International Journal of Social Robotics (2023) 15:2005–2018
https://doi.org/10.1007/s12369-023-01005-0

Ethics and Robot Democratization: Reflecting on Integrative Ethics Practices

Jesse de Pagter[1]

Accepted: 21 April 2023 / Published online: 22 May 2023

Abstract
This paper aims to address the need for new integrative approaches to aid the further development of robot ethics as a research field. It presents observations gathered in a collaborative robotics project that pursued an integrative ethics approach. The paper first contextualizes the increasing attention to ethical approaches in robotics development, design and implementation. Several existing considerations and attempts to further integrate ethics into the field of robotics are reviewed, while also providing context with respect to new trends that generate new possibilities for integrating ethical approaches into robotics and human-robot interaction research. Subsequently, the paper elaborates on the specificities of the collaborative robotics project that pursued an integrative approach. The collaborative robot is presented as a specific type of technological artifact, and insight is given into the project's aim to "democratize" this technology. Furthermore, practices required to integrate ethics are reviewed based on the author's observations and experiences in the project. Finally, to contribute to the further development of integrative approaches in robot ethics, these observations are developed into three constitutive components that should aid the further development of integrative robot ethics approaches, namely: increased attention for the way interdisciplinarity is defined in the field of robot ethics, the role and nature of ethical critique, and finally the chances integrative approaches offer for the development of more thorough anticipatory concepts.

Keywords Collaborative robots · Democratization · Interdisciplinary integration · Integrative robot ethics · Research practices

1 Introduction

Current technological advancements in robotics and artificial intelligence (AI) have led to increasing worries over the societal impact of these technologies. Potential ramifications of robots' evolving interactive capabilities constitute a major concern in this realm. In order to mitigate potential negative ramifications of autonomous systems and reduce public anxiety, the need to implement ethical approaches has become an increasingly mainstream narrative in academic, policy and corporate contexts. Even though it is a hopeful and encouraging development that ethics is now a prominent component of the debate around interactive autonomous systems, it is important to remain critical about the actual influence and application of ethical approaches. Whereas ethical thinking is increasingly contributing to the definition of high-level concepts and ideas for the governance of robotics and AI, it is imperative to further solidify the ideas and practices that have emerged as a part of this new discourse [1].

The field of robot and AI ethics is heavily vested in the further establishment of research that can help integrate ethics into the development, design and implementation processes of social robots and other types of autonomous systems. In line with such developments, this paper follows the robot and AI ethics community's mission to establish new pathways for further integrating ethical approaches. A helpful step in fulfilling this goal is to carefully reflect on ongoing attempts to develop new kinds of integrative ethics approaches, while simultaneously building a community of ethicists that have thorough knowledge and experience when it comes to the practices, habits and methods affiliated with the interdisciplinary integration of robot ethics. This paper aims to contribute to these efforts by reporting on

✉ Jesse de Pagter
jesse.de.pagter@tuwien.ac.at

[1] Institute for Management Science, TU Wien, Theresianumgasse 27, Vienna, Austria

the author's involvement in an interdisciplinary project that sought to "democratize collaborative robot technology". After reflection on the project activities, the paper presents the author's observations concerning the particularities of an integrative ethics approach. Based on this, several perspectives for the future of integrative ethics approaches and ideas are developed.

The paper is structured in the following manner. Section 2 contextualizes the need for integrative ethics approaches. This is done by reviewing the increasing attention to ethics in robotics, while examining several existing perspectives on robot ethics that can serve as an inspiration to the paper. Section 3 zooms in on the collaborative robot project central to the observations presented in this paper. The particularities of collaborative robots (or: cobots) are presented, together with the project's intention to "democratize" this technology. The opportunities and challenges that arose during this process are reviewed. Section 4 synthesizes these observations by developing them into a structure comprising three fundamental components for the future of successful integrative ethics approaches: an emphasis on *interdisciplinary methodologies* in integrative ethics, the way integrative ethics can engage in *critique*, and the new modes of *anticipation* that integrative ethics approaches offer. Finally, the conclusion will provide remarks about the future of integrative ethics and its methods, tools and practices in social robotics.

2 Contextualizing the Integration of Ethics

2.1 Ethics and the Anticipation of Future Robots

In order to contextualize the need for further integration of ethics approaches into robotics, it is useful to start with a rather general question: why are the ethics of robotics so widely discussed at this moment? In this regard, it is important to emphasize that the predominant rhetoric around robotics as a technology emphasizes its new, unknown character [2]. Whereas robotic artifacts have a long history of capturing the human imagination in many different ways, the rise of new robotic technologies in recent years has induced expectations of fast-paced and widespread impact in the near future [3, 4]. This has fed the narrative that autonomous robotic systems are likely to exhibit unprecedented properties and abilities, giving them a potentially revolutionary and transformative potential [5]. Accordingly, better futures with robots (e.g. in terms of efficiency or comfort) are often an important component of this narrative, while simultaneously there are a multitude of contradictory and contested understandings of the societal impacts that robotic futures might entail [6].

Thus, many of the hopes, fears and expectations around robotics are related to anticipatory notions that engage with robots' potential societal ramifications and include a wide range of uncertainties with regards to their (future) impact. Prominent examples of this narrative include concerns about the impact of the potential leap in automation on the job market, potential lacunae in the legal system when it comes to robots' autonomy, social robots' ability to deceive their users and so on [7–9]. Even though there are still quite a few types of anticipated robots with interactive capabilities that have not yet been implemented on a wide scale, the possible consequences of new types of interactive robots already yield a range of different concerns. Part and parcel of these anticipatory narratives is the public discourse regarding robots' impact, often combined with concerns about the development of AI technologies [10, 11].

With the rise of these anticipatory notions, the policy-making context has also exhibited substantial attention to the potential impact of interactive robotic systems. In recent years, the governance of autonomous systems with interactive qualities has become an issue of explicit governmental concern [12]. In many cases, this has led to the development of new policy plans involving strategies to facilitate the smooth implementation of such robots in society. When it comes to the potential societal impact of robots' interactive capabilities, the establishment of novel forms of artificial agency has become an increasingly prominent prospect and has started to become defined as a specific policy issue that has already led to proposals for new regulations [13]. Furthermore, this development includes the building of a new consensus on strategies to safeguard the development and implementation of AI and robotic technologies in a human-centered and trustworthy fashion [14].

It is in such a context that the implementation of ethical approaches in the development, design and implementation of robots functions as a prominent solution to the problems and issues that have emerged. Ethical approaches have provided important pathways for establishing and solidifying the idea that robots need to be developed in a sustainable, resilient way in a context where interactions between humans and robots are expected to increase rapidly [15]. Furthermore, value-based, ethical approaches to robotic and AI technologies are often mentioned as useful frameworks that can provide new types of constructive solutions to issues that arise in the discussion on the impact of robots on society. Furthermore, inherent to this narrative is an increasing focus on the regulation, legislation and standardization of robotic and AI technologies, while emphasizing important broader ethics-related topics integral to democratic, open societies, such as human rights, inclusivity, and participation [16].

Nevertheless, even though intentions are laudable and the momentum for ethics is building, there are still many questions regarding the actual implementation and effectiveness of ethical thinking. While it seems clear that governments like the EU are poised to invest in human-centered, ethical robotics and AI, an often-heard criticism of ethical approaches is that their current prominence does not necessarily imply anything about their effectiveness. For instance, AI and robot ethics have been described as potentially "toothless" [17]. In the same spirit, it has been argued that ethical approaches could easily turn into platitudes if the concepts and content of the narratives behind them are not continuously discussed and criticized [18]. Others have gone even further by arguing that ethics itself has become "big business" and that ethical commitments risk culminating into different forms of "ethics washing" [19–21]. Therefore, the current commitment to ethical approaches does not yet guarantee the ethical use, ethical implementation, or ethical behavior of robots in interactive settings.

