

Trust in robot futures: The role of sociotechnical potential

The aim of this paper is to develop an approach that conceptualizes the role of sociotechnical potential in the discourse around trust in emerging technology. Thereby it focuses on robotics as an example of an emerging technology which is subject to many expectations concerning its future. The paper first provides a general overview of the thinking on trust in philosophy of technology. In the section after that, the paper argues for the importance of adding the perspective of technological potential by emphasizing how the discourse around emerging technologies like robotics is often referring to a world-to-come. Based on this idea, the final section argues for an approach to trust in technology that is based on the perspective of multiplicity of technology narratives. As such, the goal is to build and foster trust in the future of robots based on a notion of technodiversity.

Keywords: *futures, robotics, speculation, technodiversity, trust*

Author Information

Jesse de Pagter, Department for Labor Science and Organization, TU Wien

How to cite this article:

Jesse de Pagter. "Trust in robot futures: The role of sociotechnical potential."

Információs Társadalom XX, no. 2 (2020): 55–69.

<https://dx.doi.org/10.22503/inftars.XX.2020.2.4>

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

Introduction

During the last decade, philosophy of technology has become firmly involved in the development of conceptual frameworks that address social and ethical issues that arise around emerging technologies. As such, the field has developed a considerable voice in discussions on policies that deal with the (future) implications of new technologies such as robotics and AI. By bringing ethical and societal issues to the foreground, an important goal of those frameworks is to represent the human factor in technology. In this way, „human” values like dignity, autonomy and equality should be safeguarded when it comes to the further introduction of new technologies to our societies (Ethics Advisory Group 2018).

In this paper, the goal is to understand future implications of emerging technology from the perspective of trust in technology. While the influence of new technologies is a continuous concern in our societies, trust in emerging technologies is also on the rise as an issue of concern. The topic of trust has therefore become increasingly important in recent discussions on technological development and has aroused a discourse around the attitude of users, consumers, publics and citizens towards automation and robotics. In this context, surveys have demonstrated that the trust in robotics and other emerging technologies is a multidimensional concept that entails many different social, economic and political aspects (Miller et al. 2020; Edelman 2020). Important to note is that when it comes to emerging technologies, many of the discussions about the societal effects of those technologies are referring to consequences that have not yet been fully exhibited. A considerable part of such discussions is therefore based on future expectations and imaginations (Suchman 2019). Even though emerging technologies are not yet fully developed and might even fail to correspond to those different expectations, this paper considers the speculative element of those discussions to be of relevance to philosophy of technology exactly because of the sociopolitical effects that expectations and imaginations have (Rowland and Spaniol 2015).

In this paper, an approach to trust in (robotic) technology’s futures will be developed in three sections. The first section will provide a short overview of the way in which trust in technology has been addressed in contemporary philosophy of technology. Several major themes will be distinguished in order to describe this relationship. The section after that, the paper will introduce the topic of sociotechnical potential in order to grasp the concept of speculation as part and parcel of technological futures. Finally in the last section, the paper will use this concept in order to discuss a narrative approach towards trust that is based on the notion of technodiversity.

Philosophy of technology & trust in technology

The topic of trust in technology is an interesting issue in philosophy of technology, especially because it can be directly connected to several major topics

that have been prominent within philosophy of technology in recent decades, as will be explained below. Several important notions and ideas about trust in technology will be described and divided into three, interrelated, subsections. The first subsection will deal with the normativity of technology, the second with the appearance of technology and the final one with the reliance on technology. The topics of those subsections are consciously chosen, but alternative distinctions would certainly be possible. Furthermore, most of the literature that is featured comes from philosophy of technology and, when useful, references to the case of robotics will be made.

Trust and normativity of technology

The study of norms and values embedded in technologies and technological objects has been very prominent in philosophy of technology for a long time. Langdon Winner's famous 1980 article 'Do Artifacts Have Politics' is an often-cited example from this tradition (Winner 1980). Winner argued that technological artefacts themselves can be the embodiment of political values. In this context, his article demonstrated that technological artefacts have traditionally been strictly exempt from understandings that recognized their important role as carriers of norms and values. In research on trust in technology, this discussion on norms and values also has an important place, because it is often (implicitly) argued that trust in technologies can be increased by looking at the norms and values behind them (Vermaas et al. 2010). The processes of designing, introducing and using technological artefacts are key in such conceptualizations of technological trust.

