**Expanding Care-Centered Value Sensitive Design: Design Care Robots with AI for Social Good Norms**

**Abstract**

The increasing automation and ubiquity of robotics deployed within the field of care boasts promising advantages. However, challenging ethical issues arise also as a consequence. This paper takes care robots for the elderly as the subject of analysis, building on previous literature in the domain of the ethics and design of care robots. It takes the value sensitive design (VSD) approach to technology design and extends its application to care robots by not only integrating the values of care, but also those specific to AI as well as higher-scale values such as the United Nations Sustainable Development Goals (SDGs). The ethical issues specific to care robots for the elderly as discussed as well as examples of specific design requirements to ameliorate those issues.

1. **Introduction**

SARS-CoV-2, amongst its more immediate medical issues, has betrayed fundamental precarities and instabilities across a host of social, economic and political domains globally. Amongst these issues, the practice of care, that is, caring for patients – both ill with the disease and those isolated from it – has been particularly troublesome as the solution(s) to such issues are not clear. Eraviperoor village in India has recently deployed a series of care robots that are equipped to provide patients in their local medical center with medicines, bedsheets and food, weighing up to 8 kg (Kuttoor 2020). Similarly, researchers studying assisted living at Heriot-Watt University in Edinburgh are employing co-design approaches to develop care robots to combat COVID care isolation (Macdonald 2020).

These examples are a few amongst many illustrations that betray a trend towards further automation and the deployment of information and communication technologies (ICTs) and robotics within the domain of care (Mordoch et al. 2013). Consequentially, the diaspora of care robots globally has and continues to raise several challenging ethical, social, cultural and political issues among others. The reliability of care robots to provide beneficial aid, their medical fidelity to ensure proper treatments, and the question of whether they can provide at least sufficient companionship and comfort to those who are isolated are some of the many ethical questions that emerge when considering deploying robots within the domain of care. To confront these issues, a significant body of literature has emerged to determine how to integrate applied ethical approaches towards the design and deployment of these types of autonomous systems to ensure beneficial ends (e.g., van Wynsberghe 2012, 2013, 2015). Intervening at the design phase has been a long-standing position in the field of responsible innovation and has recently been the focus of various multinational governance and funding bodies (United Nations 2019; van den Hoven and Jacob 2013; van Lente et al. 2017).

The aim of this paper is to focus primarily on care robots for the elderly, providing a conceptual investigation of the ethical issues and human values that emerge within the framework of value sensitive design (VSD), a principled approach to the design of technologies *for* human values (Friedman and Hendry 2019). In doing so, this paper builds off of the Care Centered VSD (CCVSD) approach proposed by van Wynsberghe (2013), expanding it to include other sources of values, such as the United Nations Sustainable Development Goals (SDGs) and norms that are specifically related to beneficial autonomous systems design, such as that AI for Social Good principles by Floridi et al. (2020) and the ethics guidelines for trustworthy AI by the High-Level Expert Group on Artificial Intelligence (HLEG AI) (High-Level Expert Group on AI 2019). By adopting the multi-tiered approach to AI design viz. VSD proposed by Umbrello and van de Poel (2020), this paper aims to provide a more thorough analysis of care robot design for the elderly that more accurate maps onto the ethics of care proposed by van Wynsberghe (2013).

Whereas previous studies have focused on solely the ethical issues of care robots (Sharkey and Sharkey 2012; Vandemeulebroucke et al. 2018) and their design using care ethics (van Wynsberghe 2016), this paper is comparatively unique in its approach as it takes a multi-tiered approach to implementing VSD by drawing on multiple sources of values both at the domain, technological and international levels as they pertain to autonomous systems. By doing so, this paper aims to contribute to the salient design of these types of artificial intelligence systems at a global level.

To do this, this paper is divided into the following parts. Section 2 discusses the previous literature on VSD, outlining the basic methodology that will be employed throughout the paper. Section 3 discusses the current state-of-the-art regarding ethics and design of care robotics, particularly focuses on the work of van Wynsberghe. Section 4 introduces the multi-tiered VSD approach of Umbrello and van de Poel (2020) and outlines ways to applying this approach to elderly care robots. Section 5 discusses in greater depth the ethical issues particular to elderly care robots and how the proposed approach can be employed to provide salient design requirements to meet those challenges. Section 6 provides some conclusions.

1. **The Value Sensitive Design Approach**

Value-sensitive design (VSD) is defined as “a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process” (Friedman et al. 2013, 2). Against the neutrality thesis which states that technological systems are neutral tools and depend on the users for their status, VSD promotes an interactional understanding of technological systems. The latter implies that the impact of technological systems on users is shaped by the features of their design, the context in which they are used, and the people involved in their use (van den Hoven et al. 2015). Therefore, the main theoretical aim of VSD is to incorporate an explicit investigation on moral and social values and a clear and coherent methodology into the overall design and implementation process of systems.

In the VSD literature, the definition of value given by Friedman and colleagues refers to ‘‘what a person or group of people consider important in life’’(Friedman et al. 2013, 2). Thus, the identification of values in VSD varies depending on the specific systems, contexts, stakeholders and application domain under analysis. Indeed, VSD was originally elaborated in the information and communication technology domain (Friedman and Kahn Jr. 2003; van den Hoven 2007), but the approach is now more widely adopted and further extended to different domains, sometimes under the different heading of Design for Values (van den Hoven et al. 2015). However, the methodology remains the same and is based on Friedman et al. (2002) tripartite methodology, which is composed of three types of iterative and integrative investigations: conceptual, empirical and technical (see Fig.1).

**Conceptual Investigations**

Values from both the relevant philosophical literature and those explicitly elicited from stakeholders are determined and investigated.

**Technical Investigations**

The technical limitations of the technology itself are evaluated for how they support or constrain identified values and design requirements

**Empirical Investigations**

Stakeholder values are empirically evaluated through socio-cultural norms and translated into potential design requirements

**Fig. 1.** The recursive VSD tripartite framework employed in this study. Source: (Umbrello 2020)

The conceptual investigation involves two primary activities, the identification of stakeholders that are or will be affected by the system and the identification and definition of values and possible trade-offs. The empirical investigation examines stakeholders’ “understandings, contexts, and experiences” (Friedman and Kahn Jr. 2002). Finally, the technical investigation is concerned with specific features of existing or new systems. Since its inception, VSD has accompanied the investigation of values with the use of a range of social science methods, and Friedman et al. (2019) recently propose 14 more specific methods that can be used: (1) stakeholder analysis; (2) designer/stakeholder explicitly supported values; (3) coevolution of technology and social structure; (4) value scenarios; (5) value sketches; (6) value-oriented semi-structured interview; (7) granular assessments of magnitude, scale, and proximity; (8) value-oriented coding manual; (9) value-oriented mock-ups, prototypes, and field deployments; (10) ethnography focused on values and technology; (11) model for informed consent online; (12) value dams and flows; (13) value sensitive action reflection model; and (14) envisioning cards.

