Würzburg Centre for Social and Legal Implications of Artificial Intelligence

Research Paper No. 2

Just Artificial Intelligence?

Luise K. Müller

Just Artificial Intelligence?

Luise K. Müller[*]

[*] TU Dresden – Institute for Political Science

Index Terms- justice, Artificial Intelligence, ethical AI, social change, machine ethics

I. INTRODUCTION

Both in the public and in the academic debate the idea has taken hold that the rapid development of Artificial Intelligence (AI) produces various ethical challenges. Especially the kind of AI with a capacity for (supervised and unsupervised) learning inspire enthusiastic predictions about its enormous potential on the one hand, and dark prophecies about its destructive capacities on the other. This type of AI learns autonomously using large amounts of data: it can detect patterns without being told previously what pattern to look for, for example. Ethical problems arise from both its relative autonomy, and the fact that AI is closely integrated into human society – with the implication that AI decision-making impacts the lives, opportunities and wellbeing of persons, and sometimes massively so. Think of AI in health care, law and courts, autonomous weapon systems or credit rating and welfare rating systems. In these areas, AI is used foremostly to optimise the efficiency and accuracy of reaching independently set goals. It can do so on the basis of more data than humans could ever consider, thereby transcending human cognitive limits. Additionally, given it is fed 'clean' data, AI is not influenced by common human weaknesses that tend to interfere with logical reasoning, like biases or personal feelings.

But sometimes the contexts in which AI is deployed requires a type of decision-making that no optimisation based on large amounts of data can solve; namely contexts that require the balancing consideration of moral values. Consider two different ways of how the use of AI can raise questions of value. The first perspective focuses on situations in which individual actors are confronted with a situation in which they must make a decision based on ethical considerations. It is exemplified by the ubiquitous autonomous driving dilemma, in which autonomous AI must make a moral decision in situations of competing values: in the widely discussed autonomous vehicle case, AI must choose between saving and sacrificing different lives. Under the heading of 'machine ethics', a body of literature examining these types of problems is developing.[1] Machine ethics is a new research field at the intersection of philosophy, computer science and robotics, and it focuses on developing ethical principles for machines. In that sense, machine ethics is different from, for example, animal ethics: while animal ethicist think about how humans ought to treat animals, machine ethicist think about how machines can be ethical agents. Machine ethicists do not develop ethical principles for humans that deal with machines, but ethical principles that machines apply in their decision-making.

As a type of applied ethics, machine ethics or AI ethics represent a specific perspective on the problem of values in AI decision-making, whereby its focus is placed on agents in specific roles making decisions in morally loaded situations. This is an important field, but I will approach the problem of value in AI decision-making from a different normative perspective: that of justice. The justice perspective shifts our focus away from individual actions, and on to social structures. Questions of justice typically evaluate the distribution of benefits and burdens within a given social scheme by applying a normative principle that embodies the ideal of justice. Where technological advancements affect human equality, wellbeing, freedom, resources and opportunities, the justice perspective asks under what conditions the use of such technology is fair. Its answers to those questions often come in the form of rights and duties, thereby creating a normative web between agents that owe something to each other. This makes justice characteristically a matter of enforceable obligations, as "we can be made to deliver what is due to others as a matter of justice, either by the recipients themselves or by third parties".[2]

The ethical and the justice perspectives cannot be distinctively divided: both perspectives refer to moral values and principles, and ask what the right thing to do is. Nevertheless, a few features characterise each perspective, and hence make them distinguishable. Justice asks about what we owe to each other in the context of specific political and social roles, and therefore