Despite these issues, from the perspective of robot ethics as a field of research, it is encouraging to see the increasing recognition and application of ethics in robotics. It is important to emphasize here that the social sciences and humanities have long-standing traditions highlighting the need to address social and cultural elements in the development and implementation of (emerging) technologies in general and of robots in particular [22]. The current attention to ethics can be seen as an invitation to further develop already existing discourses about how to foster ethical approaches to social robots, while also motivating the search for integrating practices that can help develop more trustworthy and human-centered robots and robot sociotechnical systems. As a result, the increasing focus on ethics can be understood as an important step that is creating incentives and commitments to construct new, inclusive narratives about the future of robotics.

2.2 Discussing Ethics Integration

Considering the above, further integrating ethics approaches and ethical thinking into the development, design and implementation of robots can be considered a pivotal factor in the successful further maturation of the robot ethics field. Scholars have already developed useful approaches and gathered interesting insights about applied ethics in general. In fact, ethicists have explicitly engaged for several decades with a range of issues and questions that accompany the implementation of ethics in research and innovation processes related to (emerging) technologies [23]. Many of these endeavors in various fields of science and technology could be discussed here, regardless of the specific area they are applied to (e.g. nanoethics, medical ethics, machine ethics) [24]. Since this would exceed the scope of the present paper, the aim is instead to provide concise insight into approaches from the field of robot ethics that can provide some initial ideas on the further integration of ethics.

The last decade has seen a rapid rise in the attention paid to ethical approaches as a prominent component of the general discussion on the socially sustainable, responsible development of robots and other types of artificial interactive agents [25–27]. Many of these ethical approaches aim to raise ethical concerns, often with regard to a specific type of robot [28]. As such, these approaches help to expose potential issues and enrich the considerations and debates around specific types of robots and their sociotechnical configurations. Such frameworks are often accompanied by principled approaches that, as the term suggests, aim to develop ethical principles and guidelines for implementing ethically informed values [29, 30].

Furthermore, in robot ethics, there is generally a strong emphasis on the role values can play in the design and implementation of new technological systems. Such value-based approaches are promising because they provide a way to exert a positive influence on the development of robots themselves, as well as on their implementation in real-world social contexts. Furthermore, these approaches provide interesting examples of the growing interdisciplinary collaborations in the realm of social robots research. From this vantage point, important contributions have been made by developing different value-based frameworks that try to improve the design process of (social) robots. Examples of tangible frameworks include Participatory Design (PD) and Value Sensitive Design (VSD), which place a strong focus on implementing different stakeholders' visions, norms and values [31, 32]. These approaches can help to understand and influence the norms and values that are inscribed into robots, while providing a useful framework for the implementation of ethically sound values [33]. This can help ensure the beneficial design and deployment of these robots and provide a way to analyze potential ethical issues around different types of social robots at an early stage [34].

Furthermore, several approaches focus on the embedding of ethics into technology assessment processes. Technology assessment (TA) is a specific type of study and evaluation of societal impacts of technology (and science) and often engages in foresight analysis. A good example is a TA study on care robots by Michael Decker, who argues that ethics can fulfill a central role in such studies [35]. Another useful example can be found in a paper on the ethics of healthcare robotics, where Coeckelbergh and Stahl aim to "directly and substantially involve ethicists", while also "involving stakeholders in innovation and practice" [36]. They maintain a strong focus on the collaborative element of this type of ethics practice, thereby arguing for the notion of "embedded

ethics", which must be established in a "dialogical and collaborative way". Others have used similar arguments in the context of social robotics [37, 38].

Furthermore, particularly interesting in the context of this paper are two approaches that capture many of the issues that will be presented in later sections. The first is the integrative social robotics approach from Seibt et al. [39–41]. This approach focuses on the development of culturally and ethically sustainable social robotics applications. In their definition of integrative robotics, the authors strongly emphasize the importance of the cultural sustainability of robots. In this context, they argue for integrating existential, ethical and cultural dimensions into the development, design and implementation processes of social robots. Importantly, they state that ethics ''include not only moral but also existential and aesthetic aspects as these matter for what philosophers call "the good life"'' [41]. In order to establish such an interpretation of ethics, they integrate approaches from fields such as anthropology, social phenomenology, social ontology, the philosophy of technology, the history of ideas and art as one large cluster of interrelated ideas, approaches and methods. As such, their goal is to establish ISR as an interdisciplinary "generic model" for implementing these ideas and approaches in social robotics and social HRI.

The second approach that is particularly useful for this paper is the "synthetic ethics" approach by Dumouchel and Damiano [42]. In a recent publication, Dumouchel explicitly contrasts this synthetic approach with other critical ethical approaches in social robotics by describing the attitude of synthetic ethics as "more modest". This modesty mainly pertains to the epistemic position of ethicists in the sense that other approaches, according to Dumouchel, ''seem convinced that they already possess the ethical knowledge that allows them to judge new situations and forms of interaction, sometimes even before they actually arise'' [43]. Instead, they advocate for an approach that is based on a notion of coevolution. That is to say, they focus on the mutual influence between the robot and the user, which can evolve in the course of their interactions. As will become clear later, this very much aligns with the integrative approach pursued in the project that this paper is based on.

2.3 Trends Towards Ethics Integration in Robotics & HRI

Apart from developments within the field of robot ethics itself, recent years have seen a growing number of collaborations where ethicists participate in robotics development, design and implementation. In fact, in research fields like social robotics and human-robot interaction (HRI), ethical considerations and related topics have increasingly become a matter of discussion [44]. Looking at the research and development practices of these fields, ethical considerations seem to fit rather well within the laboratory-based studies that are often used to analyze the interactions between humans and (social) robots. For instance, several attempts have been made to implement ethics as part of autonomous agents' behavior. The term "ethical" in this case often entails the implementation of ethical control and reasoning systems which can be made part of the robot's characteristics and tested as such in interaction research [45–47]. Furthermore, ethical standards and guidelines as a way of assessing and addressing ethical issues also fit well within existing methodological frameworks that are used to study new forms of robots' interactive capabilities.

Moreover, it is important to consider that interactions between humans and robots in real-world scenarios are likely to increase rapidly. Because of that, there are currently more possibilities for implementing new research methodologies and approaches in social robotics and HRI research. These fields are calling for new insights and ideas that can aid the analysis of new, unprecedented types of interactions between humans and robots [48]. Whereas laboratory-based studies are likely to stay relevant in many ways, their limitations are increasingly addressed by roboticists and HRI researchers. For example, Dautenhahn writes that such work cannot address "how *real* people, in *real-world* environments, would interact face to face with a *real* robot" [49]. In that spirit, it is increasingly discussed how one can conduct studies in social robotics and HRI using more qualitative, hermeneutic approaches such as ethnographies, narrative approaches and so on [50–52]. It is exactly in such contexts that robot ethics as a research field enjoys great chances of becoming more thoroughly involved in the creation of new concepts, theories and methodologies in order to contribute to the study of social robots and their interactions. In what follows, the paper will zoom in on a specific case that exemplifies this type of involvement.

