ideas are compared to the ones presented in previous publications. The aim of this study is to find solutions to the issues presented in scientific literature and, also, to find new issues for consideration and further studies. The study is qualitative, and a theme interview was utilized as the main method for acquiring knowledge. The study is a part of the SHAPES Horizon 2020 project. From the perspective of SHAPES, the aim of the study is to provide useful knowledge for the project, which would in part promote the goal of SHAPES, i.e., the development of an international healthcare ecosystem. Based on the results of the study, it can be argued that the issues presented in previous academic publications regarding the ethics and accountability of robots in practical healthcare work are not relevant. Both the legislation and the logic of the AI algorithms used by care robots prevent those situations presented in previous academic discussions in which robots would presumably be forced to make decisions demanding ethical consideration. The results also point toward the fact that current legislation does not limit the development of healthcare robots more than it limits healthcare work in general. Thus, the considerations of ethics regarding care robots should rather be focused on the threshold values used by robots, when making interpretations, as well as the data used for the purpose of machine learning. These were identified as potential subjects for further research.

Keywords: care robot, ethics, accountability

1. Introduction

Robots can make everyday life easier, create a sense of security, and perform a variety of tasks. Their main classes are industrial and service robots depending on whether they are used for the benefit of industry or to perform tasks useful to humans (Fosch-Villaronga & Mahler, 2021). Service robots are able to perform partially or completely independently services that are beneficial to human well-being, or the environment; social robots complement, increase, or replace human social interaction (Särkikoski et al. 2020). Care robots can perform tasks related to physical or mental care independently or semi-automatically such as assisting in daily tasks, rehabilitation, or mental care (Van Aerschot & Parviainen, 2020; Särkikoski et al. 2020). Care robots can help with daily activities and provide companionship and a sense of security. However, they cannot completely replace human work in helping the elderly in their daily activities (Van Aerschot & Parviainen, 2020). Because care robots are directly involved in human life, they raise crucial ethical problems for our society (Tzafestas, 2018).

This qualitative study highlights stakeholders' current views on the ethical and responsibility aspects of care robotics through thematic interviews. The study is a part of the SHAPES Horizon 2020 project, and the aim is to provide useful knowledge for the project, which would in part promote the goal of SHAPES, i.e., the development of an international healthcare ecosystem.

After this introduction, Section 2 reviews some previous literature on the ethics and accountability of care robots. Section 3 describes the applied research methods and interviewed persons. The themes covered in the interviews emerged from the literature review, and Section 4 presents the results of the interviews. Finally, Section 5 summarizes the literature review as well as the interviews, discusses the results, and concludes the paper.

2. Literature review

For this article, previous research and publications were studied to identify the essential questions, viewpoints, and concerns regarding the ethics and accountability of care robots. For the research, other fields where robotics implementation and development have been present were excluded from the study, and only the healthcare sector and related artificial intelligence (AI) and robotics development were assessed. According to Nevanperä, Rajamäki and Helin (2021), when researching the ethical aspects of AI development, it is sensical for the sake of

the integrity of the research and its results, to narrow the perspective down to include only a specific field or industry instead of generalization.

Gerritsen, Kool & Van Est (2017) have brought up in their report various situations, where an autonomically functioning robot could cause problems concerning accountability. Elderly care includes tasks that can generally be speaking be viewed as beneficial to the patients even though the patients themselves might oppose them. These include e.g., binding the patient to hold them in place, pressuring them to take medication, or force-feeding. In these types of situations, a large amount of trust is placed on the developer of the robot, that they have been mindful during the development process, which is the best operational model from the point of view of the end-user. At worst, the robot's actions can be seen as borderline autocratic and in conflict with the autonomy of the patients themselves. Recent studies have, however, found that such risks have not been realized but, instead, the patients' attitudes towards, for example, a robotic feeding assistant have been positive, because the use of robots, in this case, enable better patient privacy. Problematic situations can also be circumvented by configuring the robots according to EU recommendations so that the robot expresses its intentions clearly and the end-user has the option to cancel the task. When considering general attitudes towards care robots, according to Cresswell, Cunningham-Burley and Sheikh (2018) stakeholders' attitudes have been mainly negative in the past. Healthcare professionals are worried about their jobs and there's a general lack of acceptance of robotics, However, this is mainly founded on low usage of robots and a general lack of interaction between people and robots and thus can be remedied with the right actions and training. (Gerritsen, Kool & Van Est 2017, 25-28; Cresswell, Cunningham-Burley & Sheikh 2018, 7-9.)