Also, tools such as the *values hierarchy* developed by van de Poel (2013) are crucial in translating values into more tangible design requirements (see Fig. 2). A values hierarchy is a structure that comprises three basic layers: 1) values, which are general values that need to be promoted and strived for their own sake; 2) norms, which are restrictions on or prescriptions for action, and 3) design requirements, as specific criteria that should be achieved as much as possible. The hierarchies can be constructed top-down as well as bottom-up. First, through the means of a non-deductive and context-dependent specification of high-level elements into lower ones. Second, through the means of *for the sake of* as a way to relate lower-level elements to higher ones, which suggests that the latter has a motivating and justifying role for the former.

Design Requirements

Norms

Values

Value

Norm

Design Requirement

Design Requirement

Design Requirement

Design Requirement

Norm

Norm

Norm

**Fig. 2.** Values hierarchy. Source: (van de Poel 2013)

Despite its fruitful applications, VSD has on the other hand also received some criticism, mainly due to the fact that such approach does not recommend or forbid the explicit commitment to a particular ethical theory (van den Hoven et al. 2015). Following the critiques’ argument, this implies that VSD lacks a clear methodology to normatively distinguish between general moral values from mere stakeholders’ preferences (Albrechtslund 2007; Manders-Huits 2011). For example, bottom-up approaches that argue that it is better to elicit values from stakeholders often lack the ability to normatively justify value prioritizations or value trade-offs (Borning and Muller 2012; Le Dantec et al. 2009). Moreover, the predominant focus on stakeholder may lead to the exclusion of other values and actors that are of ethical importance and should be included in VSD. Other approaches provide a list of values, as in the case of Friedman et al. (2017) that propose a list of 13 values related to the design of information systems. However, these top-down approaches are often too indeterminate to critically assess specific contexts and systems.

1. **Care Robots and Care Centered VSD**

The VSD approach in addressing issues pertaining to design shows how the evaluation of systems cannot be disentangled from the role and tasks they are expected to fulfil. This is a fundamental aspect in the health care domain, where systems may shape the decision-making processes, practices and behaviours of vulnerable persons. One of the definitions of care robots is: “Carebots are robots designed for use in home, hospital, or other settings to assist in, support, or provide care for the sick, disabled, young, elderly or otherwise vulnerable persons” (Vallor 2011, 252). In the literature of robot ethics, the discussions surrounding care robots focused on the autonomy and vulnerability of patients as care receivers (Pirni et al., 2017; Sharkey and Sharkey 2012; Sorell and Draper 2014; Sparrow and Sparrow 2006) or, alternatively, on caregivers and standards of care (Sharkey 2014; Vallor 2011). While these contributions were right in claiming that the introduction of robots inevitably change healthcare practices, the necessity of a normative framework coupled with a design-oriented perspective has been only very recently recognized by the works of van Wynsberghe on Care Centred Value Sensitive Design (CCVSD).

In this scenario, CCVSD has been developed as a particular response to the issues related to care robots and to the criticisms against VSD we mentioned in the previous section. Indeed, van Wynsberghe promoted the CCVSD approach to demonstrate how it may be used with a twofold aim. First, as a new framework specifically tailored to evaluate care robots and practices even outside the health care domain, given also the lack of standards provided by the International Organization for Standardization (ISO 2011). Second, as a means to overcome the lack of a transparent and explicit normative grounding in VSD (van Wynsberghe 2012, 2013b, 2013a, 2016).

According to van Wynsberghe, the manner through which this task may be accomplished implies the integration of the traditional VSD approach with normative criteria and elements from a care ethics perspective. This serves to better identify and establish whichvalues should be promoted in the VSD process (van Wynsberghe 2013a). Following the care ethicist Joan Tronto (1993), van Wynsberghe identifies four fundamental value of care to be promoted in the design of systems: 1) attentiveness, as the capacity to recognise the needs of the care-receiver; 2) responsibility, which implies caregiver’s concern for meeting the needs of the care-receiver; 3) competence, as the capacity of executing an action to fulfil the needs of the care-receiver; and 4) responsiveness or reciprocity, as the capacity of the care receiver to guide the caregiver and the instauration of a reciprocal interaction (van Wynsberghe 2012, 2013b, 2013a, 2016).

Van Wynsberghe insists that these four elements are crucial in any care practice that impact both on caregivers and care receivers, due to the ethical importance they assign to the relationship and distribution of roles and responsibilities between them (Tronto 2010). Specifically, van Wynsberghe’s approach is intended to help ethicists and designers in their investigations with the means of two tools: 1) a care-centred framework, which consists of five components that require attention in a care analysis: context, practice, actors, type of robot, manifestation of the four moral elements (see Fig. 3); 2) a specific CCVSD methodology, which using the framework proactively guide ethicists and designers through a series of steps, from data collection and value analysis of the care practice (with and without the robot) and of robot’s capabilities, *to scenario comparison and recommendation for design.*



**Fig. 3** *Care-Centered Framework*. Source: (van Wynsberghe 2013a, 420)

In summary, CCVSD relies on the concept of care practice as a response to the needs of ‘the other’ (van Wynsberghe 2016) to determine the four values to be included in the design of systems. However, if on one side this approach has the merit of stressing the importance of a normative evaluation in VSD, on the other it dismisses in its methodology other potentially relevant values that may play a pivotal role in care practices. Thus, other approaches have already tried to expand and develop CCVSD further. For example, Santoni de Sio and van Wynsberghe (2016) rely on a nature-of-activities approach for a CCVSD of care robots. In giving a more explicit philosophical foundation and more fine-grained descriptions to the nature of care activities, this approach extends the CCVSD approach by leaving space for the inclusion of other values and other philosophical traditions, such as individual autonomy that is traditionally associated with traditions other than care ethics.

