



Universiteit
Leiden
The Netherlands

Gathering expert opinions for social robots' ethical, legal, and societal concerns. Findings from Four International Workshops

Fosch Villaronga, E.; Lutz, C.; Tamò-Larrieux, A.

Citation

Fosch Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Gathering expert opinions for social robots' ethical, legal, and societal concerns. Findings from Four International Workshops. *International Journal Of Social Robotics*, 12(2), 441-458.
doi:10.1007/s12369-019-00605-z

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/92271>

Note: To cite this publication please use the final published version (if applicable).



Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops

Eduard Fosch-Villaronga¹ · Christoph Lutz² · Aurelia Tamò-Larrieux³

Accepted: 1 November 2019 / Published online: 20 November 2019
© The Author(s) 2019

Abstract

Social robots, those that exhibit personality and communicate with us using high-level dialogue and natural cues, will soon be part of our daily lives. In this paper, we gather expert opinions from different international workshops exploring ethical, legal, and social (ELS) concerns associated with social robots. In contrast to literature that looks at specific challenges, often from a certain disciplinary angle, our contribution to the literature provides an overview of the ELS discussions in a holistic fashion, shaped by active deliberation with a multitude of experts across four workshops held between 2015 and 2017 held in major international workshops (ERF, NewFriends, JSAI-isAI). It also explores pathways to address the identified challenges. Our contribution is in line with the latest European robot regulatory initiatives but covers an area of research that the latest AI and robot governance strategies have scarcely covered. Specifically, we highlight challenges to the use of social robots from a user perspective, including issues such as privacy, autonomy, and the dehumanization of interactions; or from a worker perspective, including issues such as the possible replacement of jobs through robots. The paper also compiles the recommendations to these ELS issues the experts deem appropriate to mitigate compounding risks. By then contrasting these challenges and solutions with recent AI and robot regulatory strategies, we hope to inform the policy debate and set the scene for further research.

Keywords Ethical, legal and societal issues (ELSI) · Social robots · Human–robot interaction · Autonomy · Privacy · Responsibility · Policymaking

1 Introduction

Social robots—those that exhibit personality and communicate with us using high-level dialogue and natural cues [1]—will soon be part of our daily environment [2, 3]. The application of social robots comes with numerous benefits. Therapeutic robots, for instance, can alleviate daily tasks for older patients or help doctors monitor their patients' recovery. Beane and Orlikowski [4] show that, under certain conditions, the introduction of telepresence robots can help physicians and nurses in post-surgical intensive care units coordinate themselves more efficiently. Other research shows that assistive robots in autism therapy facilitate communication between physicians and patients [5] and provide promising results for patient wellbeing [6].

On the other hand, social robots come with a range of challenges. As European and US surveys on the perception of artificial intelligence and robots illustrate [7–9], such challenges range from liability issues [10] to privacy concerns [11, 12], discrimination [13] and the replacement

✉ Eduard Fosch-Villaronga
e.fosch.villaronga@law.leidenuniv.nl

Christoph Lutz
christoph.lutz@bi.no

Aurelia Tamò-Larrieux
aurelia.tamo-larrieux@uzh.ch

¹ eLaw Center for Law and Digital Technologies, Leiden University, Steenschuur 25, 2311 ES Leiden, The Netherlands

² Nordic Centre for Internet and Society, Department of Communication and Culture, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway

³ Center for Information Technology, Society, and Law (ITSL), University of Zurich, Rämistrasse 74/57, 8001 Zurich, Switzerland

of jobs [14]. In its latest resolution, the European Parliament (EP) alerts on the possibility that robots used in care settings may dehumanize caring practices by reducing human–human interaction [15].

We contribute to the existing literature by exploring the main ELS challenges and recommendations on how to address them, derived from the *Workshop Series on the Ethical, Legal and Social Issues of Robots in Therapy and Education*.¹ Starting as a standalone workshop in 2015 under the *NewFriends First International Conference on Social Robots in Therapy and Education*,² the workshop turned into a workshop series that has brought together researchers from different disciplines and countries over three years to discuss the most pressing ELS issues of social robots in a solution-oriented roundtable format. We organized workshops in Barcelona, Spain, under the *NewFriends 2nd International Conference on Social Robots in Therapy and Education* in November 2016; in Yokohama, Japan, under the *International Symposia on Artificial Intelligence (isAI)*, organized by the Japanese Society for Artificial Intelligence (JSAI) (also JSAI-isAI 2016) in November 2016; and at the Edinburgh Center for Robotics, Scotland, as part of the *European Robotics Forum* on March 2017. The workshops focused on the benefits, risks, and possible solutions to the ELS conflicts that social robots pose. Researchers from different disciplines as well as practitioners actively engaged in the discussions and shared their thoughts on crucial challenges in an interdisciplinary manner. We present the results of the discussions, addressing two main questions:

- What are the key ELS challenges of social robots, as perceived by experts?
- What recommendations can be determined to overcome such challenges?

Our paper contributes in several ways to the existing literature on ELS and social robots (e.g., [16] and the articles compiled in this special issue). First, we provide a one-stop-shop to the topic that spans areas and disciplines. In contrast to literature that looks at specific ELS challenges, often from a certain disciplinary angle, our contribution intends to show current ELS discussions in a holistic fashion. While our overview certainly cannot cover all significant ELS challenges of social robots, we capture a broad overview of essential debates. Second, and connected to the first point, our perspective is shaped by active deliberation with a multitude of experts across four workshops. By listening to the voices of these experts, we provide a more grounded understanding, powered by the collective intelligence of the

participants [17]. The experts pointed us in new directions we would have otherwise ignored, for example, the importance of creating interdisciplinary living labs that can inform policies, as in the Japanese Tokku zones [18]. Third, we explore the challenges and solutions that recent AI and robot strategies have scarcely covered [15, 19–22].

Because robotics terminology is still developing and is not consistent across research and regulatory stakeholders [22], the second section includes relevant definitions. These definitions may be essential to establish a common ground for interdisciplinary discussions and future policymaking, especially if the European Institutions are uncertain about whether we need a legal definition for robot technology [15, 23]. In the third section, we describe the structure of the workshops, including the methods used, participants, and content. Section 4 focuses on answering the research questions established above. We approach each major topic subsequently, with three steps within each topic. First, we capture the challenges discussed by the workshop participants for each topic. Second, we present the recommendations brought forward to address the challenges. Third, we embed the challenges and recommendations into broader discourses on the topic, as per recent human–robot interaction, and ELS literature. The section concludes with a summary of the key discussion points and introduces two meta-challenges: uncertainty and responsibility. Finally, Sect. 5 of the paper summarizes the main implications of the paper and identifies future areas of discussions serving as a road-map for policymakers and roboticists working on the use and development of social robots.

2 Relevant Definitions: Social, Healthcare and Therapeutic Robots

In this paper, we identify a robot as a “movable machine that performs tasks either automatically or with a degree of autonomy” [24]. Since the legal instruments governing robots depend on the context of use (e.g., [25, 26]), we distinguish between industrial robots, which are used for “industrial automation applications,” and service robots, or those robots that “perform useful tasks for humans” (ISO 8373:2012). Among service robots, there are “systems able to perform coordinated mechatronic actions (force or movement exertions) on the basis of processing information acquired through sensor technology, with the aim of supporting the functioning of impaired individuals, medical interventions, care and rehabilitation of patients and also individuals in prevention programmes” [27]. Such robots are named *healthcare robots*.

Since some of the ELS challenges arise from the ways in which robots interact with humans, we distinguish between those robots that interact with us physically, either actively

¹ Cfr. <https://legalrobotics.wordpress.com>.

² Cfr. <https://legalaspectsofsocialrobots.wordpress.com>.

(a powered lower-limb exoskeleton, for instance), passively (e.g., a robotic wheelchair), or socially (i.e., *social robots* or *socially interactive robots*) [28, 29]. Social robots express and perceive emotions, communicate through high-level dialogue, learn and recognize models of other agents, establish and maintain social relationships, and use natural cues such as gaze or gestures [1].

Such robots are mostly used in individual situations rather than team situations, so that the level of shared interaction is generally low, meaning that robots communicate with humans but not with other robots. In Yanco and Drury's taxonomy, such robots refer to type A (human–robot), type D (human-to-human-and-to-robot), and type E (human–robot–human) in terms of the level of shared interaction [30]. The interaction role we are interested in is that of the operator or teammate. The type of human–robot physical proximity is on the high end of the taxonomy spectrum (approaching and touching, rather than avoiding, passing and following). Within the space/time dimension of HRI, such robots afford synchronous and collocated interaction, and the autonomy of such robots is low to medium so that human intervention is still required to support the operation of the robot successfully [30]. If the robot gives aid or support to a human user socially, we identify them as *socially assistive robots* [31].

3 Methods

The workshop series had multiple objectives, and they were organized accordingly:

- Collecting different opinions from all the involved stakeholders, including therapists, roboticists, industry representatives, academics, and legal practitioners working on the use and development of social robots;
- Identify all the ELS concerns, problems, and difficulties concerning social robots, particularly socially assistive robots;
- Enabling discussions on the ELS aspects in an interdisciplinary fashion and on a multicultural scale;
- Providing a comprehensive roadmap of the most pressing issues in this specific domain of research;
- Providing the relevant stakeholders with an informed vision of innovative recommendations to the challenges identified.

