The full authenticated version is available online at https://doi.org/10.1007/978-3-030-66151-9_3

The Quest for Actionable AI Ethics

Emma Ruttkamp-Bloem[1, 2] [0000-0003-0299-6406]

[1] Department of Philosophy, University of Pretoria, Pretoria, South Africa
[2] Centre for AI Research (CAIR), South Africa

emma.ruttkamp-bloem@up.ac.za

Abstract. In the face of the fact that AI ethics guidelines currently, on the whole, seem to have no significant impact on AI practices, the quest of AI ethics to ensure trustworthy AI is in danger of becoming nothing more than a nice ideal. Serious work is to be done to ensure AI ethics guidelines are actionable. To this end, in this paper, I argue that AI ethics should be approached 1) in a multi-disciplinary manner focused on concrete research in the discipline of the ethics of AI and 2) as a dynamic system on the basis of virtue ethics in order to work towards enabling all AI actors to take responsibility for their own actions and to hold others accountable for theirs. In conclusion, the paper emphasises the importance of understanding AI ethics as playing out on a continuum of interconnected interests across academia, civil society, public policy-making and the private sector, and a novel notion of 'AI ethics capital' is put on the table as outcome of actionable AI ethics and essential ingredient for sustainable trustworthy AI.

Keywords: AI Ethics, Virtue Ethics, Multi-disciplinary Research, AI Ethics Capital, Trustworthy AI.

1. Introduction

In this paper, I argue that in order to ensure that AI ethics is actionable, the approach to AI ethics should change in two ways. AI ethics should be firstly approached in a multi-disciplinary manner focused on concrete research in the discipline of the ethics of AI and secondly as a dynamic system on the basis of virtue ethics in order to work towards enabling all AI actors to take responsibility for their own actions and to hold others accountable for theirs. In conclusion, the paper emphasises the importance of understanding AI ethics as playing out on a continuum of interconnected interests across academia, civil society, public policy-making and the private sector (including private sector companies ranging from start-ups to small-and medium enterprises to

large transnational companies). In addition, a novel notion of 'AI ethics capital' is put on the table as a core ingredient of trustworthy AI and an outcome of actionable AI ethics.

In the face of the relative ineffectiveness of a host of recent policy guidelines, including inter-governmental policies, national policies, professional policies, and policies generated in the private sector, there is a growing call from the AI community to increase the effectiveness of AI ethics guidelines[1]. Luciano Floridi (2019a) highlights the risks of not actionalising AI ethics guidelines in his article *Translating Principles into Practices of Digital Ethics*. He identifies five dangerous practices that may take root in a context in which AI ethics remains idealistic and removed from the every day working reality of the technical community, and which ultimately may work against actionable AI ethics: (1) Ethics shopping: there is confusion given the almost 100 sets of AI ethics policies available at present (Algorithm Watch 2020) rather than having "clear, shared, and publicly accepted ethical standards" (Floridi 2019a); (2) ethics bluewashing: pretending to work, or working superficially together towards establishing trustworthy AI instead of establishing "[p]ublic, accountable, and evidence-based transparency about good practices and ethical claims" (ibid.) and ensuring AI and AI ethics literacy of all AI actors (including board members of private sector companies and government officials); (3) ethics lobbying: promoting self-regulation instead of introducing enforceable ethical and legal norms; (4) ethics dumping: "the export of unethical research practices to countries where there are weaker … legal and ethical frameworks and enforcing mechanisms" (ibid.) as opposed to establishing a culture of research and consumption ethics; and (5) ethics shirking: weak execution of ethical duties given a perception of low returns on ethical adherence, instead of establishing clear lines of responsibility.