 As a matter of fact, even van Wynsberghe recognises that the list of values in the CC framework is not exhaustive and allows for additions (Fig.3). Moreover, in her recent work she acknowledges that when a care robot has a sufficient level of AI, it can engage in a degree of attentiveness, competence and reciprocity that may lead to the creation of an unprecedented care practice and to the impossibility of scenario comparisons. Such type of robot is considered a “replacement” robot to which human role and responsibilities are fully delegated and a reciprocal partner for the care receiver (van Wynsberghe 2016). This, however, leaves unexplored the most interesting consequences of this scenario on CCVSD that she delegates to potentially fruitful areas of future research.

Health systems are now using a variety of care robots, from physical robots to embodied AI and chatbots (van Wynsberghe and Li 2019), whose advances are bound to have a radical impact on how we approach values. In the case of these AI-driven systems, the manner of manifestation of moral elements is not independent from actors and care practices but nonetheless suggests that there are external criteria to the relationship between caregivers and care receivers to be considered when assessing the role of care robots. This warrants the consideration of norms such as *explicability*, *autonomy*, *nonmalificence*, and *fairness*, among others, implicated in emerging autonomous systems underlying care robots, which could constitute a first step for moving beyond CCVSD.

1. **A Multi-Tiered Approach to AI Design with VSD**

The preceding section laid out what the CCVSD approach consists of and what merits it as a unique flavour of the more general VSD methodology. Rightly so, the CCVSD methodology that van Wynsberghe (2013b) formulated adapts itself to the values of care that umbrella the domain in which care robots find themselves, and this adaptability to discrete design programs is a fundamental directive of VSD (Friedman and Hendry 2019; Friedman and Kahn Jr 2002). Given that the artefacts brought into question in this paper are of a similar nature, using CCVSD as a starting point makes sense. However, where the CCVSD falls short is its almost exclusive focus on the values of care for care robots rather than symbiotically sourcing others values and principles that are directed at the autonomous systems underlying those robots. By combining CCVSD with other norms and values that are specifically geared towards autonomous systems such as AI4SG principles (Floridi et al. 2020) and the values of the HLEG AI (High-Level Expert Group on AI 2019) not only can the design of AI systems avoid doing harm, but can be directed towards social good, even beyond the deployment domain. Umbrello and van de Poel (2020) argue along these lines for an explicit orientation to the UN Sustainable Development Goals (SDGs) as a strong approximation of what can be collectively believed to be valuable societal ends (Umbrello and van de Poel 2020, 1). The following subsections describe their approach which is then preliminarily operationalized in section 5.

**4.1 VSD for AI**

Various considerations need to be taken into account when considering the design of AI systems. There is no longer doubt as to whether or not AI systems currently or will continue to have significant and lasting sociocultural, economic and ethical impacts (Baum 2016; Khakurel et al. 2018). Likewise, many of the ethical impacts that AI systems have implicated are not explicitly accounted for in the original value litanies that VSD scholars have proposed for other ICTs (Friedman and Kahn Jr. 2002; Umbrello 2019). Values of the HLEG AI (High-Level Expert Group on AI 2019), for example, provide excellent starting points for considering values that are markedly implicated by AI systems. Having litanies of AI specific values are useful for ensuring a certain level of top-down alignment when engaging in AI design programs, despite almost certain need for bottom-up stakeholder engagement and value elicitations to make AI alignment robust and holistic (Umbrello and van de Poel 2020).

 In doing so, the design of these types of AI systems need not only avoid doing harm, but also contribute to social good. Umbrello and van de Poel (2020) propose the UN Sustainable Development Goals (SDGs) as a larger set of values for social good to design AI systems *for* (discussed further in the proceeding sub-section). Regardless, the approach they propose is fundamentally predicated on three sources of values (Fig. 4.): (1) avoiding harm, which should be construed as boundary conditions or design constraints, (2) are values that should be designed *for* as much as possible and can be construed as design requirements and criteria, and (3) are context-specific values that can take the form of avoiding harm and/or doing good. The three can overlap in many cases, albeit deserving different and explicit attention in the design process.

Avoiding Harm (Design Constraints)

Doing Good (Design Requirements)

Contextual Values

**Fig. 4.** Three sources of values for VSD for AI4SG

 While the third source of contextual values can differ dramatically based on many varying contextual factors, Umbrello and van de Poel (2020) argue that it is nonetheless useful to use litanies of values for (1) and (2):

* For the first tier of values that should be taken into account in any application of AI and that ensure that AI does not do not do more harm than good, they propose to have recourse to the values articulated by the HLEG AI and translated through the more concrete AI4SG norms into technical design requirements.
* The second tier of values that they actively seek to promote in order to contribute to social good through AI, they propose to use the SDGs as first-order operationalisations of what it means to contribute to social good through AI. Here the idea is that the SDGs to which an AI application contributes will be specific for that application.

(Umbrello and van de Poel 2020, 5)

The following sub-section discusses the SDGs (tier 2) with the subsequent sub-section discussing the AI4SG meanings and factors (tier 1).

**4.2 SDGs**

With the ultimate aimed geared towards the attainment of global peace, the United Nation drafted the proposal of objectives that are to be designed and implemented for a safe and sustainable future (United Nations 2018). The foundation of this proposal is built on 17 actionable sustainable development goals (SDGs) (Fig. 5). The goals are presented as being necessarily combinatory and complimentary rather than rank-ordered or prioritized. The ultimate objective is a synergistic and symbiotic approach to achieving all of these SDGs.

**Fig. 5.** United Nations Sustainable Development Goals. Source:(United Nations 2019)

 The UN’s underlying philosophical approach framed technologies in an interactional way, arguing that technologies co-vary with their societal and cultural contexts rather than purely deterministic artefacts or instrumental tools. This institutional orientation permits the SDGs to be tackled holistically rather than haphazardly and likewise envisions technologies not only as a potential problem exacerbating these SDG issues, but also as potential solutions (Umbrello and van de Poel 2020). Umbrello and van de Poel (2020) use the SDGs as a higher-order source of values in the VSD for AI systems (alongside others such as those by the HLEG AI as mentioned above) given that there is a marked global orientation to a set of common goals. Given that technology is explicit as a key force toward the exacerbation as well as amelioration of the issues that the SDGs are proposed to address, they provide a salient set of higher-order guidelines to design *for*.