In this respect, a call for papers targeting robot engineers, legal practitioners, psychologists, ethicists, philosophers, and cognitive scientists was released. Various means were used for this: social media (Twitter, ‘The iii’ blog, LinkedIn, Facebook), university channels (BI Norwegian Business School, University of Twente, ETH, University of Zurich,

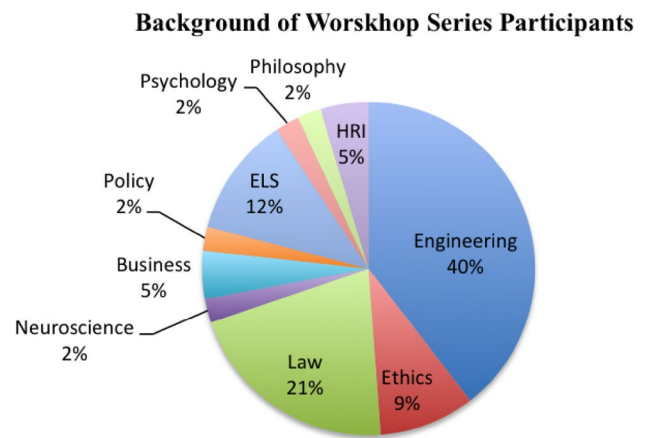


Fig. 1 Background of the workshop series participants

EMJD LAST Programme), and some miscellaneous channels, including euRobotics mailing lists, the ECREA and AoIR mailing lists, LSN, and College of Europe. The organizers created two websites for this purpose.

The workshop organizers aimed at collecting papers covering different ELS issues, such as dignity, privacy, and liability, to discuss these issues in an open workshop format. In total, 43 participants, including the three moderators, took part across all workshops. Except for one participant and the organizers, all remaining participants took part only in one workshop of the series, rather than in multiple workshops. The organizers carefully selected participants according to the relevance of the proposed abstract and its contribution to the field. Several co-authored papers were submitted, and every author that wanted to participate in the workshop was invited to join. However, typically, one author per abstract participated in the workshop, and this was the only person counted. Due to the lack of an official publication opportunity provided by the conferences in which the workshops were held, the participants' papers were not published as a workshop proceeding paper.

The experts' background ranged across various fields of expertise (Fig. 1), and the experts were based on a broad range of countries (Fig. 2).

The topics of the participants' papers referred mostly to healthcare. Concerning the sub-field of research, the most significant percentage of papers related to ‘care’ (21%), although the participants used other keywords (Fig. 3).

The development of the workshops differed every year. In 2015 and 2017, the workshops were organized in a traditional way, where each participant identified what ELS challenges they considered most relevant and pressing. The participants presented their abstract with a particular idea in front of the audience. Afterwards, a moderator discussed recommendations in an open format with the participants (Mattias Scheutz for the workshop in 2015; and Sanja

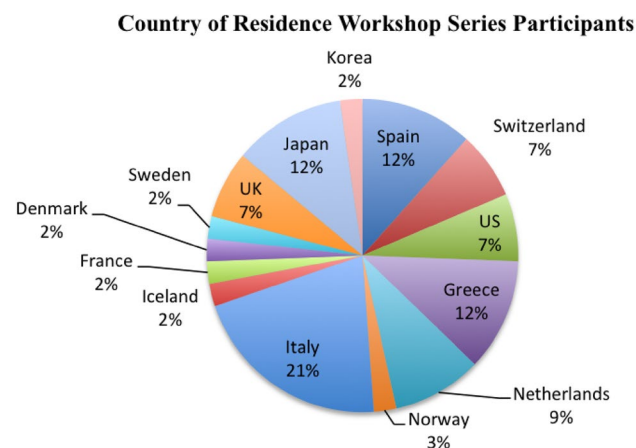


Fig. 2 Country of residence of the workshop series participants

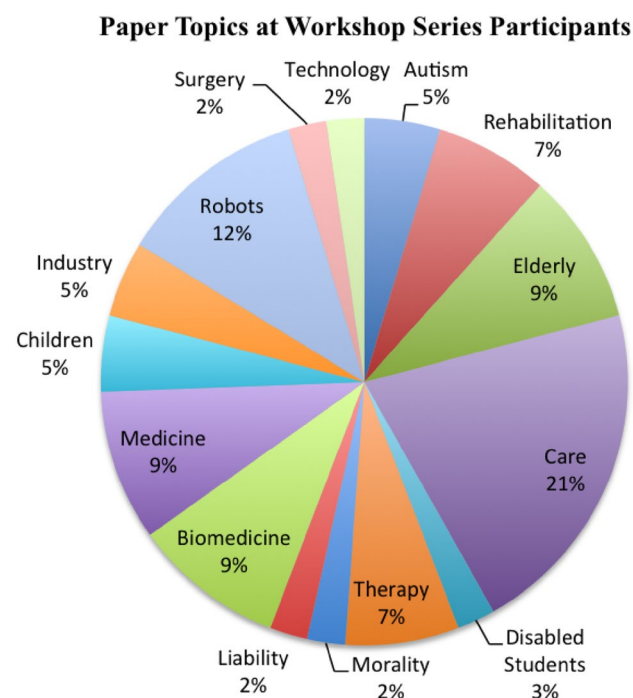


Fig. 3 Paper topics at the workshop series participants

Dogramadzi for the one in 2017). The organizers took notes during the discussion.

The development of the workshop was different during the workshops held in November 2016 (New Friends and JSAI-isAI). They were both half-day workshops, covering four hours with a coffee break of 30 min in the middle. After introducing the case studies to the audience, the workshop organizers asked the participants in a roundtable to elaborate more on their thoughts of the case studies. The three organizers moderated the discussions, including questions, comments, and challenges to the participants. At the same

time, everyone (participants and organizers) was asked to collect thoughts and ideas with sticky notes. The first half of the workshop focused on challenges and issues, and the second half focused on recommendations and future directions. These two workshops were audio-recorded and, after all the participants had consented, included on the workshop website. The recordings were not transcribed and analyzed with a specific methodological approach such as grounded theory. Instead, we opted for a discursive summary approach. We synthesized the findings from the workshops by going through the materials collected throughout the workshops (audio-recordings, personal notes taken during the conversations, pictures of the sticky notes used for brainstorming sessions), discussing and structuring them between authors and categories. This discursive summary resulted in a report that builds the critical document for categorization into the five fields discussed below. Given the lack of a systematic transcription and analysis of the workshop conversations, our paper should not be read as an in-depth qualitative study with ample information and context about the participants. We instead frame it as a position paper, where we also bring in our voices along with the perspectives from the workshop participants. Thus, our paper is similar to other articles in the area of ethics and information technology that relied on a workshop methodology [17, 32].

The participants that attended the workshops in 2016 received a booklet in advance with reading materials and background information. The materials included relevant definitions, facts, and figures of healthcare robot trends, and relevant ethical guidelines and regulation (including the General Data Protection Regulation or the, at that time, the draft resolution of the European Parliament 2015/2103 (INL) on Civil Law Rules on Robotics [15]). The booklet also introduced three case studies to be explored in the workshop: one on an assistive robot for older adults, one on a robot for cognitive therapies, and one on a flying robot for care purposes. A mix of real and fictional information inspired the case studies, which revolved around different legal and ethical issues, including privacy and responsibility. Finally, the booklet also included the extended abstracts of the participants. Altogether, these materials established the grounds for the discussion at the workshop.

The organizers asked each participant to explain their abstract and integrate it with one of the case studies from the booklet. Based on their expertise, the rest of the participants reacted and elaborated on their thoughts. They also classified their thoughts and propositions into different categories. Over the different workshops, the participants identified various challenges, which can be grouped into the following categories: (1) security and privacy, (2) legal uncertainty, (3) autonomy and agency, (4) impact of robots on employment in healthcare, and (5) interaction with humans and dehumanizing practice.

4 Results and Discussion

In the following section, we discuss the five challenges that emerged in the workshops and explain the various sub-challenges identified within those broader topics. As stated, the workshops aimed to discuss ways to address the identified challenges. The organizers asked the participants to deliberate over different approaches that could remedy some of the challenges they brought forward during the debate. They also encouraged participants to consider older technologies and fields of applications, and to reflect on how associated challenges were addressed in the past. After introducing the five challenges, we summarize the recommendations that participants proposed. Finally, for each of the five challenge and recommendation areas, we provide a contextualization that embeds the topic into broader discussions in the human–robot interaction and ELS literature. This section is called “Discussion” for each of the five topics or challenges. Thus, the first two sub-sub-sections (i.e., “Challenges” and “Recommendations”) in each of the five sub-sections reflect the participant voices, giving justice to their experiences and opinions. By contrast, the third sub-sub-section (i.e., “Discussion”) in each of the five sub-sections connects the participant voices to the broader literature, including current approaches to these topics. By separating the participant voices from the broader literature, we can give more voice and room to the emerging and spontaneous themes from the workshop, while reflecting these themes in light of recent scholarship.

4.1 Privacy and Security

4.1.1 Challenges

During the workshops, particularly the workshops at New Friends 2015 and 2016, participants tried to find a common understanding of the concept of privacy. As privacy is a protean concept, the discussion soon switched to the ways social robots can move around, monitor their environment, record their interactions with individuals, and register their daily routines and preferences. In this sense, social robots may affect the “physical and informational” spheres that surround humans. The participants highlighted that the ability to physically reach out into the world in an autonomous fashion enables robots to surveil individuals across places and gain access to personal rooms, which was impossible at this scale before.

Second, and linked to the informational privacy aspect, during NewFriends 2015 a participant pointed to transparency (avoiding secretive systems), access to individuals’ records and their uses as well as privacy controls: How can

we prevent information about oneself from unwanted purposes without consent? Is there an informed consent bias?

The participants also raised their concerns on their ability to control the collected and processed data, mainly referring to the data that robots can collect without human perception (e.g., muscle changes in exoskeleton devices): Is this data also theirs? What control would we have of our personal information? Would we have the ability to decide for ourselves if we want the robot to process such data? Participants discussed these questions during the workshops, especially at NewFriends 2015 and 2016. The discussions often revolved around the concepts of integrity, the ability to correct or amend the data, but also around the question how data misuse can be prevented.

Some participants in the JSAI-isAI workshop brought up the issue of what type of permission social pet robots employed in cognitive therapeutic settings or hospitals require. The participants highlighted the difficulties they had with group therapies, where some of the children had parental permission to give images and data to the researchers and other children did not.