In his turn, Brent Mittlestadt (2019) warns in an article entitled *Principles Alone cannot Guarantee Ethical AI*, that the "real" work of AI ethics only starts now that we are faced with a multitude of policies. This work is "to … implement our lofty principles, and in doing so to begin to understand the real ethical challenges of AI" (ibid.). Thilo Hagendorff (2020), in an article entitled *The Ethics of AI Ethics: An Evaluation of Ethical Guidelines*, concurs, and mentions the lack of mechanisms AI ethics has to "reinforce its own normative claims" (ibid, 99), the view of AI ethics guidelines as coming from "'outside' the technical community" (ibid., 114)[2], and the lack of "[d]istributed responsibility in conjunction with a lack of knowledge about long-term or broader societal technological consequences [causing] software developers to lack a feeling of accountability or a view of the moral significance of their work" (ibid) as serious obstacles towards realising the 'lofty principles' of current AI ethics.

[1] See, e.g., Crawford & Calo 2016, Campolo et al 2017, Wachter, Mittelstadt, & Floridi 2017, Taddeo & Floridi 2018, Pekka et al 2018, Green 2018, McNamara et al. 2018, Morley et al 2019, Floridi 2019a, Mittelstadt 2019 and Hagendorff 2020; as well as Spielkamp et al 2019, Winfield 2019, and Jobin et al 2019 for discussions from various points of view of the current state of affairs of AI ethics.

[2] For instance, 79% of tech workers would like practical guidance with considering, implementing and adhering to ethical guidelines (Miller & Coldicott, 2019).

Some suggestions have been made to address the current 'inactive' status of AI ethics. These include advocating for and suggesting hands-on concrete suggestions for ethical machine learning from within the machine learning community itself[3] in terms of technical methods of addressing concerns around bias, transparency and accountability[4]; warnings about the consequences of ineffective ethical guidelines[5]; considering whether guidelines are converging on a global set of guidelines[6] and whether or not that would somehow increase the punching power of AI ethics guidelines; developing tools or templates to evaluate compliance with ethical guidelines[7]; and collating best practice examples[8]; among many others. This paper contributes to this debate from a *philosophical* perspective. Based on a virtue ethics approach, it suggests a dynamic and participatory model for AI ethics that is informed by multi-disciplinary research in the quest for actionable AI ethics.

More specifically, in what follows, in §2, a call for multi-disciplinary research as mechanism for grounding AI ethics and as a counter to the alienation of an increasingly commercially driven technical community is defended. In particular, it is suggested that the growing multi-disciplinary nature of the discipline of the ethics of AI, given the involvement of the technical community in research in the field, can enhance understanding among members of this community of the moral and ethical implications of the societal impact of AI technologies on human lives. Consequently, it is argued that, if AI ethics concerns and regulations were scaled down to the more concrete level of the ethics of AI such that the latter's state of the art multi-disciplinary content informs AI ethics, this would contribute to the action-ability of AI ethics.

In §3, the need to involve every AI actor across the spectrum ranging from government to civil society, to the private sector and academia in the AI ethics project is considered. Here, 'AI actor' means any entity involved in or impacted on by AI technologies in at least one stage of the AI lifecycle[9]. The term can refer both to natural and legal persons, and as such can refer to inidviduals such as researchers, programmers, engineers, data scientists, and end-users, and to large technology companies, small and medium enterprises, startups, universities, and public entities, among others.[10] It is argued that a virtue ethics approach to an AI ethics model as a

---

[3] Acknowledgment of the work of the ethics and society branch of Deepmind, the Open AI initiative, and the FAT ML association is important in this regard.

[4] See, e.g., Diakopoulos 2015, Taddeo & Floridi 2018, and Morley et al 2019.

[5] See, e.g., Wachter, Mittelstadt, & Floridi 2017, Green 2018, Floridi 2019a, Mittelstadt 2019, and Hagendorff 2020.

[6] See, e.g., Royakkers et al 2018, Jobin et al 2019, and Floridi & Cowls 2019.

[7] See, e.g., Alshammari & Simpson 2017, Abdul et al 2018, Kroll 2018, Anabo et al 2019, and Floridi 2019b.

[8] See, e.g., Taddeo & Floridi 2018, Morley et al 2019, and Mittelstadt 2019.

[9] The AI system lifecycle is taken to range at least from research, design, development, deployment to use ("including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly, and termination" (UNESCO 2020)

[10] This definition is based on the one given in the UNESCO First Draft of the Recommendation on the Ethics of AI (UNESCO 2020).

complex and dynamic system of values has the potential to allow for all AI actors to participate in the AI ethics project and to take responsibility for their actions and hold others accountable for theirs, which would contribute to the action-ability of AI ethics.