There is currently no consensus on how accountability, legal or moral, should be divided among stakeholders when using care robots. According to Cresswell, Cunningham-Burley and Sheikh (2018) accountability in problematic situations has been determined case-by-case (Cresswell, Cunningham-Burley & Sheikh 2018, 11-12.). According to Anton, Karpov, Lahti & Saukkonen (2021) moral and legal accountability, stakeholder rights and uncertainties regarding these cause challenges in the use of AI and currently stakeholders are likely to try to avoid taking responsibility for the result of AI decision making, especially if the decision is regarded as being wrong or unjust (Anton, Karpov, Lahti & Saukkonen 2021, 186-187). According to Leenes, Palmer, Koops, Bertolini, Salvini & Lucivero (2017) regulations regarding robotics is contradictory in general, because regulations and legislation should keep up with the development of robotics without obstructing it and anticipatory legislature can even cause significant risks business-wise to robotics developers, if they have no clear idea on how the legislature is going to affect their business in the future (Leenes, Palmer, Koops, Bertolini, Salvini & Lucivero 2017, 1-44). According to Voss (2021) in the EU, the lack of consensus concerning legislature hinders the development of care robots, for each country have their independent legislative institutions as well as enforcement, which makes EU a high-risk market for robot developers (Voss 2021, 21-25). However, it is worth noting that the Finnish patient insurance law (948/2019) clearly states that a healthcare service provider is accountable in accordance with law for the services they provide and must have sufficient insurance. In cases of accident, legal accountability is divided among different stakeholders in accordance with the stake they had in the cause of the accident.

3. Research method

This study is qualitative in nature and the thematic interview is the main method for acquiring knowledge. These methods were chosen due to the complex and broad nature of the topic, general lack of similar previously conducted studies as well as the aim of the study being acquiring new knowledge. According to Hirsjärvi, Remes & Sajavaara (2007) the aim of qualitative research is to depict real life and its diverse phenomena. In qualitative research, the aim is to holistically study the research subject and the material is gathered from studying real and naturally occurring events in their actual form. Qualitative research shouldn't be used as a method to ratify already existing hypotheses.

As for research methodology, when researching new fields which have not yet been thoroughly studied, a more flexible method for acquiring knowledge can yield greater results than a rigid questionnaire. Allowing the subjects to freely express their thoughts can more easily lead to new ideas and viewpoints. According to Hirsjärvi, Remes & Sajavaara (2007) an interview is a versatile and flexible method suitable for many purposes. The interview can be done in direct interaction with the subject and thus it is optimal for uncovering any hidden ulterior motives the subjects might have. According to Hirsjärvi & Hurme (2014), when conducting a thematic interview, the interviewer should define the theme of the interview beforehand. However, it is possible to define

a different set of themes separately for each individual interviewee. According to Kankkunen & Vehviläinen-Julkunen (2013), content analysis is frequently used to study qualitative research results. The aim of content analysis is to condense the gathered knowledge using a variety of models, concepts, and classifications and to create descriptions, and uncover meanings as well as cause-effect-relations (Hirsjärvi, Remes & Sajavaara 2007, 152-155, 198; Hirsjärvi & Hurme 2014, 47-48).

For this paper, interviews with five different stakeholders were conducted in order to obtain their opinions on the topics presented in the literature. Table 1 presents the stakeholders and the time of the interviews.

Table 1: Interviewed stakeholders

Nickname	Stakeholder	Date of the interview
'Lawyer'	Master of Laws with court training	2nd January 2022
'Nurse'	Practical Nurse	13th January 2022
'Robot supplier'	CEO of a Finnish robotic systems development company	22nd January 2022
'Relative'	A relative of an elderly person	26th February 2022
'Consultant'	Senior Manager of a global information technology services and consulting company	4th March 2022

The aim was to find potential solutions to the previously brought up ethical conundrums as well as potentially uncover new subjects for future research. The study was conducted as part of the international SHAPES project and the aim was to produce useful information for enabling the project's goal of developing an international healthcare ecosystem.

4. Results

This section presents the results of the interviews. The interviewees are referred to as described in the previous section: the Lawyer, the Nurse, the Robot supplier, the Relative, and the Consultant.

4.1 Healthcare legislature

According to the Lawyer, when considering the Finnish legislature regarding care robots either product accountability legislation or patient insurance legislation is applied. If a robot is supplied to the patient's own home, product accountability law is applied, and if the robot is used as a part of health care service production, patient insurance law is applied. The Lawyer mainly concerns robots used in elderly care services and due to the expensive nature of robotics, it is not reasonable to presume that elderly folks living in their own homes would have sufficient financial means to either acquire or finance the development of a care robot.