Concerning security, multiple participants were worried about surveillance and security breaches, especially during ERF and NewFriends 2015 and 2016: What are the consequences of robots that are hacked? As the Internet of Things (IoT) community had already identified such events, participants feared that external (and often malicious) agents could control the correct functioning of therapeutic robots. Participants stressed the importance of security in this regard, especially in such sensitive contexts.

4.1.2 Recommendations

One of the participants of NewFriends 2016 believed that the issue of privacy and data protection is overly dramatized and does not reflect reality. The participant argued that we, as end users, need to choose among three attitudes towards the processing of personal data. These attitudes go from permissive to restrictive:

1. We (end users) permit robots (and their manufacturers) to collect our data without any control on our side;
2. We indicate which data can be collected and require robots to make these data anonymous (very much in line with the spirit of the current legislation on privacy in the European Union) as soon as robots collect it; or
3. We prohibit social robots (and their manufacturers) from collecting any data.

While transparency from the data controller would lead users to choose scenario 1—because then the threat to privacy disappears as the users agree on the use of the collected data—the participant argued that the second scenario is the

most likely to happen. She argued that the idea of informed consent follows a more typical data protection approach. According to the participant, scenario 3 is entirely the opposite, as it drives the impression that we should prohibit all data collection at any time. In line with scenario 2, the participants discussed the possibility of pre-defining the risks and establishing processes to identify and mitigate them, in the sense of a more fact-based than a fear-based approach. Concerning 3, beyond the fact that the processing of the data outside of the primary purpose of the robot should be prohibited, some of the participants suggested that the terms and conditions of companies that collect data should be revised to include the principle of data minimization.

If the technology allows for it, one of the organizers mentioned across different workshops that the robot should be privacy-friendly. For example, removing the cameras for drone navigation would reduce privacy impacts; or using vibration detection of the floor when a patient has fallen, instead of recording an image of them, could allow for a more privacy-friendly notification system. Participants in the JSAI-isAI workshop referred to the noise that cameras make in Korea when a person takes a picture. With some modification, this could be adopted for social robots to spread awareness among the users of when and how data is collected. The problem with this solution—and referring to the Korean camera noise—is that there are available apps that allow users to take a picture without making noise. Research is needed to know whether this could still be feasible in robot technology, especially for robots that do not incorporate a robot app store.

Some forward-thinking comments on existing concepts also took place during JSAI-isAI and NewFriends 2016. For instance, some participants suggested the inclusion of an evolutionary modulation of the privacy concept, to reflect the times in which we are living. Others reflected upon the creation of dynamic consent forms: a user could be asked more frequently whether they consent with specific data being processed and would be re-asked for permission when a specific purpose changes.

4.1.3 Discussion

As highlighted by the participants, social robots affect the privacy of individuals differently than current technologies [12, 33–35]. To understand the privacy implications of social robots, it is necessary to distinguish the privacy aspects or types that apply within this context [36]. In that regard, we can differentiate between physical and informational privacy (concerns). Physical privacy refers to the notion of non-interference and the idea of having a private space without prying, surveilling eyes. Social robots challenge physical privacy due to their ability to enter personal spaces [11]. Informational privacy refers

to the ability to control information disclosure and aligns more with a central ideal of European data protection law, namely the concept of informational self-determination. With social robots, not only the collection of personal information becomes ‘murky’—with social robots constantly processing data in the background [12]—but also its processing and dissemination is hard to grasp for an individual user. In light of the anthropomorphic effect of social robots [37], the ability to grasp the data processing and disclosure abilities of a social robot are challenged. Informational privacy concerns not only relate to the interaction between the user and the robot itself but also to the interaction between individuals through a robot, for example when a robot is hacked or surveillance takes place through a telepresence robot [11].

Since the workshops took place, from 2015 to 2017, the European data protection framework has been revised and the General Data Protection Regulation (GDPR) entered into force in May 2018. The GDPR governs the processing of personal data in Europe, whereby both terms “processing” and “personal data” have been interpreted broadly by legal scholars and by courts such as the European Court of Justice [38–40]. While the GDPR has been called “the law of everything” [40] because of its aim to tackle all challenges arising from digitization, it becomes clear when reading its recitals and articles that policymakers’ main concerns when drafting the Regulation were web-based applications, cookies, profiles (also of children), and automated decision-making. The impact of social robots on the “fundamental rights and freedoms, in particular their right to the protection of personal data” (cf. Recital 2 of the GDPR) were not at the forefront of the policy discourse.

Nonetheless, the GDPR contains provisions that will shape how social robots collect and process data of European citizens. First, the GDPR contains *information duties* for data controllers, that is, information they have to provide to the data subject prior to the processing of their personal information (cf. Art. 13 and 14 of the GDPR). The information duties are an ex-ante transparency requirement, unlike ex-post mechanisms such as explanations [41]. Second, the GDPR provides data subjects (i.e., users of social robots) *individual rights* to access the data that the social robots process (Art. 15) or demand erasure of the data that is processed about them (Art. 17). Third, the GDPR codifies the principles of *privacy-by-design* and *privacy-by-default*, which state that systems processing personal data must be designed in a way that fulfills all the requirements set forward in the GDPR. Data controllers must include technical and organizational measures that ensure that the processing principles of the GDPR are met, including security measures such as encryption of data. However, privacy-by-design is not entirely or purely technical, but there are always organizational measures to take into account [42, 43]. The privacy

culture in a company should be extended to cover the life-cycle of the robot, from its creation until its deployment.

4.2 Legal Uncertainty

4.2.1 Challenges

Many of the discussions across the workshops frequently came back to the uncertain legal landscape around social robots. Importantly, the participants highlighted the difficulty of defining what a robot is in legal terms. Because there is no legislation governing the use and development of robot technologies, but there are regulatory initiatives, participants added another legal uncertainty: What is the actual applicable law in this transition period? Some pointed to existing legislation that could apply, although they also acknowledged that this might largely vary depending on the robot type, context of use, and other variables. Relevant legislation ranged from the Directive 2001/95/EC on general product safety to the directive 85/374/EEC on liability for defective products, the recent regulation 2017/745 on medical devices, the low voltage directive 2014/35/EU or even the 2014/30/EU electromagnetic compatibility directive or the 2014/53/EU on radio equipment.

Policymakers have started discussing the legal implications of robots in general but not specifically for social robots, left aside for concrete applications such as therapy or education [15]. Instead, there has been a focus on ethical guidelines. While such an approach is desirable as it promotes an extensive discussion without blocking innovation in a still-emerging field [44], participants, in general, felt that many of their questions would need to be addressed within a more concrete, enforceable format. Not only the enforceability of guidelines and regulations in the field of social robots but also a global understanding of how to create more certainty in dealing with robots was seen as a challenge at JSAI-isAI. In particular, participants on different continents felt that cultural differences might inhibit global guidelines, although they all agreed that guidelines at worldwide scale were an intermediate step towards legal certainty in the field of robots in healthcare and therapy.

At NewFriends 2016, another new uncertainty discussed was whether robots could be certified as helpers or not. This discussion arose because one of the case studies introduced the question: “*What can happen if a robotic system and competent nurse have contradictory approaches to a particular situation?*” During that discussion, questions revolved around whether a robot or an AI system could become a certified nurse or doctor and know what is best for the patient to the same extent as an actual nurse or doctor. The participants highlighted that this would break the hierarchical decision-making currently existing in hospitals, where, most of the times, the doctor decides how to proceed. If

technology advances as the current forecast predict, we may have robots and artificial technologies that surpass human knowledge and that are better at making medical decisions.

4.2.2 Recommendations

Concerning legal certainty, participants agreed on the necessity to harmonize definitions of *robots*, *impact*, *risk*, and legal issues concerning social robots. In particular, they agreed that it is vital to incorporate knowledge about domestic laws when implementing robots in different countries, including knowing the habits and cultures of the place. At NewFriends 2015, one participant mentioned the “regulation-by-design” principle to refer not only to the embedding of a hybrid top-down/bottom-up regulatory model into the system to avoid violations with domestic and international laws, but also to a future compliance-by-default model, where both technical and organizational measures are put in place to make robots comply with the law. The Japanese participants mentioned that the Tokku approach could be a model to follow. This approach connects living labs, where robot technology is tested, with the policymaking process to achieve policies based on the evidence collected in the lab.

The participants also sustained that the creation of an agency that could deal with robots as it occurs in Japan could help improve the management of all the ELS issues. This agency could be responsible for the development of standards and codes of conduct for engineers and other relevant stakeholders. An international expert committee could be created to draft group reports too, and to conduct research on the legal and regulatory implications of such technologies. The agency could also have some certification agencies that could test whether robots are compliant with existing laws, and certify them accordingly.

Both at JSAI-isAI and NewFriends 2016, participants agreed that there should be some purpose limitation. For instance, all the participants at both of these workshops agreed that the use of robots in the military field should be banned, as doing the contrary would increase violence and unfortunate scenarios. In the context of healthcare robotics, however, such clear-cut recommendations were missing. Participants reported positive attitudes towards the future applications of robots and AI technologies in the healthcare domain. The workshop organizers highlighted, however, that this opinion could be biased since all participants worked in related fields.

One participant at NewFriends 2016 wondered what barriers the current regulatory framework posed to the introduction of robots into society. The participant did not believe that the actual legal order can respond to the potential violations of children’s civil rights and suggested the use of multilateral trade agreements like the Trans-Pacific Partnership to change this and provide protection for these children.

Up to now, standards concerning robot technology only focused on physical human–robot interaction, although we interact with robots in many various forms, including cognitively [44]. In a much quicker way than public policymaking, the industry has started to develop standards that address the moral implications of robot technology, including the BS 8611:2016 Guide to the Ethical Design and Application of Robots and Robotic Systems, and the IEEE Ethically Aligned Design 2017 from the IEEE Global Initiative and Standard Association. Participants at NewFriends 2016 also highlighted in this respect that, because every robot is different (due to their machine learning characteristics), it is going to be very difficult to standardize the ELS aspects concerning robot technology.