In the conclusion of the paper, based on the arguments in the previous sections, a novel notion of 'AI ethics capital' is suggested as the outcome of actionable AI ethics and essential ingredient for sustainable trustworthy AI. It is argued that the concept of AI ethics capital should be understood as a subset of the newly suggested concept of national AI capital (Momčilović 2020), and may be measured by means of a potential global AI ethics adherence index.

2. Grounding AI Ethics through Multi-disciplinary Research

In order for AI ethics to have practical impact on AI practices, it is clear that AI research, design and development should not take place in "closed-door industry settings", where "frictionless functionality that supports profit-driven business models" (Campolo et al 2017, 31 ff.) is the only name of the game due to fierce commercially driven competition for the best AI technology (see, e.g., Floridi et al 2018 and Hagendorff 2020). In response, in this section, state of the art (current) multi-disciplinary research is suggested as a counter to commercial values driving advancement in AI technologies on the one hand, and feelings of alienation from AI ethics among members of the technical community on the other.

In general, there are many reasons for placing multi-disciplinary research at the centre of discussions of increasing the impact of AI ethics guidelines. I highlight two here. First, a practical reason, already put firmly on the table by Morley et al (2019, 2), is that "[e]nabling the so-called dual advantage of 'ethical ML' – so that the opportunities are capitalised on, whilst the harms are foreseen and minimised or prevented … – requires asking difficult questions about design, development, deployment, practices, uses and users, as well as the data that fuel the whole life-cycle of algorithms …". And, what is needed to respond to these questions is input from "multi-disciplinary researchers, innovators, policymakers, citizens, developers and designers" (ibid., 3).

Secondly, current human reality is not reflecting or living up to the goals, values and principles of AI ethics, and thus, the data generated and collected in this reality are far removed from the lofty ideals of AI ethics, which is a problem, given that data 'fuels' the AI system lifecycle. This is one of the most simple reasons (apart form more technical ones) why AI is at risk of being biased, or built on unequal knowledge systems and unequal cultural, geographical, gender and age representation and has the potential to cause serious social harm and even political instability. So, the point is that ethical problems do not (only) lie within the technical aspects of the research, design, development, deployment and use of the AI based systems, but also (mostly,

albeit seldomly acknowledged) already in the 'real' world giving rise to the outcomes generated by such systems and in which such outcomes are applied. This is an essential motivation for concretising AI ethics by viewing it as a microcosm of our lived realities (a notion unpacked in more detail in the next section), and highlights the need for multi-disciplinary research – ranging from technical disciplines to social sciences – to inform AI ethics as well as to translate the long term ethical consequences of AI technologies into concrete terms for the technical community (the focus of this section).

In particular, the suggestion in this section is that the multi-disciplinary explosion of the scope of the discipline of the ethics of AI reflects the potential impact of AI technologies on human societies and political stability in a manner to which the technical community may be more open as it is more concrete to them because this discipline includes them. Members of the technical community themselves contribute to the ethics of AI in various roles, from software developing, to robotics, to computer engineering and data management. Thus, the argument here is that if state of the art multi-disciplinary knowledge of the growing scope of the discipline of the ethics of AI informs AI ethics guidelines, these will be in step with current AI technology advancement as well as more actionable, and, as such, being closer to the lived experiences of members of the technical community, also more effective to counter commercial interests.

Many calls for multi-disciplinarity have been heard in AI ethics conversations (e.g., Crawford & Calo 2017, Morley et al 2019) but none has specifically focused on the multi-disciplinary nature of the ethics of AI to clarify the long term societal consequences of actions of AI technologies in a manner that draws the technical community (and civil society I might add) into the AI ethics fold. I therefore suggest here that the multi-disciplinary nature of the ethics of AI should be reflected in the content of AI ethics in order to keep the ideals of AI ethics grounded and inclusive both of technical and social dimensions. Furthermore, the ethics of AI is built on respect for disciplines as diverse as philosophy and computer science, as anthropology and statistics, as political and legal sciences and mathematics. If this same mutual multi-disciplinary respect can drive the technical community's response to AI ethics, there is a much better chance of AI ethics being actionable, as there would be mutual respect for the understanding of the full impact of AI technologies on society, and thus mutual commitment to AI ethics.

