robotics

MDPI

Article

Beyond the Metal Flesh: Understanding the Intersection between Bio- and AI Ethics for Robotics in Healthcare

Auxane Boch [1,*], Seamus Ryan [2], Alexander Kriebitz [3], Lameck Mbangula Amugongo [1] and Christoph Lütge [1,3]

[1] Institute for Ethics in AI, Technical University of Munich, 80333 Munich, Germany; lameckmbangula.amugongo@tum.de (L.M.A.); luetge@tum.de (C.L.)
[2] School of Computer Science and Statistics, Trinity College Dublin, D02PN40 Dublin, Ireland; ryans58@tcd.ie
[3] Peter Löscher Chair of Business Ethics, Technical University of Munich, 80333 Munich, Germany; a.kriebitz@tum.de
* Correspondence: auxane.boch@tum.de

Abstract: As we look towards the future of healthcare, integrating Care Robots (CRs) into health systems is a practical approach to address challenges such as an ageing population and caregiver shortages. However, ethical discussions about the impact of CRs on patients, caregivers, healthcare systems, and society are crucial. This normative research seeks to define an integrative and comprehensive ethical framework for CRs, encompassing a wide range of AI-related issues in healthcare. To build the framework, we combine principles of beneficence, non-maleficence, autonomy, justice, and explainability by integrating the AI4People framework for a Good AI Society and the traditional bioethics perspective. Using the integrated framework, we conduct an ethical assessment of CRs. Next, we identify three key ethical trade-offs and propose remediation strategies for the technology. Finally, we offer design recommendations for responsible development and usage of CRs. In conclusion, our research highlights the critical need for sector-specific ethical discussions in healthcare to fully grasp the potential implications of integrating AI technology.

Keywords: care robots; bioethics; AI ethics; healthcare

Citation: Boch, A.; Ryan, S.; Kriebitz, A.; Amugongo, L.M.; Lütge, C. Beyond the Metal Flesh: Understanding the Intersection between Bio- and AI Ethics for Robotics in Healthcare. *Robotics* **2023**, *12*, 110. https://doi.org/10.3390/robotics12040110

Academic Editor: Sylwia Łukasik

Received: 3 June 2023
Revised: 27 July 2023
Accepted: 29 July 2023
Published: 1 August 2023

1. Introduction

Social robots (SRs) are defined by Fox and Gambino [1] as human-made artificial intelligence (AI) technologies, presented in a digital or physical form, with some degree of human or animal-like attributes. According to a recent review study, there are five main areas where SR technology could potentially be adopted: companionship, healthcare, education, social definition, and social impact [2]. The authors detail that the expected qualities of SRs lie in their abilities to make decisions, have conversations, and react to social cues. Interestingly, when it comes to decision-making, SRs do not seem expected to be moral, but to make the most efficient decision regardless of social implications [3]. Currently, the research surrounding SRs as reported by Lambert et al. [2], focuses mainly on personalisation and social awareness of the tool with the aim to create adaptable social agents with abilities to recognise social cues and mimic emotions.

With the advancement of AI, built using techniques such as machine learning (ML), robots are increasingly being adopted in the healthcare sector [4]. While their main use is reported to be in surgery and rehabilitation units, other areas of deployment includes assistive care with dementia patients [5]. This type of SR application is called care robots (CRs). CRs exhibit conventional communication skills in their abilities to comprehend natural language, display emotions, as well as mimic conversation and understanding social cues [2]. In healthcare, CRs aim to monitor patients' well-being, assist with difficult tasks and proactively avert potential health deterioration [6]. In their work, Lambert et al. [2] highlight the main areas of applications to be in assisted living, monitoring of physical and

mental well-being, and enhancer for social learning experience for patients with autistic spectrum disorders [2]. While CRs refer to all types of assistive care robotics, in the context of study we use the term "care robots" to refer to a specific type of SR created to assist or, in certain cases, replace human caregivers when providing care to vulnerable populations [7]. This type of SR is also referred to as a "socially assistive robot" in the literature [8]. CRs are believed to have the potential to benefit the health care systems around the world by helping to fulfil an increasing care demand and improving the quality of care services provided [9]. While this type of technology comes with potential positive improvements for prospective patients and carers, it also sparks ethical concerns when considering the patients' best interests [10,11]. For instance, patients presenting with physical, cognitive, or emotional impairments due to their condition are considered as a vulnerable population and could see benefits from using CRs in their activities of daily living (e.g., grooming, feeding, moving). On the other hand, discussions have to be had when considering their rights, and understanding of the implications in regards to, e.g., privacy and autonomy when adopting such technology. The most common ethical considerations discussed in the literature to date relate to deception, independence, and informed consent as it relates to data governance and privacy, as well as autonomy [8,10,12–14]. Furthermore, current frameworks do not explicitly integrate both perspectives of AI and bioethics for practical implementation for CRs [15–17].

To facilitate the actionability of ethics for AI in healthcare, and especially for CRs, we propose to bridge this gap. This normative research thus aims to define an integrative and comprehensive ethical framework for CRs, to enable an ethical analysis that encompasses a wide range of issues relevant to AI in health care, and, based on the framework, to present guidelines for the development and use of CRs.

2. Methodology

Recognising the unique nature of the medical sector and its strong tradition of ethics centred on human values, the integration proposed aims to comprehensively address the ethical implications of technology in healthcare, to finally reach comprehensive guidelines for the precise case of CRs [8]. The normative approach taken is defined as "A theoretical, prescriptive approach (...) that has the aim of appraising or establishing the values and norms that best fit the overall needs and expectations of society" [18]. Moreover, we build on past work such as Van de Ven's [19], Edgett's [20], and others [21–24] to build ethical frameworks and recommendations based on inductive conceptual discussions supported by empirical arguments present in the literature.

2.1. Reconciling Both Perspectives

The study advocates for integrating the perspectives of bioethics and AI ethics to propose a sector-specific approach to ethical discussions in the healthcare domain. This reconciliation will be carried out by conceptually discussing the integration of both AI and bioethics principles from Jones' [25] framework, and the AI4People framework for a Good AI Society [26].

This choice of initial material is motivated by the AI4People framework's similarity in terminology to the field of bioethics [26], considering the bioethics field is the area of applied ethics most resembling digital ethics through its ecosystem approach of patients, agents, and environment. For bioethics itself, the normative framework we build on is the one presented by Jones, which integrates different views to reach a consensus on the principles for the ethics of the field [25]. The integration of both frameworks encompasses five principles: beneficence, non-maleficence, autonomy, justice, and explainability. These principles provide a foundation for ethical guidelines and are derived from both bioethical and AI ethical perspectives.

2.2. Ethical Assessment of CRs

We will then propose a "proof of concept" application of the framework for the use case of CRs through the discussion of an ethical assessment principle by principle. This proof of concept is necessary in the case of normative approach to ensure the relevance of our recommendations, and legitimise our approach such as suggested by Väyrynen [27]. As per Edgett's [20] methodology, we will support our arguments with the existing literature to provide a sufficient background from which we will then be able to draw recommendations, or guidelines, for an ethical development and deployment of CRs.

2.3. Trade-Offs Deliberation

Further, we will discuss three ethical deliberations ensuing from the use of CRs in the general population through the lance of our proposed integrated framework, and, once again, in presenting arguments proposed in the literature. The purpose is to justify the relevance and applicability of the proposed integrated framework by analysing the ethical implications and providing insights on ethical decision-making. The three proposed trade-offs discussions relate to patient centricity versus profit centricity, autonomy versus dependence, and data privacy versus efficiency.

2.4. Practical Recommendations for the Integration of Ethics in CRs Lifecycle

The final part of the paper focuses on delivering practical recommendations for the integration of ethics in the lifecycle of CRs. These recommendations are derived from our conceptual and practical discussions regarding CRs building on the literature presented in our arguments, the ethical assessment and analysis of three trade-offs. They aim to guide the ethical development, implementation, and use of CRs in healthcare settings.

3. Integrating Bio- and AI Ethics

In the next section, we take a brief look at what the two areas of literature say on the ethical implications of using technologies and care methods in health.

3.1. AI Ethics

The discussion surrounding the ethics of AI has been focusing on the ethical consequences of the technology, particularly with regards to normative principles such as human autonomy, human rights, non-discrimination, and privacy, as AI could have significant impacts on these concepts [28,29]. One key concern originates in the black box nature of some AI algorithms, rendering it difficult to understand how these algorithms make decisions, due to the complexity of the model, for example, deep learning-based models [30]. Furthermore, the literature has pointed out that AI solutions might amplify existent patterns of discrimination, owing to the standardising effect AI solutions unfold when put onto the market.

The Western approaches used to gauge the ethical effects of AI are based on considerations pertaining to virtue ethics, deontology, and utilitarianism [31]. The utilitarian approach considers an act as moral if, compared to possible alternatives, it provides a better outcome to a greater number of persons. It can thus be understood as consequential. On the other hand, the deontological theory judges actions over consequences. Thus, no matter how morally good or bad the implications of a behaviour, or of a decision, some choices are morally forbidden. Based on the insights of this discourse, scholars and policymakers have articulated ethical frameworks [26], applications of existing normative frames such as Human Rights [32], soft laws [33], sector-specific standards and legal approaches such as the "European Union 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" [34] to mitigate risks posed by the under-, over-, and misuse of the nascent technology.

Most of these frameworks have been addressing rather abstract features of AI. Owing to the plethora of AI use cases in human resources, finance, and health, and their different

normative implications, breaking down general moral implications of developing and deploying AI presents a key challenge within the field of AI ethics. Meanwhile, this research gap has been partly addressed by more recent scholarly contributions [35,36]. In particular, autonomous driving has been discussed intensively, particularly due to dilemmas occurring in unavoidable crash situations [37]. Furthermore, the literature has examined AI used in human resources, facial recognition technologies and finance [38,39]. However, one quintessential sector needs to be studied further as its integration of AI technology grows, namely AI in the healthcare sector and its ethical limitations. A recent paper proposed relational ethics to rethink and ground AI ethics in healthcare [40]. The paper further highlighted "the need for non-Western ethical approaches to be utilised in AI ethics more broadly".

3.2. Bioethics

The health sector is dominated by a specific moral tradition dating back to antiquity, such as depicted in the well-known Hippocratic oath (400 B.C.). The founding document of the discipline of bioethics is the oath of Hippocrates that calls on practitioners in health to "help patients" (beneficence), to "do no harm" (non maleficence), and practice medical confidentiality [41]. The notion of patient centricity seeks to secure that treatment methods and clinical practices are in the interest of the patient, especially in situations characterised by conflicts of interest between the best of the patient and the personal interests of the practitioner [42]. Patient autonomy constitutes another key value of medicine, implying that the will of the patients is decisive for the course of action adopted by practitioners [43]. Furthermore, the notion of justice suggests not discriminating between patients due to personal and individual characteristics [44,45]. These values has been encapsulated in the bioethics principles: beneficence, non-maleficence, autonomy, and justice [25]. These four principles are action-guiding for the clinical treatment of patients and aim to prevent breaches of ethical standards or acts of omission [46]. Nevertheless, there are still situations where the normative implications arising from this set of principles remain unclear. In dilemmatic situations, practitioners might have to decide between different pillars of bioethics, for example in triage situations or in situations when the will of the patient remains unclear.