3 Ethics Integration in a Context of "Cobot Democratization"

In Sect. 2, it was argued that the quest to increasingly integrate ethics into robotics in different ways can be seen as a vital and promising development for the further development of robot ethics as a field of research. The specific case presented here is based on my participation in an interdisciplinary project that aimed to ''democratize collaborative robots''. This project focused on the implementation of collaborative robot (cobot) technology with the imperative to "democratize" this technology for lay users. It involved collaboration with a consortium of partners from industry, different fields of academia and the non-profit sector and

had therefore a trans- and interdisciplinary character, pursuing both technical and social research goals. In short, researchers from different fields as well as robotics instructors collaborated to achieve the common aim of increasing cobots' accessibility. Specifically, cobot programming and simulation tools were developed and new use cases were implemented.

Building upon the infrastructure and networks of the industry and non-profit project partners, these programming and simulation tools were used in workshops where a wide range of potential users were provided with the ability to play and work with cobots. Crucially, many of these workshops took place in the open, accessible context of a makerspace. Makerspaces are collaborative environments where members can work on DIY projects and share ideas and practical information with others. Crucially, makerspaces often have a range of different relatively expensive machines available (e.g. different kinds of 3D printers, laser cutters, etc.), which makes them very attractive for artists, do-it-yourself enthusiasts, so-called "hackers", "makers" and so on. Often, the term "makerspace" is affiliated with other terms like "hackerspaces" and "fablabs". Finally, alongside the makerspace, other cobot workshops took place in the context of a factory training facility, where trainees had the ability to take cobot courses to increase their familiarity with the field of robotics.

My mission as an ethicist and social scientist was to study the practices and interactions that emerged during the project. The results of the first project stage, focusing on the exchanges between different technical cultures, have already been published elsewhere [53]. This paper addresses the second mission, namely to understand cobot democratization as a vehicle for the further integration of ethical notions as well as a critical view on the roles integrative ethics can play in these kinds of projects. In what follows, I present observations gathered during an ethnographic study conducted via participation in many of the hands-on cobot workshops and project brainstorming sessions, as well as a range of conversations with individual members of the project [54, 55]. Furthermore, it is useful to note that the project in general pushed me towards a closer involvement with the robotics and HRI community (e.g. through participation in conferences).

3.1 The Cobot and its Promises

One of the most interesting issues in the context of this paper's general goal was the fact that I became embedded in an environment characterized by good intentions. That is to say, with reference to the considerations regarding integrative social robotics by Seibt et al. and their notion of ethics as "the good life" [41], it is interesting to note that the project's core members were genuinely engaged with some version of this idea of "the good life" in relation to cobots. They were strongly engaged with questions concerning the future of robots and the opportunities and risks connected to that future. Furthermore, in general they were very aware of the controversies around robotic technologies, as they were confronted with these controversies on a regular basis.

In other words, this was not a project in which robotics engineers and designers develop and test a specific robot in a closed, laboratory-based environment, while the ethicist figures out the ethical risks and opportunities. Instead, it was a case of the aforementioned development in robotics characterized by new robotic technologies moving into real-world interactive environments. Furthermore, most of the project's laudable values were already defined before I was invited to join the project. This meant that in the project itself, there was not necessarily that much to add from the conventional position of an ethicist who describes ethical issues and prescribes certain ethical values. As a consequence, my observations entail almost no application of theoretical ethical notions about robots or cobots, but rather provide a descriptive account rooted in a situational and contextual approach that looks at ways to improve the ethicist's integration of the larger project.

Having said that, in order to further report on the intentions of the project and the values it represented, it is useful to first describe the central artifact of the project: the cobot. In short, cobots are types of robotic arms, designed specifically to work alongside humans in shared workspaces [56] (see Fig. 1). The project's imperatives were heavily rooted in an understanding of cobots as a prominent example of new developments in robotics, since they combine the technology of traditional industrial robotic arms with novel, advanced sensor technologies. Cobots are designed to perform tasks while directly interacting with humans in so-called "collaborative modes" [56]. Furthermore, current cobot developments point in a direction where cobots are able to engage in new kinds of interaction in order to improve these modes of collaboration. For instance, the *Sawyer* cobot was designed to have eye gaze, in order to use this to improve collaboration during handovers between humans and cobots [57]. In general, the use of social cues in order to establish and maintain interactions with human co-workers is an important area of research in robotics and HRI and with the further development of these technologies [58, 59]. For instance, some point in a direction where cobot technology increasingly merges with applications of social robotics technologies in care settings. In that context, social robots have already been branded as "an evolution of the collaborative robot" [60]. However, to explain the role of the collaborative robot arm in this specific project in a bit

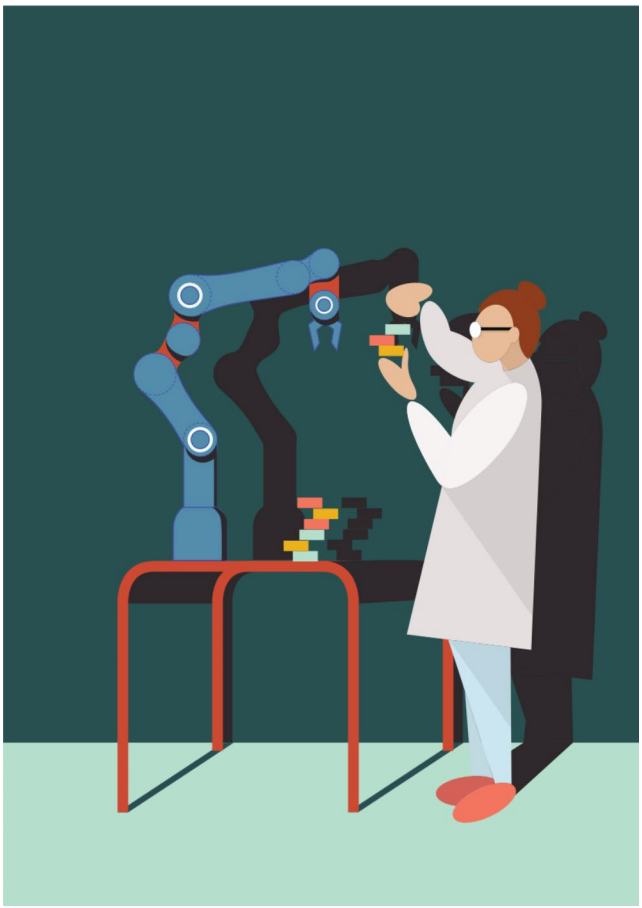

Fig. 1 Human / cobot collaboration concept image

more detail, it is useful to provide a more detailed description of the differences between cobots and conventional industrial robots:

Industrial robots have been known to be highly capable in specific settings for a long time (approximately 60 years). The most well-known example of their capability and utility can be found in manufacturing facilities, where robots have long executed pre-programmed motions. When humans are involved, these types of conventional industrial robots are strictly isolated by safety fences and other protection measures based on rigid safety regulations. This pre-programmed and highly protected setting makes industrial robots suitable for quite specific kinds of industrial production, especially in cases where high levels of accuracy and repeatability are key. Furthermore, it is important to note that industrial robots are expensive and require high levels of expertise in order to be introduced in a feasible manner.

In contrast, the promise of cobot as a specific type of robotic artifact lies in its versatility and safety, which gives it potential for an unprecedented variety of collaborations between humans and collaborative robotic arms [56]. First and foremost, this is rooted in their ability to safely engage in interactive scenarios. As a consequence, they can be implemented in a wider range of settings. Simultaneously, new developments have made it possible to develop types of cobots that are considerably cheaper than conventional industrial robots. In other words, the cobot can be seen as a type of robotic arm that almost perfectly meets the increasingly prominent goal of accessibility in robotics. This was a crucial component of the project, and the roboticists and robot trainers would emphasize these qualities over and over again when discussing the wider implications of cobot technologies.