**4.3 AI for Social Good Factors**

Umbrello and van de Poel (2020) argue that the most comprehensive and streamlined condensation of the AI4SG factors are those more recently produced by Floridi et al. (2020). The seven factors that are particularly relevant for the design of AI towards social good are: (1) *falsifiability and incremental deployment*; (2) *safeguards against the manipulation of predictors*; (3) *receiver-contextualized intervention*; (4) *receiver-contextualized explanation and transparent purposes*; (5) *privacy protection and data subject consent*; (6) *situational fairness*; and (7) *human-friendly semanticisation* (Floridi et al. 2020, 3). Like the SDGs, although these seven factors are discussed discretely, they are nonetheless co-dependent and co-vary with one another, making them inextricably linked for AI4SG to be achieved. Umbrello and van de Poel (2020) argue that the:

seven factors each relate, in some way, to at least one of the four ethical principles that EU High-Level Expert Group on AI lays out: respect for human autonomy, prevention of harm, fairness and explicability. This mapping on to the more general values of ethical AI is not insignificant, any divergences from these more general values has potentially deleterious consequences. What the seven factors are meant to do then is to specify these higher-order values into more specific *norms* and design requirements (Umbrello and van de Poel 2020, 8) [Fig. 6].

Human Autonomy

*Receiver-contextualized intervention (AI4SG #3)*

*Privacy protection and data subject consent (AI4SG #5)*

*Human-friendly semanticisation (AI4SG #7)*

Prevention of Harm

*Falsifiability and incremental deployment*

*(AI4SG #1)*

*Privacy protection and data subject consent (AI4SG #5)*

Fairness

*Safeguards against the manipulations of predictors (AI4SG #2)*

*Situational fairness (AI4SG #6)*

Explicability

*Receiver-contextualized explanation and transparent purposes (AI4SG #4)*

*Human-friendly semanticisation (AI4SG #7)*

**Fig. 6.** Relationship between higher-order values of the EU HLEG on AI and AI4SG norms.Source: (Umbrello and van de Poel 2020)

This paper does not discuss the definitions of the seven factors that Floridi et al. (2020) lay out for the purposes of length, however, Umbrello and van de Poel (2020) are explicit that the AI4SG factors function like *norms* as per van de Poel’s (2013) characterization of norms as being framed as ‘maximizing’ or ‘minimizing’ certain value or design requirements, thus bridging the gap between abstract values (e.g., HLEG AI, UN SDGs) and concrete design requirements. This is discussed further in the proceeding sub-section.

**4.4 AI4SG-VSD Process**

As mentioned, the aim of this paper is to draw from the methodology of designing AI4SG using VSD that Umbrello and van de Poel (2020) propose, using care robots for the elderly as the use case. The UN SDGs and HLEG AI principles are used as the aims for which more specific values can be derived for *doing good*, while the normative AI4SG principles are used as the basis for *avoid harm*. Figure 7 illustrates how engineers can initiate their investigations in their design program. Albeit differing from one project to another, the proposed framework provides the general outline that practitioners can follow to ensure they touch on the fundamental points proposed in this framework.

****

**Fig 7.** AI4SG-VSD design process. Source: (Umbrello and van de Poel 2020)

The four-stage iterative process that Umbrello and van de Poel (2020) composed of (1) context, (2) value identification, (3) formulating design requirements, and (4) prototyping.

**4.4.1 Context**

The sociocultural contexts in which a technology is being developed is crucial to salient design and deployment. Empirical investigations central to the VSD methodology become particularly useful here to enroll stakeholders and elicit their values to ensure more symbiotic mapping of values to design requirements (phase 3) and outcomes.

**4.4.2 Value Identification**

A starting list of values can be useful to determining a more cohesive and symbiotic set of values. Umbrello and van de Poel (2020) propose three sources of values:

(1) Values that are to be promoted by the design (i.e., UN SDGs).

(2) Values that should be respected, particularly those relevant to AI: respect for human autonomy, prevention of harm (nonmaleficence), fairness and explicability (i.e., HLEG AI).

(3) context-specific values that are not covered by 1) and 2) but which derive from the analysis of the specific context in the first phase, in particular values held by stakeholders (e.g., emotional attachment in the context of care robot design).

It is during this phase that the values considered are interpreted and defined as per the methods of conceptual investigations specific to VSD. A normative approach to upholding these values throughout the design process becomes explicit.

**4.4.3 Formulating Design Requirements**

Using the values derived in the previous two steps, more concrete design requirements must be distilled from those values. The values hierarchy illustrated in Figure 2 is one such method for undertaking this type of value-to-design requirement translation. The different sources of values are translated in different ways. The SDGs for example are to be considered throughout the design as being *designed for* as much as possible, and provides a higher-level aim for design. Whereas the HLEG AI principles are construed as boundary conditions or constraints that provide what can be understood as the minimum necessary conditions for acceptable design. Regarding the context for design, stakeholder elicitations and *a priori* value lists based on context provide an important way for how any of the uncovered values are translated into design requirements. VSD has several established methods for undertaking this type of translation from stakeholder elicitations and other empirical investigations of stakeholder values such as value scenarios (Nathan et al. 2007), value sketches (Woelfer et al. 2011), value-oriented coding manual (Kahn Jr et al. 2003), value hierarchies (Longo et al. 2020; van de Poel 2013), value-oriented mock-up, prototype, or field deployment (Czeskis et al. 2010), value dams and flows (Denning et al. 2013), and value sensitive action-reflection model (Yoo et al. 2013).

**4.4.4 Prototyping**

Directly aligned with *value-oriented mock-up, prototype, or field deployment* that Friedman and Hendry (2019) discuss, prototyping is the fourth stage where design requirements can be tested. More specifically, it is the

development, analysis, and co-design of mockups, prototypes and field deployments to scaffold the investigation of value implications of technologies that are yet to be built or widely adopted. Mock-ups, prototypes or field deployments emphasize implications for direct and indirect stakeholders, value tensions, and technology situated in human contexts

(Friedman and Hendry 2019, 62)

The first AI4SG norm rings a similar bell saying that: “AI4SG designers should identify falsifiable requirements and test them in incremental steps from the lab to the “outside world”” (Floridi et al. 2020, 7). The point here is that unforeseen or emergent values may come into play post-deployment, despite alignment with all the requisite design requirements, norms, and values (van de Poel 2016). If such emergent factors do come into play during this step, another iteration of the four-stage cycle may be needed to integrate and align the design (Umbrello and van de Poel 2020, 15).

1. **AI4SG and Care Centered VSD**

In this section, our aim is to integrate and further extend Care Centered VSD with the AI4SG framework outlined above. In order to do that, we provide an example of a specific system that has been recently developed in healthcare contexts for the assistance of elderly people.

