



## The ethical issues of social assistive robotics: A critical literature review

Júlia Pareto Boada<sup>a,b,\*</sup>, Begoña Román Maestre<sup>b</sup>, Carme Torras Genís<sup>a</sup>

<sup>a</sup> Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC, C/Llorens i Artigas 4-6, 08028, Barcelona, Spain

<sup>b</sup> Facultat de Filosofia, Universitat de Barcelona (UB), C/Montalegre 6, 08001, Barcelona, Spain

### ARTICLE INFO

#### Keywords:

Artificial intelligence  
Care  
Ethics  
Healthcare  
Human-robot interaction  
Social assistive robotics  
Justice  
Well-being

### ABSTRACT

Along with its potential contributions to the practice of care, social assistive robotics raises significant ethical issues. The growing development of this technoscientific field of intelligent robotics has thus triggered a widespread proliferation of ethical attention towards its disruptive potential. However, the current landscape of ethical debate is fragmented and conceptually disordered, endangering ethics' practical strength for normatively addressing these challenges. This paper presents a critical literature review of the ethical issues of social assistive robotics, which provides a comprehensive and intelligible overview of the current ethical approach to this technoscientific field. On the one hand, ethical issues have been identified, quantitatively analyzed and categorized in three main thematic groups. Namely: Well-being, Care, and Justice. On the other hand –and on the basis of some significant disclosed tendencies of the current approach–, future lines of research and issues regarding the enrichment of the ethical gaze on social assistive robotics have been identified and outlined.

### 1. Introduction

Plausibly, the branch of social robotics devoted to the development of assistive robots is the one that most clearly embodies the European ideal of an intelligent technology at the service of humans' well-being [1]. Indeed, engineering artificial intelligence (AI) tools for coping with the ontological condition of human vulnerability seems to be the highest exponent of a human-centric technology model aimed at prioritizing the empowerment of individuals for a higher quality of life.

This field is commonly known as “socially assistive robotics” (SAR). In general terms, it is focused on providing artificial intelligent robotic systems for aiding end-users with (physical or cognitive) special needs<sup>1</sup> in their daily activities [2]. These vulnerable subjects include elderly adults, individuals with dementia, children with autistic spectrum disorders, convalescent patients, and people with other kinds of functional diversity needs. This places SAR as a technoscientific field generally aimed at contributing to the practice of care [3]. Specifically, socially assistive robots (SARs) are designed to support tasks in a broad range of care activities –like healthcare, physical and cognitive rehabilitation or therapy, domestic daily life and special education– and thus to be used in different settings –hospitals, elder-care facilities, homes and schools–.

However, its defining particularity does not only lie in the type of tasks that they undertake (which are ultimately related to assistance<sup>2</sup>). It also consists in the way in which they carry them out: by means of socially interacting with humans –in virtue of which they can assume roles as coaching, motivating or providing company in ecosystems of care–. This is why SARs tend to be taxonomically understood as an intersection set between assistive robots, focused on assistive functions, and socially interactive robots, intended to interact with the human in a social way [4].

Despite the fact that “socially assistive robotics” has become the prevailing terminology, there are sound reasons to designate this field in broader terms as “social assistive robotics”. First, the former label may not be inclusive enough [5]: according to its original meaning [4], it leaves out the class of robots that, even if assisting through social interaction, may also involve some form of physical contact with users.<sup>3</sup> Second, the term “socially” is tautological, insofar as the so-called “socially assistive robots” are a subset of social robots. By definition, a social robot is an artificial intelligent entity (humanoid or other) that interacts in an “interpersonal manner” to achieve the predefined and domain-specific goals of the practical context in which it serves [6]. Thus, the concept of “social assistive robotics/robot” is comprehensive

\* Corresponding author. Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC, C/Llorens i Artigas 4-6, 08028, Barcelona, Spain .

E-mail addresses: [jpareto@iri.upc.edu](mailto:jpareto@iri.upc.edu) (J.P. Boada), [broman@ub.edu](mailto:broman@ub.edu) (B.R. Maestre), [torras@iri.upc.edu](mailto:torras@iri.upc.edu) (C.T. Genís).

<sup>1</sup> Although ultimately serving the needs of vulnerable end-users, SAR products are commonly conceived as tools for caregivers (primarily).

<sup>2</sup> For an insightful remark on the narrowed conception of assistance that is here at stake, see Ref. [90].

<sup>3</sup> This is the case, for instance, of the cognitive robotic system for assisting mild dementia patients in a brain-training exercise developed by Ref. [91].

enough with the main acceptance that the term “socially” originally meant to stress about these intelligent systems: the act of assisting, primarily, through social (rather than physical) interaction. Therefore, we will employ the concept of “social assistive robotics” to refer to the branch of social robotics focused on assistance and “social assistive robots” to designate its products,<sup>4</sup> alongside the widespread acronym SAR and SARs, respectively.

So far, the primary domains where SAR research is being applied belong to the healthcare field [7]. SAR mostly supports tasks aimed at helping vulnerable people in processes of restoration or health maintenance. Such tendency is in line with the European interest in healthcare as a major application area for robotics, AI, and digitalization development [8,9]. Although SARs’ widespread implementation is (still) far from being a reality, significant European research initiatives [10] and pilot projects already being launched evince institutional prospects to incorporate these technologies within (healthcare) assistive contexts. An example of this is Barcelona’s City Council pilot project to improve the quality of life of senior citizens through SARs [11]. This European goal can be explained as a response to the increasing populations with special needs [12]; a challenging phenomenon to a great extent due to the increase of ageing populations worldwide [13], for which SARs are seen as a promising technological solution [9,14]. Indeed, aged care is becoming a SARs’ central domain of application [15], which is why current ethical reflection on these intelligent robots focuses predominantly on their use in the context of elderly care.

In spite of its expectable significant contributions to care practice, SAR raises a considerable number of ethical challenges, given its disruptive potential for the organization (and conception) of our community life. Therefore, alongside its development, SAR is becoming a focus of growing ethical attention [15], already constituting a distinctive area of reflection within Roboethics [16]. Nevertheless, despite the widespread proliferation of ethical discussion on SAR, it is difficult to obtain a clear and comprehensive overview of the current debate. There is no way to straightforwardly get a complete outlook of the different ethical issues addressed in academic literature, nor to make outright sense of this fragmented ethical landscape –that is, to get an intelligible global picture of the concerns responding to some conceptual order.

There are several reasons behind the tangled ethical approach to SAR:

- (1) Ethical reflection is dispersed throughout a heterogeneous body of literature. In part, this has to do with the inherent complexities of social robotics as an object of ethical reflection [17]. For instance, the diversity of spheres of action involved in this technoscientific field implies the existence of multiple roots for SAR ethical issues, which may differ according to each developmental stage. Likewise, since the instrumental character of social robotics requires the contextualization of ethical reflection in the specific field of technological application, it fosters a diversification of the total body of SAR identifiable ethical issues into literature that concerns itself with different practical fields of

assistance. In line with the current development trend, ethical reflection on SARs is mostly developed in healthcare robotics literature [18].

- (2) There is a generalized terminological ambiguity regarding SARs, since they are very often designated with other concepts such as “care”, “medical”, or just “social robots” [19,20]. Even if used interchangeably, these terms have relevant connotative differences, so this tendency hinders mapping the specific ethical issues of SAR.
- (3) A general and thorough synthesis of SAR ethical issues so far addressed in literature is still lacking. Even though a valuable systematic review of ethics literature on SAR has been already conducted [21], it is narrow in terms of considered publications (fully elaborated argument-based literature), end-users (the elderly), and technological development stage (use).

This current scenario of the ethical debate on SAR is certainly problematic, since it may devalue the practical strength of ethics and its relevance for the normative guidance of technology’s disruptive force already from the early and throughout all different stages of development.<sup>5</sup> To engage in a (much needed) fruitful and inclusive ethical dialogue [22] for the legitimate reconfiguration of human activity through SAR developments, it is necessary to start by putting in common, conceptually ordering and analyzing the ethical issues arguably associated with this branch of social robotics. Besides, having a whole understandable picture of the different ethical issues can foster the identification of potential future lines of research on SAR for human well-being.

To address this need, we conducted a literature review of SAR ethical issues, which had a twofold goal. On the one hand, to identify and analyze the different existing ethical concerns on SAR, in order to obtain an informed knowledge of the current landscape of scholar ethical reflection on this field of robotics. Thus, our first objective was to get a comprehensive view of the state of the art of ethical thinking on SAR, by seeking to answer two main questions: which are the ethical issues generally associated with SAR, and which of these are the most frequently addressed ones?

On the other hand, and on the basis of the detected tendencies in the academic ethical debate, the second goal was to identify and outline some lines of thought and issues that need to be developed further in future research to deepen and complete the ethical approach to SAR. Accordingly, we structured our research in two main stages: an extensive descriptive one –for which a quantitative and qualitative analysis was undertaken– and a germinal critical-reflective one devoted to assessing the literature review’s results and sketching future directions for enriching ethical reflection on SAR.

## 2. Methodology

To identify and examine the main ethical issues associated with SAR, we conducted a literature review through the international bibliographic database of scientific journals Scopus<sup>6</sup> in July 2020. We focused

<sup>4</sup> A remark on the object of our review is here in order. Although some conversational AI-based systems which are already in the market, such as Amazon Alexa or Google Home, could develop some assistive tasks close to those of SARs, there are at least two significant differences between both kinds of artifacts that make it necessary to draw a distinction between them and grant a specific ethical attention to the latter, as done in this work. First, the embodiment of AI in a robotic entity entails an artificial agency whose ability to perform tasks within the physical world raises specific ethical challenges besides those posed by virtual conversational agents. Second, and more importantly, whereas the purpose of an AI conversational software product such as Amazon Alexa or Google Home can be defined by its user, SARs are conceived for and implemented within a practical context with its own (accepted and objective) ends. This requires of a very particular kind of ethical reflection, namely an exercise of applied ethics.

<sup>5</sup> The accelerated pace of technological development makes it highly important to foster not (only) a *post-facto* ethical reflection –thus focused on already designed products and its use–, but a proactive ethics instead [92], engaged in all the different levels of the process and committed to key questions arising from the very same moment of conception, such as the teleological ones (why and what for). Whereas the former kind of ethical thinking would foster narrowing ethics to an exercise of impact assessment, the latter is crucial for the so-called Responsible Research and Innovation [18], and it is an indispensable element for the recently advocated “positive ethics” [93].

<sup>6</sup> Although the informational source of this literature review is limited to Scopus, this database is one of the major and most comprehensive ones [94]. This is why we consider our findings to be broad and representative enough for an overview of the current state of ethical scholarly thinking on SAR.

exclusively on scholarly publications with the intent of offering an outlook on a well-supported ethical discussion in which concerns are grounded on up-to-date knowledge.