Considering the legislature concerning care robots, based on the research it can be argued that the current legislation does not limit care robot actions, or the development of care robots more than it limits health care workers' actions or service development in general. According to the Nurse, patients always have full autonomy unless they are under legal guardianship. The patient has full control over whether they want to be treated or not and even if the healthcare worker does not agree with the decision of the patient, it is not a valid reason to limit the patient's autonomy. Even persuasion can be regarded as questionable from the point of view of the legislature and thus healthcare workers need to be mindful even when persuading patients.

The mental health legislature in Finland includes a few exceptions to this rule, but even in those exceptional circumstances, it is crucial to initially use the lightest possible methods. According to the Nurse, de-escalation methods must always be the first tool in difficult situations and physical intervention is only to be used as the last resort, and even then, the healthcare employee should have ultimate control over the actions of the robot. De-escalation by talking is built upon the trusting relationship between the patient and the nurse, which is based on a long-lasting interaction and coexistence. Robots cannot take into account personal situations or history. Patients need to be able to feel safe and confident that they are properly taken care of and pre-defined parameters and answers will not be enough to achieve this. According to the Relative, it is important to get the patients to take their medicine and take care of their hygiene even if they have mental health issues. However, from the perspective of a relative, even then talking should always be the first method to be used. According to the Nurse and Relative, it is important to remember that demented patients can get scared of new situations more easily.

4.2 Role of care robots in elderly care

According to the Relative, humane treatment must be at the core of geriatric care services. Even if robots are present in providing parts of the service, the patients must be treated as human beings and communicated with. It is the responsibility of the service provider to be aware of the actual allocation of accountability among the employees and to communicate this allocation to the patients and relatives. Moreover, even if robots are utilized to provide some parts of the service, a human being must always bear the ultimate accountability when it comes to the actions of robots. The Consult agrees that a human must be accountable for any AI-based decisions. Thus, according to the Relative and the Consult, it is sensible to have transparency in the AI decision-making models.

According to the Robot supplier, the possibilities of AI and robots to make patients' life easier are currently rather limited. There are plenty of smart solutions, such as surveillance technology, different types of sensory equipment as well as simple robotic applications providing patients with for example physical support, which can be easily utilized and are effective. Developing a complex astronomical care robot, which would take care of all the fore mentioned things, would be a tremendously harder and more expensive task. The same result can be achieved using a multitude of simpler solutions.

4.3 Ethical considerations related to AI algorithms and machine learning

One question demanding ethical consideration with care robots regards the accuracy of decisions made by the artificial intelligence algorithm used by the robots. According to the Consultant, 100 % accuracy is practically impossible to achieve in all situations and compromise must be achieved in setting acceptable error rates. According to the Robot supplier, the AI or robot should never end up in a situation where complex decision-making is expected, but rather they should only be allowed in situations where binary decisions are possible. In those cases, the question of the decision being "right" or "wrong" really comes down to the sensory signals the robot receives and the accuracy of the situational analysis of its surroundings the robot makes based on those signals. Thus, should a robot fail to do the right decision in a certain situation, it is due to the robot failing in observing its surroundings, not a failure in the decision-making process itself. As for the situational analysis and observing the surroundings, the real ethical question regards setting the threshold values for the AI used by the robot. An example of this could be a situation where a care robot should interpret based on the sensory signals it receives how likely it is that the patient is having a medical emergency and whether action must be taken. If the threshold for the likelihood of the emergency is set too high, the signs might be dismissed too easily by the robot and the patient could suffer because of it. However, if the threshold is set too low, the robot would be falsely prompted into action even when there's no medical emergency thus rendering the robot useless. For this threshold value margin of human error could be used.

Another ethical question concerns the data used in machine learning. According to the Robot supplier, most AI solutions used in healthcare are machine learning based. Thus, the data used for machine learning plays an integral role. According to the Consultant, ethnic profiling based on given data could be cynical, but it might not be ethical. Also, for example, excluding subjects' income or sex as a factor in AI decision-making might be ethically more sustainable, even though the correlation was present. For example, breast cancer is more frequent in women and there is a correlation between income and medical problems within the musculoskeletal system. Still, it is possible that profiling based on the aforementioned factors can lead to extreme cases where an automatically made profile based on diagnosis causes exclusion of other important factors which could lead to a false diagnosis. Thus, some factors are frequently and purposefully excluded from the data used for machine learning.