These suggestions relate strongly to Hagendorff's (2020, 111) argument that "[i]n order to analyze [AI ethics] in sufficient depth, ethics has to partially transform to 'microethics'. This means that at certain points, a substantial change in the level of abstraction has to happen … On the way from ethics to 'microethics', a transformation from ethics to technology ethics, to machine ethics, to computer ethics, to information ethics, to data ethics has to take place" (ibid.). And, I argue, it is in this transformation to more concrete levels that AI ethics becomes accessible to the technical community. The reason for this, as alluded to above already, is that every sub-discipline of the ethics of AI (data ethics, robot ethics, machine ethics, information ethics, computational ethics, etc.) is informed by a different combination of disci-

plines such as computer science, mathematics, sociology, philosophy, anthropology, political sciences, law, etc.

But what is the ethics of AI? It may be divided into machine and computing ethics issues on the one hand and the impact of AI advances on society on the other hand (e.g., Asaro 2006, Müller 2020), although the lines are not exclusive (e.g., Wallach & Allen 2009, Lin et al 2012). Sometimes the ethics of AI is referred to as computer ethics, as in Moor's (1985, 266) description of computer ethics as the "analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of technology", thus, at least from the perspective of this paper, incorporating AI ethics into the discipline of the ethics of AI. In order to ground AI ethics and make it more accessible to members of the technical community (and civil society), I suggest AI ethics here as the domain focused on policymaking based on the concerns raised in each of the subfields of the ethics of AI.

Part of why there are different approaches to defining the discipline of the ethics of AI is the fact that it has crystallised into at least the (non-exclusive) subfields of machine ethics, data or algorithm ethics, robot ethics, information ethics, and neuroethics. Machine ethics focuses on the ethics of the design of artificial moral decision making capacities and socio-moral analyses of the concept of artificial morality.[11] Gunkel (2012: 101) distinguishes between computer ethics and machine ethics: "computer ethics … is concerned […] with questions of human action through the instrumentality of computers and related information systems. In clear distinction from these efforts, machine ethics seeks to enlarge the scope of moral agents by considering the ethical status and actions of machines". In these terms, machine ethics is concerned with "ethics *for* machines, for 'ethical machines', for machines as subjects, rather than for the human use of machines as objects" (Müller 2020), as the latter is the focus of robot ethics and also relates to computer ethics as defined above (see also Siau & Wang 2020). Another option (Segun 2020) is to refine machine ethics into thinking separately about technical aspects of computational tractability (computational ethics) and thinking about the ethics of machines with moral agency (machine ethics).

Robot ethics, or also known as the ethics of social robots, is focused on the impact of social robots on society (e.g., Royakkers et al 2015), on human-robot interaction (HRI), on the anthropomorphisation of robots and the objectification of humans, and robot rights[12] and also may be broken into focusing separately on AI-AI interaction, AI-human interaction and AI-society interaction (see Siau & Wang 2020). Furthermore, the ethics of social robots may also be incorporated into robo-ethics, which is "concerned with the moral behaviour of humans as they design, construct, use and interact with AI agents" (ibid.)[13]. In his turn, Asaro (2006, 10) argues that the field which he calls 'robot ethics', is focused on the ethical systems built into robots (focuses on robots as ethical subjects and relates to machine ethics and thus sometimes

[11] See, e.g., Allen, Varner, & Zinser 2000; Moor 2006; Wallach & Allen 2009, Anderson & Anderson 2007, Bostrom & Yudkowski 2014, and Brundage 2014.

[12] See, e.g., Sharkey & Sharkey 2010, Asaro 2012, Bekey 2012, Gunkel 2012, Boden et al 2017, and Danaher 2019.