The bibliographical search was carried out through the following four terminological entries: (1) “ethics” AND “assistive robot\*”, (2) “ethics” AND “care robot\*”, (3) “ethics” AND “social robot\*”, (4) “ethics” AND “human-robot interaction”. The reason for this search was to broaden the scope of possible publications setting out the ethical issues of SAR, given the intersection of concepts that define the latter. First, “social assistive robot\*” constitutes a searching subset of “assistive robot\*”, so the latter offers a richer literature niche regarding our aims. Second, given that “assistance” falls under the umbrella of “care practice” and that both terms are usually used in an overlapping sense in the academic literature on robotics, “care robot\*” was a necessary search for the review’s exhaustiveness. Third, as SAR constitutes a subset of social robotics, an entry regarding this broader field was appropriate. Lastly, SARs’ distinctive feature is to support an assistive-related task through social interaction, which may have specific ethical implications that are not to be found in other assistive technologies.

The review process of this literature search (see Fig. 1), data extraction, and quantification of ethical issues was completed through five steps.

First, for each one of the four searches’ outcomes, we started by an initial screening of the publications’ titles and abstracts in order to select and compile eligible literature for later full-text reading. Our preselection criterion was based on the fact that such sections had to indicate some consideration for the ethical implications of technology within the papers’ content. To reach a broad overview on SAR ethical issues, the specific kind of SARs on which publications (might have) focused was not an exclusion criterion. Since social robotics is a quite recent technoscientific field of intelligent robotics [23], we did not apply any restrictive criterion regarding the publication period of potential papers, which oscillated, at the most, between 2004 and the first half of the year 2020. Only publications written in English were considered.

Secondly, we proceeded to a full-text reading of the preselected publications to carry out a double task: to refine the selection of relevant papers that fitted our research goal, and second, to inductively identify any ethical problems brought up by each of them and extract this information into a table.

Regarding the screening process by full-text reading, the applied inclusion criterion was that publications had to (1) address ethical issues<sup>7</sup> (2) related to social (assistive) robotics (3) with an *indirect* focus on healthcare. We interpreted the first requirement in a non-restrictive sense, meaning that the mere mention or reference to ethical issues was a criterion for inclusion. Thus, we did not limit the definitive selection of publications exclusively to papers that argumentatively engaged in ethical matters, exposing and arguing for a particular stance thereby. An enumeration or overview of ethical issues was also taken as a reason for inclusion. The second requirement responded to the fact that the object of ethical study is, taxonomically speaking, an intersection set. Thence, challenges may arise from the different aspects involved. The source of SAR ethical issues is a ramified one, rooted in the assistive dimension (and its conceptually related practice of care) and the robots’ interactive functioning. We dismissed papers that explicitly

excluded consideration to social robots (exclusion criterion 1), but not those focused on generic groups potentially involving them (as “assistive” or “care robots”). The reason for the third requirement is that we aimed at identifying a broad scope of ethical issues regarding SAR. Therefore, we did not want to narrowly restrict the search to a specific practical context of SARs’ application (aged care, nursing care, mental or physical therapeutic care); nor to limit the end-users of the assistive practice to a particular profile of vulnerable individuals. However, since ethical reflection on technology requires a contextualized approach, we opted to select the major field of healthcare as the broad indirect focus of our search, that is, as the main field of assistance. By “indirect”, we mean that even if we did not explicitly restrict the search to this field of application –thereby including papers that were unspecific about the practical context of assistance, and thus having a general focus on social settings–, we excluded publications explicitly centered on a field of application other than healthcare, like education (exclusion criterion 2). Our choice is consistent with the current research trends and expectations on SAR for healthcare, which makes it urgent to reflect on this matter ethically.

Papers dealing with research ethics were excluded (exclusion criterion 3), except those including reflection on other stages of activity regarding social (assistive) robotics. We limited the “snowball method” to an occasional use, specifically for papers where the ethical issues being introduced were directly enumerated from a secondary document included in their bibliography. For the data extraction in the table, duplicates were removed.

Thirdly, from the table of ethical issues relative to each selected publication, we proceeded to make a comprehensive index of all the different ethical concerns raised in the reviewed academic literature. Although most shared worries were exposed in the same or similar terms, we had to undertake a basic categorization to obtain a final list that was inclusive without overlapping implied connotations.

Fourthly, we compiled the indexed ethical issues on a spreadsheet alongside the corpus of selected publications (classified in four different subgroups according to their relative terminological search). This allowed us to use the index as an analytic tool for a second full-text reading of the papers. Through this, we re-identified the ethical issues they introduced, and marked them in the spreadsheet.

Lastly, each “SAR-associated” ethical issue was numerically quantified for the total corpus of the selected publications. To obtain an intelligible overall picture of the current landscape of ethical concerns on SAR, we categorized the identified issues in three thematic groups: Well-being, Care, and Justice. To that effect, we undertook a conceptual analysis on the content of the exposed issues.

### 3. Results

In our literature review, 56 publications were included for data extraction (Table 1), through which we identified a total of 26 ethical issues currently associated with SAR. These issues were very heterogeneous. Thus, for the quantitative analysis to be truly illustrative of the ethical debate on SAR, it had to be complemented with a categorization of the compiled concerns.

The heterogeneity of ethical issues has to do with the variety of angles from which it can be critically reflected upon SAR. These giving rise to different types of concerns. Some of these angles configuring the ethical approach to SAR are the following: (1) ethical perspective [21]; (2) ontological assumptions –whether the focus is on the robot as an object or a “subject”–; (3) source of concerns –robot’s particularities from which ethical problems arise, regarding both its technical elements (cameras, sensors, mobility, ...) and functionalities or roles (specific tasks, social interactivity, autonomous decision-making, ...)–; (4) contextualization of ethical reflection –whether reflection is led by ethical criteria belonging to a sole context (for instance: bioethical principles, happiness, or trust regarding the fields of healthcare, domestic life or institutions where technology is introduced), or to an intersection of

<sup>7</sup> The notions of ethical issues/challenges/concerns/problems/conflicts are indistinctively used in the current landscape of normative-oriented reflection on SAR. Here, we will embrace them all as synonyms under the broad concept of “ethical issues”. However, we consider important to caution against the indiscriminate use of the notion “dilemma”, which is also quite frequent in the ethics literature on SAR. Connotatively, this is a very limited word: it is too dualistic, since it captures an ethical issue in polarized terms. Not only is this an unusual form in which ethical challenges arise, but, most importantly, approaching reflection on SAR in these terms may impoverish the ethical scope of meaning of the issues at hand and the quality of our normative reflection regarding them.

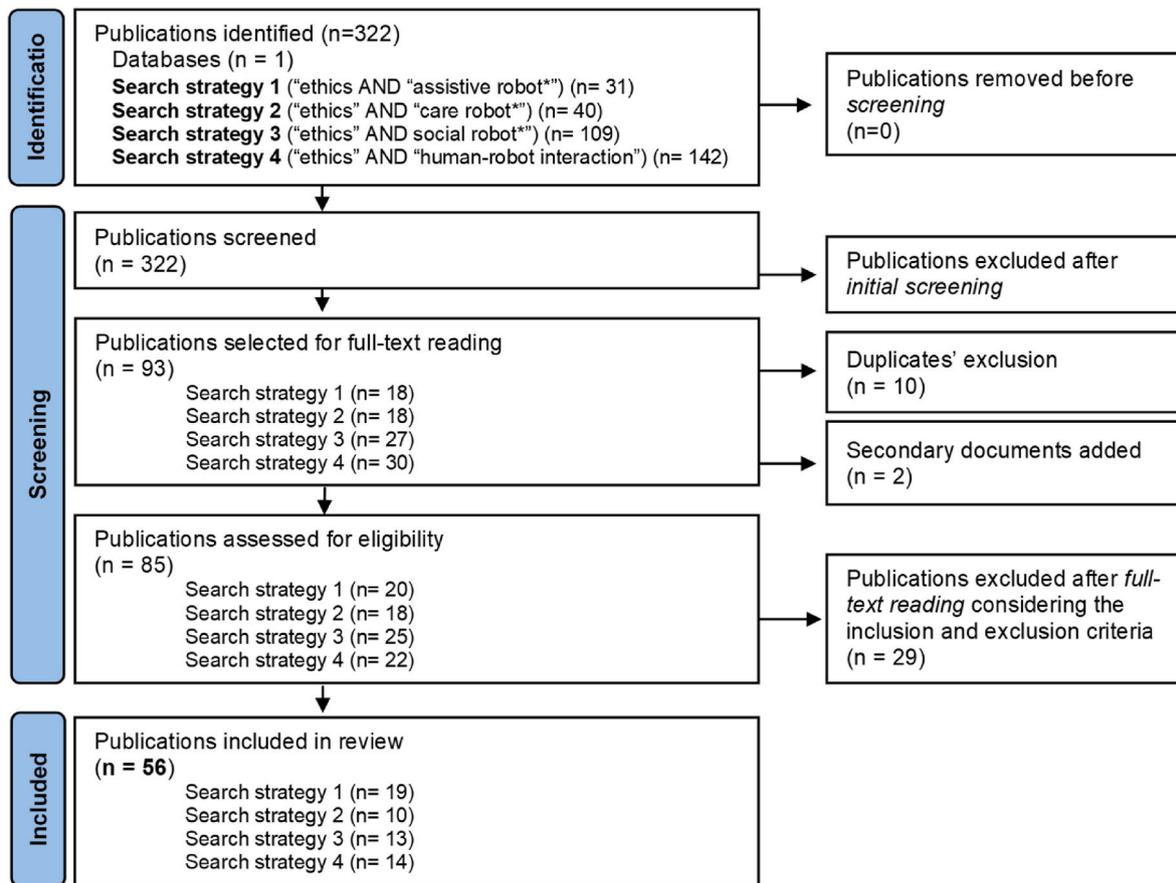


Fig. 1. Literature search and selection process. Source: Own elaboration based on PRISMA 2020 [24].

contexts instead–; (5) stage of technological development –design, research, implementation or use.

We categorized the ethical issues in three thematic groups, which we have labeled as Well-being, Care, and Justice according to the ethical dimension of human life to which these are (allegedly) linked: individual, practice-related, and sociopolitical, respectively. That is, we classified them depending on the sphere of human life that is considered as the primary focus of SAR implications. For this classification, we remained faithful to the literature’s underlying viewpoint, arranging the ethical issues in these groups according to how they are understood in the literature. This does not mean that these perspectives should not be further critically discussed: as set out in section 4, it is important to broaden the meaning (and thus the ethical dimension) of some current concerns.

This categorization was deemed pertinent because it integrates the principal spheres of action coming into play with the introduction of social robots in assistive practical contexts, namely: (1) intersubjective/human-robot interaction; (2) (specific) human practice; (3) sociopolitical activity. The chosen (unrefined) terminology of Well-being, Care, and Justice aims at the inclusion of all these spheres of activity, thus respectively encompassing SAR ethical issues regarding its implications for (1) the individuals for which this technology is provided (users), (2) the practice in which it is introduced and (3) society in its political structuring. These three categories are therefore related to three levels of ethical reflection.