For that reason, activities that try to (re)conceptualize and characterize technologies and technological artefacts constitute an important field of study. A central idea in this focus on technological objects as normative elements in sociotechnical systems is often referred to as the concept of the „black box” (Pinch 1992). An important notion behind this concept is that a lack of transparency in design-processes and artefacts can harm the fundamental open character of democratic societies. They allow for monopolization of power through technology while discouraging co-design practices. Moreover, artificial agents, such as robots and AI could pose existential threats to humanity since their strategic and practical advantages can lead to a whole new series of values: robo-defined values instead of human-defined ones (Danaher 2019). In this way, a lack of insight into the technological object itself has been connected to a lack of trust in that technology, since blackboxing of technology renders the value conflicts within the design process invisible (Pasquale 2015). Arguing from this rationale of transparency versus black boxing, the development of trustworthy technology can be achieved by rendering the technology's design less opaque (European Commission AI HLEG 2019). While understanding technology as a culturally constructed phenomenon, especially research

from science and technology studies (STS) has made a strong case for opening those black boxes. Technological artefacts are in that regard to be understood as elements of sociotechnical systems where they are part of a co-production process in which multiple stakeholders are participating (Sabanovic 2010). By analysing the social practices and cultural values that constitute such socio-technical systems, technological objects can be rendered more transparent and become subject to deliberation.

In relation to the opening of black boxes, some philosophers of technology have been arguing that the engagement with design and engineering practices themselves can foster trust in technology (Cook 2010). Arguing from this point of view, philosophers of technology have emphasized the important role of responsible research and innovation for the sake of increasing trust in robotics (Stahl and Coeckelbergh 2016). In order to develop trustworthy robots, the idea is that open and responsible design processes can help to understand and influence the norms and values that are inscribed into them. As such, the development of the notion of trust in technology has been evolving together with initiatives to include ethics as a part of the design of technologies (Dignum et al. 2018). Arguing from that point of view, technology ethics has been providing important contributions to different approaches that try to improve the design process of new technologies. A good example is the approach of Participatory Design (PD), where the main goal is to involve the stakeholders into the process of designing new technological artefacts (van der Velden and Mörtberg 2014). Another prominent example is Value Sensitive Design (VSD), which tries to further an approach where norms and values become directly embedded in design and engineering practices (Umbrello 2019). By doing so, the aim of those and other approaches is to democratize the design processes and enable users to gain a voice in those processes (Kensing and Greenbaum 2012).

Trust and appearance of technology

The insight into the role of technologies' and technological artifacts' appearance is an area of research that spreads over many different disciplines. Robots are a particularly interesting type of technology in that regard, because their appearance has many different technocultural connotations, as is for instance demonstrated in the widely used concept of the uncanny valley (Misselhorn 2010). Especially when it comes to humanoid embodied robots, the conceptualization of the artefact's appearance has become an important part of the discussion on the trust in them (Graaf, de and Malle 2017). Philosophy of technology has mainly been contributing to this issue through its conceptual work on the understanding of such appearances behind the perceived agency of technological artefacts (Coeckelbergh 2012). An important issue in that regard concerns the current (Western) conception of (moral) agency, which is increasingly turning out to be problematic with regards to new technologies such as robots. The main

issue being that on several levels (e.g. moral, organisational, legal), it is becoming theoretically tangible and practically useful to treat robots at least partly as morally responsible agents (Gunkel 2018; Sullins 2011). In this way, robots problematize categorizations of human morality and extend the class of entities that can be potentially involved in moral situations (Floridi and Sanders 2004).

Those considerations concerning the agency of technological artefacts have important implications for the discussion on trust as they have become a fruitful and widely used model for the analysis of robots' positions in our societies (Gunkel 2012). In order to deal with the moral implications of the changing agential status of technologies, the concepts around their morality help to guide the discussion on the ways in which those technologies can be controlled (Bryson and Kime 2011). A successful example of one of those concepts is that of „artificial moral agents” (AMA's) (Wallach and Allen 2009). Whereas it is in this case not really a discussion if those AMA's possess consciousness and/or sentience or not, the question is rather how appearances of robots can lead to derived forms of interpersonal trust (Nickel, Franssen, and Kroes 2010). Moreover, the issue regarding moral patiency of robots is an important one, because it asks to what extent robots (and other non-humans) are constituting an „other”, to whom moral duties and responsibilities should be appropriated (Allen and Wallach 2012). This is an important debate because it conceptualizes robots as AMA's that enforce new ethical dynamics in our societies. In that regard it is important to understand the new ways in which robots can become defined within such a society. This is for instance explicated by looking at the ways in which robots can be held responsible for their actions (Pagallo 2010).

