Building trust in robots: A narrative approach

Jesse de Pagter

Abstract

This contribution proposes a narrative approach to trust-building with regards to robots. This should serve as a complementary interpretation in order to find new ways of theorizing and studying the trust-building process. The first aim of the paper is to distinguish between already existing notions of trust-building in relation to robots. I provide an overview arguing that with respect to building trust, robots are currently conceptualized as agentic interaction partners, as artifacts in sociotechnical systems that can be altered based on novel engineering and design processes, and finally as a type of technology that can potentially disrupt existing normative and legal conventions. From this overview, this paper proposes the complementary approach based on a narrative conceptualization of robots. This conceptualization focuses on the way that robots capture the imagination of many, arguing that this is fruitful to take into account when theorizing and studying the process of building trust in robots. The paper then discusses how this conceptualization can be developed in interdisciplinary research in the social sciences by evaluating and analyzing future imaginaries, developing anticipatory concepts, and facilitating access to sociotechnical potential.

Keywords

Anticipatory ethics, Robot ethics, Sociotechnical potential, Technological imaginaries, Technology narratives, Trust in robots

1 Introduction

Arguments that emphasize the need to build people's trust in robots have become increasingly prominent in recent years [European Commission 2019; Glikson and Woolley 2020; Ryan 2020]. A main reason for this perceived need to build trust is the expected increase in the use of robotic technologies across a wide array of domains. Examples of this range from the application of robots for personal assistance to self-driving cars and new types of robots in the workplace. Rapid advances are being made in technologies pivotal to this development, such as sensing technologies, machine vision, and machine learning. These technologies have granted robotic artifacts with increasing abilities to act autonomously and safely in real-world environments and expectations are that this trend will continue in the near future. This means that people's encounters with robots are likely to increase, as is public attention to the question of robots' impact on people's lives [Yang et al. 2018]. Given this context, it is not a surprise that the question of building trust in robots has attracted increased attention: lack of trust in an emerging technology like robots can have disruptive effects both on technological development itself and also on general trust in society [Frewer 1999].

The question of building trust in robots has become prominent in a wide variety of contexts. For instance, the need to build public trust in robotic technology has become more prominent due to the lessons learned regarding the societal impact of different emerging technologies in recent decades [Bunde et al. 2022;

https://doi.org/10.34727/2022/isbn.978-3-85448-052-5_12

Edelman 2020; Ethics Advisory Group 2018]. Further impetus can be found in the expected increase in interactions with robots, raising questions regarding how trust can be built in these interactions [Lewis et al. 2018; Naneva et al. 2020]. Furthermore, the growth of the use of robots and other forms of automation in the workplace may be a source of much fear of replacement and other forms of distrust toward robots [McClure 2018].

These examples indicate the importance of paying attention to the processes behind trust-building, but they also demonstrate that these processes can be interpreted and applied in different ways. Therefore, in the section that follows, this paper broadly distinguishes between different interpretations of the process of building trust in robots. Furthermore I also explain what this means in terms of the way robots themselves are conceptualized under these different interpretations. Building on this overview, Section 3 proposes a complementary notion that proposes a complementary approach that draws attention to the role of narratives. This approach considers robots to be a prominent example of a technology that is surrounded by many different narratives that often have imaginative and speculative content. I argue that this should be taken into account when theorizing and researching the process of building trust in robots. On this basis, Section 4 explains how such an approach can be developed in the social sciences. Finally, a short conclusion is presented to discuss how this notion of a trust-building process can be of use in the further development of interdisciplinary research.

2 Building trust in robots: Different interpretations

As noted, it is challenging to define the process of trust-building in a straightforward manner, amid the various interpretations of how it can be theorized and studied. To develop proper insight into the particularities of trust-building with regards to robots, a distinction among three interpretations of the trust-building process are developed below. It should be kept in mind that other ways of distinguishing these interpretations are possible; moreover, they often complement each other in actual research practice. Each of the following subsections indicates how a given interpretation theorizes the trust-building process and how the interpretation can be of use in research on trust in robots. In this way, the subsections provide the central ideas that define the different interpretations. Furthermore, this is accompanied by a description of the ways that robots themselves are defined and portrayed by this interpretation. As a consequence, this section does not stick to one single definition of the robot, but rather presents definitions of robots in relation to the respective approaches. Finally, the subsections investigate the research contexts and the domains in which these concepts and ideas are developed and deployed.

2.1. Behavior, appearance, and interaction

First, a set of prominent interpretations in this domain incorporate the idea of trust-building in relation to adjustments and refinements to the appearance and behavior of robots. Due to the emphasis on appearance and behavior, this perspective on trust usually aims to analyze human perceptions and experiences that are produced in interactions with robots [Hancock et al. 2011b; Li et al. 2010; Van den Brule et al. 2014]. For that reason, such research keeps a strong focus on gathering empirical insight from human-robot interaction experiments to identify and explain the mechanisms that support people's trust in robots during such interactions. Notions associated with this understanding of trust-building are often based on adjustments to the concept of interpersonal trust [Billings et al. 2012]. This concept involves the development of trust by one person (the trustor) in another (the trustee). Interpersonal trust has been studied extensively in fields like psychology and sociology, and it has been deployed in many different contexts. As such, the concept plays an important role in many theories of trust [Bachmann and Zaheer 2006; Simpson 2007]. Regarding the application of this notion of trust in relation to robots: in case one perceives technological artifacts as displaying forms of intelligence, they can also potentially enter into agentic relationships with humans [Elofson 2001; Nyholm 2018]. Hence, if artificial agency or intentionality emerges in interactive situations involving robots, notions derived from interpersonal trust can begin to play a role [De Graaf and Malle 2017].

Under this interpretation, robots are often defined as autonomous agents: (perceived) autonomous behavior and trust are thus seen as connected phenomena, as trust is generally considered to be an important element in relationships between humans as autonomous social beings. It is necessary to adjust this concept of trust to make it applicable to robots. In this setup, the robot takes on the role of the trustee in the interaction or relationship [Lewis et al. 2018]. In other words, although social agents are normally considered to be human, scholars in robotics-related research fields have argued that robots, when they are experienced and/or perceived by the trustor as an (intelligent) autonomous agent, can also be conceptualized as a trustee [Coeckelbergh 2012; Hancock et al. 2011a]. As such, robots fit into a wider discussion about trust in artificial agents—a discussion that also includes other types of agents, such as virtual bots, software programs, and so on [Andras et al. 2018; Glikson and Woolley 2020; Rossi 2018]. Nevertheless, it is crucial in this context that robots are embodied agents. The embodied character of robots opens up specific research areas that identify how trust can be built, based on the attitudes that this embodied appearance evokes in interactions [Nomura 2006]. For instance, the idea that robots have an embodied humanoid appearance is often considered to have a significant effect on the experience of trust [Alesich and Rigby 2017; van Pinxteren et al. 2019]. Thus, this

element of (anthropomorphic) embodiment makes appearance a very important feature and extends it to a wide variety of aspects that concern human beings' life and work with robots [Dumouchel 2022; Jones 2021].

It should come as no surprise that many of the concepts and methods that are based on this interpretation originate in psychological research. Approaches based on the notions of interpersonal trust and associated research into mental models have long been a topic of inquiry in several fields of psychology [Simpson 2007]. Human-robot interaction (HRI) is a prominent area of academic research that has successfully incorporated such concepts and methods to apply them to the study of trust in robots [Ullman and Malle 2018]. The interdisciplinary methods and approaches adopted in HRI generally focus on the development of experiments to measure trust-related attitudes. These experiments are often based on Wizard of Oz techniques, in which robots imitate agentic behavior [Riek 2012]. In this context, it is common to use validated questionnaires to gain insight into the experiences and attitudes of the human participants related to trust, while also providing directions on how trust can be built. Many outcomes of such research are then incorporated into the development of new robots, and robotics engineers often collaborate with HRI researchers in this context. Finally, several notions and theories developed in the context of interdisciplinary ethics research have also revolved around this interpretation of trust-building [Bartneck et al. 2021]. These notions and theories have been deployed to establish the field of robot ethics itself, but ethical concepts have also been implemented and tested as part of robots' behavioral cues [Malle 2016; Malle et al. 2019].