3.3. The Quest for Reconciling Both Perspectives

In this paper, we reconcile both ethical approaches using CRs as a use case. The discourse surrounding SRs, including CRs, goes beyond the realm of traditional bioethics and is embedded within a broader academic, technological, and legal discourse concerning the advancement and implementation of artificial intelligence in our societies [47,48]. The use of CRs constitutes one of the applied cases in which AI and bioethics have to meet to consider fully important morals arising from the use of the technology. We thus argue the need for a framework that integrates and reconciles both perspectives, as has been conducted in other applications of AI in health [49].

The medical sector, which deals with personal and essential human issues such as survival, health, and well-being, has a long tradition of ethics that reflects its human-centric nature. We thus argue that to fully comprehend the ethical implications of integrating AI into the medical sector, it is necessary to respect and integrate the traditional vision of bioethics into technological ethics discussions. This approach enables a comprehensive understanding of the implications and ensures that the view fits the context. On the other hand, the AI ethics perspective looks into the societal and systemic impacts of AI implementations. This integration thus allows for a necessary multi-level view, from micro to macro, of the possible impacts of AI integration in the healthcare sector [8].

The AI4People framework for a good AI society [26] is an example of a framework that builds on bioethics terminology to address problematic AI ethics issues: beneficence, non-maleficence, justice, and autonomy. The framework has been developed based on four bioethics principles terms, while acknowledging the difference in interpretation as it

pertains to the bioethics perspective. We here propose a short introduction to each of the four common terminologies as defined by both perspectives, as well as a reconciliation of their interpretation to create an integrative framework [26,50].

3.3.1. Beneficence

The bioethics principle of beneficence encompasses the obligation to contribute to a person's welfare, entailing the responsibility to take actions that promote the well-being of others [25]. When designing interventions and provisions, the primary goal should be to directly benefit the patient. This approach emphasises actively engaging in activities that positively impact another individual's welfare, rather than simply refraining from causing harm. It necessitates proactive measures aimed at providing assistance and support.

On the other hand, the AI ethics perspective defines the principle as the need for creating AI prominently benefiting the well-being of people and the planet though positive economical impact, sustainability promotion, and safeguarding, as well as the empowerment of populations [26].

By acknowledging the common goal of promoting well-being and considering the specific contexts of bioethics and AI ethics, a reconciliation can be achieved. This entails designing interventions and provisions that actively contribute to the welfare of individuals while ensuring AI technologies have a positive impact on society, the environment, and human empowerment.

3.3.2. Non-Maleficence

Non-maleficence is defined in bioethics as the obligation not to inflict harm on other persons, referring to the responsibility of individuals to refrain from causing injury or negative consequences to others [25]. It emphasises the importance of avoiding or minimising harm to the best of one's ability. This principle serves as a foundational belief in the mission statements of medical professionals, often exemplified by the Hippocratic Oath.

On the other hand, the AI ethics perspective offers non-maleficence as a principle that highlights the importance of avoiding harm or negative consequences when developing and utilising AI technologies [26]. While the goal is to create AI systems that have positive impacts, it is crucial to be cautious about potential risks and misuse, such as issues related to personal privacy, security, and accountability.

The reconciliation of perspectives on non-maleficence involves a shared commitment to avoiding harm and negative consequences. In bioethics, it pertains to the responsibility of individuals, particularly medical professionals, to minimise harm and injury to others. In AI ethics, it extends to the development and use of AI technologies, emphasising the need to prevent harm and ensure positive impacts. By recognising this common commitment and applying the principle in both contexts, a reconciliation can be achieved, prioritising the minimisation of harm in healthcare and responsible AI development and deployment.

3.3.3. Autonomy

In bioethics, autonomy is defined as the respect for persons, entailing the recognition of the inherent worth and dignity of individuals, emphasising that humans should be treated as ends in themselves rather than mere instruments or tools [25]. This principle encompasses the fundamental right to autonomy, including the freedom to make decisions about one's own body and personal choices.

The AI ethics perspective understands autonomy as finding a balance between the decision-making power retained by individuals and that which is delegated to artificial agents [26].

The reconciliation of perspectives on autonomy involves recognising the common foundation of respecting individuals as ends in themselves. In bioethics, autonomy emphasises the inherent worth and dignity of humans, acknowledging their right to make decisions about their own bodies and personal choices. From the AI ethics perspective, autonomy involves striking a balance between individual decision-making and delegation

to artificial agents. By acknowledging the importance of individual agency and finding a harmonious equilibrium between human and machine decision-making, a reconciliation can be achieved, honouring both the principles of respect for persons and the dynamic relationship between humans and AI.

3.3.4. Justice

In bioethics, justice refers to the fair and equitable distribution of both health outcomes and health care services. It necessitates the careful consideration of prioritisation and rationing [25]. Allocating resources in a just manner does not have a single universal approach, as different systems employ multiple prioritisation strategies in combination to strive for a fair distribution.

In AI ethics, justice encompasses the fair and equitable distribution of AI's decision-making power and its consequences, considering the societal disparities in autonomy. It emphasises the promotion of equity, elimination of discrimination, shared benefit, shared prosperity, and equal access to AI's good doings [26]. Justice also addresses concerns related to biased data sets, defending solidarity in systems like social insurance and healthcare, and rectifying past wrongs through AI, such as eliminating unfair discrimination and promoting diversity.

The reconciliation of both definitions lies in the shared value of fair distribution, whether in health outcomes and healthcare services or AI's decision-making power. Both emphasise equitable distribution, considering societal disparities, avoiding discrimination, and employing various strategies for fairness. They strive for equity, equal access to benefits, and address issues of bias and solidarity. By recognising this common thread, adapting principles to specific contexts, and upholding fairness and equity, reconciliation is possible.

3.3.5. Explainability

In addition, the AI ethics framework expands to include the dimension of explainability, which addresses the black box character of AI solutions [26]. If chosen to be followed, the principle of explainability requires developers of AI solutions to, as an example, explain the general rationale and methodology behind an AI solution, the data used to train the model, and the data governance decisions and actions surrounding them. This principle is particularly relevant in the healthcare setting, where patients have a legitimate reason to know how AI has detected a specific illness and on what factors a health-relevant recommendation has been based [51].

3.3.6. The Reconciliation

In conclusion, reconciling the perspectives of bioethics and AI ethics requires recognising the common goals and values shared by both approaches. The integration of these perspectives becomes crucial when considering specific use cases such as CRs and the broader deployment of artificial intelligence in society. By combining the traditional vision of bioethics, which focuses on individual well-being and human-centric ethics, with the systemic and societal considerations of AI ethics, a comprehensive understanding of the ethical implications can be achieved. We argue this approach to be forceful proposition for future ethical consideration of AI in the medical sector.

4. Ethical Assessment

In the following, we develop a non-comprehensive ethical assessment for CRs based on the proposed integrative framework.

4.1. Beneficence

Beneficence implies that a technology or treatment method is conducive to conditions or situations perceived as desirable such as the promotion of well-being, economic gains or considerations relating to environmental sustainability [26]. However, when looking closer at the matter, bioethics and AI ethics offer different perspectives on the principle

of beneficence, or at least tend to emphasise different aspects. From the perspective of AI ethics, beneficence is often interpreted in the sense of wider gains for humankind and associated with frameworks such as the United Nations Sustainable Development Goals (UN SDGs), sustainability or human rights [32]. On the other hand, the bioethical interpretation of beneficence revolves around the patient in terms of the reduction of physical pain, higher life satisfaction, or the general improvement of the physical or mental condition of the patient [52,53]. Thus, the implementation of CRs should here be understood as needing to be beneficent for the patient ecosystem, and society as a whole (AI ethics perspective), but also with a strong enhancement of care offered to the individual patients (bioethics perspective).

A major concern as it relates to the future of care is the ageing of society. Indeed, as life expectancy increases and fertility rates fall worldwide, aged care services are under increasing strain. Globally, the number of older people 60 years and older is expected to double to 2.1 billion by 2050. In the same period, the number of people older than 80 years or older is expected to triple to 426 million [54]. This point is not just a concern for the individual patient, but also for the economic system and other stakeholders in the health sector such as caregivers. As it happens in parallel to the expectation of a workforce crisis linked to the global shortage, ageing, and burnout among physicians, and to the coming increasing demand for chronic care, the implementation of technological solutions is required to face the incoming crisis healthcare systems will soon face [55]. This situation can lead to a change in the way people are cared for and increase the demand for CRs [56]. Such solutions could be deployed to help with the high demand for care, ensuring that patients are timely attended to, reducing the pressure on healthcare facilities and services [57]. Moreover, without quick and implementable solutions to reduce the pressure on health systems and practitioners, the cost of healthcare will increase. This will require a paradigm shift in how care is delivered. Thus, the implementation of CRs fulfilling tasks in a similar way to, or supporting human carers by, saving scarce resources such as time and allowing for a better arrangement of the labour force in care, while increasing productivity, and quality of service for the well-being of all [4].

On the patient side, a successful and beneficial example of CRs can be found in the implementation of Paro [58] and NAO [59]. In this sense, patient centricity suggests streamlining the development and deployment of CRs to improve access to individualised care. Nevertheless, the precondition is here that CRs present an improvement of the status quo from the perspective of the individual patient.

The quintessential challenge of CRs lies therefore in the following. Owing to the immense pressure to adapt traditional caregiving to ageing societies, the deployment of CRs needs to prioritise patients over other stakeholders. This is suggested by the traditional interpretation of bioethics with its focus on patient centricity. Gains in other areas such as working conditions for caregivers are relevant too and are likely to enhance the treatment of patients. Nevertheless, improvements for other stakeholders or the realisation of other considerations such as economic or financial ones through the introduction of CRs must not lead to a deterioration of patient care. In other words, beneficence understood from the wider AI ethics perspective must not hollow out the established principle of patient centricity.

4.2. Non-Maleficence

Non-maleficence implies that a course of action or technology deployed should not create harm or risks for human beings [26]. Again, we can observe that bioethics and AI ethics create different implications here for the deployment of technology.

Bioethics specifies non-maleficence as the prevention of risks and harms to patients. This definition covers not only physical harm but also traumatic experiences that might be created by a specific treatment method or therapy. In the context of CRs, negative impacts or primarily are discussed in the context of psychological effects created by the loss of human interaction. Moreover, malfunctioning CRs might unfold adverse effects

on vulnerable groups. One example would be individuals with dementia that are more likely to be affected by deception [60]. While this argument has been discussed widely in traditional care, earlier studies have shown that the elderly in specific are often subject to cyber criminality [61], a phenomenon that could be exacerbated by the increased use of CRs.