This discussion on robot appearance and the mechanisms behind this appearance also has effects on the way in which robot design can lead to more trustworthy artefacts. Their societal contribution might be improved if we equip them with norms that increase their perceived moral competence (Malle 2016). Such discussions have therefore led to an increasing call for more transparent robotic artefacts. Especially in fields like Human Robot Interaction (HRI) and other fields this has increased initiatives that aim to make robots more transparent with regards to their appearance and behavior towards users (Wortham and Theodorou 2017). Especially in a governance context, those approaches have been urged in order to foster trust in technology (Winfield and Jirotko 2018). An example of a governance context where this has happened is the legal notion of “electronic persons” which has been proposed by the European Parliament in 2017 (EP 2017, § 59f).

Trust and reliance on technology

Finally, within philosophy of technology the question of technology's ambivalent position when it comes to reliance on technology versus mastery and control through the use of technology has been present for a long time. This theme

can already be found in the work of authors like Heidegger and Mumford, where it is used to discuss how technology is shaping human practices and perceptions of its environment (Coeckelbergh 2015). Important in this regard is the understanding of technology as an empowering extension of human faculties, while simultaneously problematizing its role as mediation between humans and their environment (Floridi 2014). As such, this perspective on technology emphasizes the notion of technological infrastructures creating novel realities which are (co-)defining the conditions under which humans live and speak (Coeckelbergh 2017).

When technology is conceptualized in this manner, the above-mentioned issue of reliance versus mastery becomes connected to the notion of trust through the issue of vulnerability. Whereas technological infrastructures need a certain amount of trust in order to function, this trust can only function if trusting agents accept the vulnerability that comes with their trust (Mcknight et al. 2011). With this constitutive character of technology in mind, trust can best be conceptualized as confidence: rather than just relying on technology, the process of technological mediation is to be understood in a sense of constitution. This entails that humans trust themselves *to* technology, whereby humans recognize how their subjectivity is partly constituted by the technologies that they rely on and entrust with authority (Kiran and Verbeek 2010). This notion of human reliance on technology is important, as it constitutes a technological risk and is therefore creating a certain degree of trust in spite of this risk (Nickel 2013). Furthermore, another important element of this discussion of reliance on technology is the idea that large and complex technological systems constitute a certain risk of breakdown (Viklund 2003). In cases of malfunction of (components of) those systems, human lives can potentially become threatened. In that regard it is important to recognize the vulnerability of complex societies, especially since the trust in technological systems can shift to an attitude of mistrust in situations that showcase vulnerability, such as disasters or other cases that expose shortcomings of technological infrastructures (Winner 2004).

Interpreting technology as constituting the environment in which humans operate while constituting a certain vulnerability is especially useful when analysing large-scale trends within society (Dierkes and Grote 2005). As technologies play an important role in our societies, the issues of trust and risk continue to be relevant to discussions on large topics such as human rights and the maintenance of democratic values. A prominent example in that regard has been the trust in digital technologies (Taddeo 2017). Also in relation to this perspective, several approaches from STS have been useful, especially by drawing attention to the perspective of trust as part of the entanglements that constitute the relationship between humans and technological artefacts (both virtual as well as physical ones) (Simon 2010). The topic of trust is in that regard a very ambivalent one, as trust is on the one hand important for technological systems to function, but can on the other hand be very misleading for

the individual users. Trustworthiness and transparency are also in this case important topics: developing technologies to be more trustworthy by making them transparent, explainable and accountable can help to analyse and expose the way in which large sociotechnical systems constitute new power-relations.

Emerging technology and the role of sociotechnical potential

After developing a short overview of the thinking on trust in philosophy of technology in general, the goal for the rest of this paper is to develop an understanding of the role of technological speculations when it comes to trust in emerging technologies. Robotics is thereby used as an interesting and useful example, since it can be understood as an exemplary case of an emerging technology that is projected to have a considerable impact in the (near) future. Following the anticipated importance of robots in the society of the future, governments and corporations have a considerable stake in the increase of their citizens' or customers' trust in robots (Miller et al. 2020). When it comes to the mechanisms behind such trust, philosophers of technology have a firm theoretical background that allows them to develop valuable insights into the societal ramifications of those developments, as the section above has shown. Next to the technophilosophical value of elaborating on such topics, they also present an opportunity for philosophy of technology to gain a direct involvement in decision-making processes around technology governance, as has already been shown with regards to robot ethics (Bösl and Bode 2018). Having said that, this section develops a complementary theme to the ones offered in the section above. This theme being the engagement with the role of speculative and imaginative elements in emerging technologies when it comes to the issue of trust. The paragraphs below will explain the consideration of this theme as a complementary perspective. After that, the last section will consider what this perspective entails for the understanding of trust in technology. By adding the element of technological futures to this discussion, the aim is therefore to enrich the conceptual framework on trust.