2.2. Research, development, and implementation

Another interpretation, focusing on the idea of human dependence on and vulnerability toward sociotechnical systems, describes the process of building trust in robots as an outcome of changes in design and engineering practices [Coeckelbergh 2013, 2015]. Taking technology to be constitutive of the environment in which humans operate and focusing on their vulnerability exposes trust as part of the entanglement that defines the relationship between humans and technological systems [Kiran and Verbeek 2010]. The implementation of this notion of trust-building draws attention to the ways in which research and innovation systems are set up, as well as the question of how they can be transformed in the direction of more open innovation in general [Geels 2004]. A good example of a framework often used in this context is the responsible research and innovation (RRI) framework [Asveld et al. 2017; van den Hoven et al. 2015]. When attention is drawn to the practices and norms that constitute sociotechnical systems, transparency and responsibility can become explicit components of (implicit) value

systems in research and engineering [Kiran et al. 2015]. For that reason, such approaches to trust-building focus on making innovation processes more open, responsible, and inclusive [Cheon and Su 2016]. Then, trust can emerge as an outcome of how characteristics such as reliance, transparency, and privacy are best managed in sociotechnical systems [Lee and See 2004; Wortham et al. 2016].

This interpretation of the trust-building process is not primarily focused on the appearance of the robotic artifact as such but rather emphasizes the notion of robotic technologies as important components of larger sociotechnical systems that (co-)define the conditions under which humans live and work [Sabanovic 2010]. Crucially, because humans construct these sociotechnical systems, they can also influence their development. Thus, it is important to consider how a technology like robotics establishes new forms of dependence and vulnerability, as well as the ways in which such issues are represented in terms of the norms and values of roboticists [Dignum et al. 2018]. Within the field of robotics itself, this perspective on the trust-building process has resulted in many calls to include norms and values that allow the needs of minorities to be recognized [Howard and Kennedy III 2020]. If design and engineering processes fail to consider and incorporate the values of different societal groups, attitudes of mistrust can arise with respect to technological systems [Howard and Borenstein 2018]. This, in turn, directly relates to overarching topics such as human rights and the maintenance of democratic values in technological design and engineering, emphasizing their importance for the way trust in robots develops in societies that have the need to mitigate the impacts of new types of robots [Torresen 2018]. An explicit openness to the deliberation on and implementation of values is in such a context considered to help ensure that societal and ethical issues are incorporated in the development processes behind robotic artifacts [Stahl and Coeckelbergh 2016]. Furthermore, the idea that robotic sociotechnical systems can establish new environments in which humans operate draws attention to the perspective of trust-building through the entanglements that constitute the relationship between humans and robotic systems [Richardson 2015]. As a central component of these sociotechnical systems, robotic artifacts can thus become more trustworthy by making their design and engineering to become more focused on issues like transparency and responsibility [Dignum 2017; Wortham et al. 2017; Wortham and Theodorou 2017].

Implicit in this idea of trust-building is the concept that existing practices in such fields can be altered to increase the general trustworthiness of robotic technologies. For instance, this can be achieved by implementing design requirements that would include the values discussed here and new types of awareness in robotics engineering and design processes [Siau and Wang 2018]. It is

important to note that many of these ideas are the subject of current discussions in the fields of robotics and HRI [Liu and Zawieska 2020; Winfield et al. 2021]. This is a crucial development, as their openness to such topics will likely have a strong effect on future developments in these pivotal fields. Critical analysis of and constructive engagement in new approaches to design and engineering practices are a prominent topic in many other academic areas as well. In philosophy, in particular, this entails the development of theories that reflect on technological design and engineering practices [Van de Poel and Royakkers 2011]. Several approaches from science and technology studies (STS) have also been crucial for drawing attention to the entanglements that constitute the (mundane) relationships between humans and technological artifacts [Maibaum et al. 2021; Rommetveit et al. 2020]. A range of topics and concepts from philosophy and the social sciences have likewise been used for interdisciplinary collaborations with roboticists, such as by creating approaches based on Participatory Design (PD) or Value Sensitive Design (VSD) [Azenkot et al. 2016; Umbrello and De Bellis 2018; Van Wynsberghe 2013].

2.3. Disruptions, rules, and regulations

The final interpretation regarding the process of building trust in robots is based on the idea that trust can be fostered with the help of rules and regulations [Nelson and Gorichanaz 2019]. Such discussions are increasingly prominent in recent years, as many proposals for rules and regulations to govern robotic and artificial intelligence (AI) technologies are currently in development [DG IPOL et al. 2016]. In close connection with this, the potential implications of the increasing prevalence of robots are a growing topic of inquiry in fields like ethics, legal studies, and governance studies [Boden et al. 2017; Leenes and Lucivero 2014; Nagenborg et al. 2008]. Beyond this, these types of interpretations of the trust-building process are generally important for the development of procedures that can help to mitigate the effects of emerging technologies on society. Ethical, legal, and regulatory schemes based on such analyses can help establish social trust in robots [Pagallo 2010]. Rules and regulations of this type can therefore function as part of a system of checks and balances that guide and govern technological developments and the implementation and use of robotics and AI, especially during a time characterized by socially disruptive technological advancements. In this context, philosophical deliberations are often concerned with new ontologies and ethical systems, while legal considerations are mostly about new rules and regulations. Both can be considered instrumental for creating a framework for further development and can help provide additional clarity for the current and future roles of robots in society [Gunkel 2012; Fosch-Villaronga and Heldeweg 2018].

In this context, robotic technologies (and AI) are largely understood and defined as a group of technologies set to disrupt existing conventions and therefore need to be guided and regulated via newly established rules and frameworks. Based on this idea, such approaches are often emphasizing the need for anticipating the potential social impact of future developments in robots' intelligence and agency. The emergence of intelligence and agency in machines is understood as a development that would potentially lead to large shifts in the issues of responsibility, liability, accountability and so on [Holder et al. 2016; Petit 2017]. If such issues are not dealt with properly, general trust in robots is likely to be compromised, which is why commitment to these issues can help to create rules and regulations to anticipate potential problems [Winfield and Jirotka 2018]. Thus, much of the work formulating ethical and/or legal arguments regarding the development of robots also takes on the current challenges and lacunae as well as those that future robots could bring about [Koops et al. 2013; Leenes et al. 2017]. In particular, with reference to concerns regarding the (im)possibilities of human control over the development and implementation of robotic technologies, ethical and legal scholars can help provide clarity to the discussion [Lin et al. 2012; Nagenborg et al. 2008].

When it comes to academic fields where trust-building of this type is a prominent topic, robot and AI ethics is a key area of research. The ethics of technology have been a topic of inquiry for many years, but it has gained importance in recent decades due to growing concerns over the social impact of other emerging technologies, such as nanotechnology or (big) data technologies [Brey 2017; Van de Poel 2008; Zwitter 2014]. In recent years, increasing interest has been seen in applying ethics to robots, and this has also become an important topic in fields investigating the governance of robotics. Hence, the meaning of the term ethics and its application have widened: according to some, ethics has even become "big business" [Richardson 2019; Sætra et al. 2021]. On a broader level, ethical considerations have repeatedly been shown to be instrumental for the exploration of potential legal and social ontologies and their consequences [Turner 2019]. In this regard, (social) robots are also becoming a subject of increasing concern in legal theory [Avila Negri 2021; Bertolini and Aiello 2018]. Furthermore, the regulation of robots and AI is now an important subject for concrete regulatory proposals, such as, for instance, in the European Union [European Parliament 2017].