4.2.3 Discussion

As workshop participants highlighted, robot legislation is overdue. As a result of the EP Resolution 2015/2103 (INL) [15], the European Union is making efforts towards providing legal certainty for robot legal compliance as “doing otherwise would negatively affect the development and uptake of robots” [23]. A recent open consultation launched by the EC [45] acknowledged that current European Harmonized Standards do not cover areas such as automated vehicles or machines, additive manufacturing, collaborative robots/systems, or robots outside the industrial environment, among others [46]. That is why the machinery directive is likely to be revised, to further consider its suitability for new areas of development in types of machinery such as robots and digitization.

The EP resolution did not provide a concrete definition for the word *robot*, although it acknowledged its importance. The EC reacted cautiously, questioning the need to define “cyber-physical systems, autonomous systems (and) smart autonomous robots” for regulatory purposes [23]. Whether a definition is required may depend on the level of uncertainty surrounding a concept, but it may be deemed necessary by legislators if such a concept has to be regulated explicitly [24]. Pointing in this direction, some legal scholars have defined robotics without having an impact on legislation [47, 48]. More recently, the High-Level Expert Group on AI (HLEG AI) has defined robotics as “AI in action in the physical world” or, in other words, as “a physical machine that has to cope with the dynamics, the uncertainties and the complexity of the physical world” [49]. However, it remains unclear whether such a definition will be adopted or whether we should focus on broadly accepted definitions found in technical, international standards (such as the definition given by ISO 8373:2012 “actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks”).

Because the EP Resolution suggested that the development of autonomous and cognitive features in a robot makes the current rules on strict liability insufficient, the EC may revise the Directive 85/374/EEC concerning the liability of defective products.³ Indeed, the EC stated, “advanced robots involve, in addition to the machinery part, complex software systems and suppose the provision of services, data transmission and possibly internet connectivity and, depending on the category of robot, a degree of artificial intelligence.” This intertwining of tangible and intangible parts may further challenge the allocation and the extent of the liability, that is, until what extent the damages are within the autonomous behavior of the robot [24, 50]. In this respect, the EC is exploring risk-based approaches that could set a safeguard baseline for current and future uses and developments of robots and AI. One example would be to use impact assessments appraising the consequences of implementing algorithm-driven technologies [51, 52], notably steered to understand how these affect the humans interacting with the robot [53]. As highlighted during the workshops, however, self-learning capabilities make each robot a unique product, which complicates their assessment—something recently highlighted by the Food and Drug Administration (FDA). Robot technologies process vast amounts of data, can learn from experience, and self-improve their performance. By doing so, they challenge the applicability of existing regulations (e.g., medical device regulations), which were not designed for progressive and adaptive AI. In addition, these capacities make the certification process more difficult.

One solution could be the use of an *obligatory insurance scheme*, as it already exists for cars [15]. However, the progressive and adaptive learning of robot technologies confronts the certainty of insurance schemes [54]. Furthermore, robot applications are vast and their embodiment as well as interaction modes differ widely from one case to another. Thus, it is intricate to understand which robots need insurance and which not, and, consequently, which companies will offer such services [54, 55].

Beyond mere standalone and provisional approaches, Japanese workshop researchers supported the creation of testing zones for robot technology connected to policy-making. The “Tokku” Japanese model could inform policies revolving around robot technology [18]. These testing zones may be in line with the EP Resolution, which also highlights the need for testing zones, although a formal communication process between the robot and regulatory developments does not currently exist in Europe [56]. New European projects directed towards establishing testing zones for exoskeletons

³ Cfr. https://ec.europa.eu/smart-regulation/roadmaps/docs/2016_grow_027_evaluation_defective_products_en.pdf.

seem to point in that direction, although their connection to policy is unclear for now.⁴

4.3 Autonomy and Agency of Robot Technologies

4.3.1 Challenges

Participants discussed lively on the responsibility behind robots' actions across the different workshops. The discussions at NewFriends 2016 and JSAI-isAI revolved around whom to blame in case of a critical event and whether robots can and should be perceived as morally responsible. The latter led to a discussion about how to think about the moral responsibility of robots, merging into an ethical and philosophical debate. Autonomy and the capacity of independent decision-making are what triggered the discussions. The participants at JSAI-isAI highlighted that autonomy is not a fixed term and that it includes different facets, calling it "the 50 shades of autonomy." They were referring to the capacity of robots to make decisions on their own or by adapting to the users' will. In this case, they highlighted that if robots adapt their personality to users, it could reduce serendipity and not challenge users' decision-making processes. However, in the case of autonomous decisions, it was uncertain whether there should be a requirement to establish a hierarchical rule order as mandatory, and whether there should be a human that supervises robot decisions.

This led to the question of whether robots should be able to override human decisions in some instances, for security or safety reasons, because the robot may possess more understanding of the issue than the human. In this respect, and beyond the general concern on how much we are still in control, participants mentioned that this scenario could create a dependency and overreliance effect that would be very difficult to overcome. Some participants stressed that there are always people who need to design and maintain the robot and also to be in charge of programming it. They insisted on the importance of determining who prepares the controls, and how these controls are determined, which could help understand who is responsible in the end. Other participants added that this is going to be exacerbated in co-participatory design projects when the end-users are part of the design of the robot.

Regarding the question of moral responsibility, we encountered contradictory positions in the JSAI-isAI workshop. One of the participants highlighted that high-agency entities are typically thought to have more moral responsibility than low-agency entities. For example, we expect moral behavior from adult humans, but not from animals. The participant argued that robots would not be considered morally

responsible unless people perceived them as agentic. Other participants mentioned that social cues (e.g., gaze, speech, or human-like appearance) influence how agentic people view robots and that we were setting higher moral standards to machines than humans. Another participant suggested that the problem with this approach is that although some discussions at European level had taken place, it is not clear whether robots are going to have agency in the first place, and what kind of agency they might have—animal-like, corporation-like, electronic agent-like type—in the second place.

4.3.2 Recommendations

Participants were very confident on what to do concerning robots' actions: clarity, transparency, and division of responsibilities between developers, manufacturers, adopters (institutions), maintenance, and end-users (particularly if the robot adapts their behavior to the end-user). More similar to a risk-based approach, participants claimed that we should make an effort to avoid unexpected consequences by predicting and modeling agents' behaviors accordingly. Participants highlighted that, in the future, we might rethink the concept of liability, mainly because it is more about prevention than about reaction. To ensure such prevention, some participants referred to the creation of protocols of actuation similar to the ones that exist for airplanes. This way we could limit unfortunate scenarios in advance. To this regard, maybe future protocols are influenced by the specific context. In the case of autonomous vehicles, for example, this might mean that lanes for autonomous cars are created, separating them from non-autonomous cars. This extra requirement could also work as a safeguard for the different types of robot technology such as unmanned ambulances and other care transport, including autonomous ground vehicles in hospitals or robotic wheelchairs.

A compensation fund schema similar to compensation funds for natural disasters for non-human faults was also highlighted at the NewFriends 2016 workshop. The question that arose afterward, regarding possible insurance schemes, was whether all robots needed insurance or not. One of the participants suggested the creation of a matrix to determine liability. According to one participant, who quoted the work of Lohmann [57], this matrix could include the different levels of autonomy of the robot and also the level of complexity of the environment.

This point is related to decision-making processes. The participants mentioned that a decision-making ruler should be established, including (1) a hierarchical model (e.g., doctor's decision primes over the nurse, but not over patients'); or (2) co-participatory with the user. Both include the robot in the loop. In the case of conflicts, some participants referred to dispute resolution systems such as the one on

⁴ See <http://eurobench2020.eu/>.

Wikipedia, although there was not much clarity on how to build that.

The participants said that as well as insurance companies have assigned different values and importance to different objects, including body parts, this approach could be useful for the case of robots, for example, whether a robotic prosthesis is considered equal to a human body part. The participants also acknowledged that this approach should be carefully adopted as it could raise some discriminatory issues in the long term.

4.3.3 Discussion

One of the main problems of roboticists is engagement. If users do not engage with the robot effectively, then it is unlikely that they will interact with the robot for a longer period of time [58] and the return of investment may stagger. Establishing a long-term engagement, however, is not an easy task. Social entities require an empathic personality to establish relationships and permanent attachment between artificial social agents and humans [58]. Personality is formed through unique, imperfect behaviors. The robots may end up being unique, and in some cases disobedient to pre-established rules [59]. The robot dinosaur PLEO, for instance, evolves differently depending on the interactions of the user and its internal non-compliant parameters. From a juridical point of view, this more realistic adaptive behavior of robots poses some questions: To what degree will these robots be in-compliant? What is going to be the tolerance level and in which contexts/circumstances is it going to be avoided? Who is going to determine the consequences for non-compliance?

Disobedient and imperfect robots that enhance long-term engagement completely clash with risk-based approaches proposed by European institutions, as this mainly challenges any certification process, apart from user trust and reliability. Non-compliant robots may also clash with by-design principles including privacy- or ethics-by-design, as changes in the internal parameters of the robot may lead to disastrous consequences in the long term. This may compromise the predictive behavior of the robot, something crucial in some therapies, for example for autism [60].

Whether these autonomous robots will be considered agents with some juridical relevance is something currently under debate. In most jurisdictions, several non-human entities have rights and obligations. Corporations are considered legal persons; nature is protected; animals are considered sentient beings; and even a person not yet conceived, but which can be taken into account for hereditary purposes (*concepturus* in Latin). On its latest proposal, the EP suggested “applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently” [15]. Such a proposal utterly shook

part of the juridical community, to the extent that an open letter was written to the EC asking to focus on the health and safety of humans, and not robot or AI technologies.⁵

Although there might be different strategies to address the responsibility gap [61, 62], the fact that a robot behaves differently from the designers’ intention should not be a reason to exempt them from being held responsible [24, 63, 64]. Such a decision is societally relevant and might have important consequences that need careful thought [23].