We understand the notion of “practice” in MacIntyre’s sense, namely, as “any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence which are appropriate to, and partially definitive of, that form of activity, with the result that human powers to achieve excellence, and

human conceptions of the ends and goods involved, are systematically extended.” [76]. Also, we regard care as the broad practice to which the activity of assistance contributes, and hence SAR. Therefore, with the use of the category of “Care” we mean to embrace ethical concerns on SAR regarding both the particularities of this relational human activity (goods, virtues, models of professional excellence ...) and its (informal or institutional) organization –that is, the implications of SAR for the practical settings of assistance (distribution of tasks, institutional legitimacy and trust ...).

The 26 identified ethical issues associated with SAR are shown in Fig. 2 alongside the results of their quantitative analysis (frequency); all of them correspondingly classified in the three main categories of Well-being, Care, and Justice. Notice that in a couple of cases (\*) the same issue appears in different thematic groups (although with different frequency), given that it relates to more than one dimension of ethical concern.

As Fig. 2 reveals, among the 26 ethical issues of the current scholarly debate on SAR, the ones appearing most frequently are Privacy/Data Control (26 refs.), Deception (25 refs.) and Autonomy (21 refs.). In addition, since the mean number of appearances for each ethical issue is 10,42 times, these concerns constitute an outstanding focus of attention. In turn, all these three major concerns belong to the thematic group of Well-being, which, as Fig. 3 shows, gathers 60% of the identified ethical issues. It is followed by Care, which integrates 22% of them, and Justice in the third place, with 18% over the total. Thus, the data shows a relevant tendency of ethical reflection on SAR, namely: that most ethical concerns have to do with SAR implications for the individual dimension of Well-being. Another significant fact highlighted in Fig. 2 is the disparity rate of ethical issues between Well-being and the other categories of Care and Justice.

**Table 1**  
List of included publications.

Reference	Title
[25]	Beyond Moral Dilemmas: Exploring the Ethical Landscape in HRI
[26]	Embedding Ethics in the Design of Culturally Competent Socially Assistive Robots
[27]	Socially Assistive Robots, Older Adults and Research Ethics: The Case for Case-Based Ethics Training
[28]	Companion Robots: The Hallucinatory Danger of Human-Robot Interactions
[29]	Nudging for good: robots and the ethical appropriateness of nurturing empathy and charitable behavior
[30]	Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition
[31]	Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics
[32]	You, robot: on the linguistic construction of artificial others
[33]	Artificial agents, good care, and modernity
[34]	A Survey of Expectations About the Role of Robots in Robot-Assisted Therapy for Children with ASD: Ethical Acceptability, Trust, Sociability, Appearance, and Attachment
[35]	Anthropomorphism in Human-Robot Co-evolution
[36]	An Ethical Evaluation of Human–Robot Relationships
[37]	Design for Values for Social Robot Architectures
[3]	Socially Assistive Robotics: Ethical Issues Related to Technology
[38]	Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy
[39]	The Dark Side of Human-Robot Interaction: Ethical Considerations and Community Guidelines for the Field of HRI
[40]	Privacy and Socially Assistive Robots - A Meta Study
[41]	The Ethical Risk of Attachment: How to Identify, Investigate and Predict Potential Ethical Risks in the Development of Social Companion Robots
[42]	Social and Assistive Robotics in Dementia Care: Ethical Recommendations for Research and Practice
[43]	Language-Capable Robots may Inadvertently Weaken Human Moral Norms
[44]	Aged Care with Socially Assistive Robotics under Advance Care Planning
[45]	Ethical challenges in the use of social service robots for elderly people
[46]	Ethical Issues for Social Robots and the Trust-based Approach
[47]	When robots care: Public deliberations on how technology and humans may support independent living for older adults
[48]	Robotics in Nursing: A Scoping Review
[49]	Robot Enhanced Therapy for Autistic Children: An Ethical Analysis
[50]	Human Rights of Users of Humanlike Care Automata
[51]	Ethical Considerations Regarding the Use of Social Robots in the Fourth Age
[52]	Robot-Care for the Older People: Ethically Justified or Not?
[53]	A complementing approach for identifying ethical issues in care robotics – grounding ethics in practical use
[54]	Robots and people with dementia: Unintended consequences and moral hazard
[5]	Towards Human-Robot Interaction Ethics
[55]	Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use
[56]	The human relationship in the ethics of robotics: a call to Martin Buber’s I and Thou
[57]	Intelligent machines, care work and the nature of practical reasoning
[58]	Social robots, fiction, and sentimentality
[59]	Robots and human dignity: a consideration of the effects of robot care on the dignity of older people
[60]	Granny and the robots: ethical issues in robot care for the elderly
[61]	Second thoughts about privacy, safety and deception
[62]	Robots in aged care: a dystopian future?
[63]	Robotics Has a Race Problem
[64]	In the hands of machines? The future of aged care
[65]	Artificial Intelligence and Robotics in Nursing: Ethics of Caring as a Guide to Dividing Tasks Between AI and Humans
[66]	Details omitted for double-blind reviewing
[16]	Roboethics: Fundamental Concepts and Future Prospects
[67]	Children’s Imaginaries of Human-Robot Interaction in Healthcare
[68]	Moral Deskillling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character
[69]	Ethical and Technical Aspects of Emotions to Create Empathy in Medical Machines
[70]	

**Table 1 (continued)**

Reference	Title
	Robots responding to care needs? A multitasking care robot pursued for 25 years, available products offer simple entertainment and instrumental assistance
[71]	The Impact of Affective Verbal Expressions in Social Robots
[15]	Ethics of socially assistive robots in aged- care settings: a socio-historical contextualization
[21]	The use of care robots in aged care: A systematic review of argument-based ethics literature
[72]	The Dark Side of Ethical Robots
[73]	Ethically Aligned Design for Assistive Robotics
[74]	Trust in and Ethical Design of Carebots: The Case for Ethics of Care
[75]	“Oh, Dignity too?” Said the Robot: Human Dignity as the Basis for the Governance of Robotics

Source: Own elaboration

### 3.1. Description of the identified ethical issues

Through our review process, we found out that, in many scholarly publications, ethical issues regarding SAR are barely developed. Ethical concerns are usually merely mentioned –without clarifying their meaning or the reasons for which they can be said to be posed by SAR– nor engaging or taking a stance in a further sustained argumentation on them. Besides, some of them are presented by using terms that stand as core ethical categories within our (Western) history of thought, the meaning of which is thus not only very rich but also disputed and permanently revisited. Therefore, in order to shed light on the current scholar ethical debate on SAR, a description of what is meant with each of the identified ethical issues is provided below according to what was concluded from our conceptual analysis of the literature.<sup>8</sup>

#### 3.1.1. Well-being

**3.1.1.1. Privacy/Data Control.** Privacy is the major issue of ethical concern in the literature on SAR. Although undefined in many publications, the concept seems to be generally understood according to Refs. [77,78]: as a right against arbitrary interference with one’s private life, implying a users’ right to be in control of their personal information. Regarding this, a special matter of concern is the right to privacy of users with cognitive deterioration, which calls into question key related notions as “informed consent”.

SARs are considered a threat to privacy mainly in virtue of their monitoring capacity, through which they can collect, store, process and access personal data. Few authors delve further into the issue, by examining different types of privacy [42] and/or the link between different robotic technical features and the threat they pose to these [40]. SARs’ “intersubjective” capacity of interaction is also taken into account as a feature posing specific risks to privacy, being the creation of user profiles occasionally mentioned as a risk to privacy too. Sometimes, concerns on privacy relate to the feeling that users may have of being observed, of not being alone. Privacy can also be challenged by SAR’s usual goal of watching over user safety, setting out a conflict of values requiring ethical ponderation.

**3.1.1.2. Deception.** Deception is a very contested issue in the literature on SAR. In broad terms, the concern lies in the deceitful intersubjective relationship that human-robot interaction (HRI) may entail. Beyond the

<sup>8</sup> In this descriptive section, no references regarding each of the ethical issues are systematically included. Whereas it could certainly provide a helpful relation of the topics and the corresponding literature, including references for each of the 26 points would entail lots of bibliographical repetitions, since most of the reviewed publications address a great number of the identified ethical issues. Inasmuch this would be quite displeasing for readers, it has been a discarded option.

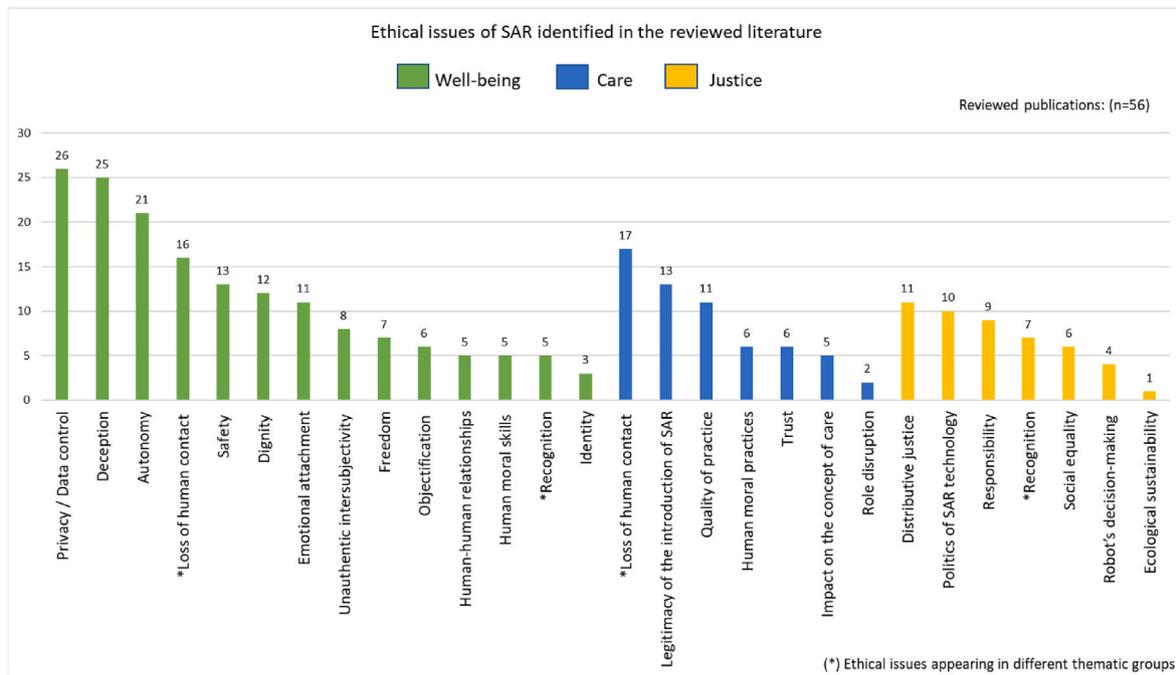


Fig. 2. Ethical issues associated with SAR in the reviewed literature. Source: Own elaboration.