Finally, whereas conventional robots require extensive (and expensive) training in order to be able to operate them, the very design of cobots fosters a situation in which working or playing with them does not require any programming background. Instead, many cobot systems have very accessible interfaces and can also be moved manually to implement the desired trajectory. This type of increased accessibility allows for a range of potential new types of use. Accordingly, besides accessibility, for this specific project the notion of creativity was central to the definition of the cobot, as it allowed for a wider range of users to engage with robotics and find new ways of using the cobot in line with their imaginative capacities.

This idea of the cobot as a tool for creative engagement fit particularly well with the makerspace context. Users did not need to possess high levels of expertise in order to work with the cobot. This opened up new forms of sociotechnical potential in terms of the possibilities that users have. Or, as one of the trainers argued: “the cobot has opened up the possibility of a playful approach to robotic arms”. For instance, the cobot was used in an art project to cut out materials. Another example was an experimental project with the cobot where someone approached the makerspace to reserve the cobot in order to program it to give neck massages.

3.2 Democratization as a Vehicle for Ethical Discussion

This brings me to a second driving component of the project: the notion of democratization. Apart from the general promises that define the cobot as a technological artifact, the intention to “democratize” these cobots was central to the project’s composition as well as its trajectory. In many ways, the term was used as a bit of a buzzword that connected different interests in the project, while also providing a framework to get project funding. As such, it also functioned as a narrative framework centered on improving the interaction and collaboration between lay persons and collaborative robots. Crucially, this was also particularly important for my work as an integrative ethicist in the situational and contextual meaning of the term, as it allowed me to explicitly deploy the concept of democratization itself

and the narratives around it as a way of eliciting useful discussions on broader ethical topics around cobots and robots in general.

Furthermore, the concept of democratization played an important role for the project itself because it facilitated the combination of technical aims (development of intuitive programming and simulation tools for cobots) and social aims (implementing and using these tools to make cobots more accessible to lay users). As such, especially for several roboticists in the team, the notion of democratization simply represented the opportunity to develop and explore more use cases for cobot technology. In this regard, democratization also meant that several hurdles had to be overcome, most prominently the re-interpretation of safety regulations around robots. An important goal in the earlier stages of the project was therefore to see how safety could be maintained in a more loose environment in order to promote more creative types of cobot use. By connecting this to the notion of democratization, new use cases for cobots could be developed that allowed for freer interaction between humans and robots. Interestingly for some of the project members, this also entailed an invitation to reprogram existing cobot software and make it openly available as a browser based interface. One member would explicitly draw connections between the notion of democratization and the idea of "hacking" cobot technology. He explained this as a way to open up the space of who "decides upon the design" of the cobot, while simultaneously bringing robot technology outside of the rigid settings of academic robotics and mechatronics by implementing it in a makerspace and seeing what novel kinds of applications would emerge from that.

During the project, it turned out that this mission of hacking the cobot was not always materializing. In the later stages of the project the notion of democratization included more references to emancipation through technological skill and increased familiarity with robotics: by learning to use cobots in this democratized context, people would be able to enjoy the advantages of cobot technology. Furthermore, it gave them access to a future with robotics, or, as one of the project members nicely stated: "cobots can show people that robotics is fun and can be creative instead of scary". Thus, for the members that were mostly engaged with using the cobot as a training resource in the makerspace and factory training center, cobots provided new possibilities for increasing the general public's familiarity with actual, physical robots.

In this regard, a prominent issue that repeatedly emerged during my observations and conversations was the trainers' frustration with people's prejudices towards the cobot. Time and time again, the teachers found that the very image of the cobot was linked to a wide range of expectations and fears people associated with robotics in general. One of the teachers explained that he/she deliberately looked for other ways to describe the cobot: "I am starting to tell people in my workshop, don't call it a robot, call it a repetitive precision positioning machine!" The idea being that the cobot was more like a sophisticated tool rather than anything resembling the science fiction imagery people had in mind. Another roboticist argued that we need to "contrast actual robotics to the fantasy of the world population: even my mother tells me now that robots will take over".

3.3 Reflecting on Ethics Integration

Having explained the central components of the project, in what follows, I reflect on the main issues that defined how ethical notions were integrated into the project.

3.3.1 The Status of Ethics Expertise

First of all, I would like to address the role of the ethicist's expertise. In general, expertise is often considered a useful instrument in interdisciplinary interaction with other fields (Collins et al. 2007). A common practice and important task for ethicists is the assessment of (potential) ethical issues related to the technology at hand. Typically, such assessments can be made on the basis of existing guidelines and principles, or can lead to the creation of new ones. Especially in the project phase in which topics are still being delineated, ethicists are generally well-equipped to provide a comprehensive overview of the current discourse with regard to ethical and social issues in a given area. These considerations can be very useful for establishing the general direction of such projects, especially ones that have both technological and social objectives.

In the project, the main issue regarding ethics expertise arose with the assessment of ethical risks. For such kinds of assessments, it is first important to consider that cobots (as well as many other robots) are rooted in the confluence of many different technologies. For instance, the interactive qualities of cobots can address many different topics, from privacy and safety, to transparency and human dignity. When it comes to interdisciplinary integration in team settings in terms of expertise, it is important to have quite specific knowledge about particular ethical issues in order to bring them up when they are relevant. This is a challenge, especially in settings that have quite a technical character, such as the development of cobot applications. What adds to the problem is that in these cases it is often not clear beforehand what to expect, simply because the ethicist (usually) does not have the technical knowledge, background or intuition to predict certain future steps in the development process.

Furthermore, I experienced many ambivalences with regard to the very definition of ethical expertise in an integrative setting. Partly because of the growing attention to robot ethics, people who work with robots are very well aware of many issues related to the potential ethical impacts of robots. On the one hand, this is fruitful and encouraging, since it makes it easier to discuss ethical topics. On the other hand, especially when there is no clear need for ethical assessment, it is rather unclear what the ethicist can deliver in terms of expertise. In conclusion, an ethics background is certainly useful for assessing certain issues, but discussions on the actual role of the ethicist in interdisciplinary contexts would constitute a clear improvement.

Finally, it is useful to keep in mind that the terms "ethics" and "ethical" have a range of different meanings. Accordingly, it turned out to be crucial not to take anything for granted, even with respect to a very basic understanding of philosophy of ethical expertise. For instance, many notions about the relation between technology and society or technology and culture, as well as how norms and values are embedded in technology, were often not shared by the other experts. Instead, I found it fruitful to engage in discussions where ethical topics were in no way predefined, but rather came up as components of larger discussions. Especially here it was important to note that many ethical notions are based on ideas that roboticists see as fantasies. As an ethicist it was therefore useful to argue with a strong awareness of different understandings between professional communities regarding the fantasies and narratives that surround robotics.

3.3.2 Disciplinary Contexts and their Difficulties

Furthermore, it is crucial to give a bit more attention to the fact that roboticists and HRI researchers often have fundamentally different disciplinary backgrounds compared to ethicists. That is to say, even though roboticists and HRI researchers also have quite different disciplinary backgrounds (ranging from engineering to psychology), they tend to work in similar lab-based environments, and as such, their activities are rather well integrated and their objectives often match relatively well. Results in these fields are therefore often presented in the form of research that at least aspires to be reproducible and is subjected to rigid, often quantifiable quality criteria. During the encounters in the context of this project, I often noticed a general belief that such rigid criteria lead to the construction of better robots that can better engage in safe, intuitive interactions.

Crucially, in comparison to robot ethics, fields like (social) robotics and HRI generally subscribe to different ideas about methodology and rigor: robots are often developed in the protected environment of the robotics lab, which means that robotics and HRI researchers mostly work in the context of quantitative laboratory studies, often based on methods derived from experimental psychology. Ethicists cannot achieve such quantified rigor in their approach and methodology. In terms of methodology, this can easily lead to fundamental incommensurabilities that can hamper the ethicist's interdisciplinary integration. In that regard, in addition to one's skills as an ethicist, it is useful to be able to engage in activities that can aid the constructive development of such projects.