4.4 (Lack of) Employment for Humans

4.4.1 Challenges

The participants of the workshops highlighted positive and negative views of the economic impact of robots. A researcher working on robotic wheelchairs highlighted that some human–robot interactions could facilitate elderly individuals to be socially active and motivate them to train and do things by themselves. In this respect, the participant argued that the introduction of specific robots in healthcare could reduce the burden of a caregiver’s workload, meaning that caregivers may have more time for other activities.

Another participant had a contrary vision to this, at least in the context of cognitive therapies. The participant explained that growing dependence on mental therapy and people’s intentional stance for computers encourage the introduction of interactive agents and robots into psychiatry. The participant argued that the introduction of computers and interactive agents for mental therapy might lead to the substitution of human narrative therapists on the one hand and the exacerbation of patient issues on the other hand. Connected to the latter, the participant argued that narcissistic people who wish to talk about themselves while leaving concealed issues hidden in their self-narratives would appreciate these agents as complementing their distorted interpretation of psychological concepts, without achieving therapeutic effects. The participant pointed out that the cultural trend of social reductionism towards psychology might impact patients when interactive agents are introduced in mental therapies.

Some participants highlighted that there is a general belief that robots are going to replace human activity gradually. A participant wondered: Does the widespread implementation and introduction of robots mark the end of our civilized humanity? In general, the rest of the participants were not very fond of the “robots are taking over” scenario, especially because some of the attendants were roboticists developing this technology. Some of the participants wondered why humans needed to surrender to the robot

⁵ See <http://www.robotics-openletter.eu/>.

taking-over scenario and who was in charge of deciding that this will be the case. A participant pointed out that it is crucially important to develop a strategy for the appropriate implementation of these technologies to avoid apocalyptic scenarios and to assure a return of investment of these technologies. Other participants mentioned that the use of robotics in therapy implies in-depth cooperation between engineers and scientists of that particular field, which would increase humans in the loop.

Other participants had a very different opinion on this. Japanese participants argued that seniors in Japan do not want to be looked after by persons that cannot speak proper Japanese, for instance, and that they did believe that robot technology could help sort out the huge problem they have, that is, the lack of workforce in care settings. Although tenderly framed, these participants agreed that the comment conveyed the impression that robot technology was elucidating ongoing soft but deeply rooted racism arguments. In other words, it was suggested that the employment of technology could not only be efficient in economic terms, but also in achieving discriminatory practices. In this respect, other participants wondered what mechanisms exist to evaluate the need to incorporate robots into the workspace globally. They stressed that rehabilitation and social robots have moved from the stage of experimentation, prototyping, and testing to be clinical and educational work tools in our society; and that it had become evident to the robotics community that the usefulness of the robots largely depends on the people who are working with them.

4.4.2 Recommendations

Most of the recommendations called for the incorporation of humans into the design processes of the robot, including end-users, nurses, and doctors in a healthcare context. The participants referred to the design of collaborative systems, the so-called cobots, and highlighted the importance of including such types of robots into the protected scopes of current harmonized standards, which is lacking at the moment.

At the workshop in Japan, some researchers presented their solution. They introduced to the group a collaborative system called “Practical Intelligent Applications (PRINT-EPS).” PRINTEPS is a platform for developing intelligent applications for the co-operation between humans and machines. Via PRINTEPS, software modules can be reconfigured and related to knowledge-based reasoning, speech dialogue understanding, image sensing, and manipulation. To promote human–robot interaction, they use different robots with different purposes: the robot NAO takes charge of imparting knowledge, Sota (another robot) takes charge of progress checking, and Jaco2 takes charge of the development of students’ interest. The researchers found that

preparing the co-operation channels in advance could promote the easy and quick use for each target purpose.

The overall idea during the workshops was the importance of keeping the human in the loop. In line with the GDPR, regarding automated decision-making, the participants were inclined to promote a more pan-human vision of robotics in sensitive contexts.

4.4.3 Discussion

Although much has been said concerning robots and labor, there is little empirical research on how new technologies, and in particular robots and artificial intelligence, affect the labor market. An exception is a recent study that investigated the effects of industrial robots from 1990 to 2007 in local labor markets in the United States concerning employment and wages [65]. The authors argue that, in a simple task-based model where the robot competes against the human in the performance of a specific task, robots have a positive impact on employment and wages because they increase productivity (what they call the productivity effect); but they also have a negative effect, which is the displacement of workers. By regressing the change in employment and wages on the exposure to robots in each local labor market, the authors proved that they can estimate the local labor market effects of robots [65]. They also highlighted that the number of job losses had been limited in the United States, mainly because in this period—1990–2007—there have been relatively few robots. In Germany, a macro-economic study found no evidence that robots lead to overall job losses [65]. However, the increasing introduction of robots shifts jobs from manufacturing to service professions. In addition, the authors found that those manufacturing workers who are most exposed to robots are the ones least likely to be replaced and that the introduction of robots might have a detrimental effect on wages [66]. Frey and Osborne [67: 261] concluded that “occupations that involve complex perception and manipulation tasks, creative intelligence tasks, and social intelligence tasks are unlikely to be substituted by computer capital over the next decade or two.” The BBC application created after this study highlighted that “social workers, nurses, therapists and psychologists are among the least likely occupations to be taken over,” mainly because assisting and caring for other people involves empathy, an essential part of their job. A growing interest in the use of emotions and empathy in human–robot interaction for healthcare purposes might change in the future.⁶ Some researchers have pointed out that by allowing the robot to show attention, care and concern for the user as well as to being able to engage in genuine, meaningful interactions,

⁶ See <https://www.affectiva.com/>.

socially assistive robots can be useful as therapeutic tools [68]. However, the latest research points out that *emotion research* is not quite there yet [69]. Although facial expressions may convey important information for social communication, the way people communicate different emotional states considerably differs between cultures, situations and people within a single situation [69].

The unclear economic impact of robots, particularly in therapy settings, aligns with the discussions in the workshops, where there were varied voices on the topic. However, most experts had a considerate and careful approach, warning apocalyptic scenarios again. Still, specific challenges in care settings such as patient preferences for robots instead of caregivers of certain national or racial backgrounds, as brought up by a participant in the workshop in Japan, merit further academic attention. A practical solution to detrimental economic effects is to consult closely with potential users, focusing on user-centered design (e.g., PRINTEPS project, [70]).

4.5 Replacement of Human Interactions

4.5.1 Challenges

During all workshops, the organizers shared with the audience the concerns of the European Parliament that the inclusion of care robots might decrease human–human interactions. Some participants at JSAI-isAI argued that robots currently only have one-to-one interactions and cannot interact well with a group of people. In this respect, human interactions might not be reduced drastically as social interactions frequently happen in interaction with more than one person.

Some participants raised inter-generational concerns connected to this discussion: How do children, middle-aged and elderly individuals perceive robot technology differently? Participants highlighted that, currently, the general population does not have access to robot technology, although they agreed on the importance to educate the population on its correct use. This comment appeared as a result of the explanation of a Japanese experiment where children beat up a robot repeatedly in a mall [71].

Moreover, participants at NewFriends 2015 pointed to a lack of research about changes in people’s attitudes towards robots over time. A researcher pointed out that there is a lack of research on the perceptions and attitudes of the middle-aged group concerning robotics. Connected to this, the question of an age limit (minimum or maximum) to use robots was discussed, as well as questions of whether therapy robots should be included at earlier stages for humans to adapt to a therapy environment led or shared with social robots.

4.5.2 Recommendations

The participants referred to different corpora that could be a sound basis for framing ELS questions, for instance, the Declaration of Human Rights of the United Nations. The participants acknowledged that although one problem is that Human Rights are interpreted differently in various nations, this could not constitute a barrier to look into it as a possible framework.

Some participants mentioned new ways of using robot technologies. One participant at NewFriends 2016 believed that the current moral damage responsibility framework has a default based on the compensation of non-material damage, which oversees actual circumstances and future consequences (e.g., psychological or social impact) and does not acknowledge the source of damage, which is critical to evaluate the moral damage. In the future, social robots could incorporate learning algorithms oriented through a pattern of the motor, verbal, and psychological, cognitive conversations, to determine whether and how much damage a user experienced.

The same participant envisaged a future where individuals may be forced to deal with therapeutic software agents and robots due to the social pressure that prescribes self-control over emotions and mental health. He suggested a more careful approach for the use of such technologies, not only because there are already some people feeling anxiety towards robots. Also, because these people may experience double-bind situations where mental problems cannot be solved regardless of whether or not they use the therapeutic system (i.e., social pressure prohibits them from removing themselves from the situation).

4.5.3 Discussion

Given the limited adoption of social robots at this time, we lack evidence of the long-term impact of healthcare and therapy robots. The central question of whether long-term interaction with robots reduces human–human interaction is thus hard to answer. However, studies with therapy and caregiver robots, such as Paro, suggest positive effects on users, particularly among specific population groups such as children with autism spectrum disorder and older adults with dementia.

Shibata and Wada [72], in a mini-review of previous research on the topic, show how Paro has the same effect as an animal treatment for seniors. Similarly, Broekens, Heerink, and Rosendal [73] discuss the positive wellbeing effects of introducing assistive social robots to older patients’, despite pointing to the limited generalizability of previous studies. Since most research has been done on select population groups (mostly young and old people), we know less about their impact more generally on social

Table 1 Overview of social robot challenges and proposed recommendations

Challenge	Key aspects discussion	Recommendations proposed
Privacy and security	No common privacy understanding Lack of consent and control over data Group settings complicate privacy	Data minimization and purpose limitation Dynamic consent Transparency and right of explanation: robot communicates data collection Empowerment: easier user access and erasure of data
Legal uncertainty	Moral responsibility Uncertainty whether existing law applies—and which law Cultural differences might inhibit global legislation Certification of (healthcare) robots	Regulation-by-design and compliance-by-default Tokku zones, testing zones and living labs Government robot agency Risk-based approaches and insurance schemes as explored by the European Commission Robot compliance seal
Autonomy and agency	Overriding of human decisions for safety Responsibility in co-participatory design settings Uncertainty on agency: animal-, corporation or agent-like	Clear decision-making rules and dispute resolution mechanisms Compensation funds Distributed liability New infrastructures
Employment effects	Individuals' employment affected Robots-taking-over scenarios prominent in the media Robots replacing low-skilled workforce could strengthen anti-immigration sentiments	Inclusion of humans in the design process/Participatory design Japan: PRINTEPS Keeping humans in the loop
Replacement of human interactions	Problem of HRI in group settings Role of religion Bias and discrimination	Human in the loop Human-centered design

interactions. The workshop participants stressed the need to find out more about middle-aged individuals. They also pointed to the limited capabilities of social robots in group settings, where the robots have to interact with more than one individual. Thus, robots might be more likely to replace one-to-one human interaction than the group or team interaction. In any case, individuals and patients should have the choice to opt out of interacting with robots, especially in double-bind situations.