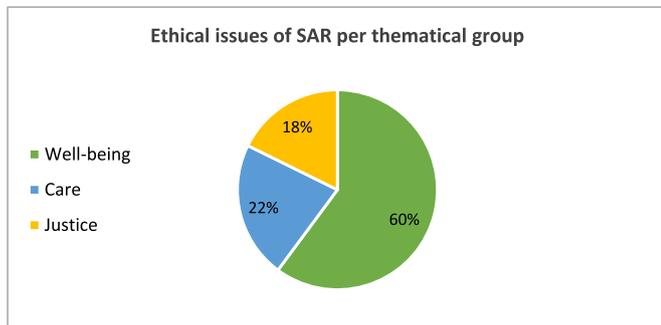


Fig. 3. Ethical issues of SAR per thematical groups. Source: Own elaboration

dispute over whether social (assistive) robots are inherently deceptive or not, there is debate on whether deception is morally wrong or not. The former stance is advocated from two main perspectives.

On the one hand, deception is conceived as morally wrong in virtue of the negative consequences that it may have for users, especially vulnerable ones. For instance, deception could lead to over-trust the robot –thus endangering the user’s safety or reasonable decision-making processes–, to foster the user’s emotional dependency on the robot –by perceiving it as having feelings or emotional states–, or to rather trigger emotional discomfort –given the robot’s incapacity for emotional reciprocity–. On the other hand, deception is argued to be intrinsically morally wrong, independently of whether upholding a false belief of the robot’s (emotional) capacities or a distorted view of the interpersonal relationship with them has positive or negative consequences for the user. This stance is linked to a comprehension of delusion as inherently wrong, from which unauthentic relationships are deemed morally unacceptable (as the one held by an affective bond with an entity like a social robot).

Lastly, some authors link the moral wrongness of deception to the fact that, in itself, it is a product of human intent, blaming it on the deliberate goal of deceiving users that lies behind a robot’s development –a deception that in many occasions is conceived as necessary in order to reach a robot’s full functionality in an assistive context with vulnerable

users–. In Kantian terms, this is a violation of human dignity because it involves instrumentalizing humans for the sake of achieving some goal.

**3.1.1.3. Autonomy.** (Human) autonomy is considered to be possibly challenged by SAR in numerous ways. From a consequential ethical perspective, an excessive (and thus inappropriate) extent or degree of assistance could cause a loss of the users’ capacities alongside a dependency on technology; thereby conflicting with the primary goal of assistance (to promote autonomy) and conversely raising new forms of vulnerability. The problem lies in the correct suitability (in terms of the kind and the quantity or proportionality) of assistance. Other vulnerabilities of human autonomy may also be fostered by SARs, such as exposure to manipulation or improper decision-making delegation.

From another perspective, the infringement with autonomy is not a (potential) result but rather a constitutive fact of SARs. This view comes from a specific understanding of the politics of SARs as a technology grounded in values and interests that are alien to that of ultimate users, which is why their introduction in the care practice equates to a violation of those users’ capacity to live according to their own reasons. This relates to the concerns on *objectification* and *informed consent*.

Technological autonomy also triggers ethical concern on human autonomy, given the potential conflict that can take place between both, as in cases where a user’s safety has to be balanced against their decision-making. This opens up the need of an ethical reflection upon the correct scope of beneficence (promotion of what is at the user’s best interest).

**3.1.1.4. Loss of human contact.** SAR could foster or even increase social isolation, which is detrimental to one’s well-being. First, SARs’ introduction may come along with a reduction in human contact, thereby potentially lessening an end-user’s opportunities for meaningful social interaction with humans. This is a problem both for individual patients and the organizational context in which the care practice is held. The shift in how the activity is carried out requires the adequate (re)articulating of the core values of the practice, all professionals involved, and that organization as a whole. Indeed, the problem has to do with the way in which robots are introduced: are they going to offer support within care practices as human replacements or as collaborative agents

instead? And to assume which type of tasks and roles? Second, the worry about robots as an isolating factor is also linked to HRI, which can foster certain relationship pathologies, such as emotional dependence on the robot or the user's seclusion to their inner world.

**3.1.1.5. Safety.** SARs pose a potential threat to a human's physical and psychological integrity. This problem is usually introduced in terms of 'safety', which globally refers to the harmful effects that robots may have for users regarding this (double) aspect of well-being. However, current concerns on safety are still mostly explicitly directed at the potential harm to a user's physical integrity. That is, they relate to the risk of accidents arising from robots sharing the same space as humans and interacting with them. Regarding this, not only actual, but also 'perceived safety' is under consideration. In addition, the worry on safety is generally associated to the compliance with the (bioethical) principle of nonmaleficence, seemingly leaving out an explicit consideration to beneficence as part of that same concern (a robot's potential contribution to a good life). Safety is thus a concern highly related to responsibility as liability for damage.

Safety is not only under ethical consideration because of the (harmful) results of robotic functioning, but because of the possible tension that may arise between the user's safety (a major goal of certain applications of SARs) and their autonomy or privacy.

**3.1.1.6. Dignity.** Dignity plays a crucial role in the ethical debate on SAR: it is both a recurrent issue of direct concern –insofar as potentially compromised by SAR for different reasons–, and also the value on which other concerns are ultimately grounded –leading to either consequential or deontological assessments of SARs' introduction in social settings of human practices–. Therefore, there is a constellation of issues revolving around the core value of dignity, such as objectification, recognition, deception and identity, among others.

From a consequentialist point of view, SAR could both enhance or negatively impact human dignity. For instance, linking dignity to the capabilities approach, one perspective in this vein argues to assess SAR depending on whether a robot's use expands or restricts the access to the basic set of capabilities to live a worthy human life. Also, dignity is under attack when these unauthentic intersubjective interactions alienate users from real life, thereby impoverishing their world. From a deontological perspective, SAR infringes upon human dignity because robots lack the basic abilities to provide assistance in an appropriate way, given that they cannot care and therefore cannot possibly fulfill the end-user's needs.

**3.1.1.7. Emotional attachment.** Basically, the reason why human emotional attachment to robots is introduced as an ethical issue is that it may generate contradictory effects with SARs' beneficial goal. For instance, it may lead to the loss of therapeutic benefits in the eventuality that the robot breaks down, does not fulfill the user's expectations or has to be shared with other users (feelings of jealousy). In the same lines, it may foster human over-trust on the robot with counterproductive results such as a robot's misuse by the user's over-delegation on it. Also, emotional attachment may promote certain pathologies of the human-robot relationship –such as a user's emotional dependency on the robot– and, in turn, human autonomy vulnerabilities.

Robot appearance is a matter of discussion closely linked to this issue, given that robotic design is considered to be a decisive element in the potential promotion of humans' emotional tie with robots.

**3.1.1.8. Unauthentic intersubjectivity.** The unauthentic intersubjectivity that characterizes HRI is a controversial subject of ethical debate, mostly regarded as morally problematic for several reasons. First, because it may reduce social life to an illusion, which comes along with the risk of an impoverishment of one's own world and life. The lack of a shared horizon of meaning between the human and the robot makes their

relationship a mirroring one (of the human with themselves). Second, because the unauthentic intersubjectivity involved in HRI may affect the proper development or exercising of human moral faculties (empathy, care ...), as well as foster an instrumentalist view of relationships, in which the "other" is not a "you" but rather an object. This endangers relations and forms of life that are intrinsically valuable and define us. Third, taken as a form of deception, unauthentic intersubjectivity is also argued to be wrong per se.

**3.1.1.9. Freedom.** The concern on freedom is grounded on a narrow understanding of this concept as "negative liberty" [79], that is, as the absence of external restrictions to one's own activity. Under this view, SARs challenge a user's freedom insofar as their goal of promoting well-being (in terms of health) may require that these robotic devices conduct in a way that somehow restricts a human's actions or decisions. This concern is related to machine ethics, because it has to do with how the robot should ponder conflicting values in the course of selecting the conducts to undertake in relational practices.

**3.1.1.10. Objectification.** This issue mainly refers to two questions. On the one hand, to the feeling that SARs' end-users may experience when being assisted and interacting with robotic devices regarding vital human needs. On the other hand, it may also refer to the attitude or moral cosmivision underlying the decision of introducing social artificial agents to carry out care-related tasks –objectification thus describing the lack of respect to human dignity involved in such delegation–, as well as to the stigmatization of user collectives that robotic design and functionality may involve (for example, infantilization).

**3.1.1.11. Human-human relationships.** The effects that social robots may have on human-human relationships are an ethical concern linked to the one of human-moral skills. On the one hand, the kind of socialization that can be fostered by interacting with machines, and the changes in opportunities for people to practice certain social skills essential for human interaction, on the other, have deep implications for human-human relationships. These could be impoverished or limited by substituting humans for machines in the social interaction practice, which is argued to endanger both the intrinsic values of human relationships and our self-understanding as human beings.

**3.1.1.12. Human moral skills.** SAR is considered to basically pose a threat to human moral skills for two main reasons. The first has to do with HRI, whereas the second relates to the adoption of social robots in care. The reason for caring about moral skills is twofold: they are purported to be essential prerequisites for developing practical wisdom and virtuous character and, at the same time, to be intrinsically valuable. This last point is linked to the idea that moral deskilling implies diminished human beings [68].

On the one hand, because of their "interpersonal" particular kind of interaction with humans, social robots may influence and shape human moral character, by potentially cultivating both vices and virtues. The source of ethical worry here is ultimately grounded on the relationship of pseudo-recognition involved in HRI. Human improper behavior and interaction with robots could foster a human's moral corruption. Even without explicit or intentional abusive behavior, the very same simulation of unconditional recognition carried out by robots could lead to a moral deskilling (for instance, it could normalize the experience of exercising control and power over what is seen as an autonomous agent with recognitive abilities). On the other hand, the adoption of SARs in care, by outsourcing practices central to human existence to non-human actors, could blind us from the awareness of the constitutive vulnerability and (inter)dependence of human life, thus threatening the cultivation of virtues essential to a flourishing society. More specifically, the new technological practices could reduce the opportunities for cultivating moral skills regarding human caregiving.

A partly adjacent concern related to human moral deskilling is the influence that language-capable robots, because of their acting as social moral agents (and given their constrained dialogue systems), may have on human moral reasoning.

**3.1.1.13. Recognition.** The issue of recognition falls both within the individual and the sociopolitical dimension of ethical concern on SAR.

In general, the problem that SAR is considered to pose in terms of recognition has its source in a perspective focused on the HRI, in which the emphasized sphere of ethical attention is the individual one. The concern mostly arises from the characteristic ontology of the robot, which makes it unable to enter a genuinely affective relationship and, therefore, deprives the human interactant of recognition, which is a fundamental element of social relationships –which are in turn indispensable for well-being–. From a more social-relational point of view, it is argued that the unconditional recognitive relationship that the robot establishes with the human is a source of a human interactant’s potential moral corruption, insofar as it makes the relational asymmetry be one of a power-relationship.

In fewer cases, SAR is understood to challenge recognition in that it may fail to respect the commitment with an individual’s equal civic rights regarding politics of welfare. This would be the case if the needs, interests and (reasonable) preferences of assistive technology users were disregarded or not equally taken into account and represented by these technological developments. Whose priorities prevail, and how problems are defined, are matters of justice that have to do with the ethical issue of recognition.<sup>9</sup> Robot appearance may also have ethical implications for the sociopolitical dimension of recognition (see *Identity*).