has to do with how individual claims are structured by political and social institutions. The emphasis on social structures could be explained by the fact that justice is fundamentally relational: its focal point lies on the principles regulating what we may legitimately require each other to do. As mentioned above, in its modern form, justice often comes in the form of enforceable rights and duties. In that sense, the perspective of justice is closely aligned with the legal form: it operates with the same (or similar) concepts, and it deals with 'what we are required to do'. In contrast, ethical perspectives tend to focus on 'how we want to live' – that is, the 'good' as opposed to the 'right'. Ethical decisions are often the result of commitments to certain values, and regard how we behave as individuals towards other individuals. Under this description, it makes sense that companies would publish guidelines on their ethical use of AI: it describes what those companies commit themselves to; what they believe 'good' AI should look like. In contrast, justice is – to a certain extent – detached from one's own ethical commitments or values; instead, it describes what others may legitimately claim from us, and what we owe to each other. Note again that the distinction I made between the ethical perspective and the justice perspective is somewhat stylised – both concepts are used in various ways in both philosophy and ordinary language. Nevertheless, it captures something intuitive: namely, that justice is not about what is good, but about what is right.
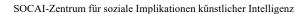
Without wanting to lessen the importance of the ethical perspective, in this paper I focus on what the ideal of justice contributes to the normative evaluation of AI. The perspective of justice presupposes that agents relate to another in a specific normative structure. This normative structure is shaped by moral and political concepts: rights and duties, moral agency and moral status, autonomy, responsibility, and citizenship are some of the core ones. Considering AI under the perspective of justice also means to view AI as more than a mere tool; a conceptual move that is not uncontroversial, but perhaps justified if machine learning is becoming more widespread.[3] Machine learning methods improve predictive and decisional models through experience. Being trained (supervised or unsupervised) on large amounts of data, machine learning algorithms can make predictions and decisions without explicit human programming or intervention. In that sense, machine learning algorithms are 'autonomous' in a functional sense of the word. This feature justifies at least the presumption that AI can plausibly be conceptualised as an agent relevant to relations of justice.

I begin by exploring in what ways core moral and political concepts – like rights, duties, responsibility, moral agency, status and citizenship – can be applied to AI (II). We will see that ascribing moral agency and/or moral status to AI changes our human normative relations profoundly. Then, I discuss the plausibility of applying justice-related normative concepts to AI, and what potential problems we need to be aware of when we do so (III). Finally, I consider how to develop normative principles for practices in which AI decision-making is used (IV). I argue that AI theorists can learn from methodological debates in political philosophy, focusing in particular on the idea of practice-dependent principles of justice.
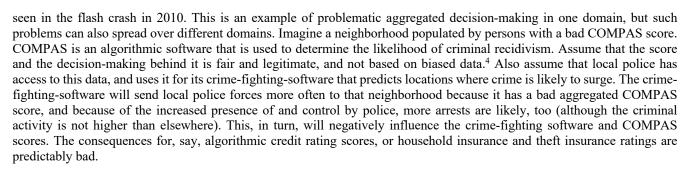
II. WHAT NORMATIVE ROLE FOR AI?

Autonomous AI decision-making may bring great benefits. The downside of the relative autonomy of AI is that its decision-making may become opaque: nobody understands just how the AI came to a particular decision, making it difficult to predict it. This can be problematic for a number of reasons, but in normative terms, the big issue that we cannot predict whether a particular decision is, or will be morally acceptable. One problem commanding significant attention at the moment is what is often called algorithmic unfairness. We frequently see algorithmic decision-making processes that are used in fields like credit rating, financial markets, security and risk assessments, law and policing, and insurance policy. Algorithmic unfairness originates in 'bad', biased, or unbalanced training data on which algorithms are trained and upon which decisions are ultimately based. One case that has made the rounds in the media is that of facial recognition, the accuracy of which in pattern detection appears to be questionable when applied in the context of different skin colours. When such technology is used, for example, for identifying suspects, this becomes a problem because its lack of accuracy may lead to the misidentification of innocent persons – a 'bug' that disproportionately affects groups with characteristics that ought to be morally irrelevant to the task of suspect identification.

Even if algorithmic decision-making operated on 'good data', and algorithms made efficient and rational decisions, one problem — we can call it the problem of aggregated injustice — still remains. When it comes to aggregated injustice, the problem is not that individual artificial agents make immoral or discriminatory decisions, but how independently reasonable decisions aggregate into a state that is unjust. Within systems of social cooperation, when a large number of presumably neutral actions are interlaced, they might amplify already existing injustices, and perhaps even create them in the first place. Aggregated algorithmic decision-making in fields like credit rating, law, and insurance, for example, can have disastrous consequences for individuals as well as social groups, and are especially unfair to those who are already in vulnerable social positions. In this case, there does not even *need* to be bias in the data used to train the decision-making algorithm: Individually rational decisions can have extremely unjust consequences when aggregated. Consider, for example, high-frequency trading, in which the decision-making of trading algorithms may lead — in their interaction — to disastrous consequences, as we have