Our example is an AI based robot, RoBear, a bear-shaped nursing care robot that can lift and move patients in and out of beds into a wheelchair, and that can help those who need assistance to stand (Dredge 2015). For several years now, Japan has been examining how robots can be used to provide care for the elderly. With its rapidly increasing elderly population, an insufficient number of caregivers and low birth rates, this issue has become of considerable importance (Elsy 2020; Wright 2019). RoBear is the successor to heavier robots RIBA and RIBA II, which are both mentioned by van Wynsberghe (2013a) as an example of autonomous or “replacement” robots for lifting. As a prototype robot, RoBear is used for conducting research on nursing care motion and on human-robot body interaction, in tasks such as standing-up assistance, integration of force-type sensors, tactile human-machine interface, and others (Jiang et al. 2015). RoBear can work by a number of sensors, cameras and microphones and is covered with soft materials, such as tactile sensors made of rubber. The latter allow RoBear to do intensive tasks without hurting care-receivers, and so it succeeds in making physical interactions with elderly people safer and more comfortable. We illustrate the design of the RoBear prototype, albeit *ex post facto* in this case, using the framework described above (i.e., Fig. 7).

**5.1 Context**

In the RoBear case, the context of use can be understood as the motivating force beyond its development, such as the need to relieve burden on caregivers and nurses or the need to promote a more comfortable treatment to elderly people in comparison with traditional mechanical lifts. The design of systems such as RoBear should be specifically targeted at trying to balance values tensions and design away the moral overload of *prima facie* conflicting values (van den Hoven et al. 2012). The prioritization of one value instead of another is strictly dependent on the context of use, being this a hospital, a home or a nursing home. For example, the analysis on context may suggest to leave aside one value in exchange for another: values of privacy and safety may be preferable to human contact and trust in domestic settings, due to the fact that in such contexts the relation of trust between caregivers and care-receivers is already established and does not need to be prioritized (see also van Wynsberghe 2013b). Therefore, the development of systems such as RoBear may assist in resolving these kinds of tensions in a way that still reduces health risks as much as possible.

**5.2 Value Identification**

**5.2.1. Values that are to be promoted by the design, in particular deriving from the SDGs**

The design of Robear can be said to be part of a large network to support #SDG 3, Ensuring healthy lives and promoting well-being at all ages. In particular, it may promote SDG target 3.8: the achievement of universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all. As a matter of fact, accessibility, accuracy and affordability are crucial values in healthcare contexts, and should be values of ethical importance especially to designers, organizations and industries that aim to develop systems that are driven by AI and machine learning, such as the case of RoBear. RoBear may help to create a more efficient care environment for caregivers, care-receivers and the whole healthcare system, avoiding unnecessary surgical interventions on vulnerable groups and encouraging personalised and meaningful care practices sustained by other service applications even outside the original context (i.e. outside hospitals and for other practices, such as walking, bathing, socializing).

**5.2.2. Values that should be respected, in particular those values that have been identified in relation to AI: respect for human autonomy, prevention of harm (nonmaleficence), fairness and explicability**

This second level of values are values that are to be promoted especially in relation to AI.

*Respect for Human Autonomy*: we increasingly interact with autonomous decision-making systems in different domains. Such systems influence our lives in various and multifaceted ways, from shaping the context in which individual decision making occurs, to altering interactions between individuals and assumptions of democratic participation. Autonomy thus refers to the capability of agents to retain full freedom of choice, in a balance with the delegation of decisions to systems. Systems in turn should be designed in such a way as to promote autonomy, avoiding those cases in which systems’ efficacy falls short in making consistent and coherent decisions on the behalf of human users (Floridi et al. 2018). Regarding care practices in CCVSD, the degree of autonomy of systems can be evaluated through the lens of attentiveness (i.e., the capability of recognizing needs of others), competence and reciprocity in the tasks. In addition, autonomy has a direct impact on what van Wynsberghe calls systems’ ‘appearance of moral agency’ (van Wynsberghe 2016, 313), which means that systems are placed into inherently ethical contexts in which they are responsible for vulnerable groups, such as elderly people. Thus, following and expanding this last suggestion, our explicit inclusion of autonomy among these moral elements may show how RoBear’s ‘appearance of moral agency’ could be investigated first and foremost through: a) contexts analysis and b) different descriptions and understandings of care practices. These latter can indeed contribute to respect caregivers and care receivers’ autonomous choices in relation to their assistance and treatments.

*Prevention of Harm or (Nonmaleficence):* this value seeks to avoid risks and harms by the understanding of systems’ capability and limitations. In the case of RoBear, harm may occur due to the specific ways in which the system is used, particularly its modality to promote the well-being and safety of users at the cost of downplaying their need for human contact and autonomy. However, there exist other non-trivial ways in which risks and harm may occur. Indeed, systems such as RoBear can raise privacy concerns, which refer to the practices used to store, archive, collect data and to monitor care receivers. In AI driven systems what can be discussed and problematized is the quality of the training data and the reliability of the algorithms used for predictions. In such cases, the delegation of tasks to systems may also lead to other risks: the disappearance of certain medical and caregiver professions, and the reallocation of expertise and responsibilities in healthcare systems.

Compared to the CC Framework, nonmaleficence can be subsumed under the value of competence, which asserts systems’ capability and limitations in executing a task. These systems capabilities may include safety, efficiency, quality of task execution, force feedback, tactile perception and other capabilities according to van Wynsberghe (2012, 111). Nevertheless, values such as efficiency and privacy are quite obfuscated in CCVSD, since these are considered as values exclusively driven by either consequences or duties and that lack the ability to take into consideration the overall development of users’ ethical characters (van Wynsberghe 2016). Conversely, we can arguably include them in the list of moral elements, especially for their relevance in increasingly AI driven systems.

*Fairness*: according to Floridi et al. (2018), the value of fairness can be framed as justice and can be defined in a tripartite way: (1) Using AI to correct past wrongs such as eliminating unfair discrimination; (2) Ensuring that the use of AI creates benefits that are shared (or at least sharable); (3) Preventing the creation of new harms, such as the undermining of existing social structures. The value of fairness in relation to care practices can be put in place through the allocation of healthcare resources and services on the basis of objective and fair health related factors. This should also imply the values of accessibility and affordability, which are strictly aligned to SDG #3 as we noted above (5.2.1).

*Explicability*: this value means that AI systems should be intelligible and not opaque, and there should be at least one agent that can be considered accountable for how the system works (Floridi et al. 2018). In AI contexts, this raises questions about public accountability gaps, which beyond the issues of information disclosure and visibility, point out that we need modalities to make systems explainable and understandable to the users and stakeholders at large (Mecacci and de Sio 2019; Pasquale 2017).