The AI ethics discourse also highlights the relevance of the personal data of patients. AI4People has mainly defined non-maleficence in the AI context as data privacy concerns [26]. This is relevant especially when considering the relevance of health data, but also the close communication between CRs and patients. According to this argument, people might not be aware of the volume of data that robots are collecting, where that data is uploaded, or how it will be used. This lack of awareness would inhibit giving informed permission [62]. While laws like the European General Data Security Regulation (GDPR) [63] provide a few layers of protection in the European environment, these types of laws and regulations have their limitations. The novelty of social machines is in their ability to sense, process, and record the entirety of a patient's environment, as well as their augmentation of daily medical or non-medical routines [62,64]. When thinking of having a home CRs, the main goals could be to ensure that a patient takes their prescriptions on time, while constantly monitoring the patient's position in space to inform emergency assistance in case of a fall. The patient may comprehend what the robot performs, but it does not mean they are aware of the continual data collecting required for the robot to function well. Therefore, it is necessary to specify how data from CRs will be gathered and processed, how much of these data should be retained or uploaded to the cloud, how to obtain consent for doing so, and how to stop unauthorised external actors from obtaining personal specific information. The development of specific regulations must be taken into consideration as a result of new technology advancements that enable robots to acquire more data about their surroundings than ever before. In Europe, the GDPR [63] could be expanded, taking into account the characteristics of ML algorithms needed in social privacy [65]. However, the specific use case of CRs will remain problematic owing to the amount of data collected. Developers and deployers of AI solutions will therefore need to take care of the management and governance of data.

The AI ethics perspective uses a wider definition of non-maleficence, which also includes other stakeholder interests. As we consider possible harm to society, the replacement of carers by robots comes as a strong concern in the literature [66], bringing scepticism towards robots and their deployment. If we not prepared for a change in the ecosystem, consequences such as unemployment, and thus a possible decrease in the quality of life for the target population, are potential negative second and third order effects of CRs adoption. But a major point to consider is the political impact such deployment could have on societies and globalisation. It is thought that robots replacing the low-skilled workforce could strengthen populism and anti-immigration sentiments [67].

Thus, the "right way" to implement CRs in healthcare systems have to be considered on a global scale, but also on an individual scale to prevent foreseeable harm to individual and societies. Solutions might involve strong strategies in the re-orientation of other tasks for persons seeing their work done by robots, but also the education and monitoring of populations on cybersecurity, data collection, and artificial intelligence, as well as the requirement for stronger regulations around data governance and reliability of CRs abilities and cybersecurity architecture.

4.3. Autonomy

The third principle of autonomy calls for the consideration of individuals' right to make enlightened decisions [26]. In bioethics, this concept focuses on the right to make informed decisions for one's medical care. When considering this definition alone, the black box issues for complex AI systems seem to be a strong case against the use of AI in healthcare as a whole and will be discussed in the explainability part of this paper. In

AI ethics, autonomy focuses on the right to decide what decision power to give to the AI system.

In the case of vulnerable populations, it may happen that the patient is not in a condition to decide on their own care settings, for example, if no physicians are available due to a shortage of time and practitioners and the only time-sensitive solution is the use of CR to support the needed care [55]. In this context, the willingness of the patient's care ecosystem to delegate decision-making tasks to AI systems for the purpose of efficiency could go against the principle of autonomy. A balance needs to be achieved to protect the individual choice of the patient to delegate care tasks concerning their own health and well-being to AI systems. The human should always have the possibility of reversing, not implementing, stopping, and starting all decisions made by the CR.

A long-term consideration concerns the potential creation of functional emotional dependency on CRs over time for the patients [13]. Considering that CRs here are a specific type of SRs, their main features and goals are to build, develop, and maintain relationships over time with people. Through physical behaviours or spoken communication, an SR can express social and emotional cues. These queues may cause attachment to develop towards such machines [68,69]. Furthermore, Boch et al. [70] argue that, the more a robot is perceived by the users as autonomous and emotional, the more their attachment towards it seems to grow [71–73]. Interestingly, attachment towards CRs does not always happen. If it is present at a high level though, it stays consistant throughout time [74]. The ensuing feelings developing towards the CRs might lead to the creation of a relationship perceived as structured, real, and evolving throughout time between the user and the CRs. This relationship is a unidirectional one, defined in other areas of science as para-social relationships [75,76]. Such relationships, experienced as genuine, might raise opportunities for better care, by, for example, enhancing users' will to listen to the robots care instructions [77,78] and possible risks regarding emotional distress if CRs are taken away [79]. It is to be considered that in this context, patients with mental health and ability impairments, as well as children, are to be seen as more vulnerable than other populations to this possibility, but all populations can develop this type of relationship [58,59]. The creation of a para-social relationship can also be linked to the notion of emotional trust [70,80]. This irrational sense of trust poses a risk when issues of responsibility arise [62]. Indeed, the consequences of such a relationship could be simple enjoyment of CRs's company and go as far as an emotional dependency. In this case, the autonomy of the patient could be at risk, as its decision-making process could be impacted by its affection for the robot, and thus create a situation in which the human loses part of its freedom of choice, especially as it relates to the robot itself. For example, as a consequence of attachment, patients could refuse to part from the CR, regardless of the beneficence it brings to their health, or the threat it can be to their privacy.

Following up on this first worry, the European Parliament expressed its concern in 2017 that robots deployed in care settings may dehumanise the action of providing care by limiting human–human interaction [81]. In order to determine whether a care service is successful and can be categorised as "preserving the meaning", key social characteristics that relate to it, including friendliness, should be thoroughly defined [82]. On the other hand, the use of CRs for specific tasks could support carers in reducing their workload and allowing them to spend more time on the human side of their relationship with patients, counterbalancing this concern and reducing the stress put on such workers [83]. As a counterpoint to this argument, the implementation of CRs could support caregivers in taking over some of the laborious work, thus giving them back some autonomy; a frequent issue when carers attempt to prioritise their daily activities [84].

Finally, the use of CRs in everyday life can help compensate for functional losses and promote everyday skills, supporting or restoring the independence—in this context understood as the ability to achieve tasks without the need for help from another human—of individuals [85]. If this point is short, it is a major one as it pertains to the daily autonomy of a human. The implications on an individual's life can thus be incomparable in allowing

them to, e.g., live at home for a longer time and avoid the crowding of nursing homes for an elderly population requiring only support and monitoring.

Thus, CRs bring possible threats to patients' autonomy, and concerns can be raised in regard to their use on an everyday basis, but their implementation could be highly beneficial for individuals' independence and bring strong support to healthcare workers. In detail, trade-offs need to be considered on a case-by-case strategy to bring more beneficence to patients and systems.

4.4. Justice

Justice, as a bioethics principle, relates widely to resource allocation, including the availability of novel and experimental therapies, the highest possible treatment quality, and ordinary healthcare [44]. In the case of CRs, it could translate into their availability and accessibility to the global population, as well as their adptability to different cultures and needs. Adding layers to this definition, AI ethics calls for the avoidance of discrimination by AI systems based on individual and personal characteristics (e.g., gender, age, ethnicity), and the creation of shared benefits on a global and individual level, while preventing the creation of new harm, or the enhancement of existing ones [26].

Starting with the bioethics perspective, it can be stated that equal access to CRs technology is unlikely to be met on a worldwide scale as SRs are entering different societies at various rates. As presented in Boch et al. [70], the fastest growing markets for SRs in the coming years (between 2021 and 2026) will be led by economically developed countries; the USA and countries in Oceania and East Asia project the highest, with Europe and Canada projecting a medium growth rate [86]. This leaves a big part of the world out of the equation; South America, the Middle East, and Africa, are all predicted to have a low rate of growth as it relates to robotic technologies and the costs associated with the development and thus use of the technology [86]. Significant disparities and inequalities can be observed in terms of regional inclusion and participation in discussions surrounding the development, design, and implementation of CRs [70]. These considerations gain importance when acknowledging the existing divide between the Global North and Global South, which already influences the development of numerous AI-driven technologies available in today's market, as discussed in the literature but also brought to light in the 2023 World Economic Forum annual meeting [87,88]. A specific example of such bias impact on accessibility is facial recognition features. Algorithms allowing for such technology allow, in CRs, for a higher level of human–machine interactions. In the case of vulnerable minorities, already facing structural bias in society, the error rate in those systems disproportionately affects them [89].

Growing this argument into the AI ethics perspective of justice, the risk of entrenching existing inequalities due to the lack of geographical and other background diversity in the creation and development team could participate in enhancing bias towards specific populations and thus generate additional issues in the case of deployment in countries or regions that are initially non-targeted [90]. Moreover, when looking at the specific implementation of AI systems in the healthcare sector during the pandemic, screaming examples of the accentuation of discrimination against specific groups have been noted even in geographical target populations [91,92]. As it relates to CRs, those concerns are highly relevant; if a CR makes health related decisions for patients towards whom algorithmic bias plays, the well-being and even the life of the patient could be at risk.

Thus, the accessibility to the beneficence of CRs technology on a global scale is not a given, and neither is its equality in performance with every type of target population. When narrowing down the scope to the individual level, the reduction of algorithmic bias against interpersonal characteristics of vulnerable populations has to be a paramount point of concern for the developers, and care ecosystem of patients to avoid harm at all costs.

4.5. Explainability

Explainability is the one principle pertaining to only AI ethics, enabling other principles to co-exist, and allowing for accountability [26]. It is a complex but foundational requirement in all AI-enabled healthcare technology [93].

The challenge of explainability exists at two separate but interconnected levels when discussing AI-enabled robotics. At one level there exists technical explainability, the ability to explain and understand the mathematical weighting and prioritisation that the underpinning advanced statistical models used to create the AI [94]. As an example, the visual system of a SR may take in and process a 360-degree field around it but may only make decisions using a small subset of this visual information. In an environment where decisions may need to be audited or reviewed, then what exact information is used in the decision-making process needs to be clearly explainable. The feasibility of this type of explainability depends on the type of statistical model and the technical decisions made during the development of AI-enabled robotics.

This challenge is further confounded when considering the types of explainability available. Some models are "White-box", where the exact approach is identifiable and clear and all of the reasoning involved can be reviewed [95]. In situations like this, an auditor would be able to give a deterministic and affirmative explanation as to why the robot did what it did. This sits in contrast to "Black-box" models where, if the decision can be explained at all, it is intuited via a set of external analysis tools [96]. In this situation, an auditor may only be able to give a probable, non-deterministic answer as to why a decision was made.

Which level of explainability required in a healthcare environment has not yet been regulated explicitly but adjacent laws such as GDPR [63] expect data processing to be handled in a transparent and explainable manner [97]. As CRs begin to see adoption there, the expected level of explainability will need to be planned for in advance and needs to be part of the requirements at the earliest stage of conception.

At the next level, there exists socio-technical explainability, the need to understand the context in which the systems are used and on which levels they affect our everyday life [98,99].

Going back to the bioethics perspective on autonomy, patients (when cognitively and physically able) have the right to make their own decisions regarding their medical care with all the information presented to them. In the current state of AI, the use of SR in healthcare cannot ensure full transparency in the decision making process of complex systems such as CRs. A strong need for "White Box" implementation and an understandable explanation for the patients' abilities should be at the heart of CR development to allow for full autonomy of the patient in understanding all technical aspects pertaining to using such tools in their medical routine.

On the other hand, users should be informed in a transparent way about the benefits and harms that might emerge from the interactions with CRs [70]. When considering accountability as a necessary point of ethical technology, explainability of the tool and processes around it are at the centre [100]. Providing a comprehensive view of potential adverse outcomes, including physical and psychological harm that may arise from the utilisation and engagement with CRs, is of the utmost importance. This necessitates a clear delineation of the circumstances in which problems might occur. For instance, the consequences of establishing a para-social relationship between users and CRs are currently uncertain. This highlights the importance of conducting further empirical research and establishing accountability mechanisms before the widespread adoption of social robots in personal care contexts. Some argue that genuine friendships can develop between humans and robots, and thus CRs [101], while others point out the issue of deception inherent in such a connection [34,102]. Thus, the potential for para-social relationships to bloom between the patient and the CRs is real and comes with both positive and negative possible consequences. For instance, users may unknowingly trust and confide in CRs based on their evolving relationship, thereby sharing more personal information and data [103].