3 Complementary interpretation: Robot narratives

In the previous section, different interpretations of the trust-building process were provided, accompanied by different definitions of robots: robots and robotic tech-

nologies were described as agentic interaction partners, as central artifacts in sociotechnical systems that are subject to alteration based on responsible engineering and design processes, and finally as a type of technology that (potentially) has the ability to disrupt existing ethical and legal conventions. I argue here for a complementary interpretation, describing a trust-building process that establishes the robot as the subject of narratives that may (and often do) contain speculative and imaginative content.

To explain this narrative perspective, it is useful to first discuss the notion of the narrative as found in social research, where it is used to analyze social life and has played that role for a long time [Nash 1994]. In social research, narratives are understood as carriers of meaning and assumptions, organized into plot-like structures [Deuten and Rip 2000]. Narratives constitute a crucial element of human social life: we think and communicate with the help of stories, which determine the limits of what we consider imaginable, knowable, and doable [Felt 2017]. In other words, narratives are instrumental for establishing meaning and structure [Czarniawska 2004]. As such, narratives can be analyzed in many different contexts, from policy documents to patient testimonies [Kirkpatrick 2008; McBeth and Lybecker 2018]. With regard to robots, the analysis of narratives can help clarify how robots become situated within shared meanings and assumptions. Thus, narratives are not simply stories: they can play a constitutive role in the development of concepts and ideas concerning the way our future with robots is to be configured. They point in certain directions, and the values implicit in them facilitate current and future development into a meaningful whole. Based on this, I argue that narratives can provide useful perspectives on the way we understand the role of robots, both in interactions with humans as well as in their larger societal context. Therefore, this paper argues for a more explicit inclusion of a narrative focus to come to grips with the way that the notion of trust-building can be further developed.

To ground the argument of the paper more securely, it will be useful to draw attention to narratives regarding robots and their imaginative and speculative elements. Why do narratives play such a crucial role for trust-building in robotics technology in particular? To provide a first answer to this question, it may be useful to provide insight into certain prominent elements from the history of robotics, as they demonstrate how the technological artifacts we call robots are surrounded by a host of speculative and imaginative narratives. The very term “robot” comes from a science fiction play, Rossum’s Universal Robots (R.U.R.), published in 1921 by the Czech writer Karel Čapek [Čapek 2004]. In this play, robots are created to work for humans, but they eventually rebel and cause the human race to go extinct. Even before this introduction of the word, autonomous non-human entities were a source of fear and fascination [Gasparetto 2016].

Depictions of and experiments with inanimate autonomous beings were part of larger (sometimes mesmerist and occultist) fascinations with automata. Such fascinations were rather widely expressed during the earlier phases of modern science and engineering [Coeckelbergh 2017; Liu 2010; Willis 2006]. The period of the Enlightenment for instance, exhibited an increasing engagement of clockmakers, mechanics at princely courts as well as other artisans with the creation of automata [Voskuhl 2013]. Furthermore, the history of fictional writing includes many examples of fascination with non-human forms of intelligence, such as the monster in Mary Shelley's Frankenstein, Henri Maillardet's Automaton, Nathanael (Nate) in E.T.A. Hoffmann's The Sandman, and many others [Cave and Dihal 2018; Selisker 2016].

In the context described above, as actual artifacts automata were mostly created in the domain of artisans, not that of engineers. With reference to the later establishment of robotics as a field of research and engineering, it is interesting to note that famous roboticists, such as Hans Moravec and Marvin Minsky, deliberately engaged in arguments that extrapolated research trends in their field toward futurist narratives. They claimed that science fiction futures that feature high levels of robot autonomy and intelligence could become a reality within a relatively short time. They explicitly referred to narratives that contained a strong fascination with the autonomy of robots. In that way, they were well aware that pop science efforts could help raise the political and cultural power of robotics as a field, which could in turn help increase their research funding [Geraci 2010].

In hindsight, it could be concluded that these early roboticists were quite successful in establishing robotics as a professional field. In this context, it is important to realize that the speculative dimension of the narratives around robots go well beyond the fictional realm. In recent decades, narratives about the further implementation of robots have continued to capture society's imagination [Heffernan 2019]. In the current moment in particular, there is a strong focus on the narrative that robots are an emerging technology that could, combined with AI technology, considerably alter the way we live and work while thoroughly changing society and the economy [Suchman 2019]. In this context, we have seen a general increase in concerns regarding the potential socially disruptive effects of the increasing implementation of autonomous systems, including robots, and their rapid technological progress. Important players like the European Union, Organization for Economic Co-operation and Development (OECD), and the United Nations have expressed the intention to maintain a strong emphasis on the need for anticipation of the future development of robotics in combination with AI technologies [European Commission 2020; OECD 2019; UNESCO 2021]. In this way, robots continue to be connected to the development of efforts to assess and predict future social and economic impact [Ford 2015; Nourbakhsh 2013]. There-

fore, due to its framing as an emerging technology, the future of robots is covered extensively in general public discourse, as well as in governance, which projects many different expectations onto its possible future development.

The speculative and anticipatory rhetoric that surrounds robotics is typical for emerging technologies, which are often characterized by high levels of ambiguity regarding their future [Asquer and Krachkovskaya 2021; Schaper-Rinkel 2013]. I argue here that the anticipation placed on (future) robots can usefully be understood and analyzed with the help of a narrative approach. As such, research and theory can treat narrative as a specific and distinct factor in the overall process of trust-building in robots. That is to say, robots' imagery and cultural status influence the way that they are portrayed and understood in the context of trust-building, in which individual robotic artifacts themselves, as well as robotics in general (as a field of research, design and implementation), play a crucial role in the emergence of narratives. Furthermore, I draw an explicit contrast to conceptions that disregard imaginative and speculative narratives about robotics as future-grasping hubris. Certainly, many solid and insightful studies exist that expose technological hubris and its distorting effects, but I argue that in relation to the process of building trust in robots, it can be insightful to explore how such narratives influence technological development and the culture that emerges around it. Expectation, imagination, and the anticipated/speculative future connected to them are thus considered major narratives that are constitutive of the ways that a culture thinks and acts with respect to robots.

4 Materializing a narrative approach: Studying trust

With a focus on narratives firmly established, it remains to describe how a narrative approach can be materialized. Here the interpretations from Section 2 are to be complemented by developing an understanding of how trust can be theorized and studied with the help of narratives. In other words, research that uses such an interpretation should be based on concepts of trust that explicitly, critically, and constructively engage with the narratives around robots. A particular focus is placed on three main components that are constitutive for a narrative approach to trust-building in robotic technologies: (1) scrutinizing existing imaginaries in the narratives about robots, (2) configuring anticipatory concepts regarding the narratives about robot futures, and (3) facilitating the emergence of new narratives around the sociotechnical potential of robots, mostly by increasing access to robots and robotics.

4.1. Scrutinizing technological imaginaries

To understand trust-building in robots using a narrative approach, it is crucial to draw attention to the technological imaginaries that are inherent to robot narratives. The concept of technological imaginaries or sociotechnical imaginaries emphasizes the entanglement of technologies in their social and cultural contexts [Jasanoff and Kim 2015]. The main idea being that these contexts define the development and implementation of technologies, as well as the norms and social and cultural practices around them. The analysis of technological imaginaries fit easily into a narrative approach, as these imaginaries can be found through the analysis of narratives. The main rationale here is that technological imaginaries drive cultural understandings and the perceptions of robots by defining and influencing arguments and concepts regarding robots' roles in our (future) societies. Thus, the imagined futures of robots should be understood as shaping the ways that societies deal with the contingencies connected to these futures through the visions and expectations that they represent. These imaginaries also shape the development of technologies connected to such visions [Jasanoff and Kim 2009]. As such, the technological imaginary should also play a constitutive role in the critical analysis of anticipatory notions surrounding robots in relation to the construction of novel social realities based on the futures of emerging technologies [Vallès-Peris and Doménech 2020]. For instance, Lucy Suchman has convincingly argued that the robot imaginary confronted at present is largely based on Euro-American notions of embodiment, emotion, and sociality. From this argument, she demonstrates that narratives of social order are reproduced in the specific technological designs of robots [Suchman 2006]. Another example is the book The Robotic Imaginary by Jennifer Rhee, which analyzes the conceptualizations and visions of humanness and dehumanization as seen in discourses on robotics [Rhee 2018].