In van Wynsberghe’s works explicability is not explicitly mentioned, however it may be linked to the general notion of trust, which is an “hybrid event between the human caregiver and the robot” (van Wynsberghe 2013a, 428). Trust is closely aligned with the value of responsibility (van Wynsberghe 2012, 2016) insofar as they are equated to the capacity to be held accountable and liable, but the former is never systematically introduced and discussed among the moral elements in the CC framework. Due to their higher degree of autonomy, AI systems replace caregivers and can introduce new forms of attentiveness and competence that may lead in the long run to the establishment of trusting bonds. In this scenario, trust is not a matter between two actors interacting, but between multiple actors, which include also groups, institutions and policy makers behind technologies. Thus, if we want to further explore trust and its link with the value of explicability, the predominant focus of our approach should not be on the reciprocal engagement of the robot compared to the human, but on how the “forum” - employing van Wynsberghe’s nomenclature - of trust has changed from the one associated to the traditional relationship between caregivers and care receivers to a new and unprecedented one.

**5.2.3 Context-specific values that are not covered by 1) and 2) by which derive from the analysis of the specific context in phase, in particular values held by stakeholders.**

We showed how the development of RoBear has been a response to societal changes, the increase of elderly population and the scarcity of human professionals such as caregivers and nurses. Many of the values and side effects in relation to the deployment of RoBear have been already discussed, such as the values of *general health and well-being* (under 5.2.1) and *autonomy*, *non-maleficence, fairness as justice,* and *explicability* (under 5.2.2). For example, RoBear may elicit a false sense of security related to the value of health, or it may increase the dependency on technological systems to the detriment of users’ autonomy. Similarly, it may lead to discriminations regarding its accessibility and affordability and thus its fairness. However, there are other values that can less clearly be subsumed under the two levels outlined above. A contextual analysis serves to elicit such classes of values, which are related to stakeholders’ values and preferences (see 5.1).

One of the possible values at play in this context-specific level may be the value of emotional attachment, which is strictly aligned to trust. Current robots are said to be incapable of giving the ‘‘real compassion and empathy or understanding’’ that is proper of human companionship (Sharkey 2014). However, RoBear can be interpreted as an attempt of semi-reciprocal interaction, also due to its friendly and nonthreatening appearance, its soft sensors and its auditory and visual capabilities. Even from a CCVSD perspective, systems such as RoBear should be designed in a way that promotes the value of responsiveness or reciprocity (van Wynsberghe 2016). Various kinds of transfer and lifting assistant robots have already been proposed previously. However, designing comfortable transfer motions for individuals has not been intensively studied due to “the difficulty of modelling physical human-robot interactions in conjunction with user preference” (Ding et al. 2017, 74). Therefore, at this stage VSD methodological tools such as envisioning cards (Friedman and Hendry 2012), which are designed to evoke considerations and discussion, may help in reconstructing stakeholders’ preferences and values. They are also useful for the identification of new values that point out which behavioural impacts these systems may have on the broader context. It is of crucial importance also for the future development of care systems and their long-term and indirect effects that these context-specific values are understood and then translated into design requirements.

**5.3 Formulating Design Requirements**

Although various instruments and methods characteristic of the VSD methodology can be adopted to help designers to distill and formalize the necessary design requirements for any given design, the values hierarchy (Fig. 2) is particularly useful as a way to illustrate and trace design requirements from norms and value, and vice versa. Fig. 8 is one example of how to visualize the translation of higher-level values, through AI4SG norms and into technical design requirements.

**Fig 8.** Translating human autonomy to design requirements through AI4SG norms.

 Fig. 8 visualizes the more abstracted value of *respect for human autonomy*. It is translated through various AI4SG norms (3, 5, and 7) which are illustrated in Fig. 6 and into more concrete design requirements. Umbrello and van de Poel (2020) construe the AI4SG factors as norms rather than as more abstract values in design given the way that Floridi et al. (2020) describe them, that being, as imperatives for designers. It should be noted that any given context of use will naturally change any given combination of values, norms and subsequent design requirements. Fig. 8 is just one illustration of how this process is done, and even given the use case of RoBear, the value of *respect for human autonomy* need not be defined as such. To this end, “there is no exclusive nor exhaustive route for satisfying a value translation” (Umbrello and van de Poel 2020, 19). Both *prevention of harm* and *explicability* for example overlap with *respect for human autonomy* given that they implicate AI4SG norms 5 and 7 respectively (see Fig. 9). For this reason, these values mutually co-vary and, in many cases, should be used to operationalize each other. Design requirements translated from the value of *explicability,* for example, can be used as a route for engaging with and operationalizing the value of *situational fairness* and *prevention of harm*.

**Fig 9.** Translating explicability to design requirements through AI4SG norms.

Functionally speaking, the AI4SG norms are apt to avoiding *most* of the harms that may emerge as a consequence of AI systems deployment, but not all. However, this does not entail that by doing so such systems will positively contribute to social good. “Global beneficence”, that is, contributing to global good comes part and parcel with engaging with higher-level values like the *actual* operationalization of the SDGs as discussed above. For this reason, this multi-tiered approach to VSD is adopted here, because by marrying that values specific to AI with stakeholder values and their treatment towards SDG attainment by means of AI4SG norms, then most of the threats posed by the ‘ethical white-washing’ that comes as a consequence of the sanctioning of AI systems that do not respect these central ethical principles can be mitigated.

**5.4 Prototyping**

As mentioned in section 4.4.4, prototyping requires building mock-ups for the technologies in question based on the design requirements distilled in the preceding sections. In doing so, the technologies move out of the conceptual space and *in situ*, thus implicating the values of the stakeholders who populate that area. Penetration testing type activities take place to determine if any of the design decisions prove to be recalcitrant that were not in the design space as well as to determine if there are any emerging technical issues as well as ethical issues that were not foreseen during the previous stages of the VSD methodology (van de Poel 2020). Because the technology is in limited deployment rather than ubiquitous rollout, it maintains the ability to be removed from *in situ* spaces and brought back into the design sphere in which more iterations of the VSD approach can be undertaken to account for those unforeseen and emerging issues. Unlike the examples employed by Umbrello and van de Poel (2020) regarding the SARS-CoV-2 contact-tracing app that have incentives for immediate deployment, RoBear instead is not motivated by similar immediate issues. Smaller scale testing and direct stakeholder mock-ups permits technologies like RoBear to be more capable of affirming the values that it implicates and avoiding the harms that could have emerged if deployed differently.