The European Parliament [15], a recent report by the Rathenau Institute [74], and part of the literature reflect on the fact that the use of robots could exacerbate social isolation and involve a loss of dignity [75]. Depending on how robots are designed, they might promote, exacerbate, or reduce the contact between humans. For instance, a robot interacting with a child under the autism spectrum disorder might end up socializing more with the robot rather than with other peers. However, if the robot is a social mediator, in the sense that it keeps encouraging the child to “ask your classmate, ask your teacher”, then the child will probably understand that to get their goal it is better to learn how to interact with another person rather than with the ‘dumb’ robot [76]. The report calls on the Council of Europe to clarify to what extent the right to respect for family life should also include *the right to meaningful contact* [74].

Engineers often develop robots primarily on the premise that the product should appeal to as many end-users as possible. However, while such a procrustean design choice seems legitimate from a return-of-investment point of view,

it might not do as much justice to end-users who are more concerned about how that device helps them in their particular situation. The lack of personalization and adaptation to user needs might impede the effective engagement and successful implementation of such technologies in the long run, thus damaging both interested parties. There are already different approaches geared towards integrating users into design processes [77, 78]. Such initiatives could facilitate access to a more personalized technology.

4.6 Summary and Meta-Challenges: Uncertainty and Responsibility

Given the breadth and complexity of ELS challenges and recommendations, we provide a summary table of the key discussion points (Table 1) before concluding the article with implications and limitations.

While the key aspects of discussion cluster roughly into the five themes outlined above, two overarching challenges or meta-challenges are worth pointing out. The first meta-challenge is *uncertainty*. It becomes apparent in all themes, yet in different forms. In privacy and security, it relates to a lack of common ground and understanding of the concept itself. In the legal area, uncertainty itself provides the name of the challenge. Uncertainty here refers to a lack of guidance on which laws apply and how the law can help manage expectations, for example through certification. Within the theme of autonomy and agency, uncertainty is tied to philosophical questions on the ontological status of social robots.

When it comes to employment effects, the actual economic impact of social robots in the future is a big uncertainty. Finally, uncertainty in human–robot interaction is apparent in at least two of the three sub-challenges in Table 1: the role of religions and HRI in group settings. Uncertainty as an overarching theme reflects a lack of ethical, legal and social standards for service robot technologies that are deployed in therapy. Such uncertainty opens up opportunities for unethical behavior that leverages information asymmetries and exploits technical or legal loopholes (*see* [79] for a good example how such loopholes have been exposed). Cross-disciplinary research efforts, including technical and non-technical scholars, and the popularization of social robots, for example their introduction in educational institutions, could help address this challenge. A critical public sphere with an engaged media system could further alleviate uncertainties and reveal malpractices, problems, and ethical loopholes.

The second meta-challenge is *responsibility*. Again, this aspect is present across the five themes. In privacy and security, responsibility refers to control, transparency, and accountability. The GDPR has clarified some aspects of responsibility in the area of data protection but many questions remain for the specific technology of social robots [80]. In the legal area and in the themes of autonomy and agency as well as human–robot interaction, responsibility connects to practical and moral questions when responsibility is not clearly attributable. Examples include co-creation between humans and social robots, co-decisions on important matters, or the use of social robots in group settings. Here, the matter of assigning responsibility for both credit (who should get the reward) and blame (who should get the punishment) is murky. Within the theme of employment effects, responsibility questions arise concerning who should be responsible for addressing tensions that arise from automatization through social robots. Recommendations for addressing the meta-challenge of responsibility are based on human-centered design, clear decision-making rules and appropriate sanctions. Across both meta-challenges, it is clear that technological, legal and social approaches need to complement each other.

5 Conclusion

5.1 Implications

In this contribution, we investigated vital ELS challenges of social robots. Drawing on rich discussions among experts across four international robotics conferences, we identified five main areas where ELS issues arise: security and privacy, legal uncertainty, autonomy and agency, employment, and the replacement of human interactions. For each of these areas, we summarized sub-challenges and issues that were

particularly pronounced in the workshop discussions. We identified two overarching challenges or meta-challenges: uncertainty and responsibility. Both showed up across themes but in different configurations and aspects.

Regarding security and privacy, the difficulty of defining privacy adequately, the notion and reliance on informed consent when data collection is pervasive and mostly autonomous, and the complications of privacy and security in group settings were addressed in depth. Concerning autonomy and agency, the issue of moral responsibility of robots came up as well as practical questions whether robots should be allowed to override human decisions when lives are at stake, for example in hospitals. The legal uncertainty and (lack of) employment implications align with recent debates in the literature about automatization [14, 67, 81]. Finally, the challenges within the area of human–robot interaction mainly revolved around social implications of healthcare robots, for example, intergenerational difficulties in robot adoption or potentially decreased human–human interaction.

Apart from raising specific ELS challenges of social robots and more general meta-challenges, this contribution provides viable recommendations to address such challenges. Again, the workshop participants were very creative and had a range of approaches in mind, from technological, to legal, and social, to mitigate the compounding risks these technologies pose. The recommendations targeted different stakeholders: end users, manufacturers, regulators, and legal experts, as well the broader society. Manufacturer-centered recommendations called for more consideration of end-user needs concerning their ethical concerns. For example, privacy-friendlier robots that make users aware about background data collection and give them more control fall within this category. However, the participants also held regulators responsible for overcoming some of the challenges. A best-case brought up were Tokku living labs, as seen in Japan, where robots are tested in realistic scenarios with concrete policy implications in mind. Specific recommendations also called for a stronger collaboration between different stakeholders. A key aspect here concerned user-integration into the design process. Collaborative and participatory design of robots could help overcome skepticism and create trust, thus addressing issues related to HRI and employment.

5.2 Limitations

Our approach comes with certain limitations, three of which are discussed in more detail. First, we chose breadth over depth when identifying the challenges and approaches to address them. Instead of discussing one challenge in great depth, we covered a broad spectrum of ELS challenges at the workshops. This openness led to interesting, sometimes surprising insights. However, it also limited the amount of

time spent on specific challenges or sub-challenges such as privacy or robot–human–interaction. We thus see our work as an encouragement for future research to deepen the five challenge areas identified through dedicated studies, both qualitative and quantitative. Second, the workshops were exploratory and held at different conferences within the robotics field. This makes comparing the workshop results problematic and, due to the small number of participants per field, also prohibits a systematic comparison of disciplines, research cultures, and career stages. However, our goal was not to provide a comparative study but rather to explore ELS challenges and recommendations from an interdisciplinary perspective. Future research from a comparative approach might want to design workshops aimed explicitly at drawing comparisons, for example through focus groups that use sampling according to the background of participants (scholars vs. practitioners), their expertise (experts vs. non-experts), or their application area (users of cognitive therapeutic robots vs. users of physiotherapeutic robots). Third and finally, even though we are confident we identified salient ELS issues and potential avenues for addressing these challenges as of today, the field is rapidly advancing, and additional issues can arise regularly. Future research is thus advised to keep an eye on new ELS issues and discuss them in workshops, through expert interviews, or industry surveys.

Despite these limitations, our research shows the importance of taking into account not only the specificities of the particular field of application but also an international and multidisciplinary approach. Our article contributes to the existing ELS literature by giving an in-depth, empirically grounded, overview of salient ELS issues for emerging healthcare robot and AI technologies and then by providing useful recommendations to these issues. We hope that these findings can inform future robot governance instruments and steer them in the appropriate direction.