Finally, it is also important to realize that people working in the field of robotics (as roboticists and/or robot trainers) often identify with narratives focused on improving the human condition via technology. Instead of taking a critical attitude towards that world, like most ethicists do, they engage with it through their investment in the development of new types of technologies. The art of an integrative ethics approach is thus to see how the embeddedness of ethical topics is implicitly recognized by developing pathways for conversation and interdisciplinary collaboration.

3.3.3 Robot Narratives and the Importance of Larger Frameworks

Finally, the assessment of ethical implications in cases like this is very much a matter of application and the scale in which robotics is implemented. The roboticists and trainers were well aware of this and continuously emphasized that they were trying to work towards solutions where cobots could function as a useful addition to society. In a context focused on integrating ethical approaches as well as the interdisciplinary integration of me as an ethicist, I noticed that it was mostly out of place to assess potential ethical issues that had not yet materialized in any way, even though they could indeed emerge in the future. An important reason for this was that interdisciplinary teams with specific objectives are looking for solutions to reach their objectives, rather than engaging in deliberations about what their technical solutions could potentially lead to in the future.

For that reason, I found it much more useful to invest in discussing common narratives about robots that could help to place the project's activities within a larger framework of societal discussions about robots. For instance, such narratives could help to embed a project in actual societal developments while also discussing possible inclusive futures with robots. In this way, buzzwords like "democratization" related to social issues of robots could function as a very useful avenue for interdisciplinary engagement since they provided a common ground for discussion. As such, they functioned as starting points for ethical discussions that were not principled but exploratory. This turned out to be useful for gaining insights into perceptions of robots and

how their futures were anticipated by relevant experts and stakeholders.

4 Discussing the Future of Integrative Practices in Robot Ethics

Departing from the experiences and observations described above, three key components for the process of ethics integration as part of the further development of robot ethics are elaborated below. The first concerns the interdisciplinary character of integrative ethics: how can the further development of interdisciplinary work and accompanying methodologies aid the process of ethics integration? Second, the critical component of ethics integration: how can integrative ethics pursue and maintain a critical attitude? Third and finally, the role of integrative ethics in the general task of reflecting on the value commitments involved in technological futures: how can a more integrative ethics help to anticipate and foster a responsible future with robots? Apart from their relevance for the further development of integrative ethics approaches as an important development within robot ethics, these three components capture crucial debates about the position of ethics and the future of the philosophy of technology in general [61–63]. Each of the following subsections connect the specific discussions in this paper to these general debates and dilemmas in robot ethics.

4.1 Interdisciplinary Methodologies

Philosophy of technology is a field in which interdisciplinarity and related notions of collaborative practice have long been a central matter of consideration and discussion. An important driving notion of the field is that technologies shape humans' practices and perceptions of their environment. By critically analyzing the socio-cultural values that constitute sociotechnical systems, the main idea is that one can trace how values emerge and understand how value-based decisions are taken. In these efforts to describe and analyze technology and society as co-constructed phenomena, researchers have made many fruitful attempts to build close affiliations with other researchers and work on the basis of dynamic exchanges with a wide range of disciplines. Within the realm of philosophy of technology, the so-called "empirical turn", "practice turn", and "policy turn" are all interesting examples of this development. More specifically, robot ethics itself is in many ways a successful case of the establishment of an interdisciplinary field that has developed many different linkages to the actual practices of robotics development, design and implementation [64, 65].

As a consequence, much work in robot ethics draws on this familiarity with other epistemic communities. Being a successful, experienced ethicist thus often entails not only a firm theoretical background in ethics, but also skills in interdisciplinary practice and communication. The idea is that the process of building interdisciplinary practices and vocabulary can foster more profound collaborations between ethicists, engineers and other relevant experts. In order to enhance and refine integrative approaches to robot ethics, it is first and foremost fruitful to simply reflect on the difficulties that arise throughout interdisciplinary work. That is to say, not only the outcomes of such endeavors should be discussed, but also the practices behind them so as to reveal the particularities of implementing integrative ethics practices. Furthermore, emphasis must be placed on the social components of this integrative process. It is therefore crucial that ethicists have specific, well-developed practices at hand to become more involved in the very process of robotics development itself.

The use of qualitative methodologies can be seen as a promising instrument for further integration. In terms of methodology, (qualitative) empirical research methodologies are usually associated with the social sciences and different fields of the humanities, but less so with ethics, as a branch of philosophy. A prominent reason for this is that ethics is often considered to be a theoretical, philosophical endeavor. Whereas philosophical analysis is certainly often based on observations, there is generally less of an urge to gather such observations in a systematized manner guided by a specific empirical methodology. The goal here is not to argue that this needs to be changed drastically, as the theoretical, conceptual work of philosophers often provides many important insights. Furthermore, a theoretical approach allows for types of fruitful, thorough conceptual analysis that become less feasible if subjected to the standards of (qualitative) empirical methodologies. Nevertheless, in terms of the practices that are involved with an integrative approach, it is useful to think about methods from related fields.

Thus, in very practical terms, with respect to the ethicist's social integration within an interdisciplinary project team, it can be useful to place more emphasis on qualitative empirical methods that help to explore novel aspects of real-world human-robot interaction. As already stated, even though qualitative approaches are not mainstream in robotics and HRI research, they are increasingly being considered useful as more robots are moving into real-world scenarios. In short, they can help to study interactions with complex, autonomous robots in unrestricted (or less restricted) contexts. By using such methods, ethicists can make direct contributions to the project goals in interdisciplinary contexts. By involving themselves in robotics

and HRI research, ethicists are likely to find more profound ways to immerse themselves in the messy, contingent forms of interaction and use those methods to look for emergent ethical themes and issues. These types of research methods can then be employed to trace, describe and criticize the values (and politics) that are either deliberately or indeliberately inscribed into robotic technologies. By observing interactions between humans and robots in real-world scenarios, integrative ethics practices can be pursued in a collaborative manner and make direct contributions in an interdisciplinary setting.

4.2 Collaboration and Technology Critique

On a general level, ethics as a branch of philosophy represents a specific type of critical reflection, seeking to promote actions that pursue moral good, while simultaneously discussing the nature of that good. Drawing on these aspirations, most ethical work done in a context of philosophy of technology has a fundamentally critical character [66]. Likewise, ethical reflection on specific technologies generally seeks to develop and define a certain critical attitude towards these technologies and the practices affiliated with them. That is to say, ethicists often provide their insights in a critical fashion, and several scholars have explicitly discussed the inherently critical component in technology ethics [67, 68]. This task of ethics is still very much alive within the empirical turn in contemporary philosophy of technology [69]. In short, an important component of ethicists' general contribution to the debate on technology is their theoretical and applied criticism, either through the deployment of existing ethical theories or by proposing novel or adjusted ethical notions and approaches. Therefore, this critique and the emancipatory ambitions affiliated with it are pivotal for robot ethics as a field and approach and will normally play a central role in its argumentations.

This critical component of ethics can lead to practical issues when it comes to ethics integration, because the very aim of integration can render it more difficult to remain critical and develop a normative stance. As already described in Sect. 3, critique is an ambivalent topic in contexts where cooperation and integration stand central. One possible solution is to strictly decouple constructive interdisciplinary cooperation and critical engagement. In that case, the ethicist would collaborate within a given context and provide specific inputs that are constructive to the collaboration and its development. Simultaneously but separately, observations from such collaborations can be used to criticize the ways in which certain realities and practices emerge in research, design and implementation processes while assessing their ethical implications. Such a decoupled approach can help to initiate debates with and within the robotics community about the artifacts they produce and the implicit values involved therein.