[13] See also Veruggio & Operto 2006.

machine ethics is viewed as a subset of robot ethics); the ethics of people who design and use robots (focuses on humans as ethical subjects and relates to robo-ethics and computer ethics); and the ethics of how people treat robots (focuses on ethical interaction and relates to what is sometimes called the ethics of social robots). Asaro (ibid., 11) argues that the best approach to robot ethics is one that addresses all three of these and that views robots as socio-technical systems.

Data ethics is centered on issues around fair, accountable and transparent machine learning or co-called 'critical machine learning', socio-technical analyses of machine learning practices and their impact on society, and responsible data governance[14]. As such, it is a "branch of ethics that studies and evaluates moral problems related to algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)" (Floridi & Taddeo, 2016:1). Information ethics, in its turn, relates to data and algorithm ethics on the one hand, and on ethical elements of media and information governance, such as the impact of miss-and disinformation on society and political stability[15], on the other. Finally, neuro-ethics is focused broadly on the hard problems of consciousness (Chalmers 1995)[16] and how they relate to the everyday folk concept of the human 'mind'. The focus is on metaphysical and ethical conditions for mind-uploading[17]; Clark and Chalmers' (1998) concept of the extended mind[18], trans-humanism[19], and cyborg rights and identity[20].

It is clear from the above that the discipline of the ethics of AI indicates concerns with issues of human dignity, human agency, consciousness, freedom of expression and the right to information, morality, personhood, personal identity, the quality and nature of social relationships, and rights such as privacy, ownership and non-discrimination, among many others. All of these concerns – and their consequences – should be mirrored concretely in AI ethics guidelines and policies and their roots in different disciplines acknowledged. And, given this multi-disciplinary scope of the ethics of AI, my argument is that the content of AI ethics will be less 'lofty', more actionable, and more concretely communicable, if informed by state of the art research on these concerns in the ethics of AI, as members of the technical community are involved in this research themselves. Moreover, the multi-disciplinary scope of the ethics of AI brings home to every AI actor, from members of civil society to software engineers, the full impact of AI technologies on human lives. Viewing AI ethics in this manner as an active domain in step with (as opposed to always lagging behind)

---

[14] See, e.g., Baracos & Selbst 2016, Floridi & Taddeo 2016, and Veale & Binns 2017.

[15] See, e.g., Couldry & Hepp 2017, Greenhill & Oppenheim 2017, and Innes et al 2019.

[16] This is basically the problem of why consciousness occurs at all, combined with the problem of explaining subjective experience, or the 'feeling what it is like'.

[17] See, e.g., Schneider 2009, Chalmers 2010, Corabi & Schneider 2012, Wiley 2014, Pigliucci 2014, and Benedikter, Siepmann, & Reymann 2017.

[18] See e.g., Clark 2005, Clark 2008, Steffensen 2009, Adams & Aizawa 2010, and Pearlberg & Schroeder 2015.

[19] See, e.g., Clark 2003 and Hansell 2011.

[20] See, e.g., Sandberg & Bostrom 2008, Walker 2011, Eliasmith 2013, and Sandberg 2013.

the advances of AI technologies reflected in the subfields of the ethics of AI, allows the multi-disciplinary research driving the latter to become both an explanation and an affirmation of the concerns covered by AI ethics and confirms calls for the latter's status as a microcosm of human reality in all its social, political, economic, and philosophical dimensions (e.g., Mittlestadt 2019). This brings us to the next section in which a participative model for AI ethics is introduced.

3. A Participative Dynamic Model for Actionable AI Ethics

While it is essential to involve the technical community in order to turn inactive AI ethics to actionable AI ethics, as argued in the previous section, involving the members of civil society is equally important. Given that one of the indicators of the current exponential growth in AI adoption is the increasing "consumer readiness to consume AI in all of its forms" (Comninos & Konzett 2018, 8), ordinary members of civil society can and should play an important role in demanding and ensuring actionable AI ethics. In this section, the focus is on suggesting a model for AI ethics that satisfies the need to involve every AI actor across the spectrum, ranging from government to civil society to the private sector and academia. Humans are vulnerable to potential new harms generated by AI technologies in every dimension of their lives and should take responsibility to protect themselves, which is why it is so important to involve every AI actor in the AI ethics project in a practical participative role. Our vulnerability is exacerbated by the fact that humans do not necessarily by default hold a central role in "the world of information and smart agency" (Floridi 2015, 10).