This highlights the need to provide understandable explanations regarding the actual functioning and data usage of CRs, and enabling users to understand the implications of their exchange with CRs regardless of their perceived relationship.

In addition, when considering the context of elderly care, trade-offs may arise as the efficacy or beneficence of CRs may affect users' autonomy. Striking a balance between a more autonomous social robot and cultural considerations is thus crucial [104]. Indeed, users' cultural environments might moderate their comfort levels in delegating tasks to CRs. Therefore, different societal contexts may require specific sets of values to be embedded in their robotic systems, and the set of values needs to be presented clearly to the patient prior for their understanding and approval of the use of the CRs.

Thus, from a technical and socio-technical perspective, explainability is of paramount importance for CRs implementation. Their use for healthcare purposes in addition to their social purpose creates a particularly challenging context when it comes to the need for transparency and understandability.

5. Ethical Trade-Offs Deliberations

In this section, we will now discuss the three main ethical trade-offs arising from the use of CRs in healthcare through the scope of our integrative framework. We will first consider patient centricity as an ethics of care requirement, versus profit centricity, which is a business requirement. We will then discuss patients' autonomy versus dependency, and how to balance the risks. Finally, the question of data privacy rights versus efficiency of the robot will be addressed.

5.1. Patient Centricity vs. Profit Centricity

The use of CRs in healthcare presents a complex ethical trade-off between patient centricity and profit centricity. Patient centricity stems primarily from the bioethics perspective of beneficence, while profit centricity can be argued as belonging to the economical "do good" vision of the principle through the AI ethics lens.

Patient centricity is considered to be more than just an ethical requirement but a social responsibility in the healthcare sector globally. Russo's [105] work on the topic partly builds on Werhane's [106] theory emphasising the importance of maximising the treatment and well-being of designated populations while also respecting the rules of the game, such as operating within a free market. However, informative asymmetries may prevent the optimal situation in which all patients involved are satisfied, highlighting the need for regulatory agencies to put a frame on the market. Borgonovi's theory [107], on the other hand, focuses on the healthcare organisation's ability to carry out its function in the best way possible, which requires efficiency, shared definition of health strategies, and environmental protection. Finally, Emanuel and Emanuel [23] propose a theory based on 'New Contractualism', which emphasises the equality of fundamental rights of all parties involved and is based on a concept of justification rooted in a social contract between stakeholders. This theory attempts to balance the economic interests of shareholders, the social impact of meeting patient needs, and the expectations of healthcare organisations, while ensuring accountability. Russo [105] thus identifies three main dimensions that are important for the social responsibility of healthcare organisations: maximising the treatment and well-being of designated populations, carrying out the healthcare organisation's function in the best way possible, and providing a justification and being held responsible for actions by another party. Finding the trade-off between all parties to obtain to a balance between profit and patient centricity is thus an ongoing discussion in the healthcare sector at the global level. Interestingly, Collins [108] found that healthcare managers tend to prioritise patient care over profit maximisation. However, the pressure healthcare managers face to produce higher results with fewer resources may inadvertently test their moral fortitude and social consciousness. Future healthcare managers may strongly focus on patient care but may still require guidance to ensure ethical and socially responsible decision-making. Thus, providing the best care for patients is at the centre of current behaviours and objectives

from a personal and organisational social responsibility perspective in healthcare, with a layer of accountability that healthcare providers have to answer to, while considering the profit of an organisation. These principles align with the principle of patient centricity requirements while integrating the profit an organisation needs to ensure survival. Those issues can be translated to the product level, such as for CRs.

CRs can promote patient centricity by providing personalised care and support that meets the unique needs and preferences of each individual, while improving the effectiveness of care [109]. Their use enhances the patient experience, improves health outcomes, and empowers patients to take an active role in their care [110]. However, the overuse of such technologies could lead to the deficiency in individualised care by humans, which might in turn reduce the quality of care received [110]. From a personal perspective, CRs can also prioritise the economic interests of the patient and of the healthcare system through the reduction of costs and increasing efficiency of care. Indeed, providing personalised care that meets the unique needs of each individual can reduce the likelihood of adverse events, hospital readmissions, and unnecessary procedures, all of which can result in significant economic costs [111]. From a group perspective, CRs can reduce the cost of care for entire systems by increasing efficiency and reducing the need for human labour. For example, robot-based systems for telemedicine have economic value and can potentially provide proper and timely medical care to patients in medically underdeveloped regions [112].

However, it is argued that the deployment of robotics technology in the field of care is increasingly focused on standardisation and selection into economic and marketable care measures, which can unintentionally produce CRs that rely less on traditional, care-intrinsic knowledge [111]. For instance, referred to as the "Silver Economy" by the European Commission [113], increasing numbers of elderly people will create a new "silver" market of consumers, and are being targeted as a new consumer group for assistive technologies.

In summary, the use of CRs in healthcare presents a complex ethical trade-off between patient centricity and profit orientation. Healthcare organisations have a social responsibility to maximise the treatment and well-being of designated populations, carry out their functions in the best way possible, and be held accountable for their actions by stakeholders. When looking at the specific technology of CRs, it is recognisable that they can promote patient centricity by providing personalised care and support, improving the patient experience, and reducing economic costs. However, the overuse of CRs and their focus on standardisation and marketability could result in the deficiency of care-intrinsic knowledge and compromise the safety and well-being of individuals. Hence, the deployment of CRs must be guided by ethical and socially responsible decision-making to balance patient needs, economic interests, and social expectations.

5.2. Autonomy vs. Dependence

Here, we deep dive into the trade-off CRs present between promoting autonomy and dependence to the tool. This trade-off pertains to the Autonomy principle as seen by both ethics as it relates to the autonomy of an individual, and their use of a technology.

We here define autonomy as the independent living ability of an individual [114], and the capabilities related to everyday life regardless of the physical impairment or ageing faced by a patient, hoping for the technology to reduce avoidable hospitalisation or institutionalised care [111]. Autonomy is moreover viewed as the individual's right to make informed decisions, based on the cultural ideal that agents are independent, rational, and self-interested. In order to be autonomous, people must be able to freely make decisions that are not influenced by coercion and that reflect their own thought processes [115]. Finally, we argue that the care ethics perspective of keeping the patient at the centre of all concerns also enhances the need for available and adequate care-giving services to fulfil patients' everyday care autonomy.

Taking into account the definition of autonomy that we have presented, the implementation of CRs may potentially enhance patients' autonomy by satisfying their requirements in three key areas: (1) the continuous demand for care, which surpasses the capacity of

human caregivers; (2) the prevalence of patient abuse in the care of others, leading to diminished autonomy and dignity; and (3) the inadequacy of current practices to meet the expected level of care, resulting in patients' compromised autonomy, dignity, and health [116]. Moreover, CRs have the potential to promote autonomy by enabling individuals to maintain their independence, agency, and control over their healthcare decisions, also due to the availability of options in regards to services, which can lead to improved quality of life and better health outcomes [117].

On the other hand, the efficacy and resilience of CRs in fulfilling its intended duties are likely to evoke heightened levels of trust from patients [118]. Trust can thus support the use of such technology, but in some cases, result in overuse of and over reliance on the CRs, which might end up creating dependency. In this case, patients would tend towards dependence on the robot rather than independence thanks to the robot, leading to a loss of agency and possible social isolation.

An instance in which individuals may experience a reduction in their ability to self-govern may arise when they follow the recommendations made by robots. Studies have indicated that people tend to be more compliant when instructed by robots, as compared to when given instructions by other humans [119]. While this feature may be advantageous in aiding patients with autism or those undergoing challenging behavioural modifications, there is a valid concern that people may be unduly influenced or coerced into carrying out actions that they would not otherwise, due to the novelty of the technology or the absence of companions to discuss alternate courses of action with [12]. In this case, social isolation is thus a risk and a factor. Furthermore, the literature suggests that there is a concern regarding the adoption of CRs for elderly patients, as it may exacerbate their sense of loneliness. It is vital for healthcare providers to acknowledge the dignity and independence of older adults, as well as their right to participate in social and cultural activities when introducing new technologies into their care [120].

Thus, a trade-off needs to be found, and might have to be case by case. In regards to care ethics, we understand that agency and autonomy of the patients go hand in hand when considering decisions of care services they want and agree to access.

Thus, when considering the use of CRs, the patient's consent—or that of their family if the patient is not able to give enlightened consent—is necessary. The use of CRs should also be stoppable and retractable without any conflict at any point [121]. Moreover, this entails that detailed information regarding the use of CRs are to be given to the patient to allow for a full understanding of the limitations and risks of the technology before making their decisions. Interestingly, in some cases, patients might be open to limitations in their own autonomy if it is necessary to ensure their safety in the long term [122]. Thus, prioritising the robot's role in promoting safety for patients seems the better road to reach an agreement and keeping the patients' interests at the centre. Therefore, the terms of robot use should be discussed and agreed upon in advance between all the involved stakeholders [122]. Finally, the most important point is to use CRs in a transparent manner to allow for the patient's autonomy to be intact. To ensure that patient autonomy is respected, guidance should be developed on how to implement applications, including when and how consent should be obtained, and how to handle matters related to vulnerability, manipulation, coercion, and privacy. Building on the Clinical Trial Regulations on informed consent proposed by the European Patient Forum [123], we emphasise here the need for detailed and clear discussions with the patient regarding their right to privacy, an adapted presentation of the limitations and risks associated with the proposed CRs solution with confirmation of understanding on the patient's part through questions or tests, but also the integration of a dynamic consent process through which the patient can access the information they require ongoingly regarding the tool. The capacity of the patient to receive and understand such information is moderated by but is not limited to their cognitive abilities, their spoken language, and other diversity characteristics to take into consideration.

In summary, the use of CRs presents a trade-off between promoting autonomy and dependence on the tool. CRs have the potential to enhance patient autonomy by providing

continuous care, preventing abuse, and offering options for improved quality of life and better health outcomes. However, over-reliance on CRs may result in dependency and loss of agency, leading to social isolation and a reduction in the ability to self-govern. To ensure patient autonomy, stakeholders need to find a balance that prioritises the patient's safety while respecting their right to make informed decisions. In other words, the use of CRs should be transparent, involve patient consent, and be accompanied by guidelines that address issues related to specific situations. Ultimately, the goal should be to keep the patient at the centre of all concerns and to provide adequate and available caregiving services that fulfil patients' everyday care autonomy.

5.3. Data Privacy vs. Efficiency

One of the major ethical concerns that arises with the use of CRs is the trade-off between data privacy and the efficiency of CRs in their tasks. This trade-off belongs to the principle of non-maleficence as perceived by both perspectives when it comes to ensuring the privacy of a user, but also touches on the principle of beneficence when it comes to "do good" and efficiency. Efficiency in this context refers to maximising CRs' accuracy and effectiveness in delivering care reliably. The collection and analysis of large amounts of patient data are crucial for training machine learning algorithms that can adapt to the specific needs of individual patients. Moreover, to interact naturally with humans, social robots rely on sophisticated algorithms and collect large amounts of data both about users and their environment [124]. For example, current generations of SR are equipped with sensors such as cameras, and GPS sensors [125]. Additionally, CRs have the ability to establish internet connections and transmit data collected by their integrated cameras and microphones to remote servers [121]. The collection of personal data enabled by the variety of sensors on CRs is necessary for their proper functioning. However, the use of sensors such as cameras and microphones can infringe on the privacy of patients. Similarly, the transmission of personal data via the internet risk potential exposure of personal data through hacking and cyber attack. Here, data privacy should be discussed to understand the acceptable balance between data privacy and the efficiency of the robot.