**3.1.1.14. Identity.** As a matter of concern, identity is introduced in the literature on SAR in two senses. First, as a self-conception, thus having to do with the respect for oneself. SAR potentially challenges identity in terms of impinging on (a user’s) comfort with one’s own image. For instance, due to its technological design –which may reinforce the image of the specific impairment to which assistance is provided– or to the way in which the artifact assists the user –which can be considered to harm people’s integrity–. Second, as an externally projected identity, thus relating to the image that third parties project onto users. In this sense, the concern has to do with the representation of an individual’s or a collective’s identity underlying SAR development, which can be an act of stigmatization not only impacting users individually, but also having ethical implications at a sociopolitical level. In turn, identity may be challenged by the inferences to which specific artifacts may lead to regarding aspects such as gender or race, which may be a collective’s discriminatory representation infringing upon human dignity and equality. Reinforcing, at the same time, narrative and structural injustices. This concern is tightly linked to a robot’s appearance.

### 3.1.2. Care

**3.1.2.1. Legitimacy of the introduction of SAR.** The legitimacy of the introduction and use of SAR in practical settings as a means of supporting their defining activity is a matter with key ethical significance at an organizational level, given that institutionalized practices must appropriately articulate the core values of professional, organizational and public ethics so as to grant a good service to citizenry. The question has to do mainly with two intertwined issues. On the one hand, with the consistency tool-task, i.e., whether (and how to ensure that) SARs are an appropriate tool for the task in which they are aimed to serve. This has to do both with the goals of SARs’ function (whether these are reached or

<sup>9</sup> Given the debate about the conceptual articulation of recognition and distribution in a theory of justice, we have decided to distinguish between both categories on this paper in order to leave it open for further exploration on how to connect them in regards to SAR ethical implications at a sociopolitical level.

not) and the values that are essential to the SAR-assisted practice (i.e., the way in which it will be held by means of SARs’ introduction). In turn, it implies an attention to the question of how to properly reshape the practical context. That is, how to redistribute the tasks or roles in order to guarantee an alignment with the specific practice’s goals and values, as well as with the core civic ones (based on Human Rights). Closely tied to this first issue, SAR’s legitimacy has also to do with the question whether a user’s needs and preferences are actually fulfilled, since the consistency of tool-task partly depend on whether the tool is user-centered. In this sense, the so-called ‘information gap’ between technology design and the end-user’s specific needs poses a great obstacle to legitimacy. Therefore, different stakeholders’ involvement in technological development is often introduced as a linked ethical issue.

Thus, the ethical issue of SAR’s legitimacy has to do with an attention to goals, values and processes of technological development and implementation, and it connects to other ethical questions such as *responsibility* and *trust*.

**3.1.2.2. Quality of practice.** The implications that SARs entail regarding the quality of the practices that they are meant to support is an important focus of ethical concern. This worry is commonly expressed in terms of “dehumanization” of care practices and their settings, a phenomenon attributed to a robot’s inability to enter in real intersubjective relationships with humans, which implies an inability to care (given their lack of moral agency and moral epistemology [65]). Ultimately, this concern is dependent on the redistribution of tasks; the central ethical question is how to reshape the traditional roles and functions of the professionals in the practice, i.e., which are the tasks that should be delegated to robots and why. Human substitution by robots could not only impact the practice quality, but also the meaning of care.

The quality of practice is also (partly) dependent on the consistency between the (presumed) tool and the task it is aimed to fulfill, which has to do both with the kind of task to be technologically assumed, and the way of carrying it out. To a large extent, such coherence depends on a proper knowledge and consideration to the context’s particularities and the involved stakeholders. The endangering inconsistency tool-task could arise from a developer’s knowledge gap about the needs and interests of the affected network of people, as well as the values and goals of the practice itself. Regarding this point, it is highly important to pay attention to the imaginaries of patients. This is mostly an unnoticed matter within ethical literature on SAR, the importance of which has been very well stressed out by Ref. [67],<sup>10</sup>.

**3.1.2.3. Human moral practices.** SAR may disrupt human moral practices that are constitutive of our societies and culture and, in turn, endanger both the internal goods of these practices, and certain human moral capacities that can only be developed and exercised through these forms of activity. SAR may erode care as a central practice of human moral life because it may reduce our engagement in such activity, thus lessening the cultivation of its associated moral skills and leading to a moral and professional deskilling. This has implications for the organizational sphere of human life, given that it challenges the core values of the exercised practice of care, which calls into question those of the whole institutional context within which is held.

**3.1.2.4. Trust.** SARs’ introduction may distort the essential element of trust inherent to care relationships, which is a problem for the organizational context of the practice, since the quality of the institution (or service sector) requires ensuring the quality of the care practice. A robot’s assumption of certain tasks that until now fell under (human) professionals’ scope of action implies a restructuration of roles and

<sup>10</sup> More recently [20], have further analyzed the role that roboticists’ imaginaries play regarding this issue too.

responsibilities that may lead users to inappropriate levels of trust, both regarding SARs and human caregivers using them as a means of support within the practice (along with the organization as a whole). A major concern is to ensure that SARs are trustworthy, so that HRI's goals can be successfully achieved. Efforts are directed towards achieving social acceptance of robots. Besides the matter of safety and responsibility for harm, the question of trust is also related to suitable knowledge about a robot's functions and capabilities (i.e., to a user's legitimate expectations), as well as to the coherence between SARs' functions and the practice's values and goals.

**3.1.2.5. Impact on the concept of care.** The new possibilities opened by SAR in the practice of care blur and call into question the meaning of such concept and its value and transcendence for human condition. In the literature, concerns on this issue relate to possible changes on social values surrounding care, as well as on society's concept of eldercare. Seemingly, the issue is linked to SARs particular ontology as 'almost-subjects', in virtue of which their inclusion in the practice of care disrupts our previous conceptualization of this activity as one exclusively entailing interpersonal relationships. Moreover, the concept of care may also change due to (new) needs either created or highlighted by these artifacts when introduced in the framework of social relations.

**3.1.2.6. Role disruption.** The introduction of SARs challenges current roles and responsibilities in care practice settings, thus threatening the quality of the practice and the essential element of trust that is constitutive of the relationship between caregivers and care recipients. Which tasks can be responsibly delegated to SARs or not in order to legitimately reshape these roles is a question with deep ethical implications at the organizational level.

### 3.1.3. Justice

**3.1.3.1. Distributive justice.** Some of the concerns on SAR revolve around distributive justice insofar as they have to do with the distribution of benefits and burdens across members of society. The matter is mainly about the fair allocation of SAR's initiatives' benefits and costs, being the latter primarily understood in terms of job impact (decrease of caregiver jobs due to the replacement of human workers by robots). The distribution of SARs and care as resources or goods is also an ethical issue falling under distributive justice concerns: who will have access to care robots? Could SAR contribute to a fairer distribution of care?

This issue is not argumentatively developed in the selected literature. The kind of individuals to whom these considerations of justice regarding SAR are meant to apply (whether among fellow citizens or rather international ones), is never explicitly stated, thus remaining unspecified whether distributive justice is contemplated within a certain political territory, or rather/also among countries, which would imply considerations of international distributive justice. Seemingly, it is the local distribution of benefits and costs that is under ethical reflection (with the exception of the question of ecological sustainability, which is presented as both a local and global matter of concern). Besides, distributive justice between generations is not mentioned in the literature, although SARs may well open up the need of reflecting upon the fair distribution of costs and benefits between contemporary and future generations (intergenerational justice).

**3.1.3.2. Politics of SAR technology.** This issue refers to the interests and values behind SAR initiatives and the question of their legitimacy. Concerns on SAR development as being driven by "technological solutionism" [80] are quite frequent, and SARs' suitability as tools for solving social problems (as the shortage of available social services for the care of the elderly) is often called into question. Ethical reflection on this matter has to do with the need of examining and grounding the reasons for SAR initiatives. Which is the problem at which they aim to

respond? Which are the underlying (economical, political, ideological) interests and how are they being, or should be prioritized? Which are the values grounding our social practices around vulnerability, and how and to which extent are they re-configured by SARs' introduction for care? The ethical issue thus revolves around the need of discussing, openly and inclusively at the societal level, the organization of the practice of care and the production of technological goods allocated to it. A usual worry regarding this issue is the prospect of the so-called "machinery of care".

**3.1.3.3. Responsibility.** SAR raises concerns on the ethical issue of responsibility, mainly in virtue of robots' technological autonomy, i.e., their ability to choose what to do based on previous information processing and regarding predefined goals, as well as their ability to behave accordingly. Therefore, the ethical worry mostly revolves around the question of liability for harm, where the latter is understood as the bad outcomes of SARs' functioning or tasks' execution. Who is ultimately responsible for the potential negative consequences of a robot's behavior, and who should be answerable to these? As a reflection linked to responsibility attribution, this issue is closely linked to matters of product safety and decision-making transparency of systems or "explainability"—which has to do with a key dimension of responsibility, namely: accountability (being able to explain and justify decisions).

**3.1.3.4. Social equality.** SAR has implications for social equality, since depending on how it is developed and implemented, it may either contribute to increase or lessen the equality of care both in terms of access and quality of treatment. Since intelligent autonomous machines are developed and trained using databases, social divide in terms of access is a big exclusion problem leading to inequality in the healthcare service [69], since the data of non-users won't be included into the databases from which the service is offered. This implies data bias by lack of representativity, which equates to an unequal (medical) treatment (or the impossibility of granting it) to such collectives.

**3.1.3.5. Robot's decision-making.** As a kind of technology to be introduced in daily life so as to autonomously carry out certain tasks within assistive practices, SARs' behavior has decisive consequences for individuals, which is normally why a robot's decision-making process is an issue of ethical concern. Guaranteeing that SARs will behave correctly according to the context goes beyond an issue of technical safety, and requires that their decision-making is aligned not only with the goals of their task, but also with human values. For it, a robot's ability to correctly assess and manage possible tensions or conflicts between different values that may arise in certain situations in real life (safety vs autonomy or privacy, for instance) seems to be a necessary (even though not a sufficient) condition. Therefore, this issue turns ethical attention to machine ethics, aimed at endowing robots with ethical reasoning capacities, so that their decision-making process is grounded on an understanding of, and an appropriate response to the moral relevant facts of each situation. Robot decision-making as a matter of ethical reflection on SAR is thus linked to concerns on *harm* and *responsibility*.

**3.1.3.6. Ecological sustainability.** The implications of SAR for both local and global sustainability are hardly ever examined in the literature. However, the supply of raw materials for robots, the energy consumption they require, and the dumping waste that these new care technologies generate are important ethical challenges of SAR. This issue is tightly related to matters of international distributive and intergenerational justice.