4.4 Transparency of AI algorithms and machine learning data

One question pondered in previous academic discourse regards the transparency of AI algorithms. Some have even suggested making the algorithms public so that anyone could have access to them. However, according to the Consultant, algorithm transparency in AI decision-making is not the main issue, but rather it should be made possible to point out the relevant and included factors on which the decisions are based. The complexity of AI algorithms makes explainability of the decision-making process even harder and explaining it in clear and understandable terms becomes more difficult when AI and dataset complexity grows larger or diversifies. In the healthcare sector, AI solutions subjects are always grouped based on certain factors, and analysis is done based on the grouping and some groups can have several dozen attributes. The grouping also helps the decision-making process more explainable and thus improves transparency.

According to the Robot supplier, it should be noted, that usually both, the algorithms and machine learning data are considered confidential information by the developer companies. However, to achieve adequate transparency, one solution could be to appoint a public institution for producing and maintaining a publicly available dataset for machine learning. The produced data could be used for benchmarking Ai algorithm simulations and with the output result, the algorithm itself can be verified to be working in accordance with any predefined set of rules and regulations, ethical or legislative in nature. This kind of process is already in use in academic circles and autonomous car development. The Consult agrees that a publicly available dataset could be a good idea, if there are no issues regarding people's personal privacy. Especially in the case of health information, privacy is crucial. The risk of data privacy breach is higher than usual since it is possible, for example, with some rare diseases or conditions to discover the person's identity even from anonymous data. There are possibilities on how to prevent this, for example, data masking.

5. Discussion and conclusions

In the literature review, presumed problems encountered by autonomous care robots were presented. Based on the interviews, it can be argued that many of these presumed problems do not actually concern, nor will they ever concern care robots. Some problems presented concern other robotic solutions such as self-driving cars. The most notable discrepancy between the theoretical issues presented in scientific literature and the reality of care robot development seems to be in the attitudes towards legislation and regulation. The literature review found that certain concerns have been highlighted in the past about situations where the legislation and regulation are too strict and thus hindering AI and robotics development as well as situations where the regulations are not strict enough which would impose risks on end-users' safety. Especially the incoherence of legislature in the EU was considered a hindering factor for AI and robotics development and thus should be reviewed and redefined. Research done for this article does not support the view.

5.1 Perceived issues concerning care robot development

The Finnish legislation and regulations for the healthcare industry are strict, but still, they aren't directly limiting the development of care robots. Care robots operate under the same rules and legislation like health care workers and service providers in general. The laws have been created to secure patient rights with the autonomy of patients being at its core and this legislative framework also functions as the regulatory basis for developing care robots. In fact, it could be argued that the clear legislative framework simplifies care robot development since the legislature doesn't have room for interpretation. The legislation also, in part, correlates with the perception of the nature of nursing work. Nursing work is based on trust between the caretaker and the patient, and the legislation supports this aspect. Even in rare cases, where limitation of a patient's autonomy is necessary, the same trust is key. Considering the current speed and trajectory of technological development, it seems rather impossible for robots and artificial intelligence solutions to replace this relationship of trust between people. Thus, any actual limitations hindering care robot development lie in the human nature of healthcare and nursing work, not in the legislature concerning it. It is also worth noting, that the current legislation clearly defines accountability for healthcare service providers and according to the research there seems to be no interest nor need to redefine these laws. A service provider is always responsible that the healthcare service quality they are providing meets the standards defined in the legislature regardless of whether robots are used to provide some part of the service or not. All stakeholders also agree that a human has to always be ultimately accountable for a robot's actions. Based on the interviews it can also be argued that developing complex care robots might not be as sensible, to begin with as directing the development efforts towards creating a more de-centralized ecosystem of different solutions. Complex autonomous care robots are expensive to develop, and the technology is still not at a level where complex tasks currently done by human workers could be achieved by robots. Moreover, robots are not yet considered trustworthy enough that they could be given full autonomy. Human intervention is still required and demanded. Combining this notion with one of the human natures of nursing work and the lack of sufficient technological advancement, it can be argued that any theoretical ethics-based issue related to complex autonomous care robots will not be realized for the reason that such robots might never come to existence in the first place.

5.2 Real ethical issues regarding care robots requiring consideration

One significant issue requiring further ethical consideration is where to draw the line for thresholds when interpretations are being made by robots and AI in general. This question directly leads to the question of acceptable error margins in AI and robotics, since 100 % accuracy is impossible to achieve in each individual

situation. In the case of care robots, a threshold value set too low could pose a risk to a patient's safety and a value set too high could cause false alarms and inadvertently render the robot useless. No consensus currently exists on these threshold values, and it is a potential subject for future studies.