The analysis of imaginaries is particularly useful when one wants to study and analyze different interpretations and controversies in narratives that are concerned with the (future) role of robots in our societies. Many other technologies and their particular imaginaries have already undergone scrutiny using analysis of this type [Jasanoff and Kim 2015; Sismondo 2020]. These studies have repeatedly demonstrated that perceptions of technologies and their futures are a crucial factor in the decision-making of governments and corporations. Furthermore, they are instrumental to the development and negotiation of novel and already present social arrangements in terms of new technologies, for instance in the context of governance [Grunwald 2018]. In relation to the process of building trust in robots in light of promises, expectations, and fears regarding robots and their futures, trust-building can be conceptually connected to the ways in which robots are presented in (speculative) narratives [Rommetveit and Wynne 2017]. In other

words, such narratives must be interpreted as drivers of debates regarding the possibilities, dangers, and challenges of robotization and automation.

Thus, narratives and their imaginaries are drivers of the establishment of social and public trust with respect to robots [Kearnes et al. 2006]. Furthermore, when used in conjunction with concepts of trust that are derived from interpersonal trust, they can help provide a deeper understanding of people's attitudes in human-robot interaction [Fortunati et al. 2015; Weiss and Spiel 2021]. In this way, concepts of trust in robots can be further refined through careful investment in inclusive and responsible imaginaries with respect to our future with robots. Thus, analyzing narratives that establish certain social imaginaries, the heavily anticipated roles of robots in society can be assessed, discussed, and criticized. Finally, the analysis of imaginaries of robotics in different domains (e.g., robot engineering, robot governance, and industrial contexts) can help establish new understandings of social and collective life with robots while recognizing the social character of such technological futures.

4.2. Configuring anticipatory concepts

In addition to the critical analysis of robot imaginaries, a second component involves actively taking part in the development and configuration of concepts that can support narratives that are engaged with the anticipation of robots the sociotechnical systems that emerge around them [Floridi 2014]. Here, philosophers and social scientists themselves can become involved in the anticipation of potential scenarios in order to develop the arguments and concepts that can be of use in the responsible implementation of robotic technologies [Brey 2012]. In comparison to the subsection above, this component also requires a critical stance toward robot futures, but simultaneously it is more strongly focused on constructive and sometimes speculative engagement with the futures of emerging robotics. The different ways in which the technological potential of robots is imagined can be assessed and refined to shape the sociotechnical systems that surround robots [Plas et al. 2010]. Although many types of robots that are anticipated are not yet in widespread use, speculative engagement with their future incarnations can be an important part of concepts of trust-building that are based on a narrative approach. The provision of new directions and concepts to guide the construction of narratives about our futures with robots can allow new roles to be allotted to them, ones that can already be anticipated [Gunkel 2022; Selkirk et al. 2018].

In general, the advancement of such anticipatory concepts can encourage reflection on notions of trust to address current challenges surrounding automation and robotics. The main emphasis should fall on creating concepts to help soci-

eties adjust in times of transformative change, times in which technological developments challenge and redefine societal norms and practices [Bratton 2017; Sardar 2010]. As such, the configuration of anticipatory concepts involves the production of a thorough overview of the meanings and interpretations that develop in the anticipated trajectories of robotic development, including its speculative elements. The aim is thus to create new concepts that can help anticipate and modify the sociotechnical ramifications of those developments [Castañeda and Suchman 2014]. Apart from analyzing technological products and innovations in their social context, the goal should be to engage in the development of new narratives that can help steer the development of future products and innovations.

I argue here for explicit commitment to the continuous (re)configuration of anticipatory concepts related to robots. This is largely an exploratory endeavor, in which it is crucial to invest in concepts that support more inclusive narratives of robots as a widely implemented technology [Grunwald 2010; Selin 2008]. Interdisciplinary work is crucial for such efforts and for developing concepts that are on the one hand speculative, but rooted in engineering reality on the other. Moreover, it is a significant platform for implementing concepts and ideas that mobilize the technoscientific imagination toward emancipatory sociotechnical systems. Thus, by facilitating novel definitions of robots and their roles in social contexts, anticipation based on speculative concepts would be instrumental to fostering novel engagement types with robots. In this way, robots can help change well-established social ontologies [Coeckelbergh 2010; Gunkel 2018].

4.3. Facilitating sociotechnical potential

Finally, within a narrative approach that is focused on trust-building, it is important to provide insights and pathways that can actively facilitate the emergence of new narratives about robots' sociotechnical potential. These narratives may be instrumental for developing ideas for the use of robots, founded on the imaginative capacity of the general public or of specific future users regarding how they conceptualize and imagine life and work with robots. Facilitating narratives around robotics' potential is therefore mostly about deliberately providing access to robots in order to allow new narratives to emerge [Chun 2019; Fischer et al. 2020]. Furthermore, this calls for critical but constructive engagements with people's concrete imaginings with respect to their use of and work with robots. An important idea in this approach is that technological artifacts such as robots are (re)defined in terms of how their use is imagined and practiced [Soljacic et al. 2022]. Thus, the identification of new forms of sociotechnical potential can enable the development of a way to allow for new understandings of the roles that robots can play in society.

Here, an important question is how the different uses of technology can be facilitated and analyzed [Cressman 2019]. The researcher's role in this process is to work to provide access to robots and connect the narrative understanding of technology to people's experiences while using and interacting with robots. Thus, it is helpful to facilitate the emergence of new narratives around the possible uses of robots for building social trust in robots in a democratic society. To create trust and implement technology in accordance with democratic values, the general public as well as individual users must be prompted to form new narratives around robots' future potential [Bijker 2010; Ionescu and Schlund 2019]. In line with this, constructive engagement with narratives that involve robots' sociotechnical potential can be developed by increasing interactions with robots and robotics. The goal of this activity is not necessarily to see how different groups and guidelines can be included in the design but rather to inquire into the ways in which people use and understand technologies in novel ways that are previously unimagined.

It is crucial to recall that this approach must be explicitly neutral to any narrative trajectory, even with respect to those trajectories that could be classified as irrationally utopian or dystopian. The goal is rather to facilitate the way that associations of this or other types lead to unanticipated mundane uses of robots. Pioneering studies in the social construction of technology have been undertaken in relation to the user as an agent of technological change [Kline and Pinch 1996]. These studies indicate the way that a certain technological artifact and its social environment evolve over time, based on actual use. Therefore, in relation to robots and building trust in them, research activities should not only critically analyze and anticipate robot futures but also focus on providing the possibilities for emancipation and democratization through imagination in narratives regarding the use of and work with robotics. In this way, the development and implementation trajectories of robots can become increasingly democratized through the emphasis on possibilities for choosing and designing different technologies [Feenberg 2002].

5 Conclusion

This paper presented an approach to the process of building trust in robots that focuses on the role that narratives can play. I have demonstrated that robotics is necessarily embedded in narratives about its own future. I argue that this necessitates a complementary view on building trust in robots, which I presented in this paper. The goal of this approach is to deploy already existing discourse on trust to generate new ideas for bringing robots into our societies in ways that, without

profoundly disturbing our economic, social, and political lives, might empower us to achieve more equal, sustainable, and desirable futures with robots.