As has been argued in the introduction, the notion of the projected rise of robotics is generating many different forms of speculative imaginaries regarding its future. Moreover, even though robotic artefacts are already quite actively deployed in different manners, many components of the policies and strategies concerning robotics are (unsurprisingly) referring to robotic futures (Bösl and Bode 2018). Crucial thereby is that the rhetorics surrounding those expectations are often based on diverging assessments regarding the future of those technologies. Some of those assessments refer to the possibility that robotic artefacts will possess a plethora of novel properties and abilities which provides them with a revolutionary and transformative future potential (Fox 2018). Other assessments however rather point at the potential effects of robotics on the labor market or the potential challenges they

pose to fundamental human rights (Koops et al. 2013; Freeman 2015). In that regard, the robots of the future are very much the objects of imaginations and projections regarding their societal impact, together with other emerging technologies such as AI, nanotechnology, and biotechnology (Heffernan 2019). Whereas the uniqueness of this new robotic wave of automation should still be viewed with a healthy amount of scepticism, the goal of this paper is not to discredit such visions of the future. Rather, the paper's goal is to understand how speculative thinking constitutes new understandings of emerging technologies and their futures. In order to do that, let's get more insight into the analysis of technology's speculative character.

The focus on the analysis of technology's speculative character as such is nothing new in philosophy of technology. There are several accounts within that point at the ideological or even eschatological character of technological speculations (Burdett 2014; Geraci 2010). Moreover, speculations concerning technological potential are within philosophy of technology often conceptualized as part of the human drive for mastery over its environment; a drive that became particularly „successful” in Modernity in which ‚modern man aims at actively controlling fate’ (Mul 2014, 18). Others have connected this speculativeness to the fictional character of expectations in contemporary societies while understanding them as an element of the ‚dynamic restlessness of capitalism’ (Beckert 2016, 90). Another interesting view on the speculative element in technology has recently been made by Daryl Cressman. Arguing for a stronger engagement with the topic of (sociotechnical) potentiality regarding the study of technological artefacts, Cressman argues for a renewed attention for dialectical philosophy of technology, thereby prioritizing the dynamic tension between ‚that-which-is and that-which-could-be’ (Cressman 2020, 4). While referring to the work of Andrew Feenberg, Cressman develops an understanding of sociotechnical potential as a principle that is based on the imaginative capacity to project a better future. In this way, Cressman mainly locates the imaginative and speculative element in the *user* of a technology when he writes that ‚users transformed the function and meaning of technology to better realize concrete potentials that were not considered in the original design’ (Cressman 2020, 8). This is an interesting approach to the issue, especially because it calls for an engagement with the potentiality in technology by emphasizing the importance of the user's imagination. However, whereas this paper subscribes to Cressman's vision on the potentiality of technological objects, it cannot focus on the user in the same way as Cressman does. While this focus is certainly not considered invalid or less valuable, it does not fully suffice for the specific case of this paper because of two main reasons.

First of all, the focus on the user is often not feasible, exactly *because* of the speculative possibilities of robots. On the one hand, many anticipated robots are not yet being used, while their potential use is already being negotiated on the other hand. They are developed or projected to be developed, but at the same time they are often already part of the people's imagination of the

future. So while their meaning and function is already being negotiated, these negotiations are not happening on the level of the actual users of the artefact. One might question the need for a study of a technology that does not yet (fully) exist, but that is exactly what this article is aiming at: as the already mentioned surveys regarding public perception of emerging technology have demonstrated, those expectations and imaginations have an effect on the trust in the future of those technologies (Miller et al. 2020; Edelman 2020). In other words, while in many cases, consumers, users and other potentially relevant personas do not yet have the ability to transform the function and meaning of technologies in their capacity as users, they do have the possibility to engage in the different imaginations concerning the consequences and possibilities of robotics and other emerging technologies.