When comparing the suggestions presented in scientific literature to the ones presented in the interviews it can also be argued that publicly produced and available datasets for AI applications for healthcare and care robots would be a sufficient way to guarantee algorithm transparency. Publishing the algorithms themselves would pose a significant risk to developers from a business perspective and even if the algorithms were public, it would not directly mean that the AI decision-making process would be any more understandable to the general public due to the process complexity. A publicly maintained dataset would potentially be an effective way to direct and affect the AI development process without taking control of the algorithm development itself. The dataset would require constant work, however, since it has to be updated and maintained. This type of dataset could also be another potential subject for further study.

Since the study is among the first of its kind, it can be argued that a more general perspective on the matter at hand obtained from different stakeholders is more useful in determining the relevance of the ethical conundrums especially from further studies perspective than concentrating on each individual stakeholder's viewpoint in excessive detail. It should also be noted that the importance of co-creation in the field of robotics is promoted in scientific literature as well. Thus, the study serves as a starting point for further, more detailed studies of the field of care robot ethics.

Acknowledgements

This work was supported by the SHAPES project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no. 857159.

References

Anandarajah, L., Monett, D. 2021. "Will Robots Take all the Jobs? Not yet." *ECIAIR 2021 Proceedings*. Computer Science, Berlin School of Economics and Law, Berlin, Germany. DOI: 10.34190/EAIR.21.029

Anton, K., Karpov, N., Lahti, W. & Saukkonen, J. 2021. Human Rights, Employee Rights and Copyrights: Parallels of AI Enablers and Obstacles Across Occupations in Human-Centric Do-mains. ECIAIR 2021 proceedings. Undergraduate program in International Business, JAMK Uni-versity of Applied Sciences, Jyväskylä, Finland. International Business, JAMK University of Ap-plied Sciences, Jyväskylä, Finland. DOI: 10.34190/EIAIR.21.003

Cresswell, K., Cunningham-Burley, S., Sheikh, A. 2018. "Health Care Robotics: Qualitative Exploration of Key Challenges and Future Directions." *Journal of Medical Internet Research* Vol 20, No 7. Referred 15.6.2022. https://www.jmir.org/2018/7/e10410/

Fosch-Villaronga, E. and Mahler, T. (2021) "Cybersecurity, safety and robots: Strengthening the link between cybersecurity and safety in the context of care robots", *Computer law & security review* (41). DOI: 10.1016/j.clsr.2021.105528

Gerritsen, J.B.A., Kool, L., Van Est, R. 2017. "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. Referred 15.6.2022. https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf

Hirsjärvi, S., Remes, P. & Sajavaara, P. 2007. *Tutki ja kirjoita*. Keuruu: Otava.

Kankkunen, P. & Vehviläinen-Julkunen, K. 2013. *Tutkimus hoitotieteessä*. 3., uudistettu painos. Helsinki: Sanoma Pro.

Mielenterveyslaki (1190/1116). Referred 2.2.2022. https://www.finlex.fi/fi/laki/ajantasa/1990/19901116#L4P22

Nevanperä, M., Rajamäki, J. Helin, J., 2021. "Design Science Research and Designing Ethical Guidelines for the SHAPES AI Developers." *Procedia Computer Science*, Volume 192, 2021. Referred 12.4.2022. https://reader.elsevier.com/reader/sd/pii/S1877050921017191?token=85945C89B3030EF6BB871D9C851CF6E8F374C9E1C3593CB5D285A3100F71C1C4FE4BCDAFC5F28AEDDBFE5579FFF9F818&originRegion=eu-west-1&originCreation=20220412141011

Särkikoski, T., Turja, T. and Parviainen, J. (2020) *Robotin hoiviin? Yhteiskuntatieteen ja filosofian näkökulmia palvelurobotiikkaan*, Vastapaino, Tampere.

Tzafestas, S. (2018) "Roboethics: Fundamental Concepts and Future Prospects", *Information,* 9 (148). DOI:10.3390/info9060148

Van Aerschot, L. and Parviainen, J. (2020) "Robots responding to care needs? A multitasking care robot pursued for 25 years, available products offer simple entertainment and instrumental assistance", *Ethics and Information Technology,* (22), pp. 247–256. DOI: 10.1007/s10676-020-09536-0

Voss, A. 2021. Draft report on artificial intelligence in a Digital Age. European Parliament Special Committee on Artificial Intelligence in a Digital Age. Referred 26.1.2022.