seen in the flash crash in 2010. This is an example of problematic aggregated decision-making in one domain, but such problems can also spread over different domains. Imagine a neighborhood populated by persons with a bad COMPAS score. COMPAS is an algorithmic software that is used to determine the likelihood of criminal recidivism. Assume that the score and the decision-making behind it is fair and legitimate, and not based on biased data.[4] Also assume that local police has access to this data, and uses it for its crime-fighting-software that predicts locations where crime is likely to surge. The crime-fighting-software will send local police forces more often to that neighborhood because it has a bad aggregated COMPAS score, and because of the increased presence of and control by police, more arrests are likely, too (although the criminal activity is not higher than elsewhere). This, in turn, will negatively influence the crime-fighting software and COMPAS scores. The consequences for, say, algorithmic credit rating scores, or household insurance and theft insurance ratings are predictably bad.
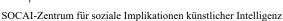
It is important to note that for the purposes of the example above, all of those decisions are, by themselves, rational in a narrow sense: they are not unduly influenced by bias (although in the real world, such decisions may raise a number of other moral concerns). It is their aggregation that produces a pattern of injustice. In such cases, it is not sufficient that algorithms not discriminate: the whole scenario is the outcome of individually fair decisions. Rather, these types of scenarios are united by two features: first, AI is used as a decision-making tool, and second, decisions take place in structures in which agents act together, and where agents' decisions and actions influence one another. Contexts such as globalised finance or social services are not simply the outcome of random actions, but constitute a practice with a goal, and with rules that structure the practice as well as the decisions and actions of its participants. That is why the eradication of biased data in algorithmic decision-making may simply not be enough to erase what makes us queasy about these scenarios. Instead, what would be needed is a rule-based coordination of artificial and biological agents, with the aim of achieving a fair distribution of the benefits and burdens effected by aggregated algorithmic decisions. I emphasise the 'aggregation' part of the argument again, because in these cases the injustice enters when a number of interdependently rational and seemingly fair actions form a pattern that leads to an unjust outcome. This is not merely a question of avoiding discrimination, but of how political institutions steer the actions of individual agents with the aim of distributing the burdens and benefits fairly between participants in the practice.

One of the ways that is currently being discussed as a solution to this problem is the development of 'artificial moral agents' (AMA). Moral agency could be defined as the capacity to take a moral standpoint, that is, to reflect and act upon moral reasons. There is no principled reason to preclude ascribing moral agency to AI, but it makes sense to distinguish between functional morality and full morality. According to this distinction, AI can only have a 'functional morality'. Hence, artificial moral agents are *moral in a reduced sense*: they may be able to make moral decisions, but unlike humans, they are not 'genuine' moral agents. Christian Neuhäuser argues that the moral agency of AI is automatic, but not autonomous in a strong sense of the term.[5] At best, they are morally competent: they recognize and evaluate moral contexts. But AI lacks a genuine moral perspective. In contrast to that, humans are full moral agents: they too recognize and evaluate moral contexts, and beyond that, they act within moral contexts and can justify their decisions and actions to others.

Even if we are (currently) only capable of building functional or 'automatic' moral agents, ascribing moral agency to AI impacts our normative relations enormously. By scaling back human intervention in algorithmic decision-making, we grant AI a sphere of autonomy that is potentially expansive and less controllable. The reasons for doing so are often convincing: AI decision-making *can potentially be* more efficient, more rational, quicker, and less biased. Nevertheless, making a moral decision in a complex context may prove more difficult than programmers and developers anticipate. There are three main reasons for that. First, the principles of morality are controversial, and their application is full of contradictions. People plainly disagree about what the right thing to do is in morally charged situations. The second reason is that often making a moral decision is not sufficient to ensure fairness, because moral contexts are social practices in which agents' actions and decisions are interdependent, and can have unintended consequences, like the patterns of injustice I described above. And directly connected to this is the third reason, which is the fact that in order to do something together with others, agents need to refer and rely on a common framework that outlines the rights and duties others must obey. These three circumstances – disagreement, coordination and reliability – explain why, in the human context, we do not rely on moral reasoning as a constraint on behavior alone. Hence, being capable of moral decision-making may be a necessary, but not a sufficient condition for deploying AI in contexts of justice.

Furthermore, moral agency entails a whole set of other-regarding normative concepts: If robots can be moral agents, can they bear responsibility for their actions? What would that mean for questions of liability and redress? Does moral agency imply that AMA's have duties, like a duty not to discriminate, or a duty to act according to principles of justice? For example, Günther Teubner argues that we need a special legal status for software agents because their deployment bears the risk of an ever-expanding responsibility gap.[6] The idea is to ascribe legal subjectivity to algorithms because of their agential properties, and to differentiate between three distinct legal statuses for different risk contexts. Other normative concepts may, however,

be useful in shaping and constraining the behavior and decision-making of AI, too. One might think of AI as being a special kind of political agent, for example, with obligations of non-discrimination toward humans.