 In sum, prototyping should not be understood as the narrower testing of the technical aspects of a technologies functioning, but also the social and ethical effects that it has and that emerge from its limited deployment in the field of use. The RoBear robot is a particularly apt case, while values such as *respect for human autonomy* can be affirmed through various design decisions such as automatic receiver recognition allowing RoBear devices to communicate with across different levels of abstraction between agents such as medical staff and patients, other values such as *situational fairness* may require *in situ* insights with stakeholders to understand the more nuanced behaviors that are effected and affect the salient design of more abstract values likes *situational fairness*. For this reason, it may be more effectual to begin with and gradually scale up small-scale mock-ups to ensure that the progressive iterations of the methodology sufficiently account for the changing and emerging values of *in situ* deployment are accounted for in design (c.f., van de Poel 2020). This type of approach can aid designers to discover new values, thus prompting another iteration of the cycle that might have not been triggered otherwise.

1. **Conclusions**

The values enrolled in the design of care robots are implicated across various levels of abstraction. This paper aimed to begin with the values central to care in the design of care robots for the elderly, and expand on them by introducing other sources of values such as the UN SDGs as well as values and norms specific to AI systems. In doing so, we adopt the value sensitive design approach and modify it according to Umbrello and van de Poel’s (2020) framework for a multi-tiered approach to the ethical design of AI systems. Drawing on the example of the RoBear care robot, this paper illustrated *prima facie* how designers can begin to direct their practices towards designing AI *for* social good.

**References**

Albrechtslund, A. (2007). Ethics and technology design. *Ethics and information technology*, *9*(1), 63–72.

Baum, S. D. (2016). On the promotion of safe and socially beneficial artificial intelligence. *AI and Society*, (July), 1–9. https://doi.org/10.1007/s00146-016-0677-0

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12*, 1125. https://doi.org/10.1145/2207676.2208560

Czeskis, A., Dermendjieva, I., Yapit, H., Borning, A., Friedman, B., Gill, B., & Kohno, T. (2010). Parenting from the pocket: Value tensions and technical directions for secure and private parent-teen mobile safety. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (p. 15). ACM.

Denning, T., Kohno, T., & Levy, H. M. (2013). A framework for evaluating security risks associated with technologies used at home. *Communications of the ACM*, *56*(1). https://doi.org/10.1145/2398356.2398377

Ding, M., Matsubara, T., Funaki, Y., Ikeura, R., Mukai, T., & Ogasawara, T. (2017). Generation of comfortable lifting motion for a human transfer assistant robot. *International Journal of Intelligent Robotics and Applications*, *1*(1), 74–85. https://doi.org/10.1007/s41315-016-0009-z

Dredge, S. (2015). Robear: the bear-shaped nursing robot who’ll look after you when you get old. *The Guardian*. https://www.theguardian.com/technology/2015/feb/27/robear-bear-shaped-nursing-care-robot. Accessed 18 September 2020

Elsy, P. (2020). Elderly care in the society 5.0 and kaigo rishoku in Japanese hyper-ageing society. *Jurnal Studi Komunikasi*, *4*(2), 435–452.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(December), 1–24. https://doi.org/10.12932/AP0443.32.4.2014

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). Designing AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics*, 1–26. https://doi.org/10.1007/s11948-020-00213-5

Friedman, B., & Hendry, D. G. (2012). The envisioning cards: A toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the 30th International Conference on Human Factors in Computing Systems - CHI ’12* (pp. 1145–1148). https://doi.org/10.1145/2207676.2208562

Friedman, B., & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: Mit Press.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A Survey of Value Sensitive Design Methods. *Foundations and Trends® in Human–Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Friedman, B., & Kahn Jr., P. H. (2002). Value sensitive design: Theory and methods. *University of Washington Technical*, (December), 1–8. https://doi.org/10.1016/j.neuropharm.2007.08.009

Friedman, B., & Kahn Jr., P. H. (2003). Human values, ethics, and design. In J. A. Jacko & A. Sears (Eds.), *The Human-computer Interaction Handbook* (pp. 1177–1201). Hillsdale, NJ, USA: L. Erlbaum Associates Inc. http://dl.acm.org/citation.cfm?id=772072.772147

Friedman, B., Kahn Jr., P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7844-3\_4

Friedman, B., & Kahn Jr, P. H. (2002). Human values, ethics, and design. In *The human-computer interaction handbook* (pp. 1209–1233). CRC Press.

High-Level Expert Group on AI. (2019). Ethics guidelines for trustworthy AI | Shaping Europe’s digital future. *European Commission*. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Accessed 25 February 2020

Jiang, C., Ueno, S., & Hayakawa, Y. (2015). Optimal control of non-prehensile manipulation control by two cooperative arms. In *2015 International Conference on Advanced Mechatronic Systems (ICAMechS)* (pp. 533–537). IEEE.

Kahn Jr, P. H., Friedman, B., Freier, N., & Severson, R. (2003). Coding manual for children’s interactions with AIBO, the robotic dog–The preschool study. *University of Washington CSE Technical Report 03-04*, *3*.

Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., & Zhang, W. (2018). The Rise of Artificial Intelligence under the Lens of Sustainability. *Technologies* . https://doi.org/10.3390/technologies6040100

Kuttoor, R. (2020, July 2). Kerala village deploys robot nurses for patient care. *The Hindu*. https://www.thehindu.com/news/national/kerala/kerala-village-deploys-robot-nurses-for-patient-care/article31966585.ece. Accessed 2 July 2020

Le Dantec, C. A., Poole, E. S., & Wyche, S. P. (2009). Values As Lived Experience: Evolving Value Sensitive Design in Support of Value Discovery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1141–1150). New York, NY, USA: ACM. https://doi.org/10.1145/1518701.1518875

Longo, F., Padovano, A., & Umbrello, S. (2020). Value-Oriented and Ethical Technology Engineering in Industry 5.0: A Human-Centric Perspective for the Design of the Factory of the Future. *Applied Sciences*, *10*(12), 4182. https://doi.org/10.3390/app10124182

Macdonald, K. (2020, July 1). New robot technology to fight Covid care isolation. *BBC News*. https://www.bbc.com/news/uk-scotland-53241556. Accessed 2 July 2020

Manders-Huits, N. (2011). What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics*, *17*(2), 271–287. https://doi.org/10.1007/s11948-010-9198-2

Mecacci, G., & de Sio, F. S. (2019). Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. *Ethics and Information Technology*, 1–13.