Implicitly, this focus on narratives involves the addition of perspectives from history, the arts, literature, and philosophy to the already rapidly growing body of research on the implications of emerging technologies such as robots. This development is far from finished and certainly is not limited to roboticists adapting or being open to these kinds of perspectives. It also means that significant efforts must still be made to bring the above-mentioned fields and disciplines closer to the field of robotics and identify ways in which the interpretations and ideas of each can be of use for the other. This is and will continue to be very challenging, not least because interdisciplinary work often necessarily encounters and must deal with long-standing preconceptions and conflicting epistemologies between disciplines [Weszkalnys and Barry 2013]. Therefore, it is crucial to continue investing in efforts to produce a deeper integration of these inter- and transdisciplinary perspectives.

Bibliography

Simone Alesich and Michael Rigby. 2017. Gendered Robots: Implications for Our Humanoid Future. *IEEE Technology and Society Magazine* 36, 2 (June 2017), 50–59. https://doi.org/10.1109/MTS.2017.2696598

Peter Andras, Lukas Esterle, Michael Guckert, The Anh Han, Peter R Lewis, Kristina Milanovic, Terry Payne, Cedric Perret, Jeremy Pitt, Simon T Powers, Neil Urquhart, and Simon Wells. 2018. Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems. *IEEE Technology and Society Magazine* 37, 4 (December 2018), 76–83. DOI:https://doi.org/10.1109/MTS.2018.2876107

Alberto Asquer and Inna Krachkovskaya. 2021. Uncertainty, institutions and regulatory responses to emerging technologies: CRISPR Gene editing in the US and the EU (2012–2019). *Regulation & Governance* 15, 4 (2021), 1111–1127. DOI:https://doi.org/10.1111/rego.12335

Lotte Asveld, Rietje van Dam-Mieras, Tsjalling Swierstra, Saskia Lavrijssen, Kees Linse and Jeroen van den Hoven (eds.). 2017. *Responsible innovation 3: a European agenda?* Springer. https://doi.org/10.1007/978-3-319-64834-7

Sergio M C Avila Negri. 2021. Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence. *Frontiers in Robotics and AI* 8, (2021). https://doi.org/10.3389/frobt.2021.789327

Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a participatory design approach. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 3–10. https://doi.org/10.1109/HRI.2016.7451727

Reinhard Bachmann and Akbar Zaheer (Eds.). 2006. *Handbook of trust research.* Edward Elgar Publishing, Cheltenham.

Christoph Bartneck, Christoph Lütge, Alan Wagner, and Sean Welsh. 2021. *An Introduction to Ethics in Robotics and AI.* Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-51110-4

Andrea Bertolini and Giuseppe Aiello. 2018. Robot companions: A legal and ethical analysis. *The Information Society 34*, 3 (May 2018), 130–140. https://doi.org/10.1080/01972243.2018.1444249

Wiebe E Bijker. 2010. Democratization of Technological Culture. In *Science and Technology Studies at Maastricht University. An Anthology of Inaugural Lectures,* Karin Bijsterveld (ed.). Maastricht University Press, Maastricht, 13–41.

Deborah Billings, Kristin Schaefer, Jessie Chen, and Peter Hancock. 2012. Human-robot interaction: Developing trust in robots. (March 2012). In *HRI '12: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction.* https://doi.org/10.1145/2157689.2157709

Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorrell, Mick Wallis, Blay Whitby, and Alan Winfield. 2017. Principles of robotics: regulating robots in the real world. *Connection Science* 29, 2 (April 2017), 124–129. https://doi.org/10.1080/09540091.2016.1271400

Benjamin H. Bratton. 2017. The New Normal. Strelka Press, Moscow and London.

Philip Brey. 2017. Ethics of emerging technology. In *The ethics of technology: Methods and approaches,* Sven Ove Hansson (ed.). Rowman & Littlefield International, 175–191. Philip Brey. 2017. Ethics of emerging technology. In The ethics of technology: Methods and approaches. 175–191.

Philip Brey. 2012. Anticipatory Ethics for Emerging Technologies. *Nanoethics* 6, 1 (April 2012), 1–13. https://doi.org/10.1007/s11569-012-0141-7

Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel H J Wigboldus, and Pim Haselager. 2014. Do Robot Performance and Behavioral Style affect Human Trust? *International Journal of Social Robotics* 6, 4 (November 2014), 519–531. https://doi.org/10.1007/s12369-014-0231-5

Tobias Bunde, Sophie Eisentraut, Natalie Knapp, Randolf Carr, Julia Hammelehle, Isabell Kump, Luca Miehe, and Amadée Mudie-Mantz. 2022. *Munich Security Report 2022: Turning the Tide – Unlearning Helplessness.* Munich Security Conference, Munich. Retrieved from https://doi.org/10.47342/QAWU4724

Karel Čapek. 2004. R.U.R (Rossum's universal robots). Penguin.

Claudia Castañeda and Lucy Suchman. 2014. Robot visions. *Social Studies of Science* 44, 3 (June 2014), 315–341. https://doi.org/10.1177/0306312713511868

Stephen Cave and Kanta Dihal. 2018. Ancient dreams of intelligent machines: 3,000 years of robots. *Nature* 559, 7715 (July 2018), 473–475. https://doi.org/10.1038/d41586-018-05773-y

EunJeong Cheon and Norman Makoto Su. 2016. Integrating Roboticist Values into a Value Sensitive Design Framework for Humanoid Robots. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*, IEEE Press, Christchurch, New Zealand, 375–382. doi: 10.1109/HRI.2016.7451775

Bohkyung Chun. 2019. Doing autoethnography of social robots: Ethnographic reflexivity in HRI. Paladyn, *Journal of Behavioral Robotics* 10, 1 (January 2019), 228–236. https://doi.org/10.1515/pjbr-2019-0019

Mark Coeckelbergh. 2010. Robot rights? Towards a social-relational justification of moral consideration. Ethics and Information Technology 12, 3 (September 2010), 209–221. https://doi.org/10.1007/s10676-010-9235-5

Mark Coeckelbergh. 2012. Can we trust robots? *Ethics and Information Technology* 14, 1 (March 2012), 53–60. https://doi.org/10.1007/s10676-011-9279-1

Mark Coeckelbergh. 2013. *Human being @ risk: enhancement, technology, and the evaluation of vulnerability transformations.* Springer, Dodrecht. https://doi.org/10.1007/978-94-007-6025-7

Mark Coeckelbergh. 2015. The tragedy of the master: automation, vulnerability, and distance. *Ethics and Information Technology* 17, 3 (September 2015), 219–229. https://doi.org/10.1007/s10676-015-9377-6

Mark Coeckelbergh. 2017. *New Romantic Cyborgs. Romanticism, Information Technology, and the End of the Machine.* MIT Press, Cambridge, MA.

Darryl Cressman. 2019. Disruptive Innovation and the Idea of Technology. *Novation: Critical Studies of Innovation* 1 (June 2019), 17–39.

Barbara Czarniawska. 2004. *Narratives in social science research.* Sage Publications, London; Thousand Oaks, Calif.

Jasper Deuten and Arie Rip. 2000. Narrative Infrastructure in Product Creation Processes. *Organization* 7, 1 (February 2000), 69–93. https://doi.org/10.1177/135050840071005

DG IPOL – Directorate-General for Internal Policies of the Union, European Parliament, Nathalie Nevejans. 2016. *European civil law rules in robotics.* Publications Office. Retrieved January 13, 2020 from https://data.europa.eu/doi/10.2861/946158

Virginia Dignum. 2017. Responsible Artificial Intelligence: Designing Ai for Human Values. *ITU Journal: ICT Discoveries* 1 (2017), 9.