Funding First author acknowledges that the writing part of this project is funded by the LEaDing Fellows Marie Curie COFUND fellowship, a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 707404. This study received funding by the Research Council of Norway (project “Fair Labor in the Digitized Economy”, Grant Number 247725) to cover the travel costs of the second author.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent Informed consent was obtained from all workshop participants to have their statements recorded and used for scientific and non-commercial purposes. No violation of any ethical standards took place.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. *Robot Autonom Syst* 42:143–166. [https://doi.org/10.1016/S0921-8890\(02\)00372-X](https://doi.org/10.1016/S0921-8890(02)00372-X)
2. Gupta SK (2015) Six recent trends in robotics and their implications. *IEEE Spectrum*. <https://spectrum.ieee.org/automaton/robotics/home-robots/six-recent-trends-in-robotics-and-their-implications>. Accessed 17 Sept 2019
3. Van den Berg B (2016) Mind the air gap. In: Gutwirth S, Leenes R, De Hert P (eds) *Data protection on the move: current developments in ICT and privacy/data protection, law, governance and technology*. Springer, Amsterdam, pp 1–24
4. Beane M, Orlikowski WJ (2015) What difference does a robot make? The material enactment of distributed coordination. *Organ Sci* 26:1553–1573. <https://doi.org/10.1287/orsc.2015.1004>
5. Atherton J A, Goodrich M A (2011) Supporting clinicians in robot-assisted therapy for autism spectrum disorder: creating and editing robot animations with full-body motion tracking. In: *Human–robot interaction: perspectives and contributions to robotics from the human sciences*. Workshop at robotics science and systems, Los Angeles, CA. <https://faculty.cs.byu.edu/~mike/mikeg/papers/AthertonFinalRSS2011.pdf> Accessed 17 Sept 2019
6. Cabibihan JJ, Javed H, Ang M Jr, Aljunied SM (2013) Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *Int J Soc Robot* 5:593–618. <https://doi.org/10.1007/s12369-013-0202-2>
7. Eurobarometer (2015) Special Eurobarometer 427: autonomous systems. Report published in June 2015. http://ec.europa.eu/public_opinion/archives/ebs/ebs_427_en.pdf Accessed 17 Sept 2019
8. PEW Research Center (2014) Global opposition to U.S. surveillance and drones, but limited Harm to America’s image. Report published in July 2014. <http://www.pewglobal.org/2014/07/14/global-opposition-to-u-s-surveillance-and-drones-but-limited-harm-to-americas-image/> Accessed 17 Sept 2019
9. Eurobarometer (2012) Special Eurobarometer 382: Public attitudes towards robots. Report published in September 2012. https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_382_en.pdf. Accessed 17 Sept 2019
10. Hubbard FP (2016) Allocating the risk of physical injury from “sophisticated robots”: efficiency, fairness, and innovation. In: Calo R, Froomkin AM, Kerr I (eds) *Robot law*. Edward Elgar, Northampton, pp 25–50
11. Calo R (2012) Robots and privacy. In: Lin P, Bekey G, Abney K (eds) *Robot ethics: the ethical and social implications of robotics*. MIT Press, Cambridge, pp 187–202
12. Lutz C, Tamò A (2015) RoboCode-ethicists: privacy-friendly robots, an ethical responsibility of engineers? In: *Proceedings of the 2015 ACM web science conference*, Oxford, pp 1–12. <https://doi.org/10.1145/2786451.2786465>
13. Asaro P (2016) Hands up, don’t shoot! HRI and the automation of police use of force. *J Hum Robot Interact* 5:55–69. <https://doi.org/10.5898/JHRI.5.3.Asaro>
14. Ford M (2015) *Rise of the robots: technology and the threat of a jobless future*. Basic Books, New York

15. European Parliament (2017) Civil law rules on robotics: European parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.pdf. Accessed 17 Sept 2019
16. Lindemann G, Matsuzaki H, Straub I (2016) Special issue on: going beyond the laboratory—reconsidering the ELS implications of autonomous robots. *AI Soc* 31:441–444. <https://doi.org/10.1007/s00146-015-0623-6>
17. Danaher J, Hogan MJ, Noone C, Kennedy R, Behan A, De Paor A, Murphy MH (2017) Algorithmic governance: developing a research agenda through the power of collective intelligence. *Big Data Soc* 4:1–21. <https://doi.org/10.1177/2053951717726554>
18. Weng YH, Sugahara Y, Hashimoto K, Takanishi A (2015) Intersection of “Tokku” special zone, robots, and the law: a case study on legal impacts to humanoid robots. *Int J Soc Robot* 7:841–857. <https://doi.org/10.1007/s12369-015-0287-x>
19. RockEU (2016) Robotics coordination action for Europe report 2016. www.eu-robotics.net/eurobotics/about/projects/2013-2016-rockeu.html Accessed 17 Sept 2019
20. RoboLaw Project (2014) Regulating emerging robotic technologies in Europe: robotics facing law and ethics. https://cordis.europa.eu/project/rcn/102044_en.html Accessed 17 Sept 2019
21. AI Now (2017) AI Now 2017 report. https://ainowinstitute.org/AI_Now_2017_Report.pdf Accessed 17 Sept 2019
22. European Foresight Monitoring Network—EFMN (2008) Roadmap robotics for healthcare. Foresight Brief No. 157. http://www.foresight-platform.eu/wp-content/uploads/2011/02/EFMN-Brief-No.-157_Robotics-for-Healthcare.pdf Accessed 17 Sept 2019
23. European Commission (2017) European Commission’s response to European Parliament’s resolution on civil law rules on robotics. <https://www.eu-nited.net/robotics/news-events/robotics-news/european-commissions-response-to-the-european-parliament-s-resolution-on-civil-law-rules-on-robotics.html> Accessed 17 Sept 2019
24. Fosch-Villaronga E, Millard C (2019) Cloud robotics law and regulation: challenges in the Governance of complex and dynamic cyber-physical ecosystems. *Robot Auton Syst* 119:77–91. <https://doi.org/10.1016/j.robot.2019.06.003>
25. European Parliament (2006) Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0042&from=EN> Accessed 17 Sept 2019
26. European Parliament (2017) Regulation 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02017R0745-20170505&from=EN> Accessed 17 Sept 2019
27. European Foresight Monitoring Network (2008) Roadmap robotics for healthcare for the definition of healthcare robot. http://www.foresight-platform.eu/wp-content/uploads/2011/02/EFMN-Brief-No.-157_Robotics-for-Healthcare.pdf Accessed 17 Sept 2019
28. Breazeal C, Scassellati B (1999) How to build robots that make friends and influence people. In: Proceedings of the international conference on intelligent robots and systems (IROS), Kyongju, South Korea, pp 858–863. <https://doi.org/10.1109/IROS.1999.812787>
29. Dautenhahn K, Billard A (1999) Bringing up robots or—the psychology of socially intelligent robots: from theory to implementation. In: Proceedings of the third annual conference on autonomous agents, Seattle, WA, pp 366–367. <https://doi.org/10.1145/301136.301237>
30. Yanco H A, Drury J (2004) Classifying human–robot interaction: an updated taxonomy. In: Proceedings of the 2004 IEEE international conference on systems, man and cybernetics, The Hague, The Netherlands, pp 2841–2846. <https://doi.org/10.1109/ICSMC.2004.1400763>
31. Feil-Seifer D, Mataric MJ (2005) Defining socially assistive robotics. In: ICORR 9th international conference on rehabilitation robotics, Chicago, IL, pp 465–468. <https://doi.org/10.1109/ICORR.2005.1501143>
32. Zevenbergen B, Mittelstadt B, Véliz C, Detweiler C, Cath C, Savulescu J, Whittaker M (2015) Philosophy meets internet engineering: ethics in networked systems research. In: GTC workshop outcomes paper. https://papers.ssrn.com/sol3/paper.cfm?abstract_id=2666934 Accessed 17 Sept 2019
33. Calo R (2016) Robots in American law. University of Washington School of Law Research Paper No. 2016-04. <http://euro.ecom.cmu.edu/program/law/08-732/AI/Calo.pdf> Accessed 17 Sept 2019
34. Felzmann H, Beyan T, Ryan M, Beyan O (2015) Implementing an ethical approach to big data analytics in assistive robotics for elderly with dementia. In: Proceedings of the 2010 ACM ETHICOMP Conference, Leicester, pp 280–286. <https://doi.org/10.1145/2874239.2874279>
35. Syrdal D S, Walters M L, Otero N, Koay K L, Dautenhahn K (2007) “He knows when you are sleeping”—privacy and the personal robot companion. In: Proceedings of the 2007 AAAI workshop human implications of human–robot interaction, Washington DC, pp 28–33
36. Lutz C, Schöttler M, Hoffmann CP (2019) The privacy implications of social robots: scoping review and expert interviews. *Mob Media Soc* 7:412–434. <https://doi.org/10.1177/2050157919843961>
37. Breazeal C (2003) Toward sociable robots. *Robot Auton Syst* 42:167–175. [https://doi.org/10.1016/S0921-8890\(02\)00373-1](https://doi.org/10.1016/S0921-8890(02)00373-1)
38. Borgesius FJZ (2016) Singling out people without knowing their names—Behavioural targeting, pseudonymous data, and the new data protection regulation. *Comput Law Secur Rev* 32:256–271. <https://doi.org/10.1016/j.clsr.2015.12.013>
39. Clifford D (2014). EU data protection law and targeted advertising: consent and the cookie monster—tracking the crumbs of online user behaviour, *JIPITEC* 5:194–212. <https://www.jipitec.eu/issues/jipitec-5-3-2014/4095> Accessed 17 Sept 2019
40. Purtova N (2018) The law of everything. Broad concept of personal data and future of EU data protection law. *Law Innov Technol* 10:40–81. <https://doi.org/10.1080/17579961.2018.1452176>
41. Felzmann H, Villaronga EF, Lutz C, Tamò-Larriex A (2019) Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data Soc* 6:1–14. <https://doi.org/10.1177/2053951719860542>
42. Koops BJ, Leenes R (2014) Privacy regulation cannot be hard-coded. A critical comment on the ‘privacy by design’ provision in data-protection law. *Int Rev Law Comput Technol* 28:159–171. <https://doi.org/10.1080/13600869.2013.801589>
43. Tamò-Larriex A (2018) Designing for privacy and its legal framework. *Data protection by design and default for the internet of things*. Law, Governance, and Technology Series, 40. Issues in Data Protection. Springer, Cham
44. Fosch-Villaronga E (2016) What do roboticists need to know about the future of robot law? In: New friends conference proceedings, new friends: 2nd international conference on social robots in therapy and education, Barcelona, Spain
45. Spiliopoulou-Kaparia M (2017) The evaluation of Directive 85/374/EEC on liability for defective products and Directive 2006/42/EC on machinery. European Stakeholder Forum—workshop on regulatory challenges for a digitizing industry, Essen, Germany. <https://ec.europa.eu/futurium/en/system/files/ged/b3-spiliopoulou-liability.pdf> Accessed 17 Sept 2019