Data privacy is a fundamental right guaranteed by Article 8 of the European Convention on Human Rights and the Universal Declaration of Human Rights [126]. Privacy concerns have been raised by users, patients, and caregivers when considering the use of CRs [127,128]. Lutz et al. [124] define privacy in social robots, which is thus applicable to CRs as defined in this paper, in four categories: (1) informational privacy: quantity and confidentiality of data, potential security breaches, the ability for third-party access, connectivity to cloud services, the opaqueness of data collection processes, and a lack of understanding on the part of users; (2) psychological privacy: psychological dependence, diminished self-reflection and human autonomy, chilling effects as a result of feeling surveilled, and specific concerns for vulnerable user groups such as children; (3) social privacy: the social connection established between the robot and the user, accompanied by feelings of fondness and trust, which may result in the disclosure of confidential information; (4) physical privacy: capacity to access areas that are private, or that users may not be able to access, and the discomfort of being too close.

Thus, social robots, and thus CRs, present unique privacy risks aligned with their collecting of sensitive information. Moreover, the possibility for users to create emotional bonds with their robots and interact with them in more open and intimate ways leads to increased privacy risks. Finally, owners tend to forget that data collection is ongoing while interacting with their robots [121]. Informed consent is therefore crucial for users as they are at risk of not being (made) aware of the variety of data collected while using CRs.

To address privacy risks, Lutz et al. [124] recommend the use of privacy by design and privacy as contextual integrity frameworks. Interestingly, a small amount of training data seems to be necessary for the high accuracy of the system in similar cases [129], and human-generated feedback could facilitate personalisation [130]. Moreover, the reuse of existing data sets seems to be a valid solution for re-collection of data in specific situations

to learn end-to-end skills [131], meaning that the collection of large amounts of data from the patient might not be necessary for every situation. The concept of data minimisation is here relevant to implement, understanding which type of data are needed for efficiency, and making sure to focus collection on those information solely. Signalling data collection could be a viable alternative, building on the concept of explainability [124]. In this case, the users might also receive an explanation and thus understand more clearly which types of data is being collected, and be in control of what they are willing to share or not, and with whom. In the context of CRs, the questions also relate to sharing medical information with family and care staff, in addition to providers and third parties. Having an honest map of which data will be used by the system to provide better care is also of paramount importance as the trade-off with efficiency is clear. The patients should thus know what will or will not help their tool better its services. In addition, ensuring the understanding of all stakeholders involved in the CRs ecosystem (e.g., patients and their direct family ecosystem, doctors and nurses) regarding the robot's data collection and governance is crucial. In this context, finding the proper balance with transparency also means understanding the target population and adapting the discourse to the patient's abilities. Only by doing so can the full consent of using the tool given.

Another potential solution to privacy is federated learning, a privacy-preserving ML technique that involves sharing the model instead of the data [132]. This implies that a base model will be sent to CRs, learn from data, and make inferences locally without sending user data to a central server. While this approach addresses concerns about data sharing, it introduces new technical limitations related to the accuracy, transparency, and security of models developed using federated learning. Therefore, caution should be exercised when considering the adoption of solutions from domains where federated learning is not a standard practice into the healthcare field, as such solutions may not be feasible within a federated environment.

In conclusion, the use of CRs presents ethical concerns related to data privacy and the efficiency of CRs in delivering reliable care. While the collection and analysis of large amounts of patient data are crucial for training machine learning algorithms and improving CRs' accuracy and effectiveness, it also raises concerns about data privacy. The collection of (personal) data allowed by the variety of sensors present on CRs is quite important and necessary for their proper functioning. However, issues regarding data privacy need to be discussed to understand the acceptable balance between data privacy and the efficiency of the robot. To address privacy risks, CRs must be designed and implemented in a way that complies with data protection regulations such as the General Data Protection Regulation (GDPR) to protect patient privacy. Technological solutions such as anonymisation, encryption, and design that signals data collection are also recommended [124]. In addition, finding the proper balance in regards to transparency and explainability means understanding the target population and adapting the discourse in regards to data governance to the patient's abilities [133].

6. Discussion

In this paper, we reconcile the theoretical frameworks of bioethics and AI ethics to create an integrated ethical framework to guide the design, deployment and use of CRs. As a proof of concept, we explored ethical trade-offs accounting for multiple ethical perspectives. Furthermore, we provide practical resolution and recommendations, including finding the adequate discourse for each patient and stakeholders to understand the technology they are using and what it entails, the establishment of regulations and standards, and prioritising the patients' interests.

In the past decade, there has been significant growth in research on the ethics of robotics, especially in the field of healthcare. For instance, Stahl and Coeckelbergh [134] argued that in addition to ethical analysis, traditional technology assessment, and philosophical speculation, it is crucial to incorporate forms of reflection, dialogue, and experimentation that closely align with innovation practices and real-world contexts of the use of

robotics in healthcare. However, there are few studies that are focused on ethics of CRs. Bradwell et al. [135] surveyed 64 adults after they have had interactions with companion robots. Their study, demonstrated disparities between ethical concerns discussed in philosophical literature and the considerations that influence the decision-making process of purchasing a companion robot. These discrepancies, observed between philosophers and end-users involved in the care of older individuals, as well as differences in the methods used to gather information, highlight the need for additional empirical research and discussion. Another study outlines the ethical challenges associated with robotic care assistants and proposes potential strategies for addressing them through their design and use [136]. Finally, a series of frameworks and recommendations have been proposed [15–17]. Those bring a lot to the conversation, but do not comprehensively review the AI ethics and bioethics approach. Rather, they give pointers to methodologies on how to integrate ethics in general in AI in healthcare, and who should be responsible for it. Moreover, existing frameworks are not tailored for the use case of CRs. Our approach is to bridge the existing gap by integrating both AI and bioethics in an ethical framework for CRs, ensuing in the proposition of practical design recommendations.

6.1. Practical Design Constraints for CRs

In terms of implementation, developers of CRs should ensure that the system adheres to AI ethical principles such as beneficence, non-maleficence, autonomy, justice, and explainability as understood in the integration of both AI and bioethics perspectives. This can be achieved by prioritising the well-being and safety of users at the same time as averting potential harm, ensuring users are not coerced into decisions, promoting fairness and equity in healthcare, being transparent and establishing the means to hold developers accountable for the decision processes of SR. It is necessary to collaborate with healthcare providers, users, and AI ethicists to develop a well-rounded robot that prioritises patients' autonomy and well-being.

To ensure the involvement of relevant parties in the early stage of CR development, multiple methodologies can be implemented. First, focus groups can provide a platform for different groups to come together and share perspectives, experiences, and expectations regarding the given technology [137]. This methodology, nevertheless, might come with its own set of challenges when facing vulnerable populations, which might be the case in the context of CRs for autism, or CRs for dementia patients [138]. A second possible methodology to involve different target populations opinions and ideas is to deploy user surveys, interviews, and observation [139]. This set of data collection allows a more personalised understanding of the target users to be obtained. Finally, public consultations or meetings can facilitate a broader engagement with the community at large. This methodology provides open dialogue, and a platform for knowledge sharing, as well as the opportunity to understand a community's values [140]. By employing these approaches, we can establish a collaborative and inclusive environment where stakeholders are actively involved in shaping the development of CRs in healthcare. Their input and insights can guide decision-making, address ethical considerations, and ensure the development of CRs that align with societal needs and values.

From a methodological point of view, developers should adopt ethics by the design approach, which ensures that ethical principles are not just an afterthought but human values are considered from the offset and maintained throughout the AI pipeline. Practically, developers of SR should consider the following:

1. Beneficence: CRs should be designed to promote the well-being and safety of the users. This could involve incorporating features that encourage healthy behaviour or providing personalised medical advice based on the user's health data. It is also important to ensure that the robot does not inadvertently cause harm, such as by providing incorrect medical advice or failing to respond appropriately in an emergency.

2. Non-Maleficence: CRs should be designed to avoid causing harm to the users. This requires careful consideration of the potential risks and benefits of the robot's actions.

For example, if the robot is providing medical advice, it should be based on accurate and up-to-date information and should be tailored to the individual needs of the user. The robot should also be programmed to recognise and respond appropriately to potentially harmful situations, such as detecting physical or mental signs of distress in the user, but also be able to understand a negative feedback from the patient. In other words, CRs should be able to understand the limits of the patient by taking explicit feedback, spoken or perceived.

3. Autonomy: CRs should be designed to respect the autonomy of the users. This means that the robot should not coerce or manipulate the users into making decisions that they do not want to make. Instead, the robot should provide information and support that enables the user to make informed decisions about their health and well-being.

4. Justice: CRs should be designed to promote fairness and equity in healthcare. This could involve incorporating features that address healthcare disparities or providing access to healthcare resources to under-served communities. It is also important to ensure that the robot does not perpetuate or reinforce biases or discrimination in healthcare.

5. Explainability: CRs should be designed to be transparent and accountable in their decision-making processes. This requires that the robot's algorithms and data sources are open and explainable to the users and healthcare providers. The robot should also be programmed to provide clear and understandable explanations for its actions and recommendations.

As acknowledged at the start of this section, there will be specific ethical requirements within the sub-field. This will be domain-, culture-, and potentially user-specific. However, the discussed practical design constraints will need to be conceptual design goals that must be considered as part of the foundational stage for robotics in healthcare. Further ethical considerations will need to build on or augment these constraints.

6.2. Limitations and Outlook

Our study is not without challenges. First, we take a normative approach to address a problem that could be understood as technical. We recognise that our contribution, if not highly technical, is still of importance in the decision making of using, designing, and implementing technologies such as CRs.

Second, this study was not conducted systematically.

Finally, our proposition rests on one case study and thus might not be generalised as is. To overcome this issue, we would offer a few pointers for the adaptation of our framework to different healthcare use cases and contexts. First, cultural sensitivity, through the understanding of cultural diversity and existing normative frameworks is paramount [40]. Second, we would encourage strong stakeholder engagement from diverse backgrounds in the definition and further deliberations attached to the framework application and its use case [141]. Thirdly, an iterative approach to the adaptation of the framework is recommended. This entails ongoing evaluation and refinement of the guidelines based on feedback, empirical evidence, and real-wold implementation. This will allow for its ongoing improvement and adaptation to the specific context [142]. It is important to note that while we provide these general strategies, the precise method of adaptation may vary depending on the specific healthcare context and cultural considerations. Therefore, further research and collaboration with stakeholders will be essential to develop and refine the adaptation process in each unique setting.

Nevertheless, we believe this paper to be a strong contribution to the conversation, and a necessary one to build towards a sector-specific understanding of ethics for the integration of AI systems. We thus believe that future research should first reproduce our integrative approach for different use cases of healthcare, integrating both bio- and AI ethics to evaluate different AI applications and their frame of use.