## 4. Reflections towards a critical approach to the ethical debate

Through the literature review, significant tendencies of the ethical approach to SAR have been disclosed, which evince a need of critically analyzing the way in which reflection is being directed. Accordingly, in

this section we will outline some topics worth examining and discussing in view of enriching the ethical gaze on SAR.

#### (1) The individual-centered focus of ethical reflection

The ongoing ethical reflection on SAR predominantly focuses on the individual dimension of human life, i.e., on the implications that this technoscientific field has for individuals, who are, in turn, principally understood as SARs' users. Ethical approach to SAR thus takes what we have categorized as (individual) well-being as the primary dimension of normative concern. This is closely linked to the tendency to exclusively narrow the attention in the dyadic interaction between humans and robots, against which some authors have already argued [81] and which has been explicitly identified as a constitutive factor of a misguided ethical approach to social robotics [17].

The individual-centered perspective comes with a disproportionately fewer attention to SAR implications from the perspective of the (care) practice in which its artifacts are used, as well as from the macro sociopolitical level of justice. Descriptively, several reasons could be found behind this tendency.<sup>11</sup> However, from a normative-oriented point of view, what matters the most is that this is an important deficiency of the current ethical approach to SAR, which shows a continuation of the individualist assumptions and the "neglect of the political" underlying the mainstream philosophy of technology and ethics of technology [82]. We contend that this tendency should be overcome.

First and foremost, because an excessively restricted ethical focus at an individual level—at the expense of the two others—is symptomatic of a loss in perspective of the decisive interrelation between all three spheres of ethical import regarding SAR. It implies overlooking the conditioning that the sociopolitical structure has regarding the configuration of care practices and thus the influence that both of these have for an individual's life. Indeed, HRI in SARs' case takes place within broader social practices that reflect values, goals and a specific cosmovision about how to organize human life. From a holistic consideration on the person, individuals must be taken in their situatedness. Hence, the constitutive interrelation between (individual) Well-being and the spheres of Care and Justice cannot justifiably be disregarded. This would mean to neglect the role of sociopolitical structure in easing or hindering, to a greater or lesser extent, the possibility of covering needs and developing personal autonomy.

Moreover, insofar as ethical reflection on SAR must primarily proceed as applied ethics understood as critical hermeneutics of human

<sup>11</sup> First, such tendency reflects the kind of world where we find ourselves. A world in which technological development takes place within the frame of market dynamics, thereby narrowing ethical reflection on technology within its boundaries. This implies leaving aside an approach that thinks at a macro level, as an approach primarily led from a sociopolitical perspective would do. Second, ethical reflection on SAR must proceed as applied ethics that contextualizes reflection according to the values and goals of the specific practical field of activity for which technology is developed. This entails to circumscribe ethical reflection primarily within this specific (and already constituted) practice, which in a way means to endorse a conservative point of departure that may well lead to neglect matters related to the sociopolitical framework on which it takes place. The discipline of bioethics is an exemplary case of this: it also began its critical activity without calling into question the aforementioned macro level. Third, the fact that ethical reflection on SAR has been mostly led by a technoscientific professional profile could explain why the focus of ethical attention revolves around the user's (individual) well-being. Indeed, given the instrumental character of the technology, the prevailing ethics among engineers is a consequentialist one. That is, their ethical approach revolves around assessing the consequences of the artifacts in terms of meeting the expected goals in the established way. Insofar as SARs assist by interacting with humans, ethics is focused on the implications of this particular kind of robotic functioning. In addition, the fact that SARs are not still implemented at a large scale may well play a role in reinforcing this ethical perspective focused at the individual level, under which Care and Justice issues are much less addressed.

activity [83], if anything, it is the (care) practice sphere the capital one. SAR should be primarily approached in light of the specific practice at which it aims to serve, within which individuals are not mere monads but members of a relational network of human activity.

Finally, an ethical approach that leaves insufficiently unattended SAR disruption potential regarding other dimensions of human life besides the (interactants') individual one, may come along with the risk of converting ethical reflection into a mere exercise of moral evaluation within an unrevised framework of values and (given) ends, by overlooking the question about the type of practices and societies we actually want, and how to accordingly (re)configure life through SAR.

#### (2) Teleological and anthropological assumptions of SAR

The current ethical approach to SAR generally lacks reflection and discussion on SAR teleology, that is, on conceptual assumptions on "assistance", "care" and other correlated notions (human well-being, human capabilities, autonomy ...) that underlie this field's development. The constellation of SAR teleological-related meanings should be examined, since it is always linked to a particular anthropology that should be carefully analyzed and further discussed. For now, SAR development hints at the specific idea of human vulnerability and fragility as an annoyance, the care of which can be delegated to technology. The background anthropology is thus a liberal one, revolving around capacitism and adultism, from which SAR is ultimately aimed at replacing capacities.

#### (3) Restricted understanding of core ethical concepts

The literature review reveals a very narrowed understanding of some ethical concepts around which SAR's problems stand, which impoverishes the ethical approach to the disruptive implications of this technoscientific field. Indeed, the limitation in such notions' scope of meaning correlates to a loss of sight of the interconnection between the three main spheres of ethical concern (individual, care practice-related, and sociopolitical).

For instance, (human) freedom is currently understood as what is philosophically known as "negative liberty" [79]. The ethical implications of SAR regarding this issue, though, could and should be broadened up by approaching the matter from a deeper understanding of this notion in its dimension of "positive liberty" [79]. That is, closer to freedom as autonomy—in which freedom has to do with self-realization, with taking over the own life—. It could also be interesting to think from the perspective of Pettit's "republican conception of freedom as non-domination" [84]. Delving into the meaning of this notion would definitely allow for a richer normative-oriented reflection on SAR that takes into account the interdependence existing between freedom and the sociopolitical structuring of human life, thus approaching SAR's power of domination both at an interpersonal and structural level.

The same happens with responsibility, which is generally understood as liability for harm and thus revolves around the distribution of duties to answer for bad outcomes, which, on top of that, are linked to the (AI-based) robot's behavior. A more "substantial" concept of responsibility [85] would enrich ethical reflection, by enabling us to think in terms of accountability for the development of SAR (teleology, interest), which implies taking a sociopolitical perspective that approaches the matter in light of justice.

Also privacy is misguidedly understood in the current debate on SAR in too individualistic terms, although the collective ethical dimension of privacy has been already well highlighted [86]. The implications of SAR for privacy are mainly thought within the frame of a robot's impact on their interactant's life. It should be considered to reframe reflection on SAR's threat to privacy from a sociopolitical point of view, that is, in terms of justice.

#### (4) Overlooked ethical issues in Well-being

Within the constellation of concerns related to (individual) Well-being, some important issues are overlooked. Since current ethical reflection on SAR is primarily focused on this sphere (due to HRI's central place within the ethically scrutinized spheres of activity impacted by SAR), engaging in those missing issues could significantly contribute to enrich the state of the art.

- (i) Distinction between privacy, intimacy and interiority. Although Privacy is the most commonly addressed ethical issue on SAR, there is no attention to other notions that are related to its semantic field [87,88] and incorporate nuances worth examining [89] regarding the implications of SAR at an individual level of the user's life.
- (ii) The individual's "possibility to be". At the level of HRI, there are serious ethical issues (besides deception) alarmingly missing in the current landscape of reflection. These are ultimately related to the implications of SAR for the individual's "possibility to be" (and not only "to do"), in virtue of the standardization of people and relations that HRI entails.

In relational practices of care, HRI may condemn the "being" by reducing human interactants to a specific existential dimension, according to the user model that the individual represents. The interactant "stops" being a person and becomes a "model" within the assistive relationship. In turn, this model further reduces the person to quantifiable and operable patrons –sentencing them to remain as according to what they are said to be–. This latter point is related to the paradigm of experience on which a robot's learning and interaction takes place, which has, until now, been disregarded in relation to SAR –although the ethical implications of the algorithmic functioning and decision-making are indeed taken into account in other application domains of AI-systems.

Some features of the current debate landscape on SAR explain the thematic oversight on the individual's "possibility to be"; such as the above highlighted restricted view on the key concepts of autonomy and freedom, as well as the scarce and poor reflection on the notions of identity and recognition, and the absence of an examination of concepts like experience, domination and difference. Therefore, a philosophical anthropology of (inter)subjectivity would be in order, since it would advance some nuclear concepts regarding the human condition (unpredictability, affectability, openness to future) that are essential to reveal and comprehensibly address the ethical challenges of HRI.

## 5. Conclusions

This literature review has provided an overall and hopefully clarifying picture of the current ethical reflection on SAR, thus tackling the until now little structured landscape of ethical issues associated to this technoscientific field of intelligent robotics. Thereby, some relevant tendencies and thematic deficits of the current approach have been identified and outlined, which should be further critically examined and addressed. Hence, this review opens up and points at new research directions on SAR ethical implications. It constitutes a basis on which to build a research agenda for widening and deepening the normative thinking on this technoscientific field of activity. To contribute to such agenda, we sketch below some of the future lines of research that may arise from our critical literature review.

- To extend the conceptual framework from which to comprehensively identify and analyze SAR ethical implications.
- To delve into particular concepts belonging to philosophical disciplines (e.g., freedom or responsibility) and integrate other notions that may also be relevant for addressing the ethical implications of SAR (e.g., experience, intimacy).

- To consider the identified ethical issues of SAR by drawing upon conceptual resources of political philosophy and philosophical anthropology.
- To examine SAR teleological and anthropological assumptions and relate them to the different existing theoretical frameworks on care. This theoretical work could latter provide practical coordinates for grounding decisions on the different stages of SAR development. For instance, regarding the question on which services and roles should be redistributed through SARs' deployment, under which conditions and how; as well as on the elements that engineers should take into account in order to serve the aims and values of the practice to which they give support to through SAR solutions.
- To integrate a critical theory perspective in the ethical approach to SAR, in order to include a teleological reflective momentum that discusses the ends, beyond the means, and to engage in the question about the type of practices, societies and lives that we want to (re)configure through the design, deployment and usage of SAR solutions.
- To delve into other ethical challenges of HRI besides deception, with an special focus on the implications that the algorithmic functioning of SARs may have for the well-being of its interactants.
- To set practical principles for SAR deployment (design, research, implementation and use) under the light of the three interrelated categories of Well-being, Care and Justice. To that end, and given that SARs are aimed at supporting practices of care, existing frameworks of applied ethics (such as the principles of bioethics) should be revisited and updated considering the novelty of SARs-assisted care practice.
- To examine the relevance of cultural aspects regarding SAR development by drawing upon the difference between the ethics of minima and maxima.

In conclusion, this literature review is a first stage of a larger research process aimed at contributing to the ethical debate on SAR by completing the scope of issues that should be taken into account, and delving further into the most normatively relevant ones.

## Funding

This work has been partially supported by the Spanish Ministry of Science and Innovation under a FPI scholarship for predoctoral contracts for the training of doctors (PRE2018-084286), and by the Spanish State Research Agency through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656).