III.   ANTHROPOMORPHISATION?

Humans are not only moral agents, but they also enjoy full moral status, meaning that aside from their capacity to act and decide in moral contexts, they are also owed moral respect. Humans are what John Rawls calls 'self-authenticating sources of valid claims': they have claims on others for their own sake, and just in virtue of being human. This particular two-dimensional model of a person – being a moral agent and having full moral status – is so entrenched in modern political and moral philosophy that one consciously recognises it only when it comes apart. This is similar to the debates about the moral status of animals where we have seen that most nonhuman animals can be moral patients (another word for having moral status), but lack moral agency. Another example is the debate about persons in a permanent coma, who are thought to have full moral status, but also lack moral agency. There is no reason to think that this constellation could not occur the other way round: AI may be morally competent and a functional moral agent, but *lack moral status* for its own sake. In a purely analytical sense, it is an open question whether being a 'functional moral agent' implies any moral status at all.

What kind of moral and legal status does AI need? Are we even required to ascribe rights to AI? Questions of rights are often determined by the question whether to ascribe moral status to AI. It stands far from settled what determines moral status, a problem philosophers have long disagreed about, and still do. The suggestions range from intelligence to sentience, from autonomy to (self-)consciousness, from having an interest in wellbeing to being capable of moral agency. Traditionally, moral status has been conceived of as an 'all-or-nothing-concept': you either have it – in which case you have just as much as others – or you don't. Most philosophers agree that humans generally have moral status – and that they have an equal moral status – but the debate about 'marginal cases' shows that there is no agreement about what to do about the very young and the very old, or those in a permanent coma. Deciding such cases really depends on the theory of the grounds of moral status. As it happens, some grounds will also give nonhumans moral status. The most discussed group here are nonhuman animals. If one argues that sentience – the capacity to feel pleasure and pain – is the ground of moral status, like Jeremy Bentham did, then many, if not most, nonhuman animals have moral status, which means, that they matter for their own sake and that we owe them moral consideration for that reason. If one however argues that moral autonomy – the capacity to reflect and act upon moral reasons – is the ground for moral status, like Immanuel Kant did, then most, if not all, nonhuman animals do not have moral status. (Note however that in this case, many humans will also not qualify for moral status.)

The case of animals shows that the concept of moral status is also useful when applied to nonhumans, and that ('genuine') moral agency and moral status can come apart – an insight we have gained only by applying moral categories to nonhuman entities. Furthermore, it shows that moral status could plausibly be conceived as a matter of degree. While contemporary philosophy is in agreement that humans share an equal moral status ('the egalitarian plateau'), it is a plausible view to hold that some animals possess a higher moral status than other animals, and that hence moral status can in principle be a matter of degree. Such a 'degree view' also makes the application of moral status to AI at least not wholly implausible. While it will be difficult to justify the attribution of a human-like moral status to AI, it might be worth contemplating whether AI may come to have a lower moral status, but a moral status nonetheless, in the future. The plausibility of ascribing any degree of moral status to AI depends on what one identifies as the status-granting property – sentience? Moral reflexivity? A mind? – and on the trajectory of AI development in the future. Again, ascribing moral status to AI has an impact on our normative relations. For example, it is the precondition for being a holder of individual rights. Even with a lower moral status than humans, Mathias Risse contemplates that AI might come up for something like human rights someday in the future, and muses whether we must then pass a 'Universal Declaration of the Rights of Full Ethical Agents'.[7] Right-bearing AI would give rise to correlating duties that humans must observe vis-à-vis machines.

Integrating AI into our normative networks by ascribing moral agency, or even moral status, to it raises the obvious question whether that is a sensible thing to do. Is AI the 'right' kind of entity to be ascribed concepts like moral agency, rights, duties, citizenship, and responsibility? Or are we anthropomorphising what is really just a tool? In other words, are we making a category mistake when we apply moral concepts to AI? Concepts like citizenship, rights or moral status are not anthropologically tied to humans as animal ethicists have shown. They delineate and illuminate complex normative relations in which humans stand with each other, but they are not limited to human relationships: again, animal ethicists have fruitfully adopted the idea of moral status. It is likely that some of those concepts also illuminate the nature and status of, and human relations to AI. Hence, the answer to the question whether AI is the 'right' kind of entity to apply our core moral and political concepts would depend on which concepts and on which AI. The important question is not whether or not we make a category

mistake if we apply core moral and political concepts to AI, but which concepts illuminate the role of AI in our practices and which ones are less useful.