Second, even if users were able to interact more closely with robotic artefacts, their very status as an emerging technology entails a notion of widespread socioeconomic effects. One of the most essential features of emerging technologies is that they are projected to have a considerable effect on people's socio-economic situation, while there are many cases in which those people are not necessarily going to be direct users or consumers (Hilgartner 2009). As is the case with the point above, those projections do have an effect on the way in which trust in those technologies is developing (Stebbing 2009). To wit, the expected socioeconomic implications of emerging technologies often reach beyond the contexts of „users” and „consumers”. Robotics is in that regard a good example, as there are many projections regarding its future applications and potentials for societal change while not being limited to its effect on direct users and consumers. Furthermore, the complexity of this problem lies in the fact that there is often a strong rhetoric that surrounds emerging technologies. The expectations that those rhetorics purport do nevertheless lead to substantial financial and sociocultural investments in the future of those emerging technologies (Hilgartner and Lewenstein 2004).

When it comes to those two points, it is particularly important to define where the potentiality resides when it comes to the trust in (robotic) artefacts as objects of speculation. As shown, Cressman locates this mainly in the (lay) users, whereas a considerable share of robotic artefacts do not allow the access as a direct user. Nevertheless, the imaginative access to the technological artefacts is happening through the representation of the robotic artefact's futures through media, public debates and the like (Geraci 2010). In this way, the potentiality of emerging technologies such as robotics is something that gets negotiated beyond the perspective of the direct user or consumer. To summarize, Cressman's understanding of potentiality in technology provides a useful insight into the issue at hand, albeit that the focus on the direct user cannot be maintained. The goal for the last section of this paper is therefore to develop an approach to trust that is based on the notion of sociotechnical potential of technology. In line with the points above, this approach will aim to reach beyond this focus on the direct user.

A narrative approach to trust in emerging technology

When it comes to the trust in the sociotechnical potential of (emerging) technology it is important to emphasize again that the rhetorics and discussions concerning their futures are widely diverging. Therefore, in order to reach beyond the perspective of the direct user, the proposal is first of all to focus on the narratives *about* robotics. Comparable to Cressman's engagement with the user's everyday experience with technology, the approach would in this case be to engage with the narratives created around the (speculative) artefact. Furthermore, while critically engaging with narratives about technological artefacts, the proposal is to maintain and promote a concept of „technodiversity” as it has been developed by Yuk Hui (Hui 2019). One of Hui's main objectives regarding this concept is to develop philosophical reflections on technology that can reconcile Eastern and Western thinking on technology. With regards to his notion of technodiversity he writes:

The fundamental question is the regrounding of technology. We have to emphasize that this is not to add an ethics to AI or robotics, since we won't be able to change the technological tendency by just adding more values. Instead we have to provide new frameworks for future technological developments so that a new geopolitics can emerge that is not based on an apocalyptic singularity but technodiversity (Hui 2019, 277).

As becomes clear, Hui uses his concept of technodiversity to plead against „adding” ethics and/or values to technology. Instead, Hui argues for a „new framework” which is based on his notion of technodiversity. While contrasting this notion of technodiversity with the synchronization and convergence of (global) capital, Hui argues that emancipatory politics are difficult to imagine when living in a world that is strongly pushing towards technological singularity. It is here where Hui's concept of technodiversity can be interpreted and employed as an argument for the inclusion of notions of technological development that aim to engage with the *multiplicity* of futures concerning the potential use of artefact (Rowland and Spaniol 2015). Therefore, this approach would aim for the active engagement with the *different* narratives of the technological future. In other words, next to negotiating the values behind technological artefacts and technological systems, this paper proposes to develop concepts that can help develop new, diverse narratives about robot futures. Therefore, following Hui's approach, the task of philosophy of technology would be to develop concepts that can stand at the basis of a multiplicity of narratives while engaging with the norms and values that are present in the possible different futures of robotics.

In the same way that Cressman argues for the study of the users' notion of technoscientific potentiality of the artefact, the argument here is for the study of the narratives regarding the future role of the artifact. Apart from engaging

with the norms, transparency and power-relations in the design of technological artefacts, the argument here is for explicit engagement with the narratives about the future of those technologies. In this way, philosophy of technology would participate in (collective) speculations and imaginations that try to make sense of the sociotechnical world as it could become in the future. This would entail conceptual frameworks and new understandings that can help to systematically develop multiple futures of emerging technologies. The approach that this paper proposes is therefore to explicitly argue for a close engagement with those possible futures, in order to develop new concepts that help to understand emerging technology and the way its multiple futures play out. In this way, narratives about new technology should hopefully be able to push a more diverse array of visions concerning the future of those technologies. In this way, it can help to develop theories and concepts in order to grasp those futures and create a society that anticipates those futures in a democratic, inclusive and trustworthy manner.