Virginia Dignum, Frank Dignum, Javier Vazquez-Salceda, Aurelie Clodic, Manuel Gentile, Samuel Mascarenhas, and Agnese Augello. 2018. Design for Values for Social Robot Architectures. *Envisioning Robots in Society – Power, Politics, and Public Space* (2018), 43–52. https://doi.org/10.3233/978-1-61499-931-7-43

Paul Dumouchel. 2022. Ethics & Robotics, Embodiment and Vulnerability. *International Journal of Social Robotics* (March 2022). https://doi.org/10.1007/s12369-022-00869-y

Edelman. 2020. *Edelman Trust Barometer 2020 special Report: Trust in Technology.* Edelman. Retrieved May 27, 2020 from https://www.edelman.com/sites/g/files/aatuss191/files/2020-02/2020%20Edelman%20Trust%20Barometer%20Tech%20Sector%20Report_1.pdf

Greg Elofson. 2001. Developing Trust with Intelligent Agents: An Exploratory Study. In *Trust and Deception in Virtual Societies*, Cristiano Castelfranchi and Yao-Hua Tan (eds.). Springer Netherlands, Dordrecht, 125–138. https://doi.org/10.1007/978-94-017-3614-5_6

Ethics Advisory Group. 2018. T*owards a digital ethics.* EDPS. Retrieved July 23, 2019 from https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

European Commission. 2019. *Building Trust in Human-Centric Artificial Intelligence.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019DC0168&qid=1630917561643

European Commission. 2020. White Paper On Artificial Intelligence - A European approach to excellence and trust. Retrieved November 11, 2020 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:0065:FIN

European Parliament. 2017. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics.

Andrew Feenberg. 2002. *Transforming technology: a critical theory revisited.* Oxford University Press, New York, N.Y.

Ulrike Felt. 2017. "Response-able Practices" or "New Bureaucracies of Virtue": The Challenges of Making RRI Work in Academic Environments. In *Responsible Innovation 3: A European Agenda?*, Lotte Asveld, Rietje van Dam-Mieras, Tsjalling Swierstra, Saskia Lavrijssen, Kees Linse and Jeroen van den Hoven (eds.). Springer International Publishing, Cham, 49–68. https://doi.org/10.1007/978-3-319-64834-7_4

Kerstin Fischer, Johanna Seibt, Raffaele Rodogno, Maike Kirkegård Rasmussen, Astrid Weiss, Leon Bodenhagen, William Kristian Juel, and Norbert Krüger. 2020. Integrative Social Robotics Hands-on. *Interaction Studies* 21, 1 (January 2020), 145–185. https://doi.org/10.1075/is.18058.fis

Luciano Floridi. 2014. Technoscience and Ethics Foresight. Philosophy & Technology 27, 4 (December 2014), 499–501. https://doi.org/10.1007/s13347-014-0180-9

Martin Ford. 2015. *Rise of the robots: technology and the threat of a jobless future.* Basic Books, New York.

Leopoldina Fortunati, Anna Esposito, Mauro Sarrica, and Giovanni Ferrin. 2015. Children's Knowledge and Imaginary About Robots. *International Journal of Social Robotics* 7, 5 (November 2015), 685–695. https://doi.org/10.1007/s12369-015-0316-9

Eduard Fosch-Villaronga and Michiel Heldeweg. 2018. "Regulation, I presume?" said the robot – Towards an iterative regulatory process for robot governance. *Computer Law & Security Review* 34, 6 (December 2018), 1258–1277. https://doi.org/10.1016/j.clsr.2018.09.001

Lynn Frewer. 1999. Risk Perception, Social Trust, and Public Participation in Strategic Decision Making: Implications for Emerging Technologies. *Ambio* 28, 6 (1999), 569–574.

Alessandro Gasparetto. 2016. Robots in History: Legends and Prototypes from Ancient Times to the Industrial Revolution. In *Explorations in the History of Machines and Mechanisms (History of Mechanism and Machine Science),* López-Cajún, C., Ceccarelli, M. (eds). Springer International Publishing, Cham, 39–49. https://doi.org/10.1007/978-3-319-31184-5_5

Frank W Geels. 2004. From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research Policy* 33, 6 (September 2004), 897–920. https://doi.org/10.1016/j.respol.2004.01.015

Robert M Geraci. 2010. *Apocalyptic AI: visions of heaven in robotics, artificial intelligence, and virtual reality.* Oxford University Press, New York.

Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals 14, 2 (March 2020), 627–660. https://doi.org/10.5465/annals.2018.0057

Maartje de Graaf and Bertram F Malle. 2017. How People Explain Action (and Autonomous Intelligent Systems Should Too). *AAAI Fall Symposia.* (2017), 8.

Armin Grunwald. 2010. From Speculative Nanoethics to Explorative Philosophy of Nanotechnology. *Nanoethics* 4, 2 (August 2010), 91–101. https://doi.org/10.1007/s11569-010-0088-5

Armin Grunwald. 2018. *Technology Assessment in Practice and Theory.* Routledge, London. https://doi.org/10.4324/9780429442643

David Gunkel. 2012. *The machine question: critical perspectives on AI, robots, and ethics.* MIT Press, Cambridge, Mass.

David Gunkel. 2018. The other question: can and should robots have rights? *Ethics and Information Technology* 20, 2 (June 2018), 87–99. https://doi.org/10.1007/s10676-017-9442-4

David Gunkel. 2022. The Symptom of Ethics: Rethinking Ethics in the Face of the Machine. *Human-Machine Communication* 4, 1 (April 2022). https://doi.org/10.30658/hmc.4.4

Peter A Hancock, Deborah R Billings, Kristin E Schaefer. 2011a. Can You Trust Your Robot? *Ergonomics in Design* 19, 3 (July 2011), 24–29. https://doi.org/10.1177/1064804611415045

Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie Y C Chen, Ewart J de Visser, and Raja Parasuraman. 2011b. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors* 53, 5 (October 2011), 517–527. https://doi.org/10.1177/0018720811417254

Teresa Heffernan. 2019. Fiction Meets Science: Ex Machina, Artificial Intelligence, and the Robotics Industry. In *Cyborg Futures*, Teresa Heffernan (ed.). Springer, Berlin, 127–140.

Chris Holder, Vikram Khurana, Faye Harrison, and Louisa Jacobs. 2016. Robotics and law: Key legal and regulatory implications of the robotics age (Part I of II). *Computer Law & Security Review* 32, 3 (June 2016), 383–402. https://doi.org/10.1016/j.clsr.2016.03.001

Jeroen van den Hoven, Pieter E Vermaas, and Ibo van de Poel (Eds.). 2015. *Handbook of Ethics, Values, and Technological Design.* Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-6970-0

Ayanna Howard and Jason Borenstein. 2018. The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science and Engineering Ethics* 24, 5 (October 2018), 1521–1536. https://doi.org/10.1007/s11948-017-9975-2

Ayanna Howard and Monroe Kennedy III. 2020. Robots are not immune to bias and injustice. *Science Robotics* 48, 5 (2020). https://doi.org/10.1126/scirobotics.abf1364

Tudor B Ionescu and Sebastian Schlund. 2019. A Participatory Programming Model for Democratizing Cobot Technology in Public and Industrial Fablabs. *Procedia CIRP* 81, (2019), 93–98. https://doi.org/10.1016/j.procir.2019.03.017

Sheila Jasanoff and Sang-Hyun Kim. 2009. Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva* 47, 2 (June 2009), 119–146. https://doi.org/10.1007/s11024-009-9124-4

Sheila Jasanoff and Sang-Hyun Kim. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power.* University of Chicago Press, Chicago and London. https://doi.org/10.7208/chicago/9780226276663.001.0001

Raya A Jones. 2021. Projective Anthropomorphism as a Dialogue with Ourselves. *International Journal of Social Robotics* (May 2021). https://doi.org/10.1007/s12369-021-00793-7

Matthew Kearnes, Robin Grove-White, Phil Macnaghten, James Wilsdon, and Brian Wynne. 2006. From Bio to Nano: Learning Lessons from the UK Agricultural Biotechnology Controversy. *Science as Culture* 15, 4 (December 2006), 291–307. https://doi.org/10.1080/09505430601022619

Asle H Kiran, Nelly Oudshoorn, and Peter-Paul Verbeek. 2015. Beyond checklists: toward an ethical-constructive technology assessment. *Journal of Responsible Innovation* 2, 1 (January 2015), 5–19. https://doi.org/10.1080/23299460.2014.992769

Asle H Kiran and Peter-Paul Verbeek. 2010. Trusting Our Selves to Technology. *Knowledge, Technology & Policy* 23, 3–4 (December 2010), 409–427. https://doi.org/10.1007/s12130-010-9123-7

Helen Kirkpatrick. 2008. A Narrative Framework for Understanding Experiences of People With Severe Mental Illnesses. *Archives of Psychiatric Nursing* 22, 2 (April 2008), 61–68. https://doi.org/10.1016/j.apnu.2007.12.002

Ronald Kline and Trevor Pinch. 1996. Users as Agents of Technological Change: The Social Construction of the Automobile in the Rural United States. *Technology and Culture* 37, 4 (October 1996), 763. https://doi.org/10.2307/3107097

Bert-Jaap Koops, Angela Di Carlo, Luca Nocco, Vincenzo Casamassima, and Elettra Stradella. 2013. Robotic Technologies and Fundamental Rights: Robotics Challenging the European Constitutional Framework. *International Journal of Technoethics* 4, 2 (July 2013), 15–35. https://doi.org/10.4018/jte.2013070102

John D Lee and Katrina A See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* (2004), 31.