## Author's contributions

All authors contributed to this study. The literature search and data analysis were performed by Júlia Pareto Boada. The first draft of the manuscript was written by Júlia Pareto Boada and all authors commented on and critically revised the previous versions of the manuscript. All authors read and approved the final manuscript.

## Declaration of competing interest

The authors declare that they have no conflict of interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.techsoc.2021.101726>.

## Authors'\* statement

All authors contributed to this study. The literature search and data analysis were performed by Júlia Pareto Boada. The first draft of the

manuscript was written by Júlia Pareto Boada and all authors commented on and critically revised the previous versions of the manuscript. All authors read and approved the final manuscript.

## References

- [1] High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI*, 2019, pp. 1–41.
- [2] A. Tapus, M. Mataric, B. Scassellati, The grand challenges in socially assistive robotics, *IEEE Robot. Autom. Mag.* 14 (2007) 35–42.
- [3] D. Feil-Seifer, M.J. Mataric, Socially assistive robotics: ethical issues related to technology, *IEEE Robot. Autom. Mag.* 18 (2011) 24–31, <https://doi.org/10.1109/MRA.2010.940150>.
- [4] D. Feil-Seifer, M.J. Mataric, Defining socially assistive robotics, in: *9th Int. Conf. Rehabil. Robot.*, IEEE, 2005, pp. 465–468.
- [5] S.M. Payr, Towards human-robot interaction ethics, in: R. Trappl (Ed.), *A Constr. Man. Robot. Ethical Syst. Cogn. Technol.*, Springer, 2015, <https://doi.org/10.1007/978-3-319-21548-8>.
- [6] C. Breazeal, A. Takanishi, T. Kobayashi, Social robots that interact with people, in: B. Siciliano, O. Khatib (Eds.), *Springer Handb. Robot.*, Springer, 2008, pp. 1349–1369, [https://doi.org/10.1007/978-3-540-30301-5\\_59](https://doi.org/10.1007/978-3-540-30301-5_59).
- [7] M.J. Mataric, Socially assistive robotics: human augmentation versus automation, *Sci. Robot.* 2 (2017), <https://doi.org/10.1126/scirobotics.aam5410>.
- [8] European Commission, *White Paper on Artificial Intelligence - A European Approach to Excellence and Trust*, 2020, <https://doi.org/10.1017/CBO9781107415324.004>.
- [9] Z. Dolic, R. Castro, R. Moarcas, Robots in Healthcare: a Solution or a Problem? Study for the Committee on Environment, Public Health, and Food Safety, 2019.
- [10] E. Fosch-Villaronga, Grau Ruiz María Amparo, Expert considerations for the regulation of assistive robotics. A European robotics forum echo, *Dilemata, Rev. Int. Éticas Apl.* (2019) 149–169.
- [11] Ajuntament de Barcelona, *Misty II the Social Robot Becomes Part of the Lives of Twenty Senior Citizens*, 2020. [https://www.barcelona.cat/infobarcelona/en/tema/senior-citizens/misty-ii-the-social-robot-becomes-part-of-the-lives-of-twenty-senior-citizens\\_907645.html](https://www.barcelona.cat/infobarcelona/en/tema/senior-citizens/misty-ii-the-social-robot-becomes-part-of-the-lives-of-twenty-senior-citizens_907645.html). (Accessed 31 July 2021). accessed.
- [12] M.J. Mataric, B. Scassellati, Socially assistive robotics, in: B. Siciliano, O. Khatib (Eds.), *Springer Handb. Robot.*, Springer, 2016, pp. 1973–1994.
- [13] United Nations, *World Population Prospects 2019*, 2019. Highlights, <http://www.ncbi.nlm.nih.gov/pubmed/12283219>.
- [14] European Parliament, *Civil Law Rules on Robotics European Parliament Resolution (2015/2103(INL))*, 2017, p. 23. [http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051\\_EN.pdf](http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf).
- [15] T. Vandemeulebroucke, B.D. Casterle, C. Gastmans, Ethics of socially assistive robots in aged-care settings: a socio-historical contextualisation, *J. Med. Ethics* 46 (2020) 128–136, <https://doi.org/10.1136/medethics-2019-105615>.
- [16] S.G. Tzafestas, Roboethics: fundamental concepts and future prospects, *OR Inf.* 9 (2018), <https://doi.org/10.3390/INFO9060148>.
- [17] J. Pareto Boada, Prolegómenos a una ética para la robótica social, *Dilemata, Revista Internacional de Éticas Aplicadas* 34 (2021) 71–87.
- [18] B.C. Stahl, M. Coeckelbergh, Ethics of healthcare robotics: towards responsible research and innovation, *Robot. Autonom. Syst.* 86 (2016) 152–161, <https://doi.org/10.1016/j.robot.2016.08.018>.
- [19] N. Vallès-Peris, M. Domènech, Robots para los CUIDADOS. LA ética de la acción MESURADA frente a la INCERTIDUMBRE, *Cuad. Bioet.* 31 (2020) 87–100, <https://doi.org/10.30444/CB.54>.
- [20] N. Vallès-Peris, M. Domènech, Roboticians' imaginaries of robots for care: the radical imaginary as a tool for an ethical discussion, *Eng. Stud.* 12 (2020) 157–176, <https://doi.org/10.1080/19378629.2020.1821695>.
- [21] T. Vandemeulebroucke, B. Dierckx de Casterlé, C. Gastmans, The use of care robots in aged care: a systematic review of argument-based ethics literature, *Arch. Gerontol. Geriatr.* 74 (2018) 15–25, <https://doi.org/10.1016/j.archger.2017.08.014>.
- [22] C. Fernández-Aller, A. Fernández de Velasco, Á. Manjarrés, D. Pastor-Escuredo, S. Pickin, J. Salgado Criado, T. Ausín, An inclusive and sustainable artificial intelligence strategy for Europe based on human rights, *IEEE Technol. Soc. Mag.* (2021), <https://doi.org/10.1109/MTS.2021.3056283>.
- [23] C. Mejia, Y. Kajikawa, Bibliometric analysis of social robotics research: identifying research trends and knowledgebase, *Appl. Sci.* 7 (2017), <https://doi.org/10.3390/app7121316>.
- [24] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ* 372 (2021), <https://doi.org/10.1136/bmj.n71>.
- [25] T. Arnold, M. Scheutz, Beyond moral dilemmas: exploring the ethical landscape in HRI, *ACM/IEEE Int. Conf. Human-Robot Interact. Part F1271* (2017) 445–452, <https://doi.org/10.1145/2909824.3020255>.
- [26] L. Battistuzzi, A. Sgorbissa, C. Papadopoulos, I. Papadopoulos, C. Koulouglioti, Embedding ethics in the design of culturally competent socially assistive robots, *IEEE Int. Conf. Intell. Robot. Syst.* (2018), <https://doi.org/10.1109/IROS.2018.8594361>, 1996–2001.
- [27] L. Battistuzzi, C. Papadopoulos, T. Hill, N. Castro, B. Bruno, A. Sgorbissa, Socially assistive robots, older adults and research ethics: the case for case-based ethics training, *Int. J. Soc. Robot.* (2020), <https://doi.org/10.1007/s12369-020-00652-x>.
- [28] P. Bisconti Lucidi, D. Nardi, Companion robots: the hallucinatory danger of human-robot interactions, in: *AIES 2018 - Proc. 2018 AAAI/ACM Conf. AI, Ethics, Soc.*, 2018, pp. 17–22, <https://doi.org/10.1145/3278721.3278741>.
- [29] J. Borenstein, R.C. Arkin, Nudging for good: robots and the ethical appropriateness of nurturing empathy and charitable behavior, *AI Soc.* 32 (2017) 499–507, <https://doi.org/10.1007/s00146-016-0684-1>.
- [30] M.L. Cappuccio, A. Peeters, W. McDonald, Sympathy for dolores: moral consideration for robots based on virtue and recognition, *Philos. Technol.* 33 (2020) 9–31, <https://doi.org/10.1007/s13347-019-0341-y>.
- [31] M. Coeckelbergh, Personal robots, appearance, and human good: a methodological reflection on Roboethics, *Int. J. Soc. Robot.* 1 (2009) 217–221, <https://doi.org/10.1007/s12369-009-0026-2>.
- [32] M. Coeckelbergh, robot You, On the linguistic construction of artificial others, *AI Soc.* 26 (2011) 61–69, <https://doi.org/10.1007/s00146-010-0289-z>.
- [33] M. Coeckelbergh, Artificial agents, good care, and modernity, *Theor. Med. Bioeth.* 36 (2015) 265–277, <https://doi.org/10.1007/s11017-015-9331-y>.
- [34] M. Coeckelbergh, C. Pop, R. Simut, A. Peca, S. Pintea, D. David, B. Vanderborcht, 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* 22 (2016) 47–65, <https://doi.org/10.1007/s11948-015-9649-x>.
- [35] L. Damiano, P. Dumouchel, Anthropomorphism in human-robot Co-evolution, *Front. Psychol.* 9 (2018) 1–9, <https://doi.org/10.3389/fpsyg.2018.00468>.
- [36] M.M.A. de Graaf, An ethical evaluation of human-robot relationships, *Int. J. Soc. Robot.* 8 (2016) 589–598, <https://doi.org/10.1007/s12369-016-0368-5>.
- [37] V. Dignum, F. Dignum, J. Vazquez-Salceda, A. Clodic, M. Gentile, S. Mascarenhas, A. Augello, Design for values for social robot architectures, *Front. Artif. Intell. Appl.* 311 (2018) 43–52, <https://doi.org/10.3233/978-1-61499-931-7-43>.
- [38] A. Fiske, P. Henningsen, A. Buyx, Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy, *J. Med. Internet Res.* 21 (2019) 1–12, <https://doi.org/10.2196/13216>.
- [39] K.S. Haring, M.M. Novitzky, P. Robinette, E.J. de Visser, A. Wagner, T. Williams, The dark side of human-robot interaction: ethical considerations and community guidelines for the field of HRI, in: *14th ACM/IEEE Int. Conf. Human-Robot Interact.*, 2019, pp. 689–690. <http://arxiv.org/abs/1504.04339>.
- [40] T. Heuer, I. Schiering, R. Gerndt, Privacy and socially assistive robots - a meta study, in: *Priv. Identity Manag. Smart Revolut.*, Springer International Publishing, 2018, pp. 265–281, <https://doi.org/10.1007/978-3-319-92925-5>.
- [41] A. Huber, A. Weiss, M. Rauhala, The ethical risk of attachment: how to identify, investigate and predict potential ethical risks in the development of social companion robots, in: *ACM/IEEE Int. Conf. Human-Robot Interact.* 2016-April, 2016, pp. 367–374, <https://doi.org/10.1109/HRI.2016.7451774>.
- [42] M. Ienca, F. Jotterand, C. Vicà, B. Elger, Social and assistive robotics in dementia care: ethical recommendations for research and practice, *Int. J. Soc. Robot.* 8 (2016) 565–573, <https://doi.org/10.1007/s12369-016-0366-7>.
- [43] R.B. Jackson, T. Williams, language-capable robots may inadvertently weaken human moral norms, in: *ACM/IEEE Int. Conf. Human-Robot Interact.* 2019-March, 2019, pp. 401–410, <https://doi.org/10.1109/HRI.2019.8673123>.
- [44] J. Koimizu, Aged care with socially assistive robotics under advance care planning, in: *Proc. IEEE Work. Adv. Robot. Its Soc. Impacts, ARSO.* 2019-October, 2019, pp. 34–38, <https://doi.org/10.1109/ARSO46408.2019.8948742>.
- [45] T. Körtner, Ethical challenges in the use of social service robots for elderly people, *Z. Gerontol. Geriatr.* 49 (2016) 303–307, <https://doi.org/10.1007/s00391-016-1066-5>.
- [46] T. Koyama, Ethical issues for social robots and the trust-based approach, in: *Proc. IEEE Work. Adv. Robot. Its Soc. Impacts, ARSO.* 2016-Novem, 2016, pp. 1–5, <https://doi.org/10.1109/ARSO.2016.7736246>.
- [47] P. Lehoux, D. Grimard, When robots care: public deliberations on how technology and humans may support independent living for older adults, *Soc. Sci. Med.* 211 (2018) 330–337, <https://doi.org/10.1016/j.socscimed.2018.06.038>.
- [48] N. Maalouf, A. Sidaoui, I.H. Elhaji, D. Asmar, Robotics in nursing: a scoping review, *J. Nurs. Scholarsh.* 50 (2018) 590–600, <https://doi.org/10.1111/jnu.12424>.
- [49] N. McBride, Robot enhanced therapy for autistic children: an ethical analysis, *IEEE Technol. Soc. Mag.* 39 (2020) 51–60, <https://doi.org/10.1109/MTS.2020.2967493>.
- [50] L.F. Miller, Human rights of users of humanlike care automata, *Hum. Right Rev.* 21 (2020) 181–205, <https://doi.org/10.1007/s12142-020-00581-2>.
- [51] C. Misselhorn, U. Pompe, M. Stapleton, Ethical considerations regarding the use of social robots in the fourth age, *GeroPsych J. Gerontopsychol. Geriatr. Psych.* 26 (2013) 121–133, <https://doi.org/10.1024/1662-9647/a000088>.
- [52] F.M. Noori, Z. Uddin, J. Torresen, Robot-care for the older people: ethically justified or not?, 2019 jt, in: *IEEE 9th Int. Conf. Dev. Learn. Epigenetic Robot.*, 2019, pp. 43–47, <https://doi.org/10.1109/DEVLRN.2019.8850706>.
- [53] S. Nylander, S. Ljungblad, J. Jimenez Villareal, A complementing approach for identifying ethical issues in care robotics - grounding ethics in practical use, *Proc. - IEEE Int. Work. Robot Hum. Interact. Commun.* (2012) 797–802, <https://doi.org/10.1109/ROMAN.2012.6343849>.
- [54] F. O'Brocháin, Robots and people with dementia: unintended consequences and moral hazard, *Nurs. Ethics* 26 (2019) 962–972, <https://doi.org/10.1177/0969733017742960>.
- [55] S.M. Rabbitt, A.E. Kazdin, B. Scassellati, Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded

- use, *Clin. Psychol. Rev.* 35 (2015) 35–46, <https://doi.org/10.1016/j.cpr.2014.07.001>.
- [56] K. Richardson, The human relationship in the ethics of robotics: a call to Martin Buber's I and Thou, *AI Soc.* 34 (2019) 75–82, <https://doi.org/10.1007/s00146-017-0699-2>.
- [57] A. Robson, Intelligent machines, care work and the nature of practical reasoning, *Nurs. Ethics* 26 (2018) 1906–1916, <https://doi.org/10.1177/0969733018806348>.
- [58] R. Rodogno, Social robots, fiction, and sentimentality, *Ethics Inf. Technol.* 18 (2016) 257–268, <https://doi.org/10.1007/s10676-015-9371-z>.
- [59] A. Sharkey, Robots and human dignity: a consideration of the effects of robot care on the dignity of older people, *Ethics Inf. Technol.* 16 (2014) 63–75, <https://doi.org/10.1007/s10676-014-9338-5>.
- [60] A. Sharkey, N. Sharkey, Granny and the robots: ethical issues in robot care for the elderly, *Ethics Inf. Technol.* 14 (2012) 27–40, <https://doi.org/10.1007/s10676-010-9234-6>.
- [61] T. Sorell, H. Draper, Second thoughts about privacy, safety and deception, *Connect. Sci.* 29 (2017) 217–222, <https://doi.org/10.1080/09540091.2017.1318826>.
- [62] R. Sparrow, Robots in aged care: a dystopian future? *AI Soc.* 31 (2016) 445–454, <https://doi.org/10.1007/s00146-015-0625-4>.
- [63] R. Sparrow, Robotics has a race problem, *Sci. Technol. Hum. Val.* (2019), 016224391986286, <https://doi.org/10.1177/0162243919862862>.
- [64] R. Sparrow, L. Sparrow, In the hands of machines? The future of aged care, *Minds Mach.* 16 (2006) 141–161, <https://doi.org/10.1007/s11023-006-9030-6>.
- [65] F. Stokes, A. Palmer, Artificial intelligence and robotics in nursing: ethics of caring as a guide to dividing tasks between AI and humans, *Nurs. Philos.* (2020) 1–9, <https://doi.org/10.1111/nup.12306>.
- [66] C. Torras, Social networks and robot companions: Technology, ethics, and science fiction, *Métode Science Studies Journal* 9 (2019) 163–169, <https://doi.org/10.7203/metode.9.12479>.
- [67] N. Vallès-Peris, C. Angulo, M. Domènech, Children's imaginaries of human-robot interaction in healthcare, *Int. J. Environ. Res. Publ. Health* 15 (2018), <https://doi.org/10.3390/ijerph15050970>.
- [68] S. Vallor, Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character, *Philos. Technol.* 28 (2015) 107–124, <https://doi.org/10.1007/s13347-014-0156-9>.
- [69] J. Vallverdú, D. Casacuberta, Ethical and technical aspects of emotions to create empathy in medical machines, in: S.P. van Rysewyk, M. Pontier (Eds.), *Mach. Med. Ethics*, Springer International Publishing, 2015, pp. 341–362, [https://doi.org/10.1007/978-3-319-08108-3\\_20](https://doi.org/10.1007/978-3-319-08108-3_20).
- [70] L. Van Aerschoot, J. Parviainen, Robots responding to care needs? A multitasking care robot pursued for 25 years, available products offer simple entertainment and instrumental assistance, *Ethics Inf. Technol.* (2020), <https://doi.org/10.1007/s10676-020-09536-0>.
- [71] A. Van Maris, A. Sutherland, A. Mazel, S. Dogramadzi, N. Zook, M. Studley, A. Winfield, P. Caleb-Solly, The impact of affective verbal expressions in social robots, *ACM/IEEE Int. Conf. Hum. Robot Interact.* (2020) 508–510, <https://doi.org/10.1145/3371382.3378358>.
- [72] D. Vanderelst, A. Winfield, The dark side of ethical robots, *AIES 2018*, in: *Proc. 2018 AAAI/ACM Conf. AI, Ethics, Soc.*, 2018, pp. 317–322, <https://doi.org/10.1145/3278721.3278726>.
- [73] Y.H. Weng, Y. Hirata, Ethically aligned design for assistive robotics, in: *2018 Int. Conf. Intell. Saf. Robot., IEEE*, 2018, pp. 286–290, <https://doi.org/10.1109/IISR.2018.8535889>.
- [74] G.C.K. Yew, Trust in and ethical design of carebots: the case for ethics of care, *Int. J. Soc. Robot.* (2020), <https://doi.org/10.1007/s12369-020-00653-w>.
- [75] L. Zardiashvili, E. Fosch-Villaronga, “Oh, Dignity too?” Said the robot: human dignity as the basis for the governance of robotics, *Minds Mach.* 30 (2020) 121–143, <https://doi.org/10.1007/s11023-019-09514-6>.
- [76] A. MacIntyre, *After Virtue. A Study in Moral Theory*, third ed., University of Notre Dame Press, 2007.
- [77] *General Assembly of the United Nations, Universal Declaration of Human Rights*, 1948.
- [78] The Member States, *Charter of Fundamental Rights of the European Union*, 2012, <https://doi.org/10.2307/j.ctt1ffjmjq.33>.
- [79] I. Berlin, Two concepts of liberty, in: *Four Essays Lib.*, Oxford University Press, 1969.
- [80] E. Morozov, *To Save Everything, Click Here. The Folly of Technological Solutionism*, Public Affairs, 2013.
- [81] A. van Wynsberghe, S. Li, A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI), *Medicolegal Bioeth.* 9 (2019) 11–21, <https://doi.org/10.2147/mb.s160348>.
- [82] M. Coeckelbergh, Technology and the good society: a polemical essay on social ontology, political principles, and responsibility for technology, *Technol. Soc.* 52 (2018) 4–9, <https://doi.org/10.1016/j.techsoc.2016.12.002>.
- [83] A. Cortina, El estatuto de la ética aplicada. *Hermenéutica crítica de las actividades humanas*, Isegoría 13 (1996) 119–134.
- [84] P. Pettit, *Republicanism. A Theory of Freedom and Government*, Oxford University Press, 2002.
- [85] H. Jonas, *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*, Herder, Barcelona, 2015.
- [86] C. Véliz, *Privacy Is Power: Why and How You Should Take Back Control of Your Data*, Bantam Press, 2020.
- [87] F.D. Schoeman (Ed.), *Philosophical Dimensions of Privacy, An Anthology*, 1984.
- [88] F. Torralba, *La Intimitat*, Pagès Editors, 2009.
- [89] M. Illa Mestre, Proposta d'una polisèmia estructurada del concepte <<Intimitat>>, *Universitat de Barcelona*, 2018.
- [90] M. Aparicio Payá, M. Toboso Martín, T. Ausín Díez, A. Monasterio Astobiza, R. Morte Ferrer, D. López Castro, Un marco ético-político para la robótica asistencial. *An Ethical-Political Framework for Assistive Robotics*, *ArtefaCTos. Rev. Estud. La Cienc. y La Tecnol.* 8 (2019) 97–117.
- [91] A. Andriella, C. Torras, G. Alenyà, *Cognitive System Framework for Brain-Training Exercise Based on Human-Robot Interaction*, *Cognitive Computation* (2020), <https://doi.org/10.1007/s12559-019-09696-2>.
- [92] A. van Wynsberghe, Service robots, care ethics, and design, *Ethics Inf. Technol.* 18 (2016) 311–321, <https://doi.org/10.1007/s10676-016-9409-x>.
- [93] M. Coeckelbergh, *AI Ethics*, MIT Press, 2020.
- [94] R. Prancutė, *Web of Science (Wos) and Scopus: the Titans of Bibliographic Information in Today's Academic World*, vol. 9, *Publications*, 2021, pp. 1–59, <https://doi.org/10.3390/publications9010012>.