The implications of this realisation of human vulnerability echoes Sherry Turkle's (1984, 2011) decades old warning that the integration of technology into human society not only alters human potential but also transforms human characteristics and consciousness. Therefore, we should not ask (or not only ask) what will technology be like in the future, but also we should consider what humans will be like in a future at least partly structured by AI technology – what are we becoming and what wil be our role in society in the future? Considering this question places a real responsibility on every human[21] to become involved and to participate in the project of AI ethics and echoes in the educational, scientific, cultural and communication and information contexts of our lives.

What to do? It is clear that we need human wisdom to guide our actions (Royakkers et al 2015), but what does that mean? We need careful, strong and rigorous philosoph-

[21] As it was made clear in the Introduction that 'AI actor' can here refer to either individuals such as designers or users, as well as to companies, this focus on individual human actors needs qualification. The focus in this section is indeed at the individual level, but the role of companies as AI actors in actionable AI ethics does not fall away, as the idea is that the participation in the AI ethics project of individuals employed by AI technology companies will 'filter up' so that companies also become involved in the AI ethics project and hold each other accountable.

ical thinking to guide us here as we have to rethink the entire project of philosophy[22] over centuries and this time there is a real urgency to this project, given the need to address the vulnerability of human society in the face of possible harm from AI technologies, while maximising the benefit of AI technology for humanity and ensuring this benefit is shared equitably.

Furthermore, it is clear that reacting positively to AI ethics guidelines does not only lie with governments, intergovernmental bodies, big tech companies or law enforcement. Specifically, the role of civil society in driving the success of actionable AI ethics has not received close to enough attention. There is not only alienation from the abstract ideals of AI ethics on the side of the technical community but also on the side of civil society, while the latter is actually as powerful a set of players in the AI ethics project as members of the technical community are. Members of civil society should play an active role in holding technical companies accountable for the systems they design, develop, and deploy, and holding government accountable for their use of AI in law enforcement, healthcare, education, and other policy areas. In addition, the public also has a responsibility to hold themselves accountable for how they use such systems.

One high level way in which to sensitise civil society to AI ethics, is to ensure that the values and ethical standards embodied in AI ethics guidelines are shared values. Focusing on 'intrinsic' values, as opposed to 'extrinsic' values may be a good beginning. Judgements of intrinsic value are evaluations of things that have value for their own sake, while extrinsic values get their value from their function or how they fit into a bigger system.[23] Intrinsic values include human life, freedom, peace, security, harmony, friendship, social justice, etc. The rationale behind emphasising intrinsic values is that such values are respected universally, given their intrinsic nature, but more importantly, that non-buy-in to these values is detrimental to everyone, and is perhaps most felt at the level of ordinary citizens as the most vulnerable of AI actors. And this is what civil society should be sensitised to grasp. Furthermore, given the international legal stature of international human rights law, principles, and standards, a human rights perspective in AI ethics guidelines may not only strengthen the potential for legal enforcement, but again is also a way in which to establish common grounds for AI ethics standards[24] and ensuring every member of civil society understands the consequences of not adhering to AI ethics guidelines. These perspectives alone are however not concrete enough.

What is needed in addition, is to bring home to civil society that the disruptiveness of AI technology impacts on every sphere of human lives, that 'being human' and enjoying fundamental freedoms are in danger of coming under increased control of AI technologies, and, perhaps most importantly, to ensure that there are safeguards against 'moral de-skilling' by technology. In an article entitled *Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character*

[22] Referring here to the project focused on the human condition and what it means to be human, taken up by philosophers of all traditions and nationalities from ancient times to the present.
[23] See, e.g., Moore 1961, Taylor 1961, and Audi 2003.
[24] See, e.g. Comninos & Konzett 2018, Latonero 2018, and Raso et al 2018.