Ronald Leenes and Federica Lucivero. 2014. Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design. *Law, Innovation and Technology* 6, 2 (December 2014), 193–220. https://doi.org/10.5235/17579961.6.2.193

Ronald Leenes, Erica Palmerini, Bert-Jaap Koops, Andrea Bertolini, Pericle Salvini, and Federica Lucivero. 2017. Regulatory challenges of robotics: some guidelines for addressing legal and ethical issues. *Law, Innovation and Technology* 9, 1 (January 2017), 1–44. https://doi.org/10.1080/17579961.2017.1304921

Michael Lewis, Katia Sycara, and Phillip Walker. 2018. The Role of Trust in Human-Robot Interaction. In *Foundations of Trusted Autonomy. Studies in Systems, Decision and Control,* Abbass, H., Scholz, J., Reid, D. (eds) 135–159. https://doi.org/10.1007/978-3-319-64816-3_8

Dingjun Li, PL Patrick Rau, and Ye Li. 2010. A cross-cultural study: Effect of robot appearance and task. I*nternational Journal of Social Robotics* 2, 2 (2010), 175–186. https://doi.org/10.1007/s12369-010-0056-9

Patrick Lin, Keith Abney, and George A. Bekey (Eds.). 2012. *Robot ethics: the ethical and social implications of robotics.* MIT Press, Cambridge, Mass.

Hin-Yan Liu and Karolina Zawieska. 2020. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. Ethics and Information Technology 22, 4 (December 2020), 321–333. https://doi.org/10.1007/s10676-017-9443-3

Lydia He Liu. 2010. *The Freudian robot: digital media and the future of the unconscious.* University of Chicago Press, Chicago.

Arne Maibaum, Andreas Bischof, Jannis Hergesell, and Benjamin Lipp. 2022. A critique of robotics in health care. *AI & Society* 37, 2 (June 2022), 467–477. https://doi.org/10.1007/s00146-021-01206-z

Bertram F Malle. 2016. Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology* 18, 4 (December 2016), 243–256. https://doi.org/10.1007/s10676-015-9367-8

Bertram F Malle, Paul Bello, and Matthias Scheutz. 2019. Requirements for an Artificial Agent with Norm Competence. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), Association for Computing Machinery, New York, NY, USA,* 21–27. https://doi.org/10.1145/3306618.3314252

Mark K. McBeth and Donna L. Lybecker. 2018. The Narrative Policy Framework, Agendas, and Sanctuary Cities: The Construction of a Public Problem. *Policy Studies Journal* 46, 4 (2018), 868–893. DOI:https://doi.org/10.1111/psj.12274

Paul K. McClure. 2018. "You're Fired," Says the Robot: The Rise of Automation in the Workplace, Technophobes, and Fears of Unemployment. *Social Science Computer Review* 36, 2 (April 2018), 139–156. https://doi.org/10.1177/0894439317698637

Michael Nagenborg, Rafael Capurro, Jutta Weber, and Christoph Pingel. 2008. Ethical regulations on robotics in Europe. *AI & Society* 22, 3 (January 2008), 349–366. https://doi.org/10.1007/s00146-007-0153-y

Stanislava Naneva, Marina Sarda Gou, Thomas L. Webb, and Tony J. Prescott. 2020. A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots. *International Journal of Social Robotics* 12, 6 (December 2020), 1179–1201. https://doi.org/10.1007/s12369-020-00659-4

Cristopher Nash. 1994. *Narrative in culture: the uses of storytelling in the sciences, philosophy, and literature*. Routledge, London; New York.

Jake Nelson and Tim Gorichanaz. 2019. Trust as an ethical value in emerging technology governance: The case of drone regulation. *Technology in Society* 59, (November 2019), 101131. https://doi.org/10.1016/j.techsoc.2019.04.007

Tatsuya Nomura, Suzuki Tomohiro and Kanda Takayuki. 2006. Altered Attitudes of People toward Robots: Investigation through the Negative Attitudes toward Robots Scale. In *AAAI 2006 workshop on human implications of human-robot interaction,* 7. The AAAI Press, Menlo Park, CA.

Illah Reza Nourbakhsh. 2013. *Robot futures.* The MIT Press, Cambridge, Massachusetts.

Sven Nyholm. 2018. Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. *Science and Engineering Ethics* 24, 4 (August 2018), 1201–1219. https://doi.org/10.1007/s11948-017-9943-x

OECD. 2019. *Recommendation of the Council on Artificial Intelligence*.

Ugo Pagallo. 2010. Robotrust and Legal Responsibility. *Knowledge, Technology & Policy* 23, 3–4 (December 2010), 367–379. https://doi.org/10.1007/s12130-010-9120-x

Nicolas Petit. 2017. *Law and Regulation of Artificial Intelligence and Robots: Conceptual Framework and Normative Implications*. Retrieved October 7, 2019 from https://papers.ssrn.com/abstract=2931339

Michelle M E van Pinxteren, Ruud W H Wetzels, Jessica Rüger, Mark Pluymaekers, and Martin Wetzels. 2019. Trust in humanoid robots: implications for services marketing.

Journal of Services Marketing 33, 4 (January 2019), 507–518. https://doi.org/10.1108/JSM-01-2018-0045

Arjanna van der Plas, Martijntje Smits, and Caroline Wehrmann. 2010. Beyond Speculative Robot Ethics: A Vision Assessment Study on the Future of the Robotic Caretaker. *Accountability in Research* 17, 6 (November 2010), 299–315. https://doi.org/10.1080/08989621.2010.524078

Ibo van de Poel. 2008. How Should We Do Nanoethics? A Network Approach for Discerning Ethical Issues in Nanotechnology. *Nanoethics* 2, 1 (April 2008), 25–38. https://doi.org/10.1007/s11569-008-0026-y

Ibo van de Poel and Lambèr Royakkers. 2011. *Ethics, Technology, and Engineering.* Wiley-Blackwell, Chichester.

Jennifer Rhee. 2018. *The robotic imaginary: The human and the price of dehumanized labor.* University of Minnesota Press, Minneapolis.

Kathleen Richardson. 2015. *An Anthropology of Robots and AI: Annihilation Anxiety and Machines. Routledge.* Routledge. https://doi.org/10.4324/9781315736426

Kathleen Richardson. 2019. The Business of Ethics, Robotics, and Artificial Intelligence. In *Cyborg Futures*, Teresa Heffernan (ed.). Springer, Berlin, 113–126.