Conclusion

The aim of this paper was to frame the development of alternative multiple technological futures as an interesting and important part of the discussion on trust in technologies and their artefacts. Furthermore, it has argued for an engagement with such futures. Crucial for this approach is to emphasize again that it contains an explicit call for a strong engagement with the narratives concerning the potential of technologies, thereby aiming for an approach that looks for new ways to foster technodiversity. Philosophy of technology is one of the fields that has a considerable knowledge base which can develop new understandings by engaging with different, alternative futures of emerging technology and its artefacts. In this way, different narratives on technologies can arise and foster new understandings of future artefacts. Thus, by exploring different, alternative scenarios, we can help set in place today factors that will increase the probability of more desirable futures happening' (Dunne and Raby 2013, 6).

References

- Allen, Colin, and Wendel Wallach. "Moral Machines: Contradiction in Terms or Abdication of Human Responsibility." In *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, Cambridge (MA), 2012, 55–68.
- Beckert, Jens. *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Cambridge, Massachusetts: Harvard University Press, 2016.

-
- Bösl, Dominik B. O., and Martina Bode. "Roboethics and Robotic Governance – A Literature Review and Research Agenda." In *ROBOT 2017: Third Iberian Robotics Conference*, edited by Anibal Ollero, Alberto Sanfeliu, Luis Montano, Nuno Lau, and Carlos Cardeira, 693:140–46. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-70833-1_12.
- Bryson, Joanna J., and Philip P. Kime. "Just an Artifact: Why Machines Are Perceived as Moral Agents." In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence: Barcelona, Catalonia, Spain, 16–22 July 2011*, 1641–46. AAAI Press, 2011. <https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-276>.
- Burdett, Michael S. *Eschatology and the Technological Future*. 1st ed. Routledge, 2014. <https://doi.org/10.4324/9781315739397>.
- Coeckelbergh, Mark. "Can We Trust Robots?" *Ethics and Information Technology* 14, no. 1 (March 2012): 53–60. <https://doi.org/10.1007/s10676-011-9279-1>.
- Coeckelbergh, Mark. "The Tragedy of the Master: Automation, Vulnerability, and Distance." *Ethics and Information Technology* 17, no. 3 (September 2015): 219–29. <https://doi.org/10.1007/s10676-015-9377-6>.
- Coeckelbergh, Mark. *Using Words and Things: Language and Philosophy of Technology*. Routledge Studies in Contemporary Philosophy 92. New York: Routledge, Taylor & Francis Group, 2017.
- Cook, S. D. Noam. "Making the Technological Trustworthy:: On Pitt on Technology and Trust." *Knowledge, Technology & Policy* 23, no. 3–4 (December 2010): 455–59. <https://doi.org/10.1007/s12130-010-9126-4>.
- Cressman, Daryl. "Contingency and Potential: Reconsidering a Dialectical Philosophy of Technology." *Techné: Research in Philosophy and Technology* 24, no. 1 & 2 (2020): 1–20.
- Danaher, John. "The Rise of the Robots and the Crisis of Moral Patency." *AI & SOCIETY* 34, no. 1 (March 2019): 129–36. <https://doi.org/10.1007/s00146-017-0773-9>.
- Dierkes, Meinolf, and Claudia von Grote. *Between Understanding and Trust: The Public, Science and Technology*. Amsterdam: Harwood Academic, 2005.
- Dignum, Virginia, Frank Dignum, Javier Vázquez-Salceda, Aurélie Clodic, Manuel Gentile, Samuel Mascarenhas, and Agnese Augello. "Design for Values for Social Robot Architectures." In *Robophilosophy/TRANSOR*, 43–52, 2018.
- Dunne, Anthony, and Fiona Raby. *Speculative Everything: Design, Fiction, and Social Dreaming*. Cambridge, Massachusetts ; London: The MIT Press, 2013.
- Edelman. "Special Report: Trust in Technology." Edelman Trust Barometer 2020. Edelman, 2020. https://www.edelman.com/sites/g/files/aatuss191/files/2020-02/2020%20Edelman%20Trust%20Barometer%20Tech%20Sector%20Report_1.pdf.
- EP. Civil Law Rules on Robotics - European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), Pub. L. No. P8_TA(2017)0051 (2017). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.
- European Commission AI HLEG. "Ethics Guidelines for Trustworthy AI," 2019. <https://ec.europa.eu/futurium/en/aialliance-consultation/guidelines>.
- Floridi, Luciano. *The 4th Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press, 2014.
- Floridi, Luciano, and J W Sanders. "On the Morality of Artificial Agents." *Minds and Machine* 14 (2004): 31.