(2015), Shannon Vallor warns that "… moral skills appear just as vulnerable to disruption or devaluation by technology-driven shifts in human practices as are professional or artisanal skills such as machining, shoemaking, or gardening. This is because moral skills are typically acquired in specific practices, which, under the right conditions and with sufficient opportunity for repetition, foster the cultivation of practical wisdom and moral habituation that jointly constitute genuine virtue. … profound technological shifts in human practices, if they disrupt or reduce the availability of these opportunities, can interrupt the path by which these moral skills are developed, habituated, and expressed" (ibid., 109).

On the one hand, this points to the need for strong campaigns driving *both* AI fundamentals and AI ethics literacy given that society "has greater control than it has ever had over outcomes related to (1) who people become; (2) what people can [or may] do; (3) what people can achieve, and (4) how people can interact with the world" (Morley et al 2019, 1). In other words, civil society should become aware and have a basic understanding of the potential of some AI technologies to threaten fundamental freedoms and change the moral fibre of societies.

On the other hand, we should ensure that trust in technology does not have the upper hand, *by ensuring that we can legitimately trust in humans and their abilities*. There is thus a responsibility that comes with protecting human dignity, human oversight and human centeredness, i.e., of fighting for 'AI with a human face'. It is an individual and universal responsibility of each and every AI actor to ensure they are the best humans they can be, and that they act – and are able to act – within the confines of regulations on the ethics of AI. Willingness to take up this responsibility is crucial to the success of any instrument of AI ethics, which brings us to considering the model of AI ethics that is suggested in this section to induce 'actionable' AI ethics. This model of AI ethics involves all AI actors in such a way that they actively participate in the quest for ethically acceptable AI rather than just react to a set of guidelines, and is not focused on technology, but on the actions of humans involved or impacted on by technology.

AI ethics should be recognised as an adaptive process and not thought of or approached in terms of technological solutions only, since it is necessary to recognise that "AI Ethics is effectively a microcosm of the political and ethical challenges faced in society" (Mittelstadt 2019), at a given time. In his turn, Hagendorff (2020, 111-112) reminds that implementation of AI ethics guidelines happen in "a widely diversified set of scientific, technical and economic practices, and in sometimes geographically dispersed groups of researchers and developers with different priorities, tasks and fragmental responsibilities" (ibid.). Thus, while it is very important to develop and constantly update technical tools to assist with the design and development of algorithms as AI technology advances, as noted for instance by Taddeo & Floridi (2018), Morley et al (2019), as well as a host of other writers recently, it is equally important to understand that the disciplinary, geo-political, and economic challenges both as generators of data fueling the AI lifecycle on the one hand, and the result of AI applications on the other hand, are diverse and also constantly changing. To deal with these contextual and temporal aspects of the AI lifecycle, ethical impact assessment instruments and also due diligence measures are crucial as they can be employed

continuously, and have the potential to ensure full participation in implementation of AI ethics guidelines, because of their potential to clearly point out the possible harms of a certain AI technology for a certain sector of society at a given time.

But this is not enough to get every AI actor involved. I argue that what is needed in addition to deal with these temporal and contextual characteristics of the lifecycle of AI technologies, is a comprehensive participative model of AI ethics that is built on responsible interconnected participation of all AI actors and that is adaptable to advances of AI technology and to social and political contexts, and that allows every individual AI actor to manage their own moral sensitivity on a continuous basis (see footnote 21 again). I believe the most promising way in which to actualise such a model of AI ethics is to extend Hagendorff's (2020) call for a move away from deontological, rule-based approaches in AI ethics (see also Mittelstadt 2019) to a virtue ethics approach.