Besides the question whether it is conceptually adequate to anthropomorphise AI by applying 'human' concepts like duties or moral agency to it, we can also ask whether it is normatively adequate to anthropomorphise AI by applying human values to AI. Is it a good idea to fashion AI behavior after human behavior, including programming our values into AI? There are two reasons to be skeptical about this. The first reason is the fact that there exists a reasonable pluralism of religious, moral and philosophical doctrines.[8] The attribute of reasonableness signifies that this pluralism is not the result of faulty reasoning or bias, but is an inevitable consequence of the exercise of human reason under free institutions. People disagree about normative questions because those questions involve complex empirical problems and require the nuanced weighing of different values. This pluralism is currently not represented in AI ethics, given that it is governed and discussed primarily by programmers, entrepreneurs, and academics. Thus, if we call for programming our values into AI, the question would be: whose values are those, anyways? The second reason we should be skeptical about programming our behavior and values into AI is the fact that human behavior and human values are imperfect. Given the state of the world, it is obvious that our values and behavior require critical analysis and assessment instead of unreflective affirmation and replication.

IV.   A Relational Methodological Approach

Nonetheless, in the last few years, we have seen a proliferation of ethics guidelines for AI.[9] Companies, political institutions, researchers and non-governmental political agents like NGO's or think tanks have published documents outlining their recommendations for the ethical development of AI. These guidelines often list abstract moral values that developers and programmers are advised to adhere to. Data and privacy protection, fairness and non-discrimination, accountability, transparency and human control are among the most often-cited examples. The idea is that AI should be designed with these values in mind to attain beneficial or 'good' AI. However, how helpful are such abstract and general values are for the design of AI? We have seen above that although AI may have the capacity to make ethically sound decisions, the aggregation of actions and the practice within which said actions take place structure the outcome of interdependent behavior, and may thereby lead to unjust AI. As Hagendorff notes in his survey of AI ethics guidelines, in AI ethics, "technical artifacts are primarily seen as isolated entities that can be optimised by experts so as to find technical solutions to technical problems. What is often lacking is a consideration of the wider contexts and the comprehensive relationship networks in which technical systems are embedded."[10]

The wider context and the comprehensive relationship networks are exactly what relational theories of justice focus on. The grounding premise of relational theories of justice is that justice applies to subjects that find themselves in a common and cooperative practice. One methodological approach that allows normative theorists to come up with more context-sensitive principles is that of practice-dependence. The practice-dependent approach begins by identifying "an existing social practice, including its point, or the goods it is meant to realise"[11]. To proceed, we also need to know who participates in the practice, and what their interests are. We assume that participants are motivated by those interests that are presumably mainly aimed at the good the practice generates. From this situation, we construct principles that govern the practice we are seeking to normatively regulate by designing rules that distribute the satisfaction of those interests among the participants in a way all participants could agree on. Importantly, all factors irrelevant to the good are blocked out from the procedure. Such a method is largely detached from pure and abstract moral ideals and is therefore driven by a consideration for context. In that sense, it seems to be the opposite from the laundry-list-approach currently dominating AI ethics.

Is such a methodological approach more helpful in determining what normative principles should guide the design and deployment of AI technology? It certainly can be. Take the example of autonomous driving. Manufacturers of self-driving cars must make weighty moral decisions when it comes to programming the decision-making algorithms determining the car's behavior in dangerous situations. Whom to protect – the people inside the car or the people potentially colliding with the car? General ethics guidelines are not of much help. Of course fairness and non-discrimination are important values, but how do we determine what decision is fair in this particular situation? A practice-dependent approach zooms out from one particular decision and takes the whole practice into account. The point is to ask how general rules governing the practice distribute the burdens and benefits of participating in that practice – instead of asking what particular decision in a potential situation is ethical or not. This gives us the advantage of conceptually capturing all relevant stakeholders, the values they likely pursue by participating in this practice, and what their advantages and disadvantages are. This transforms the relevant question from a trolley-case-type of problem into a matter of social structures and practices. Here is an example how autonomous driving might be analysed with the help of the practice-dependent approach: The practice we are looking at is obviously traffic, and its point is to let people get from A to B quickly and safely. The participants in the practice are drivers of cars, but also cyclists, people on foot, and so on – anyone who crosses public space on whatever vehicle. We assume that