However, when it comes to ethics integration in an interdisciplinary setting, there is one major issue with this decoupled approach: by developing critique in this way, ethicists are mostly assigning moral responsibility to the designers and developers of the technological artifacts, less so to themselves. That is to say, ethicists would be taking a distanced, critical position that exempts them from responsibility for the outcomes of the very activities they collaborated and participated in. For that reason, it is not an approach that can be seen as truly integrative, since the locus of critical engagement is explicitly placed outside of the integrative setting. Therefore, in addition to this distanced critical stance, more collaborative strategies are needed for the advancement of integrative robot ethics practices. Such practices are likely to be more valuable and necessary at a time when ethicists are being asked to provide insights into situations where robots are increasingly entering real-world settings.

First of all, the ability to discuss critical views and implement them within the integrative context itself will always be a crucial skill. For that, it is important to not only focus on clearly delineated ethical issues, but also to include notions referring to the wider political economy of automation. Moreover, an approach that conceptualizes the sociotechnical configurations of a project can in itself be framed as a form of societal critique. In this specific case, the notion of cobot democratization with the associated technologies and practices was in several cases a critique towards the way robotics development and implementation is configured at present. Preferably, these sociotechnical configurations would be accompanied by well-developed and reflective narratives about how societal critique is established through such a specific configuration. In the project at hand, the topic of democratization was a useful example of this, since it functioned on the one hand as a simple buzzword, but also had the potential to be developed into a narrative that facilitated more critical concepts of cobot development and implementation in the larger context of the political economy of autonomous systems. As the observations above demonstrate, this requires the ethicist to have a high level of familiarity with what other experts in a project are doing in order to develop critical concepts that can guide robotics research, design and implementation.

4.3 Strengthening Anticipatory Concepts

Finally, in the broader context of technology criticism, anticipation of technological developments and their impact are generally an important topic of consideration and have been the subject of repeated debates [70, 71]. An important

point is whether or not and in what way technology criticism should be engaged in anticipating future technological developments, with the position of ethics playing a central role in such discussions [72]. One of the major difficulties of anticipatory notions is that they often have a speculative character. Important to consider in this regard is that a range of historical examples demonstrate that the futuristic hype surrounding emerging technologies often fails to materialize [73]. Given that one could credibly argue that robotic technologies are subject to similar hype, the question is how their potential impact on the economy and society should be approached.

In this regard, speculative thinking can be very fruitful for crafting new, emancipatory narratives about the future of robotic technologies in their sociotechnical context. Furthermore, extensive research into the role of sociotechnical imaginaries has already demonstrated their usefulness for exposing and questioning the values implicit in the futures of emerging technologies [74, 75]. Anticipation of emerging technologies can therefore be a fruitful endeavor, simultaneously recognizing the sociotechnical potential of robots and critically engaging with the hubris and buzzwords associated with them [76, 77]. Much is still unclear about the possibilities and sociotechnical potential of implementing specific robotic technologies in real-world interactive settings. Furthermore, as was argued at the beginning of the paper, robotics in general is a field that has attracted a wide range of expectations surrounding future technological developments. Many of these expectations lead to a mystification of robotics futures, which can hamper the current decision-making processes that are so important for the further development and implementation of robotics [5].

One of the promises of integrative ethics approaches is that they can make substantial contributions to develop new modes of demystified, constructive anticipation of interactive robotics technologies and their future development. The idea being that more profound levels of interdisciplinary integration can help ethicists develop more sophisticated, better informed narratives around technological artifacts and their futures. Thus, through integrative ethics practices, the hope is to come to better-informed anticipatory concepts that are simultaneously both constructive and critical. An important task for philosophers and social scientists here is to become involved in the construction of shared narratives that explore the sociotechnical potential of robot futures [78]. Other studies have already argued that narrative strategies encouraging individual and collective constructions of the meanings of a technology can help to increase societies' capacity to develop common solutions to the highly contingent challenges of new technological systems [79, 80].

Integrative approaches should therefore aim for active engagement with the practicalities of robot development, design and implementation in order to collaboratively construct more sophisticated and informed notions that anticipate the future of robotics. This is not only a contribution to robot ethics in a narrow sense, but also in a broader sense that can help to understand and analyze the political economy of robots, thereby using ethical integration to foster new frameworks of societal discussion on robots. By developing well-informed (speculative) concepts, integrative ethics can help bring about diverse, inclusive narratives concerning the future of interactive robotic technologies. It is thus important to focus on issues of collective imagination that involve the systematic development of alternative futures with robots based on deliberative consensus [81].

5 Conclusion

This paper contends that an integrative approach entails an emphasis on the establishment of the ethicist's presence and the interdisciplinary integration of ethical approaches in design, development and implementation processes regarding social robots. The main reason for using the term "integrative" is that this paper maintained a strong focus on the practices involved in this endeavor. In this regard, it was demonstrated how the ethicist's role is not yet clearly delineated in such contexts. The role of expertise is crucial here: in a professional setting, the presence of the ethicist and the interdisciplinary integration affiliated with it requires a type of expertise in which the ethicist looks beyond already existing pathways of integrating ethical expertise. This relates to the second emphasis, which is on emergence: topics and issues related to ethics must be identified in the messy, contingent, highly variable process of human-robot interaction. In the end, the integration of ethics is clearly unfinished and will probably never be conducted in a completely flawless manner. Discussions, whether over the effect, (inter) disciplinarity, or intentions of ethics, will continue to take place. However, many signs indicate that ethics can gain from the current state of affairs. From regulation to robotics practitioners: there is an increasing recognition of the need to include ethical insights and approaches. Therefore, frameworks need to be developed in which integrative and potentially also experimental aspects of ethics can be further explored and implemented.

One crucial thing to state once more is that the arguments made in this paper are almost solely focused on providing insights into the process of interdisciplinary integration. This has implications in the sense that most statements followed from observations in a context where such interdisciplinary integration was a relevant and intuitively sound endeavor. However, there are many other contexts where it is better to pursue other aims. In such cases, there is likely

to be less of an emphasis on integrating ethics, but rather on a confrontational use of ethics as a tool and narrative for pursuing more disruptive avenues of action in order to achieve long-term change [82]. For instance, when ethical approaches concern basic human rights in the context of artificial agents that are close to being implemented on a massive scale (e.g. in the context of Big Tech), it is much more likely that a detached, confrontational approach can be useful to elicit a society-wide debate on ethical topics. In such cases, assessment activities based on solid knowledge of the moral and legal foundations of those human rights will probably lead to more tangible results than an integrative approach.

Finally, in terms of future research, there are several pathways that could lead to useful insights. First, for the ethics community as a whole and technology ethics more specifically, the discussion on the integration of ethics can lead to new types of ethics frameworks. Second, the notion of democratization itself could be developed further in this context, in a critical but constructive manner. That is to say, the term democratization, as explained above, was strongly focused on bringing robots into new, real-world contexts. Instead of very controlled environments like the robotics laboratory or the factory setting, the idea behind democratization was to allow for more free interaction between people and robots. In such contexts, integrative ethical approaches have great potential to provide a larger framework that can analyze and foster the socio-ethical implications of such developments.

Acknowledgements The author would like to thank the members of the CoMeMak Project for the interesting collaborations. Furthermore the author would like to thank Rumena Trendafilova for designing Fig. 1.

Funding Open access funding provided by TU Wien (TUW). This work was conducted in the context of the CoMeMak Project (FFG Project Number 871459) and was also supported by the TrustRobots Doctoral College, TU Wien.