Second, we would encourage the creation of quantifiable characteristics to evaluate the adherence to ethical principles for the integration of AI in healthcare, building on our proposed integration of perspectives.

We also encourage work towards the standardisation of social AI systems, building on psychology and other human sciences knowledge, depending on the specific context and sector of use. The current frame regarding regulation and standardisation for technology is rapidly evolving, with, amongst other regulatory efforts, the upcoming "Regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (AI Act) and amending certain union legislative act" [123]. This proposed regulation calls for a fundamental rights assessment, and brings with it the possibility of the creation of an EU AI Office to provide guidance and coordination in the implementation of the Act from a legal perspective [143]. On the other hand, regulation is not the only way to go when it comes to ethics. Certifications and standards developed both by governmental and non-governmental organisations can contribute to the accountability mechanisms by introducing specific technical and tangible criteria to reach ethical standards with design technologies. The ongoing work of ISO in developing standards is a notable example, or the existing work of IEEE regarding standardisation on this topic [33,144]. Finally, internal auditing and guidelines can also establish a proactive approach to ensure ethical and legal compliance. Google, for instance, has developed an end-to-end internal auditing framework that offers guidance for responsible implementation [145]. By incorporating these strategies, we can establish robust mechanisms to monitor and ensure compliance with regulations and standards for the ethical use of CRs in healthcare, if adequate standards and regulations are adopted and developed for this specific context. These approaches can promote transparency, accountability, and responsible innovation while safeguarding patient well-being and societal trust. We believe in the need to clearly understand the context (e.g., culture and frame of use) to build and use appropriate technologies.

7. Conclusions

The implementation of CRs in the healthcare sector is seen as a plausible solution to address the impending demographic and workforce challenges. The adoption of CRs, however, gives rise to various ethical concerns and opportunities, which require a holistic analysis from both the AI ethics and bioethics perspectives. Integrating both approaches allows for a multi-level analysis of the situation, where bioethics focuses primarily on the patient care practice, while AI ethics examines the implications of technology for groups and society. By taking this approach, we can ensure that ethical and technical trade-offs are adequately defined to meet performance expectations while safeguarding patients and the healthcare ecosystem they belong to. In discussing those trade-offs, some major points for the reduction of risks are put forth: (1) finding the adequate discourse for each patient and stakeholders to understand the technology they are using and what it entails; (2) the creation of guidance through regulations and standards on the state of the art to follow, accompanied by a clear accountability system for developers, providers, and users; and (3) always keeping the patient's interests at the centre of all deliberations. In addition to ethical considerations, practical design constraints for social robots in healthcare must be taken into account. Developers should adhere to ethical principles such as beneficence, non-maleficence, autonomy, justice, and explainability. Incorporating these principles into the design process and adopting an ethics-by-design approach can help prioritise the well-being and safety of users, avoid harm, respect users' autonomy, promote fairness and equity in healthcare, and ensure transparency and accountability in decision-making processes. It is important to acknowledge that specific ethical requirements may vary depending on the domain, culture, and users involved. However, the discussed practical design constraints serve as foundational goals that must be considered in the development of robotics in healthcare. Further ethical considerations should build upon and augment these constraints, taking into account the specific context and needs of CRs in providing care. Finally, we conclude that a sector-specific approach to ethical discussions is indeed needed to provide a complete understanding of the potential implications of integrating AI technology into healthcare systems.

Author Contributions: Conceptualisation, A.B., S.R., A.K., L.M.A. and C.L.; methodology, A.B., S.R., A.K., L.M.A. and C.L.; project administration, A.B.; supervision, C.L.; validation, A.B., S.R., A.K., L.M.A. and C.L.; visualisation, A.B., S.R., A.K., L.M.A. and C.L.; writing—original draft, A.B., S.R., A.K. and L.M.A.; writing—review and editing, A.B., S.R., A.K., L.M.A. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: Seamus Ryan's contribution has been supported in part by Science Foundation Ireland under Grant number 18/CRT/6222.

Data Availability Statement: No new data were created for this research.

Acknowledgments: This work was supported by the Institute for Ethics in Artificial Intelligence (IEAI) at the Technical University of Munich.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fox, J.; Gambino, A. Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychol. Behav. Soc. Netw.* **2021**, *24*, 294–299. [CrossRef]
2. Lambert, A.; Norouzi, N.; Bruder, G.; Welch, G. A Systematic Review of Ten Years of Research on Human Interaction with Social Robots. *Int. J. Hum.–Computer Interact.* **2020**, *36*, 1804–1817. [CrossRef]
3. Malle, B.F.; Scheutz, M.; Arnold, T.; Voiklis, J.; Cusimano, C. Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In Proceedings of the 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Portland, OR, USA, 2–5 March 2015; IEEE: Piscataway, NJ, USA; pp. 117–124.
4. Niemelä, M.; Heikkinen, S.; Koistinen, P.; Laakso, K.; Melkas, H.; Kyrki, V. *Robots and the Future of Welfare Services—A Finnish Roadmap*; Aalto University: Otaniemi, Finland, 2021.
5. Morgan, A.A.; Abdi, J.; Syed, M.A.; Kohen, G.E.; Barlow, P.; Vizcaychipi, M.P. Robots in healthcare: A scoping review. *Curr. Robot. Rep.* **2022**, *3*, 271–280. [CrossRef]
6. Broadbent, E.; Garrett, J.; Jepsen, N.; Ogilvie, V.L.; Ahn, H.S.; Robinson, H.; Peri, K.; Kerse, N.; Rouse, P.; Pillai, A.; et al. Using robots at home to support patients with chronic obstructive pulmonary disease: Pilot randomized controlled trial. *J. Med. Internet Res.* **2018**, *20*, e8640.
7. Vallor, S. Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. In *Machine Ethics and Robot Ethics*; Routledge: Milton Park, UK, 2020; pp. 137–154.
8. Boada, J.P.; Maestre, B.R.; Genís, C.T. The ethical issues of social assistive robotics: A critical literature review. *Technol. Soc.* **2021**, *67*, 101726. [CrossRef]
9. Dawe, J.; Sutherland, C.; Barco, A.; Broadbent, E. Can social robots help children in healthcare contexts? A scoping review. *BMJ Paediatr. Open* **2019**, *3*, e000371. [CrossRef]
10. Wagner, E.; Borycki, E.M. The Use of Robotics in Dementia Care: An Ethical Perspective. In *Informatics and Technology in Clinical Care and Public Health*; IOS Press: Amsterdam, The Netherlands, 2022; pp. 362–366.
11. Riek, L.D. Healthcare robotics. *Commun. ACM* **2017**, *60*, 68–78. [CrossRef]
12. Fiske, A.; Henningsen, P.; Buyx, A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *J. Med. Internet Res.* **2019**, *21*, e13216. [CrossRef]
13. de Graaf, M.M.A.; Allouch, S.B.; van Dijk, J.A.G.M. Long-term evaluation of a social robot in real homes. *Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst.* **2016**, *17*, 461–490. [CrossRef]
14. Fosch-Villaronga, E.; Poulsen, A. Sex care robots. Exploring the potential use of sexual robot technologies for disabled and elder care. *Paladyn J. Behav. Robot.* **2020**, *11*, 1–18. [CrossRef]
15. Vallès-Peris, N.; Domènech, M. Caring in the in-between: A proposal to introduce responsible AI and robotics to healthcare. *AI Soc.* **2021**, *38*, 1685–1695. [CrossRef]
16. McLennan, S.; Fiske, A.; Tigard, D.; Müller, R.; Haddadin, S.; Buyx, A. Embedded ethics: A proposal for integrating ethics into the development of medical AI. *BMC Med. Ethics* **2022**, *23*, 6. [CrossRef] [PubMed]
17. Naik, N.; Hameed, B.; Shetty, D.K.; Swain, D.; Shah, M.; Paul, R.; Aggarwal, K.; Ibrahim, S.; Patil, V.; Smriti, K.; et al. Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Front. Surg.* **2022**, *9*, 266. [CrossRef] [PubMed]
18. Normative Approach. 2023. Available online: https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100238783;jsessionid=F2BC2B6AF0277F7B5FCC93F914EC5FC8 (accessed on 30 June 2023).
19. Van de Ven, B. An ethical framework for the marketing of corporate social responsibility. *J. Bus. Ethics* **2008**, *82*, 339–352. [CrossRef]
20. Edgett, R. Toward an ethical framework for advocacy in public relations. *J. Public Relations Res.* **2002**, *14*, 1–26. [CrossRef]
21. King, S.A. Researching Internet communities: Proposed ethical guidelines for the reporting of results. *Inf. Soc.* **1996**, *12*, 119–128. [CrossRef]