Laurel D Riek. 2012. Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (July 2012), 119–136. https://doi.org/10.5898/JHRI.1.1.Riek

Kjetil Rommetveit, Niels van Dijk, and Kristrún Gunnarsdóttir. 2020. Make Way for the Robots! Human- and Machine-Centricity in Constituting a European Public–Private Partnership. *Minerva* 58, 1 (March 2020), 47–69. https://doi.org/10.1007/s11024-019-09386-1

Kjetil Rommetveit and Brian Wynne. 2017. Technoscience, imagined publics and public imaginations. *Public Understanding of Science* 26, 2 (February 2017), 133–147. https://doi.org/10.1177/0963662516663057

Francesca Rossi. 2018. Building trust in artificial intelligence. *Journal of international affairs* 72, 1 (2018), 127–134.

Mark Ryan. 2020. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics* 26, 5 (October 2020), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Selma Sabanovic. 2010. Robots in Society, Society in Robots: Mutual Shaping of Society and Technology as a Framework for Social Robot Design. *International Journal of Social Robotics* 2, 4 (December 2010), 439–450. https://doi.org/10.1007/s12369-010-0066-7

Henrik Skaug Sætra, Mark Coeckelbergh, and John Danaher. 2021. The AI ethicist's dilemma: fighting Big Tech by supporting Big Tech. *AI Ethics* (December 2021). https://doi.org/10.1007/s43681-021-00123-7

Ziauddin Sardar. 2010. Welcome to postnormal times. *Futures* 42, 5 (June 2010), 435–444. https://doi.org/10.1016/j.futures.2009.11.028

Petra Schaper-Rinkel. 2013. The role of future-oriented technology analysis in the governance of emerging technologies: The example of nanotechnology. *Technological Forecasting and Social Change* 80, 3 (March 2013), 444–452. https://doi.org/10.1016/j.techfore.2012.10.007

Cynthia Selin. 2008. The Sociology of the Future: Tracing Stories of Technology and Time. *Sociology Compass* 2, 6 (November 2008), 1878–1895. https://doi.org/10.1111/j.1751-9020.2008.00147.x

Scott Selisker. 2016. *Human programming: brainwashing, automatons, and American unfreedom.* University of Minnesota Press, Minneapolis.

Kaethe Selkirk, Cynthia Selin, and Ulrike Felt. 2018. A Festival of Futures: Recognizing and Reckoning Temporal Complexity in Foresight. In *Handbook of Anticipation,* Roberto Poli (ed.). Springer International Publishing, Cham, 1–23. https://doi.org/10.1007/978-3-319-31737-3_107-2

Keng Siau and Weiyu Wang. 2018. Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal* 31, 2 (2018), 8.

Jeffry A. Simpson. 2007. Foundations of interpersonal trust. In *Social psychology: Handbook of basic principles,* A. W. Kruglanski & E. T. Higgins (Eds.). The Guilford Press (2007).

Sergio Sismondo. 2020. Sociotechnical imaginaries: An accidental themed issue. *Social Studies of Science* 50, 4 (August 2020), 505–507. https://doi.org/10.1177/0306312720944753

Fran Soljacic, Meia Chita-Tegmark, Theresa Law, and Matthias Scheutz. 2022. *Robots in healthcare as envisioned by care professionals.* https://doi.org/10.48550/arXiv.2206.00776

Bernd Carsten Stahl and Mark Coeckelbergh. 2016. Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems* 86, (December 2016), 152–161. https://doi.org/10.1016/j.robot.2016.08.018

Lucy Suchman. 2006. *Human–Machine Reconfigurations: Plans and Situated Actions.* Cambridge University Press.

Lucy Suchman. 2019. Demystifying the Intelligent Machine. In *Cyborg futures: cross-disciplinary perspectives on artificial intelligence and robotics*, Theresa Heffernan (ed.). Springer, Berlin, 35–61.

Jim Torresen. 2018. A Review of Future and Ethical Perspectives of Robotics and AI. *Frontiers in Robotics and AI* 4, (January 2018), 75. https://doi.org/10.3389/frobt.2017.00075

Jacob Turner. 2019. *Robot rules: regulating artificial intelligence.* Palgrave Macmillan, Basingstoke.

Daniel Ullman and Bertram F Malle. 2018. What Does it Mean to Trust a Robot?: Steps Toward a Multidimensional Measure of Trust. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18,* ACM Press, Chicago, IL, USA, 263–264. https://doi.org/10.1145/3173386.3176991

Steven Umbrello and Angelo Frank De Bellis. 2018. A Value-Sensitive Design Approach to Intelligent Agents. In A*rtificial Intelligence Safety and Security* (2018), Roman Yampolskiy (.ed). CRC Press. Retrieved October 23, 2021 from https://papers.ssrn.com/abstract=3105597

UNESCO. 2021. *DRAFT Recommendation on the Ethics of Artificial Intelligence.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000379920

Núria Vallès-Peris and Miquel Doménech. 2020. Roboticists' Imaginaries of Robots for Care: The Radical Imaginary as a Tool for an Ethical Discussion. *Engineering Studies*, 12(3), (2020), 157–176 https://doi.org/10.1080/19378629.2020.1821695

Adelheid Voskuhl. 2013. *Androids in the Enlightenment: Mechanics, Artisans, and Cultures of the Self.* University of Chicago Press. https://doi.org/10.7208/chicago/9780226034331.001.0001

Astrid Weiss and Katta Spiel. 2021. Robots beyond Science Fiction: mutual learning in human–robot interaction on the way to participatory approaches. *AI & Society* (April 2021). https://doi.org/10.1007/s00146-021-01209-w

Gisa Weszkalnys and Andrew Barry. 2013. Multiple Environments: accountability, integration and ontology. In *Interdisciplinarity: Reconfigurations of the social and natural sciences,* Andrew Barry, Georgina Born (Eds.). Routledge, 194–224.

Martin Willis. 2006. *Mesmerists, Monsters, and Machines: Science Fiction and the Cultures of Science in the Nineteenth Century.* Kent State University Press.

Alan F T Winfield and Marina Jirotka. 2018. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Scieinces* 376, 2133 (November 2018), 1–13. https://doi.org/10.1098/rsta.2018.0085

Alan F T Winfield, Katie Winkle, Helena Webb, Ulrik Lyngs, Marina Jirotka, and Carl Macrae. 2021. Robot Accident Investigation: A Case Study in Responsible Robotics. In *Software Engineering for Robotics*, Ana Cavalcanti, Brijesh Dongol, Rob Hierons, Jon Timmis and Jim Woodcock (eds.). Springer International Publishing, Cham, 165–187. https://doi.org/10.1007/978-3-030-66494-7_6

Robert H Wortham and Andreas Theodorou. 2017. Robot transparency, trust and utility. *Connection Science* 29, 3 (July 2017), 242–248. https://doi.org/10.1080/09540091.2017.1313816

Robert H Wortham, Andreas Theodorou, and Joanna J Bryson. 2016. What does the robot think? Transparency as a fundamental design requirement for intelligent systems. In *Proceedings of the IJCAI Workshop on Ethics for Artificial Intelligence: International Joint Conference on Artificial Intelligence.* IJCAI 2016 Ethics for AI Workshop.

Robert H. Wortham, Andreas Theodorou, and Joanna J Bryson. 2017. Robot Transparency: Improving Understanding of Intelligent Behaviour for Designers and Users. In *Towards Autonomous Robotic Systems TAROS 2017. (Lecture Notes in Computer Science)*, Gao, Y., Fallah, S., Jin, Y., Lekakou, C. (eds). Springer International Publishing, Cham, 274–289. https://doi.org/10.1007/978-3-319-64107-2_22

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

Guang-Zhong Yang, Jim Bellingham, Pierre E Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, Bradley J Nelson, Brian Scassellati, Mariarosaria Taddeo, Russell Taylor, Manuela Veloso, Zhong Lin Wang, and Robert Wood. 2018. The grand challenges of Science Robotics. Science Robotics 3, 14 (January 2018), eaar7650. https://doi.org/10.1126/scirobotics.aar7650

Andrej Zwitter. 2014. Big Data ethics. *Big Data & Society* 1, 2 (July 2014), https://doi.org/10.1177/2053951714559253