Data Availability Due to privacy concerns, the observation data cannot be made openly available.

Declarations

Conflict of Interest The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. De Pagter J (2023) From EU Robotics and AI governance to HRI Research: implementing the Ethics Narrative. Int J of Soc Robotics. https://doi.org/10.1007/s12369-023-00982-6
2. Heffernan T (2019) Cyborg futures: cross-disciplinary perspectives on artificial intelligence and robotics. Springer, Berlin
3. Gasparetto A (2016) Robots in History: Legends and Prototypes from Ancient Times to the Industrial Revolution. In: López-Cajún C,Ceccarelli M (eds) Explorations in the history of Machines and Mechanisms. Springer International Publishing, Cham, pp 39–49
4. Murphy RR (2021) Robots have grasped and manipulated the imagination since 1839. Science Robotics 6:. https://doi.org/10.1126/scirobotics.abi9227
5. Suchman L (2019) Demystifying the Intelligent machine. In: Heffernan T (ed) Cyborg futures: cross-disciplinary perspectives on artificial intelligence and robotics. Springer, Berlin, pp 35–61
6. Heffernan T (2019) Fiction meets Science: Ex Machina, Artificial Intelligence, and the Robotics Industry. In: Heffernan T (ed) Cyborg Futures. Springer, Berlin, pp 127–140
7. Bryson JJ, Diamantis ME, Grant TD (2017) Of, for, and by the people: the legal lacuna of synthetic persons. Artif Intell Law 25:273–291. https://doi.org/10.1007/s10506-017-9214-9
8. Gunkel D (2019) How to survive a Robot Invasion: Rights, Responsibility, and AI
9. Manyika J, Lund S, Chui M, et al (2017) Jobs lost, jobs gained: workforce transitions in a time of automation. McKinsey Global Institute
10. Operto S (2019) Evaluating public opinion towards robots: amixed-methodapproach. Paladyn, Journal of Behavioral Robotics 10:286–297. https://doi.org/10.1515/pjbr-2019-0023
11. Zhang B, Dafoe A (2019) Artificial Intelligence: American Attitudes and Trends. Social Science Research Network, Rochester, NY
12. Firlej M, Taeihagh A (2021) Regulating human control over autonomous systems. Regulation & Governance 15: 1071–1091. https://doi.org/10.1111/rego.12344
13. Gahnberg C (2021) What rules? Framing the governance of artificial agency. Policy and Society 40:194–210. https://doi.org/10.1080/14494035.2021.1929729
14. European Commission (2019) COMMUNICATION Building Trust in Human-Centric Artificial Intelligence
15. AIHLEG (2019) DELIVERABLE Ethics Guidelines for Trust worthy AI
16. European Commission (2020) WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust
17. Rességuier A, Rodrigues R (2020) AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society 7:2053951720942541. https://doi.org/10.1177/2053951720942541
18. Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nature Machine Intelligence 1:501–507. https://doi.org/10.1038/s42256-019-0114-4
19. Richardson K (2019) The business of Ethics, Robotics, and Artificial Intelligence. In: Heffernan T (ed) Cyborg Futures. Springer, Berlin, pp 113–126
20. Bietti E (2020) From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. In: Proceedings of the

2020 Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, NY, USA, pp 210–219
21. Phan T, Goldenfein J, Mann M, Kuch D (2021) Economies of Virtue: the circulation of 'Ethics' in Big Tech. Science as Culture 0:1–15. https://doi.org/10.1080/09505431.2021.1990875
22. Brey P (2012) Anticipatory Ethics for Emerging Technologies. Nanoethics 6:1–13. https://doi.org/10.1007/s11569-012-0141-7
23. Bunge M (1977) Towards a Technoethics. The Monist 60:096–107. https://doi.org/10.5840/monist197760134
24. Reijers W, Wright D, Brey P, et al (2018) Methods for Practising Ethics in Research and Innovation: A literature review, Critical Analysis and Recommendations. Sci Eng Ethics 24:1437–1481. https://doi.org/10.1007/s11948-017-9961-8
25. Boden M, Bryson J, Caldwell D, et al (2017) Principles of robotics: regulating robots in the real world. Connection Science 29:124–129. https://doi.org/10.1080/09540091.2016.1271400
26. Bösl DBO, Bode M (2018) Roboethics and Robotic Governance – A Literature Review and Research Agenda. In: Ollero A, Sanfeliu A, Montano L, et al (eds) ROBOT 2017: Third Iberian Robotics Conference. Springer International Publishing, Cham, pp 140–146
27. Winfield AFT, Jirotka M (2018) Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philos Trans A Math Phys Eng Sci 376:1–13. https://doi.org/10.1098/rsta.2018.0085
28. 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
29. Friedman B, Hendry DG, Borning A (2017) A Survey of Value Sensitive Design Methods. Found Trends Hum-Comput Interact 11:63–125. https://doi.org/10.1561/1100000015
30. Stahl BC (2021) Concepts of Ethics and their application to AI. In: Stahl BC (ed) Artificial Intelligence for a better future: an ecosystem perspective on the Ethics of AI and emerging Digital Technologies. Springer International Publishing, Cham, pp 19–33
31. van den Hoven J, Vermaas PE, van de Poel I (2015) Design for values: an introduction. In: van den Hoven J, Vermaas PE, van de Poel I (eds) Handbook of Ethics, values, and Technological Design: sources, theory, values and application domains. Springer Netherlands, Dordrecht, pp 1–7
32. Wynsberghe A (2013) A method for integrating ethics into the design of robots. Industrial Robot: An International Journal 40:. https://doi.org/10.1108/IR-12-2012-451
33. van Wynsberghe A (2013) Designing Robots for Care: care centered value-sensitive design. Sci Eng Ethics 19:407–433. https://doi.org/10.1007/s11948-011-9343-6
34. Umbrello S, Capasso M, Balistreri M, et al (2021) Value Sensitive Design to achieve the UN SDGs with AI: a case of Elderly Care Robots. Minds & Machines 31:395–419. https://doi.org/10.1007/s11023-021-09561-y
35. Decker M (2008) Caregiving robots and ethical reflection: the perspective of interdisciplinary technology assessment. AI & Soc 22:315–330. https://doi.org/10.1007/s00146-007-0151-0
36. Stahl BC, Coeckelbergh M (2016) Ethics of healthcare robotics: towards responsible research and innovation. Robotics and Autonomous Systems 86:152–161. https://doi.org/10.1016/j.robot.2016.08.018
37. Battistuzzi L, Sgorbissa A, Papadopoulos C, et al (2018) Embedding Ethics in the Design of Culturally Competent Socially Assistive Robots. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp 1996–2001
38. McLennan S, Fiske A, Tigard D, et al (2022) Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Medical Ethics 23:6. https://doi.org/10.1186/s12910-022-00746-3
39. Seibt J (2016) "Integrative Social Robotics"-A New Method Paradigm to Solve the Description Problem And the Regulation Problem? What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016 290:104
40. Fischer K, Seibt J, Rodogno R, et al (2020) Integrative Social Robotics Hands-on. Interaction Studies 21:145–185. https://doi.org/10.1075/is.18058.fis
41. Seibt J, Damholdt MF, Vestergaard C (2020) Integrative social robotics, value-driven design, and transdisciplinarity. Interaction Studies 21:111–144. https://doi.org/10.1075/is.18061.sei
42. Dumouchel P, Damiano L (2018) From Moral and Lethal Machines to Synthetic Ethics. In: Living with Robots. Harvard University Press, pp 170–206
43. Dumouchel P (2022) Ethics & Robotics, Embodiment and Vulnerability. Int J of Soc Robotics. https://doi.org/10.1007/s12369-022-00869-y
44. Wullenkord R, Eyssel F (2020) Societal and ethical issues in HRI. Curr Robot Rep 1:85–96. https://doi.org/10.1007/s43154-020-00010-9
45. Malle BF (2016) Integrating robot ethics and machine morality: the study and design of moral competence in robots. Ethics Inf Technol 18:243–256. https://doi.org/10.1007/s10676-015-9367-8
46. Malle BF, Scheutz M (2017) Moral competence in Social Robots. In: Machine Ethics and Robot Ethics. Routledge
47. Bringsjord S, Arkoudas K, Bello P (2020) Toward a General Logicist Methodology for Engineering Ethically Correct Roborts. pp 291–297
48. Park CH, Ros R, Kwak SS, et al (2020) Editorial: towards Real World Impacts: Design, Development, and Deployment of Social Robots in the Wild. Frontiers in Robotics and AI 7:. https://doi.org/10.3389/frobt.2020.600830
49. Dautenhahn K (2018) Some brief thoughts on the past and future of Human-Robot Interaction. J Hum-Robot Interact 7:4:1–4:3. https://doi.org/10.1145/3209769
50. Chun B (2019) Doing autoethnography of social robots: Ethnographic reflexivity in HRI. Paladyn, Journal of Behavioral Robotics 10:228–236. https://doi.org/10.1515/pjbr-2019-0019
51. Hasse C (2019) The multi-variation approach: cross-case analysis of ethnographic fieldwork. Paladyn, Journal of Behavioral Robotics 10:219–227. https://doi.org/10.1515/pjbr-2019-0017
52. Weiss A, Spiel K (2021) Robots beyond Science Fiction: mutual learning in human–robot interaction on the way to participatory approaches. AI & Soc. https://doi.org/10.1007/s00146-021-01209-w
53. Ionescu TB, de Pagter J (2022) "Meet your personal cobot:" framing Participatory Research in Makerspaces as a trading zone. The Journal of Peer Production 1–18
54. Van Maanen J (2011) Tales of the field: on writing ethnography, 2nd ed. University of Chicago Press, Chicago
55. Woods DD, Hollnagel E (2006) Joint Cognitive Systems: patterns in Cognitive Systems Engineering. CRC Press, Boca Raton
56. International Federation of Robotics (2018) Demystifying Collaborative Industrial Robots. Frankfurt, Germany
57. Faibish T, Kshirsagar A, Hoffman G, Edan Y (2022) Human preferences for Robot Eye Gaze in Human-to-Robot Handovers. Int J of Soc Robotics. https://doi.org/10.1007/s12369-021-00836-z
58. Ende T, Haddadin S, Parusel S, et al (2011) A human-centered approach to robot gesture based communication within collaborative working processes. In: 2011 IEEE/RSJ International Conferenceon Intelligent Robots and Systems. pp 3367–3374
59. Terzioğlu Y, Mutlu B, Şahin E (2020) Designing Social Cues for Collaborative Robots: The Role of Gaze and Breathing in Human-Robot Collaboration. In: Proceedings of the 2020 ACM/IEEE International Conferenceon Human-Robot Interaction. Association for Computing Machinery, New York, NY, USA, pp 343–357