- Fox, Stephen. "Cyborgs, Robots and Society: Implications for the Future of Society from Human Enhancement with In-The-Body *Technologies*." *Technologies* 6, no. 2 (May 19, 2018): 1–11. <https://doi.org/10.3390/technologies6020050>.
- Freeman, Richard B. "Who Owns the Robots Rules the World." *IZA World of Labor*, May 1, 2015. <https://doi.org/10.15185/izawol.5>.
- Geraci, Robert M. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. New York: Oxford University Press, 2010.
- Graaf, de, Maartje, and Bertram F. Malle. "How People Explain Action (and Autonomous Intelligent Systems Should Too)," 2017, 8.
- Gunkel, David J. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. Cambridge, Mass: MIT Press, 2012.
- Gunkel, David J. "The Other Question: Can and Should Robots Have Rights?" *Ethics and Information Technology* 20, no. 2 (June 2018): 87–99. <https://doi.org/10.1007/s10676-017-9442-4>.
- Heffernan, Teresa. *Cyborg Futures: Cross-Disciplinary Perspectives on Artificial Intelligence and Robotics*. Berlin: Springer, 2019.
- Hilgartner, Stephen. "Intellectual Property and the Politics of Emerging Technology: Inventors, Citizens, and Powers to Shape the Future" 84 (2009): 29.
- Hilgartner, Stephen, and Bruce V. Lewenstein. "The Speculative World of Emerging Technologies." Report, 2004. <https://ecommons.cornell.edu/handle/1813/36320>.
- Hui, Yuk. *Recursivity and Contingency*. Media Philosophy. London ; New York: Rowman & Littlefield, 2019.
- Kensing, Finn, and Joan Greenbaum. "Heritage. Having a Say." In *Routledge International Handbook of Participatory Design*, edited by Jesper Simonsen and Toni Robertson, 21–36. New York: Routledge Handbooks Online, 2012. <https://doi.org/10.4324/9780203108543.ch2>.
- Kiran, Asle H., and Peter-Paul Verbeek. "Trusting Our Selves to Technology." *Knowledge, Technology & Policy* 23, no. 3–4 (December 2010): 409–27. <https://doi.org/10.1007/s12130-010-9123-7>.
- Koops, Bert-Jaap, Angela Di Carlo, Luca Nocco, Vincenzo Casamassima, and Elettra Stradella. "Robotic Technologies and Fundamental Rights: Robotics Challenging the European Constitutional Framework." *International Journal of Technoethics* 4, no. 2 (July 2013): 15–35. <https://doi.org/10.4018/jte.2013070102>.
- Malle, Bertram F. "Integrating Robot Ethics and Machine Morality: The Study and Design of Moral Competence in Robots." *Ethics and Information Technology* 18, no. 4 (December 2016): 243–56. <https://doi.org/10.1007/s10676-015-9367-8>.
- Mcknight, D. Harrison, Michelle Carter, Jason Bennett Thatcher, and Paul F. Clay. "Trust in a Specific Technology: An Investigation of Its Components and Measures." *ACM Transactions on Management Information Systems* 2, no. 2 (June 1, 2011): 1–25. <https://doi.org/10.1145/1985347.1985353>.
- Miller, Catherine, Hannah Kitcher, Kapila Perera, and Alao Abiola. "People, Power and Technology: The 2020 Digital Attitudes Report." London: Doteveryone, 2020. https://www.doteveryone.org.uk/wp-content/uploads/2020/05/PPT-2020_Soft-Copy.pdf.
- Misselhorn, Catrin. "Empathy and Dyspathy with Androids: Philosophical, Fictional, and (Neuro)Psychological Perspectives." *Konturen* 2, no. 1 (October 11, 2010): 101. <https://doi.org/10.5399/uo/konturen.2.1.1341>.