Mordoch, E., Osterreicher, A., Guse, L., Roger, K., & Thompson, G. (2013). Use of social commitment robots in the care of elderly people with dementia: A literature review. *Maturitas*, *74*(1), 14–20.

Nathan, L. P., Klasnja, P. V, & Friedman, B. (2007). Value Scenarios: A Technique for Envisioning Systemic Effects of New Technologies. In *CHI ’07 Extended Abstracts on Human Factors in Computing Systems* (pp. 2585–2590). New York, NY, USA: ACM. https://doi.org/10.1145/1240866.1241046

Pasquale, F. (2017). Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society. *Ohio St. LJ*, *78*, 1243.

Pirni, A., Esposito, R., Carnevale, A., & Cavallo, F. (2017). Sostenibilità etica dei personal care robot. Linee per un inquadramento preliminare. *Nuova Corrente, 159* (1), 133–151.

Sharkey, A. (2014). Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics and Information Technology*, *16*(1), 63–75.

Sharkey, A., & Sharkey, N. (2012). Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology*, *14*(1), 27–40. https://doi.org/10.1007/s10676-010-9234-6

Sorell, T., & Draper, H. (2014). Robot carers, ethics, and older people. *Ethics and Information Technology*, *16*(3), 183–195.

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, *16*(2), 141–161.

Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethic of care*. Psychology Press.

Tronto, J. C. (2010). Creating caring institutions: Politics, plurality, and purpose. *Ethics and social welfare*, *4*(2), 158–171.

Umbrello, S. (2019). Beneficial Artificial Intelligence Coordination by Means of a Value Sensitive Design Approach. *Big Data and Cognitive Computing*, *3*(1), 5. https://doi.org/10.3390/bdcc3010005

Umbrello, S. (2020). Meaningful Human Control over Smart Home Systems: A Value Sensitive Design Approach. *Humana.Mente Journal of Philosophical Studies*, *13*(37), 40–65.

Umbrello, S., & van de Poel, I. (2020). *Mapping Value Sensitive Design onto AI for Social Good Principles*. https://doi.org/10.13140/RG.2.2.34841.85607

United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. , Pub. L. No. A/RES/70/1 (2018). United Nations. https://sustainabledevelopment.un.org/post2015/transformingourworld. Accessed 26 January 2020

United Nations. (2019). Sustainable development goals. *GAIA*. https://doi.org/10.14512/gaia.28.2.1

Vallor, S. (2011). Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century. *Philosophy & Technology*, *24*(3), 251. https://doi.org/10.1007/s13347-011-0015-x

van de Poel, I. (2013). Translating Values into Design Requirements. In D. P. Michelfelder, N. McCarthy, & D. E. Goldberg (Eds.), *Philosophy and Engineering: Reflections on Practice, Principles and Process* (pp. 253–266). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7762-0\_20

van de Poel, I. (2016). An Ethical Framework for Evaluating Experimental Technology. *Science and Engineering Ethics*, *22*(3), 667–686. https://doi.org/10.1007/s11948-015-9724-3

van de Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines*. https://doi.org/10.1007/s11023-020-09537-4

van den Hoven, J. (2007). ICT and Value Sensitive Design. In P. Goujon, S. Lavelle, P. Duquenoy, K. Kimppa, & V. Laurent (Eds.), *The Information Society: Innovation, Legitimacy, Ethics and Democracy In honor of Professor Jacques Berleur s.j.: Proceedings of the Conference ``Information Society: Governance, Ethics and Social Consequences’’, University of Namur, Belgium 22--23 May 20* (pp. 67–72). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-72381-5\_8

van den Hoven, J., & Jacob, K. (2013). *Options for Strengthening Responsible Research and Innovation*. https://doi.org/10.2777/46253

van den Hoven, J., Lokhorst, G. J., & van de Poel, I. (2012). Engineering and the Problem of Moral Overload. *Science and Engineering Ethics*, *18*(1), 143–155. https://doi.org/10.1007/s11948-011-9277-z

van den Hoven, J., Vermaas, P. E., & van de Poel, I. (2015). *Handbook of ethics, values, and technological design: Sources, theory, values and application domains*. (J. van den Hoven, P. E. Vermaas, & I. van de Poel, Eds.)*Springer Reference*. Springer Netherlands. https://doi.org/10.1007/978-94-007-6970-0

van Lente, H., Swierstra, T., & Joly, P. B. (2017). Responsible innovation as a critique of technology assessment. *Journal of Responsible Innovation*, *4*(2), 254–261. https://doi.org/10.1080/23299460.2017.1326261

van Wynsberghe, A. (2012). *Designing Robots With Care: Creating an Ethical Framework for the Future Design and Implementation of Care Robots*. University of Twente. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract

van Wynsberghe, A. (2013a). Designing Robots for Care: Care Centered Value-Sensitive Design. *Science and Engineering Ethics*, *19*(2), 407–433. https://doi.org/10.1007/s11948-011-9343-6

van Wynsberghe, A. (2013b). A method for integrating ethics into the design of robots. *Industrial Robot: An International Journal*, *40*(5), 433–440. https://doi.org/10.1108/IR-12-2012-451

van Wynsberghe, A. (2015). *Healthcare robots: Ethics, design and implementation*. Farnham, UK: Routledge. https://www.routledge.com/Healthcare-Robots-Ethics-Design-and-Implementation/Wynsberghe/p/book/9781472444332

van Wynsberghe, A. (2016). Service robots, care ethics, and design. *Ethics and Information Technology*, *18*(4), 311–321. https://doi.org/10.1007/s10676-016-9409-x

Vandemeulebroucke, T., de Casterlé, B. D., & Gastmans, C. (2018). The use of care robots in aged care: A systematic review of argument-based ethics literature. *Archives of gerontology and geriatrics*, *74*, 15–25.

Woelfer, J. P., Iverson, A., Hendry, D. G., Friedman, B., & Gill, B. T. (2011). Improving the Safety of Homeless Young People with Mobile Phones: Values, Form and Function. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1707–1716). New York, NY, USA: ACM. https://doi.org/10.1145/1978942.1979191

Wright, J. (2019). Robots vs migrants? Reconfiguring the future of Japanese institutional eldercare. *Critical Asian Studies*, *51*(3), 331–354.

Yoo, D., Huldtgren, A., Woelfer, J. P., Hendry, D. G., & Friedman, B. (2013). A value sensitive action-reflection model: evolving a co-design space with stakeholder and designer prompts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 419–428).