those people are motivated by the value of safety, and that their interests in quickness and safety are all equally important. Neither the gender nor the fact that someone is very wealthy should in any way influence the distribution of the satisfaction of those values being simply irrelevant from a moral point of view. What does seem to be relevant is a person's deliberate choice of vehicle and the potential damage it can cause. What principles are appropriate to ensure that everyone's interests are considered equally?

As I said above, what those principles should not do is favor someone due to irrelevant factors. For example, someone disproportionately wealthy should not be advantaged by the rules in the sense that she can get significantly more of the good of safety just because she is able to afford it. Also, others should not be disadvantaged by that person's choice of a potentially more destructive vehicle. Obviously, this is all very rough and quick, and there are a number of important details in need to be discussed further, but I submit that this analysis does quite a bit to help to determine our initial question: namely, who needs to be protected in case of a dangerous collision? What is clear is that just regulative principles would disallow preferential protection of the people inside the car: this would lead to an unequal distribution of safety based on the financial ability to buy such a car. Based on our (admittedly very rough) considerations, it even seems that the choice to use a potentially destructive car (the origin of the risk) requires that principles favor those with a less destructive traffic choices – after all, they did not choose to introduce the risk in the first place. The details of this case are certainly ripe for more discussion, but what I mean to show with the example of autonomous vehicles is that a practice-dependent approach gives us a more systematic and fine-grained, and hence useful procedure to design and evaluate normative principles for AI. It offers a 'holistic' lens to identify the normative problems of a given practice.

V.   CONCLUSION

In this paper, I have explored in what ways the rise of AI technology impacts human society in terms of justice. Beyond moral dilemma situations, which are foremostly discussed in the literature about ethical AI, I argue that the perspective of justice widens our view to identify problems beyond the usual ones by asking how our social practices — which artificial agents contribute to and shape — can be structured in a manner that benefits all, and to do so with the help of political institutions.

AUTHOR

Author – Luise K. Müller, Dr. phil.; Postdoctoral Research Associate, Legal and Constitutional Theory, TU Dresden, SOCAI Associate Member, luise.k.mueller@tu-dresden.de

ACKNOWLEDGEMENTS

---

[1] For an excellent overview, see Misselhorn, Catrin (2018): Grundfragen der Maschinenethik, Reclam.

[2] Miller, David (2017): Justice, in: Zalta, E (ed), The Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/fall2017/entries/justice/

[3] Although I follow a quite different approach to political theory, I would like to point out that the view that AI is not merely a tool, but possesses agential status is widely accepted among posthumanist theorists like Karen Barad or Bruno Latour. Many thanks to Ann-Kathrin Koster for spotting this unlikely commonality.

[4] In reality, this is however not the case, see Julia Angwin/Jeff Larson, Surya Mattu, Lauren Kirchner, „Machine Bias", appeared in Pro Publica (May 23, 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[5] Neuhäuser, Christian (2013): Roboter und moralische Verantwortung, in: Hilgendorf, Eric (ed.), Robotik im Kontext von Recht und Moral, Nomos.

[6] Teubner, Günther (2019): Digitale Rechtssubjekte? Haftung für das Handeln autonomer Softwareagenten, in: Verfassungsblog 2019/9/30, https://verfassungsblog.de/digitale-rechtssubjekte-haftung-fuer-das-handeln-autonomer-softwareagenten/, DOI: https://doi.org/10.17176/20190930-112523-0

[7] Risse, Mathias (2019): Human Rights, Artificial Intelligence and Heideggerian Technoskepticism: The Long (Worrisome?) View, in: CCDP 2019-02, p. 31.

[8] Rawls, John (1993): Political Liberalism, Harvard University Press.

[9] For a comprehensive overview, see Hagendorff, Thilo (2020): The Ethics of AI Ethics. An Evaluation of Guidelines, in: Minds and Machines, 1-22.

[10] See Hagendorff 2020, p. 4.

[11] James, Aaron (2005): Constructing Justice for Existing Practice: Rawls and the Status Quo, in: Philosophy & Public Affairs, Vol. 33, No. 3, 281-316.