60. Giansanti D (2021) The Social Robot in Rehabilitation and Assistance: what is the future? Healthcare 9:244. https://doi.org/10.3390/healthcare9030244
61. Kline R (2013) Teaching social responsibility for the Conduct of Research. IEEE Technology and Society Magazine 32:52–58. https://doi.org/10.1109/MTS.2013.2259331
62. Nordmann A (2020) The Advancement of ignorance. In: Maasen S, Dickel S, Schneider C (eds) Techno Science Society: Technological Reconfigurations of Science and Society. Springer International Publishing, Cham, pp 21–33
63. Rommetveit K (2022) Post-truth imaginations: New starting points for Critique of Politics and Technoscience. Taylor & Francis
64. Scheutz M (2013) What is Robot Ethics? [TC spotlight]. IEEE Robotics Automation Magazine 20:20–165. https://doi.org/10.1109/MRA.2013.2283184
65. Veruggio G, Operto F (2017) Roboethics: a bottom-up Interdisciplinary Discourse in the field of Applied Ethics in Robotics. In: Machine Ethics and Robot Ethics. Routledge
66. Rainsford D, Woods T(1999) Critical Ethics. Palgrave Macmillan, London; New York
67. Stahl BC, Doherty NF, Shaw M, Janicke H (2014) Critical theory as an approach to the ethics of information security. Science and Engineering Ethics 20:675–699
68. Waelen R (2022) Why AI Ethics is a critical theory. Philos Technol 35:9. https://doi.org/10.1007/s13347-022-00507-5
69. Brey P (2010) Philosophy of technology after the empirical turn. Techné: Research in Philosophy and Technology 14:36–48. https://doi.org/10.5840/techne20101416
70. Grunwald A (2010) From speculative nanoethics to Explorative Philosophy of Nanotechnology. Nanoethics 4:91–101. https://doi.org/10.1007/s11569-010-0088-5
71. Nordmann A (2007) If and then: a critique of speculative NanoEthics. Nanoethics 1:31–46. https://doi.org/10.1007/s11569-007-0007-6
72. Nordmann A (2014) Responsible innovation, the art and craft of anticipation. Journal of Responsible Innovation 1:87–98. https://doi.org/10.1080/23299460.2014.882064
73. Edgerton D (2008) The shock of the old: technology and global history since 1900. Profile Books, London
74. Roberson TM (2021) On the Social Shaping of Quantum Technologies: an analysis of emerging expectations through grant proposals from 2002–2020. Minerva. https://doi.org/10.1007/s11024-021-09438-5
75. Selin C (2014) On not forgetting futures. Journal of Responsible Innovation 1:103–108. https://doi.org/10.1080/23299460.2014.884378
76. Bensaude Vincent B (2014) The politics of buzzwords at the interface of technoscience, market and society: the case of 'public engagement in science.' Public Underst Sci 23:238–253. https://doi.org/10.1177/0963662513515371
77. Conceição CP, Ávila P, Coelho AR, Costa AF (2020) European action plans for science–society relations: changing Buzzwords, changing the agenda. Minerva 58:1–24. https://doi.org/10.1007/s11024-019-09380-7
78. Bain PG, Hornsey MJ, Bongiorno R, et al (2013) Collective futures: how projections about the future of Society are related to actions and attitudes supporting Social Change. Pers Soc Psychol Bull 39:523–539. https://doi.org/10.1177/0146167213478200
79. Miller CA, O'Leary J, Graffy E, et al (2015) Narrative futures and the governance of energy transitions. Futures 70:65–74. https://doi.org/10.1016/j.futures.2014.12.001
80. Tavory I, Eliasoph N (2013) Coordinating futures: toward a theory of Anticipation. American Journal of Sociology 118:908–942. https://doi.org/10.1086/668646
81. Owen R, Macnaghten P, Stilgoe J (2012) Responsible research and innovation: from science in society to science for society, with society. Science and Public Policy 39:751–760. https://doi.org/10.1093/scipol/scs093
82. Sætra HS, Coeckelbergh M, Danaher J (2021) The AI ethicist's dilemma: fighting Big Tech by supporting Big Tech. AI Ethics. https://doi.org/10.1007/s43681-021-00123-7

Jesse de Pagter is a PhD candidate in the TrustRobots Doctoral College at the TU Wien, Vienna. He is located at the Institute for Management Science. His work is focused on gaining a better understanding of the relation between society and technology, with a focus on Robotics and AI technologies.