22. Borgatti, S.P.; Molina, J.L. Toward ethical guidelines for network research in organizations. *Soc. Netw.* **2005**, *27*, 107–117. [CrossRef]
23. Emanuel, E.J.; Emanuel, L.L. What is accountability in health care? *Ann. Intern. Med.* **1996**, *124*, 229–239. [CrossRef]
24. Kass, N.E. An ethics framework for public health. *Am. J. Public Health* **2001**, *91*, 1776–1782. [CrossRef]
25. Jones, A.H. Literature and medicine: Narrative ethics. *Lancet* **1997**, *349*, 1243–1246. [CrossRef]
26. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef] [PubMed]
27. Väyrynen, P. Normative explanation and justification. *Noûs* **2021**, *55*, 3–22. [CrossRef]
28. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
29. Heilinger, J.C. The ethics of AI ethics. A constructive critique. *Philos. Technol.* **2022**, *35*, 61. [CrossRef]
30. Holzinger, A.; Kieseberg, P.; Weippl, E.; Tjoa, A.M. Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable AI. In Proceedings of the International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Hamburg, Germany, 27–30 August 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–8.
31. Bauer, W.A. Virtuous vs. utilitarian artificial moral agents. *AI Soc.* **2020**, *35*, 263–271. [CrossRef]
32. Kriebitz, A.; Lütge, C. Artificial intelligence and human rights: A business ethical assessment. *Bus. Hum. Rights J.* **2020**, *5*, 84–104. [CrossRef]
33. *IEEE Std 7010-2020*; IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being. IEEE: Piscataway, NJ, USA, 2020; pp. 1–96. [CrossRef]
34. EU. *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*; European Parliament, Council of the European Union: Strasbourg, France, 2021.
35. Max, R.; Kriebitz, A.; Von Websky, C. Ethical considerations about the implications of artificial intelligence in finance. *Handb. Ethics Financ.* **2021**, 577–592.
36. Drage, E.; Mackereth, K. Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference". *Philos. Technol.* **2022**, *35*, 1–25. [CrossRef]
37. Bostrom, N.; Yudkowsky, E. The ethics of artificial intelligence. In *Artificial Intelligence Safety and Security*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018; pp. 57–69.
38. Kriebitz, A.; Max, R.; Lütge, C. The German Act on Autonomous Driving: Why ethics still matters. *Philos. Technol.* **2022**, *35*, 1–13. [CrossRef]
39. Bonnefon, J.F.; Shariff, A.; Rahwan, I. The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view]. *Proc. IEEE* **2019**, *107*, 502–504. [CrossRef]
40. Amugongo, L.M.; Bidwell, N.J.; Corrigan, C.C. Invigorating Ubuntu Ethics in AI for Healthcare: Enabling Equitable Care. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; Association for Computing Machinery: New York, NY, USA, 2023; FAccT '23, pp. 583–592. [CrossRef]
41. Jotterand, F. The Hippocratic oath and contemporary medicine: Dialectic between past ideals and present reality? *J. Med. Philos.* **2005**, *30*, 107–128. [CrossRef]
42. Robbins, D.A.; Curro, F.A.; Fox, C.H. Defining patient-centricity: Opportunities, challenges, and implications for clinical care and research. *Ther. Innov. Regul. Sci.* **2013**, *47*, 349–355. [CrossRef] [PubMed]
43. Surbone, A. Telling the truth to patients with cancer: What is the truth? *Lancet Oncol.* **2006**, *7*, 944–950. [CrossRef] [PubMed]
44. Häyry, M. Roles of justice in bioethics. *Roles of Justice in Bioethics. Elements in Bioethics and Neuroethics*; Cambridge University Press: Cambridge, UK, 2022.
45. Takala, T. What is wrong with global bioethics? On the limitations of the four principles approach. *Camb. Q. Healthc. Ethics* **2001**, *10*, 72–77. [CrossRef]
46. Lawrence, D.J. The four principles of biomedical ethics: A foundation for current bioethical debate. *J. Chiropr. Humanit.* **2007**, *14*, 34–40. [CrossRef]
47. Colonna, L. Legal Implications of Using AI as an Exam Invigilator. *Fac. Law Stockh. Univ. Res. Pap.* **2021**, *91*, 13–46. [CrossRef]
48. Hagerty, A.; Rubinov, I. Global AI ethics: A review of the social impacts and ethical implications of artificial intelligence. *arXiv* **2019**, arXiv:1907.07892.
49. Morley, J.; Machado, C.C.; Burr, C.; Cowls, J.; Joshi, I.; Taddeo, M.; Floridi, L. The ethics of AI in health care: A mapping review. *Soc. Sci. Med.* **2020**, *260*, 113172. [CrossRef]
50. Di Nardo, M.; Dalle Ore, A.; Testa, G.; Annich, G.; Piervincenzi, E.; Zampini, G.; Bottari, G.; Cecchetti, C.; Amodeo, A.; Lorusso, R.; et al. Principlism and personalism. Comparing two ethical models applied clinically in neonates undergoing extracorporeal membrane oxygenation support. *Front. Pediatr.* **2019**, *7*, 312. [CrossRef]
51. Sand, M.; Durán, J.M.; Jongsma, K.R. Responsibility beyond design: Physicians' requirements for ethical medical AI. *Bioethics* **2022**, *36*, 162–169. [CrossRef]
52. Varkey, B. Principles of clinical ethics and their application to practice. *Med. Princ. Pract.* **2021**, *30*, 17–28. [CrossRef] [PubMed]
53. Beauchamp, T.L.; McCullough, L.B. Medical ethics: The moral responsibilities of physicians. *Pers. Forum* **1985**, *1*, 46–51.
54. WHO. *Ageing and Health*; WHO: Geneva, Switzerland, 2022.

55. Meskó, B.; Hetényi, G.; Győrffy, Z. Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Serv. Res.* **2018**, *18*, 545. [CrossRef] [PubMed]
56. Sparrow, R.; Sparrow, L. In the hands of machines? The future of aged care. *Minds Mach.* **2006**, *16*, 141–161. [CrossRef]
57. Robinson, H.; MacDonald, B.; Broadbent, E. The Role of Healthcare Robots for Older People at Home: A Review. *Int. J. Soc. Robot.* **2014**, *6*, 575–591. [CrossRef]
58. Calo, C.J.; Hunt-Bull, N.; Lewis, L.; Metzler, T. Ethical implications of using the paro robot, with a focus on dementia patient care. In Proceedings of the Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 7–11 August 2011.
59. Shamsuddin, S.; Yussof, H.; Ismail, L.; Hanapiah, F.A.; Mohamed, S.; Piah, H.A.; Zahari, N.I. Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. In Proceedings of the 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, Malacca, Malaysia, 23–25 March 2012; IEEE: Piscataway, NJ, USA; pp. 188–193.
60. Tan, S.Y.; Taeihagh, A.; Tripathi, A. Tensions and antagonistic interactions of risks and ethics of using robotics and autonomous systems in long-term care. *Technol. Forecast. Soc. Chang.* **2021**, *167*, 120686. [CrossRef]
61. Age, U. *Only the Tip of the Iceberg: Fraud against Older People*; Age UK: London, UK, 2015.
62. Fosch-Villaronga, E.; Lutz, C.; Tamò-Larrieux, A. Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops. *Int. J. Soc. Robot.* **2020**, *12*, 441–458. [CrossRef]
63. Commission, E. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*; European Parliament, Council of the European Union: Strasbourg, France, 2016. Available online: https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32016R0679 (accessed on 2 June 2023).
64. Denning, T.; Matuszek, C.; Koscher, K.; Smith, J.R.; Kohno, T. A spotlight on security and privacy risks with future household robots: Attacks and lessons. In Proceedings of the 11th International Conference on Ubiquitous Computing, Orlando, FL, USA, 30 September–3 October 2009; pp. 105–114.
65. Müller, V.C. *Ethics of Artificial Intelligence and Robotics*; Stanford Encyclopedia of Philosophy, Stanford University: Stanford, CA, USA, 2020.
66. Ford, M. The rise of the robots: Technology and the threat of mass unemployment. *Int. J. HRD Pract. Policy Res.* **2015**, *1*, 111-112.
67. Frey, C.B.; Berger, T.; Chen, C. Political machinery: Did robots swing the 2016 US presidential election? *Oxf. Rev. Econ. Policy* **2018**, *34*, 418–442. [CrossRef]
68. Darling, K. 'Who's Johnny?'Anthropomorphic framing in human-robot interaction, integration, and policy. In *Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy (March 23, 2015). ROBOT ETHICS*; Oxford University Press: Oxford, UK, 2015; Volume 2.
69. Corretjer, M.G.; Ros, R.; Martin, F.; Miralles, D. The maze of realizing empathy with social robots. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 1334–1339.
70. Boch, A.; Lucaj, L.; Corrigan, C. A robotic new hope: Opportunities, challenges, and ethical considerations of social robots. *Tech. Univ. Munich* **2020**, *1*, 1–12.
71. Turkle, S. In good company?: On the threshold of robotic companions. In *Close Engagements with Artificial Companions*; John Benjamins: Amsterdam, The Netherlands; Philadelphia, PA, USA, 2010; pp. 3–10.
72. Scheutz, M. The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots. *Robot. Ethics Ethical Soc. Implic. Robot.* **2011**, *1*, 205–221.
73. Darling, K. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In *Robot Law*; Edward Elgar Publishing: Cheltenham, UK, 2016.
74. Van Maris, A.; Zook, N.; Caleb-Solly, P.; Studley, M.; Winfield, A.; Dogramadzi, S. Designing ethical social robots—A longitudinal field study with older adults. *Front. Robot. AI* **2020**, *7*, 1. [CrossRef] [PubMed]
75. Schiappa, E.; Allen, M.; Gregg, P.B. Parasocial relationships and television: A meta-analysis of the effects. In *Mass Media Effects Research: Advances through Meta-Analysis*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2007; pp. 301–314.
76. Perse, E.M.; Rubin, R.B. Attribution in social and parasocial relationships. *Commun. Res.* **1989**, *16*, 59–77. [CrossRef]
77. Coeckelbergh, M.; Pop, C.; Simut, R.; Peca, A.; Pintea, S.; David, D.; Vanderborght, B. A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: Ethical acceptability, trust, sociability, appearance, and attachment. *Sci. Eng. Ethics* **2016**, *22*, 47–65. [CrossRef] [PubMed]
78. Birnbaum, G.E.; Mizrahi, M.; Hoffman, G.; Reis, H.T.; Finkel, E.J.; Sass, O. What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Comput. Hum. Behav.* **2016**, *63*, 416–423. [CrossRef]
79. Sharkey, N.; Sharkey, A. The crying shame of robot nannies: An ethical appraisal. *Interact. Stud.* **2010**, *11*, 161–190. [CrossRef]
80. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
81. Directorate-General for Internal Policies, Policy Department, Citizens's Riths and Costitutional Affairs European Civil Law Rules on Robotics. 2016. Available online: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf (accessed on 2 June 2023).