-
- Mul, Jos de. *Destiny Domesticated: The Rebirth of Tragedy out of the Spirit of Technology*. Albany: SUNY Press, 2014.
- Nickel, Philip J. "Trust in Technological Systems." In *Norms in Technology*, 223–37. Springer, 2013.
- Nickel, Philip J. Maarten Franssen, and Peter Kroes. "Can We Make Sense of the Notion of Trustworthy Technology?" *Knowledge, Technology & Policy* 23, no. 3–4 (December 2010): 429–44. <https://doi.org/10.1007/s12130-010-9124-6>.
- Pagallo, Ugo. "Robotrust and Legal Responsibility." *Knowledge, Technology & Policy* 23, no. 3–4 (December 2010): 367–79. <https://doi.org/10.1007/s12130-010-9120-x>.
- Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.
- Pinch, Trevor J. "Opening Black Boxes: Science, Technology and Society." *Social Studies of Science* 22, no. 3 (August 1, 1992): 487–510. <https://doi.org/10.1177/0306312792022003003>.
- Rowland, Nicholas J., and Matthew J. Spaniol. "The Future Multiple." *Foresight* 17, no. 6 (November 9, 2015): 556–73. <https://doi.org/10.1108/FS-02-2015-0014>.
- Sabanovic, Selma. "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, no. 4 (December 2010): 439–50. <https://doi.org/10.1007/s12369-010-0066-7>.
- Simon, Judith. "The Entanglement of Trust and Knowledge on the Web." *Ethics and Information Technology* 12, no. 4 (December 1, 2010): 343–55. <https://doi.org/10.1007/s10676-010-9243-5>.
- Stahl, Bernd Carsten, and Mark Coeckelbergh. "Ethics of Healthcare Robotics: Towards Responsible Research and Innovation." *Robotics and Autonomous Systems* 86 (December 1, 2016): 152–61. <https://doi.org/10.1016/j.robot.2016.08.018>.
- Stebbing, Margaret. "Avoiding the Trust Deficit: Public Engagement, Values, the Precautionary Principle and the Future of Nanotechnology." *Journal of Bioethical Inquiry* 6, no. 1 (March 2009): 37–48. <https://doi.org/10.1007/s11673-009-9142-9>.
- Suchman, Lucy. "Demystifying the Intelligent Machine." In *Cyborg Futures: Cross-Disciplinary Perspectives on Artificial Intelligence and Robotics*, edited by Theresa Heffernan, 35–61. Berlin: Springer, 2019.
- Sullins, John P. "When Is a Robot a Moral Agent?" In *Machine Ethics*, edited by Michael Anderson and Susan Leigh Anderson, 151–61. Cambridge: Cambridge University Press, 2011. <https://doi.org/10.1017/CBO9780511978036.013>.
- Taddeo, Mariarosaria. "Trusting Digital Technologies Correctly." *Minds and Machines* 27, no. 4 (December 2017): 565–68. <https://doi.org/10.1007/s11023-017-9450-5>.
- Umbrello, Steven. "Imaginative Value Sensitive Design: Using Moral Imagination Theory to Inform Responsible Technology Design." *Science and Engineering Ethics*, April 10, 2019. <https://doi.org/10.1007/s11948-019-00104-4>.
- Velden, Maja van der, and Christina Mörtberg. "Participatory Design and Design for Values." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel, 41–66. Dordrecht: Springer Netherlands, 2014. https://doi.org/10.1007/978-94-007-6994-6_33-1.
- Vermaas, Pieter E., Yao-Hua Tan, Jeroen van den Hoven, Brigitte Burgemeestre, and Joris Hulstijn. "Designing for Trust: A Case of Value-Sensitive Design." *Knowledge, Technology & Policy* 23, no. 3–4 (December 2010): 491–505. <https://doi.org/10.1007/s12130-010-9130-8>.

- Viklund, Mattias J. "Trust and Risk Perception in Western Europe: A Cross-National Study." *Risk Analysis* 23, no. 4 (August 2003): 727–38. <https://doi.org/10.1111/1539-6924.00351>.
- Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford, New York: Oxford University Press, 2009.
- Winfield, Alan F. T., and Marina Jirotko. "Ethical Governance Is Essential to Building Trust in Robotics and Artificial Intelligence Systems." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 376, no. 2133 (November 28, 2018). <https://doi.org/10/gfc8dw>.
- Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus*, 1980, 121–36.
- Winner, Langdon. "Trust and Terror: The Vulnerability of Complex Socio-technical Systems." *Science as Culture* 13, no. 2 (June 2004): 155–72. <https://doi.org/10.1080/0950543042000226594>.
- Wortham, Robert H., and Andreas Theodorou. "Robot Transparency, Trust and Utility." *Connection Science* 29, no. 3 (July 3, 2017): 242–48. <https://doi.org/10.1080/09540091.2017.1313816>.