46. Simmonds P, Brown N, Rentel M (2017) Evaluation of Directive 2006/42/EC on Machinery. Final Report. Technopolis Group. <https://publications.europa.eu/en/publication-detail/-/publication/n/2b213537-25a8-11e8-ac73-01aa75ed71a1/language-en/format-PDF/source-68663524> Accessed 17 Sept 2019
47. Bertolini A, Palmerini E (2014) Regulating robotics: a challenge for Europe. *Upcoming Issues of EU Law*: 94–129. <http://web.jus.unipi.it/summer-lisbon/wp-content/uploads/sites/3/2014/06/Paper-Bertolini-Palmerini.pdf> Accessed 17 Sept 2019
48. Richards NM, Smart WD (2016) How should the law think about robots? In: Calo R, Fromkin AM, Kerr I (eds) *Robot law*. Edward Elgar Publishing Northampton, MA, pp 3–22
49. High-Level Expert Group on AI (2019) A definition of Artificial Intelligence: main capabilities and scientific disciplines. <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> Accessed 17 Sept 2019
50. Weng YH, Zhao STH (2011) The legal challenges of networked robotics: from the safety intelligence perspective. In: Palmirani M, Pagallo U, Casanovas P, Sartor G (eds) *AI approaches to the complexity of legal systems: models and ethical challenges for legal systems, legal language and legal ontologies, argumentation and software agents*. Springer, Berlin, pp 61–72
51. Fosch-Villaronga E (2015) Creation of a care robot impact assessment. *Int J Soc Behav Edu Econ Bus Ind Eng* 9:1913–1917
52. Reisman D, Schultz J, Crawford K, Whittaker M (2018) Algorithmic impact assessment: a practical framework for public agency accountability. AI Now Institute. <https://ainowinstitute.org/aiareport2018.pdf> Accessed 17 Sept 2019
53. Melkas H, Hennala L, Pekkarinen S, Kyrki V (2016) Human impact assessment of robot implementation in Finnish elderly care. In: *International conference on serviceology*, Tokyo, pp 202–206. http://roseproject.aalto.fi/images/publications/JCServe_Melkas.pdf Accessed 17 Sept 2019
54. Bertolini A, Salvini P, Pagliai T, Morachioli A, Acerbi G, Cavallo F, Dario P (2016) On robots and insurance. *Int J Soc Robot* 8:381–391. <https://doi.org/10.1007/s12369-016-0345-z>
55. Fosch-Villaronga E (2019) *Artificial intelligence, healthcare, and the law: Regulating automation in personal care*. Routledge, London
56. Fosch-Villaronga E, Heldeweg M (2018) “Regulation, I presume?” said the robot—towards an iterative regulatory process for robot governance. *Comput Law Secur Rev* 34:1258–1277. <https://doi.org/10.1016/j.clsr.2018.09.001>
57. Lohmann MF (2016) Liability issues concerning self-driving vehicles. *Eur J Risk Regul* 7:335–340. <https://doi.org/10.1017/S1867299X00005754>
58. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2014) Empathic robots for long-term interaction. *Int J Soc Robot* 6:329–341. <https://doi.org/10.1007/s12369-014-0227-1>
59. Konok V, Korcsok B, Miklósi Á, Gácsi M (2018) Should we love robots? The most liked qualities of companion dogs and how they can be implemented in social robots. *Comput Hum Behav* 80:132–142. <https://doi.org/10.1016/j.chb.2017.11.002>
60. Fosch-Villaronga E, Albo-Canals J (2018) Robotic therapies: Notes on governance. In: *HRI 2018 workshop on social robots in therapy: focusing on autonomy and ethical challenges*, Chicago, IL
61. Gunkel DJ (2017) Mind the gap: responsible robotics and the problem of responsibility. *Ethics Inf Technol*. <https://doi.org/10.1007/s10676-017-9428-2>
62. Matthias A (2004) The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics Inf Technol* 6:175–183. <https://doi.org/10.1007/s10676-004-3422-1>
63. Johnson DG (2015) Technology with no human responsibility? *J Bus Ethics* 127:707–715. <https://doi.org/10.1007/s10551-014-2180-1>
64. Bryson JJ, Diamantis ME, Grant TD (2017) Of, for, and by the people: the legal lacuna of synthetic persons. *Artif Int Law* 25:273–291. <https://doi.org/10.1007/s10506-017-9214-9>
65. Acemoglu D, Restrepo P (2017) Robots and jobs: evidence from US labor markets. NBER working paper 23285. <https://www.nber.org/papers/w23285.pdf> Accessed 17 Sept 2019
66. Dauth W, Findeisen S, Südekum J, Woessner N (2017) German robots: the impact of industrial robots on workers. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039031 Accessed 17 Sept 2019
67. Frey CB, Osborne MA (2017) The future of employment: how susceptible are jobs to computerisation? *Technol Forecast Soc Change* 114:254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
68. Shukla J, Cristiano J, Amela D, Anguera L, Vergés-Llahí J, Puig D (2015) A case study of robot interaction among individuals with profound and multiple learning disabilities. In: *International conference on social robotics*, Paris, France, pp 613–622. https://doi.org/10.1007/978-3-319-25554-5_61
69. Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD (2019) Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol Sci Publ Int* 20:1–68. <https://doi.org/10.1177/1529100619832930>
70. Yamaguchi T (2015) A platform preprints to develop practical intelligent applications. In: *Adjunct proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 ACM international symposium on wearable computers*, Osaka, pp 919–920. <https://doi.org/10.1007/10.1145/2800835.2815383>
71. Nomura T, Kanda T, Kidokoro H, Suehiro Y, Yamada S (2017) Why do children abuse robots? *Interact Stud* 17:347–369. <https://doi.org/10.1075/is.17.3.02nom>
72. Shibata T, Wada K (2011) Robot therapy: a new approach for mental healthcare of the elderly—a mini-review. *Gerontology* 57:378–386. <https://doi.org/10.1159/000319015>
73. Broekens J, Heerink M, Rosendal H (2009) Assistive social robots in elderly care: a review. *Gerontechnology* 8:94–103. <https://doi.org/10.1159/000319015>
74. Rathenau Institute (2017) Human rights in the robot age. Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality. <https://www.rathenau.nl/en/digitale-samenleving/human-rights-robot-age> Accessed 17 Sept 2019
75. Sharkey A, Sharkey N (2012) Granny and the robots: ethical issues in robot care for the elderly. *Ethics Inf Technol* 14:27–40. <https://doi.org/10.1007/s10676-010-9234-6>
76. Werry I, Dautenhahn K, Ogden B, Harwin W (2001). Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy. In: *International conference on cognitive technology*, Coventry, UK, pp 57–74. https://doi.org/10.1007/3-540-44617-6_6
77. Chammas A, Quaresma M, Mont’Alvão C (2015) A closer look on the user centred design. *Proc Manuf* 3:5397–5404. <https://doi.org/10.1016/j.promfg.2015.07.656>
78. ISO 9241-210:2010 Ergonomics of human-system interaction—part 210: human-centered design for interactive systems
79. Greenberg A (2017) Watch hackers hijack three robots for spying and sabotage. *Wired*. <https://www.wired.com/story/watch-robot-hacks-spy-sabotage/> Accessed 17 Sept 2019
80. Felzmann H, Fosch-Villaronga E, Lutz C, Tamo-Larriex A (2019) Robots and transparency: the multiple dimensions of transparency in the context of robot technologies. *IEEE Robot Autom Mag* 26:71–78. <https://doi.org/10.1109/MRA.2019.2904644>

81. Brynjolfsson E, McAfee A (2014) *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company, New York

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Eduard Fosch-Villaronga, is a Marie Skłodowska-Curie Postdoctoral Researcher at the eLaw Center for Law and Digital Technologies at Leiden University (NL). Eduard is the co-leader of the Ethical, Legal and Societal Aspects Working Group at the H2020 Cost Action 16116 on Wearable Robots. Eduard investigates the legal and ethical aspects of healthcare robot and AI technologies and has recently published his Ph.D. research in the book 'Robots, Healthcare, and the Law. Regulating Automation in Personal Care' (Routledge, 2019). Eduard is interested in responsible innovation, human-robot interaction, and the future of law and worked previously as a researcher at the Microsoft Cloud Computing Research Center at Queen Mary University of London (the UK, 2018) and as a postdoctoral researcher at the University of Twente (NL, 2017). Eduard holds an Erasmus Mundus Joint Doctorate (EMJD) in Law, Science, and Technology coordinated by the University of Bologna (IT, 2017), an LL.M. from University of Toulouse (FR, 2012), M.A. the Autonomous University of Madrid (ES), and an LL.B. from the Autonomous University of Barcelona (CAT, 2011). Eduard is also a qualified lawyer in Spain.

Christoph Lutz is an associate professor at the Nordic Centre for Internet & Society (BI Norwegian Business School, Oslo). He conducts research in the field of new communication technologies, especially social media. Within this field, Christoph's research interests include

participation, digital inequality, privacy, mobile dating, the sharing economy, and social robotics. In 2016, Christoph obtained his Ph.D. from the University of St. Gallen with a cumulative dissertation on online participation and participation divides in Germany. In 2015, he was a visiting researcher at the Oxford Internet Institute. His background lies in sociology, management & economics, and media science and Christoph has published in a broad range of journals in communication, information systems, management and the social sciences of the Internet (e.g., *New Media & Society*, *Information Communication & Society*, *Journal of Management Information Systems*, *JASIST*, *Mobile Media & Communication*, *Social Media + Society*, *International Journal of Communication*, *First Monday*, *Computers in Human Behavior*, *Journal of Business Research*, *Academy of Management Discoveries* etc.).

Aurelia Tamò-Larrieux is a postdoctoral researcher at the Center for Information Technology, Society, and Law (ITSL) and at the Digital Society Initiative (DSI) at the University of Zurich. Her research interests include privacy, especially privacy-by-design, data protection, social robots, automated decision-making, and artificial intelligence. Aurelia has published her Ph.D. research on the topic of data protection by design and default for the Internet of Things in the book *Designing for Privacy and its Legal Framework* (Springer, 2018). Previously, Aurelia was a guest researcher at the Institute for Pervasive Computing at ETH Zurich and at the Berkman Center for Internet & Society at Harvard University. Her overall research interests include various topics within the interdisciplinary field of law and technology such as privacy or copyright and regulatory developments in these fields. More recently she became interested in robotics and the impact of in particular social robots on the current social and ethical structures. Aurelia has a background in law and economics and holds a Master's degree from the University of St. Gallen.