82. Decker, M.; Dillmann, R.; Dreier, T.; Fischer, M.; Gutmann, M.; Ott, I.; genannt Döhmann, I.S. Service robotics: Do you know your new companion? Framing an interdisciplinary technology assessment. *Poiesis Prax.* **2011**, *8*, 25–44. [CrossRef]
83. Robert Koch Institute. Gesundheit in Deutschland. 2015. Available online: https://www.gbe-bund.de/pdf/gesber2015.pdf (accessed on 2 June 2023).
84. Jacobs, K.; Kuhlmey, A.; Greß, S.; Klauber, J.; Schwinger, A. *Pflege-Report 2019: Mehr Personal in der Langzeitpflege-Aber Woher*? Springer Nature: Berlin, Germany, 2020.
85. Bendel, O. *Pflegeroboter*; Springer Nature: Berlin, Germany, 2018.
86. Mordor Intellegence. Social Robots Market Size, Share, Growth, Trends: 2022–2027. 2021. Available online: https://www.mordorintelligence.com/industry-reports/social-robots-market (accessed on 2 June 2023).
87. Arun, C. *AI and the Global South: Designing for Other Worlds*; The Oxford Handbook of Ethics of AI; Oxford University Press: Oxford, UK, 2019.
88. The 'AI Divide' between the Global North and the Global South. 2023. Available online: https://www.weforum.org/agenda/2023/01/davos23-ai-divide-global-north-global-south/ (accessed on 30 June 2023).
89. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the Conference on Fairness, Accountability and Transparency, PMLR, New York, NY, USA, 23–24 February 2018; pp. 77–91.
90. West, S.M.; Whittaker, M.; Crawford, K. Discriminating systems. *AI Now* **2019**. https://ainowinstitute.org/wp-content/uploads/2023/04/discriminatingsystems.pdf (accessed on 2 June 2023).
91. Leslie, D.; Mazumder, A.; Peppin, A.; Wolters, M.K.; Hagerty, A. Does "AI" stand for augmenting inequality in the era of COVID-19 healthcare? *BMJ* **2021**, *372*, n304. [CrossRef]
92. Delgado, J.; de Manuel, A.; Parra, I.; Moyano, C.; Rueda, J.; Guersenzvaig, A.; Ausin, T.; Cruz, M.; Casacuberta, D.; Puyol, A. Bias in algorithms of AI systems developed for COVID-19: A scoping review. *J. Bioethical Inq.* **2022**, *19*, 407–419. [CrossRef] [PubMed]
93. Pawar, U.; O'Shea, D.; Rea, S.; O'Reilly, R. Explainable AI in Healthcare. In Proceedings of the 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Dublin, Ireland, 15–19 June 2020; IEEE: Piscataway, NJ, USA; pp. 1–2. [CrossRef]
94. European Parliamentary Research Service. *Understanding Algorithmic Decision-Making: Opportunities and Challenges*; European Parliamentary Research Service: Brussels, Belgium, 2019.
95. Wanner, J.; Herm, L.V.; Heinrich, K.; Janiesch, C.; Zschech, P. White, Grey, Black: Effects of XAI Augmentation on the Confidence in AI-based Decision Support Systems. In Proceedings of the 41st International Conference on Information Systems, ICIS 2020, Making Digital Inclusive: Blending the Locak and the Global, Hyderabad, India, 13–16 December 2020; George, J.F., Paul, S., De', R., Karahanna, E., Sarker, S., Oestreicher-Singer, G., Eds.; Association for Information Systems: Atlanta, GA, USA, 2020.
96. London, A.J. Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Cent. Rep.* **2019**, *49*, 15–21. [CrossRef] [PubMed]
97. Brkan, M.; Bonnet, G. Legal and technical feasibility of the GDPR's quest for explanation of algorithmic decisions: Of black boxes, white boxes and Fata Morganas. *Eur. J. Risk Regul.* **2020**, *11*, 18–50. [CrossRef]
98. Ryan, S.; Nurgalieva, L.; Doherty, G. Perceived Fairness Concerns Within Pandemic Response Technology. *Interact. Comput.* **2022**, iwac040. [CrossRef]
99. Nurgalieva, L.; Ryan, S.; Balaskas, A.; Lindqvist, J.; Doherty, G. Public Views on Digital COVID-19 Certificates: A Mixed Methods User Study. In Proceedings of the CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April–5 May 2022; ACM: New York, NY, USA; Volume 1, pp. 1–28. [CrossRef]
100. Boch, A.; Hohma, E.; Trauth, R. *Towards an Accountability Framework for AI: Ethical and Legal Considerations*; Institute for Ethics in AI, Technical University of Munich: Munich, Germany, 2022.
101. Danaher, J. The philosophical case for robot friendship. *J. Posthuman Stud.* **2019**, *3*, 5–24. [CrossRef]
102. Nyholm, S.; Frank, L.E. From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible? In *Robot Sex: Social and Ethical Implications*; MIT Press: Cambridge, MA, USA, 2017.
103. Reig, S.; Carter, E.J.; Tan, X.Z.; Steinfeld, A.; Forlizzi, J. Perceptions of Agent Loyalty with Ancillary Users. *Int. J. Soc. Robot.* **2021**, *13*, 2039–2055. [CrossRef]
104. Vanderelst, D.; Willems, J. Can we agree on what robots should be allowed to do? An exercise in rule selection for ethical care robots. *Int. J. Soc. Robot.* **2020**, *12*, 1093–1102. [CrossRef]
105. Russo, F. What is the CSR's Focus in Healthcare? *J. Bus. Ethics* **2016**, *134*, 323–334. [CrossRef]
106. Werhane, P.H. Business ethics, stakeholder theory, and the ethics of healthcare organizations. *Camb. Q. Healthc. Ethics* **2000**, *9*, 169–181. [CrossRef]
107. Borgonovi, E. La responsabilità sociale in medicina. *Mecosan* **2005**, *14*, 3–9.
108. Collins, S.K. Corporate social responsibility and the future health care manager. *Health Care Manag.* **2010**, *29*, 339–345. [CrossRef] [PubMed]
109. Lee, H.; Piao, M.; Lee, J.; Byun, A.; Kim, J. The purpose of bedside robots: Exploring the needs of inpatients and healthcare professionals. *CIN Comput. Inform. Nurs.* **2020**, *38*, 8–17. [CrossRef] [PubMed]
110. Liang, H.F.; Wu, K.M.; Weng, C.H.; Hsieh, H.W. Nurses' views on the potential use of robots in the pediatric unit. *J. Pediatr. Nurs.* **2019**, *47*, e58–e64. [CrossRef] [PubMed]
111. Maibaum, A.; Bischof, A.; Hergesell, J.; Lipp, B. A critique of robotics in health care. *AI Soc.* **2022**, *37*, 1–11. [CrossRef]

112. Jang, S.M.; Lee, K.; Hong, Y.J.; Kim, J.; Kim, S. Economic evaluation of robot-based telemedicine consultation services. *Telemed. e-Health* **2020**, *26*, 1134–1140. [CrossRef] [PubMed]
113. European Commission; Directorate-General for Communications Networks, Content and Technology; Worthington, H.; Simmonds, P.; Farla, K.; Varnai, P. *The Silver Economy: Final Report*; Publications Office: Technopolis Group: Brighton, UK, 2018. [CrossRef]
114. World Health Organisation (WHO). Active Ageing: A Policy Framework. 2014. Available online: https://extranet.who.int/agefriendlyworld/wp-content/uploads/2014/06/WHO-Active-Ageing-Framework.pdf (accessed on 2 June 2023).
115. Killackey, T.; Peter, E.; Maciver, J.; Mohammed, S. Advance care planning with chronically ill patients: A relational autonomy approach. *Nurs. Ethics* **2020**, *27*, 360–371. [CrossRef]
116. van Wynsberghe, A.L. Designing Robots with Care: Creating an Ethical Framework for the Future Design and Implementation of Care Robots. Ph.D. Thesis, University of Twente, Enschede, The Netherlands, 2012.
117. Herstatt, C.; Kohlbacher, F.; Bauer, P. *"Silver" Product Design: Product Innovation for Older People*; Technical Report, Working Paper; Institute for Technology and Innovation Management, Hamburg University of Technology (TUHH): Hamburg, Germany, 2011.
118. Hancock, P.A.; Kessler, T.T.; Kaplan, A.D.; Brill, J.C.; Szalma, J.L. Evolving trust in robots: Specification through sequential and comparative meta-analyses. *Hum. Factors* **2021**, *63*, 1196–1229. [CrossRef]
119. Broadbent, E. Interactions with robots: The truths we reveal about ourselves. *Annu. Rev. Psychol.* **2017**, *68*, 627–652. [CrossRef]
120. Coco, K.; Kangasniemi, M.; Rantanen, T. Care personnel's attitudes and fears toward care robots in elderly care: A comparison of data from the care personnel in Finland and Japan. *J. Nurs. Scholarsh.* **2018**, *50*, 634–644. [CrossRef]
121. De Swarte, T.; Boufous, O.; Escalle, P. Artificial intelligence, ethics and human values: The cases of military drones and companion robots. *Artif. Life Robot.* **2019**, *24*, 291–296. [CrossRef]
122. Jenkins, S.; Draper, H. Care, monitoring, and companionship: Views on care robots from older people and their carers. *Int. J. Soc. Robot.* **2015**, *7*, 673–683. [CrossRef]
123. European Patient Forum. *Clinical Trials Regulation: Informed Consent and Information to Patients*; European Patient Forum: Brussels, Belgium, 2016. Available online: https://www.eupatient.eu/globalassets/policy/clinicaltrials/epf_informed_consent_position_statement_may16.pdf (accessed on 2 June 2023).
124. Lutz, C.; Schöttler, M.; Hoffmann, C.P. The privacy implications of social robots: Scoping review and expert interviews. *Mob. Media Commun.* **2019**, *7*, 412–434. [CrossRef]
125. Abney, K.; Bekey, G.A.; Lin, P. Robots and privacy. In *Robot Ethics: The Ethical and Social Implications of Robotics*; The MIT Press: Cambridge, MA, USA, 2014; pp. 187–201.
126. United Nations. *Universal Declaration of Human Rights*; United Nations, New York, NY, USA, 1948.
127. Pino, M.; Boulay, M.; Jouen, F.; Rigaud, A.S. "Are we ready for robots that care for us?" Attitudes and opinions of older adults toward socially assistive robots. *Front. Aging Neurosci.* **2015**, *7*, 141. [CrossRef]
128. Draper, H.; Sorell, T. Ethical values and social care robots for older people: An international qualitative study. *Ethics Inf. Technol.* **2017**, *19*, 49–68. [CrossRef]
129. Lockhart, J.W.; Weiss, G.M. The benefits of personalized smartphone-based activity recognition models. In Proceedings of the 2014 SIAM International Conference on Data Mining, SIAM, Philadelphia, PA, USA, 24–26 April 2014; pp. 614–622.
130. Tsiakas, K.; Abujelala, M.; Makedon, F. Task engagement as personalization feedback for socially-assistive robots and cognitive training. *Technologies* **2018**, *6*, 49. [CrossRef]
131. Ebert, F.; Yang, Y.; Schmeckpeper, K.; Bucher, B.; Georgakis, G.; Daniilidis, K.; Finn, C.; Levine, S. Bridge data: Boosting generalization of robotic skills with cross-domain datasets. *arXiv* **2021**, arXiv:2109.13396.
132. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The future of digital health with federated learning. *NPJ Digit. Med.* **2020**, *3*, 119. [CrossRef]
133. Markus, A.F.; Kors, J.A.; Rijnbeek, P.R. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J. Biomed. Inform.* **2021**, *113*, 103655. [CrossRef]
134. Stahl, B.C.; Coeckelbergh, M. Ethics of healthcare robotics: Towards responsible research and innovation. *Robot. Auton. Syst.* **2016**, *86*, 152–161. [CrossRef]
135. Bradwell, H.L.; Winnington, R.; Thill, S.; Jones, R.B. Ethical perceptions towards real-world use of companion robots with older people and people with dementia: Survey opinions among younger adults. *BMC Geriatr.* **2020**, *20*, 244. [CrossRef]
136. Johnston, C. Ethical Design and Use of Robotic Care of the Elderly. *J. Bioethical Inq.* **2022**, *19*, 11–14. [CrossRef] [PubMed]
137. Molewijk, B.; Hem, M.H.; Pedersen, R. Dealing with ethical challenges: A focus group study with professionals in mental health care. *BMC Med. Ethics* **2015**, *16*, 1–12. [CrossRef]
138. Owen, S. The practical, methodological and ethical dilemmas of conducting focus groups with vulnerable clients. *J. Adv. Nurs.* **2001**, *36*, 652–658. [CrossRef] [PubMed]
139. Park, J.; Han, S.H.; Kim, H.K.; Cho, Y.; Park, W. Developing elements of user experience for mobile phones and services: Survey, interview, and observation approaches. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2013**, *23*, 279–293. [CrossRef]
140. Harrison, S.; Mort, M. Which champions, which people? Public and user involvement in health care as a technology of legitimation. *Soc. Policy Adm.* **1998**, *32*, 60–70. [CrossRef]

141. Magelssen, M.; Pedersen, R.; Miljeteig, I.; Ervik, H.; Førde, R. Importance of systematic deliberation and stakeholder presence: A national study of clinical ethics committees. *J. Med. Ethics* **2020**, *46*, 66–70. [CrossRef]
142. Stevenson, F.A.; Gibson, W.; Pelletier, C.; Chrysikou, V.; Park, S. Reconsidering 'ethics' and 'quality'in healthcare research: The case for an iterative ethical paradigm. *BMC Med. Ethics* **2015**, *16*, 1–9. [CrossRef]
143. Releases, P. AI Act: A Step Closer to the First Rules on Artificial Intelligence. 2023. Available online: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence (accessed on 30 June 2023).
144. ISO. ISO IEC JTC 1 SC 42 Artificial Intelligence. 2023. Available online: https://www.iso.org/committee/6794475.html (accessed on 30 June 2023).
145. Raji, I.D.; Smart, A.; White, R.N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; Barnes, P. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 27–30 January 2020; pp. 33–44.