Virtue ethics points to a lifelong journey of striving to be the most virtuous person one can be – this implies being acutely aware of what one is becoming (think back to Turkle's warning). Aristotle, in Book III of the *Nichomachean Ethics*, makes clear that he differs from Plato on the nature of virtue: Virtue is not the result of abstract understanding of what is truly good for us, rather it is the result of training and habit. A virtue ethics approach to AI ethics is thus not focused on universal codes of conduct or abstract guidelines (Hagendorff 2020), but on the individual level at which everyone in society has a duty to ensure that they themselves, as well as everyone else and the companies that employ them and that are AI actors in their turn, *are able* to be / to become the best possible moral version of themselves. Moreover, virtuous actions involve the rational consideration of and deliberation about consequences, rather than some form of external justification (such as commercial gain). To make a decision after deliberating over it, implies one has taken into account all of the possibilities that the outcome could be in a given context, and this forces one to dig deep into one's beliefs, to consider all consequences of one's beliefs and to decide whether or not - *and why* - one is going to follow certain beliefs and not others.

Hagendorff (2020, 112) makes clear that the value of such an approach in the context of AI ethics is that it is focused on "situation-specific deliberations, on addressing personality traits and behavioral dispositions on the part of technology developers … The technologists or software engineers and their social context are the primary addressees of such an ethics, not technology itself"[25]. But, it is more than that. Such an approach also offers a way in which to pick up on the responsibility pointed to above of every individual AI actor to be the best person they can be and take responsibility for their design, development and use of AI systems, while holding private sector software companies and governments accountable for their deployment of AI technologies.[26] This is so, since on Aristotle's model, a virtue ethics approach implies that everyone in society – families, schools, communities, as well as business (ibid., 113)

[25] See also e.g., Leonelli 2016 and Ananny 2016.

[26] Compare Floridi's (2016) argument that every actor who is "causally relevant for bringing about the collective consequences or impacts in question, has to be held accountable" (Hagendorff 2020, 113).

– should work to cultivate a virtuous life. In the context of AI ethics this implies "generating the motivation [among all AI actors] to adopt and habituate practices that influence technology development and use in a positive manner" (ibid.). Important here is Shannon Valor's (2016) work on the kind of techno-moral virtues humans need to cultivate in order to ensure they flourish as a result of emerging technologies, rather than simply adapting passively to such technologies.

Furthermore, the focus in Aristotelian virtue ethics is precisely on how to harness intellectual and moral virtues together to ensure a virtuous life. Thus, the focus is on "techno-moral virtues such as honesty, justice, courage, empathy, care, civility" (Hagendorff 2020, 113). This picks up on Morley et al's (2019) call for a movement from 'what' to 'how' that addresses "practice, the good and the just" (ibid. 2019, 11), and also on calls by Crawford & Calo (2016) for AI ethics to focus "as much on people than on code" (Morley et al 2019, 2). Approaching AI ethics as a virtue ethics therefore brings together the necessary focus on "technical discourses" (Hagendorff 2020, 114) as well as the "genuinely social and personality-related aspects" (ibid.) of adhering to AI ethics guidelines. It also addresses the danger of moral de-skilling (Vallor 2015, 2016), as every AI actor is actively involved in taking up their own ethical responsibilities and constantly works on bettering their rational decision-making abilities in the moral context. The fact that the right thing to do is the result of rigorous and honest rational deliberation on a case-by-case basis also means that this approach can deal with the fluidity of changing societal and political structures as well as the pace of AI technological advancement. In this way, AI ethics is then less about disciplining AI actors to adhere to ethical guidelines, and more about positive self-realisation of moral responsibilities as this model "emancipate[s AI actors] from potential inabilities to act self-responsibly on the basis of comprehensive knowledge, as well as empathy in situations where morally relevant decisions have to be made" (ibid., 114).

Only if every AI actor understands *why* regulating the life cycle of AI systems is necessary and sees their own role in this process, can the AI ethics project hope to be successful. The potential for meeting these objectives within a participatory virtue ethics approach to AI ethics as a dynamic ethical system should be clear.

4. Conclusion

The call for addressing the lack of impact of AI ethics on tech communities is real. In this paper, a novel participatory model for AI ethics based on a virtue ethics approach to AI ethics and underpinned by state of the art multi-disciplinary research and collaboration concretely anchored in research in the discipline of the ethics of AI has been suggested. Such an approach may do much to change the negative conception of AI ethics as stifling innovation by "broadening the scope of action, uncovering blind spots, promoting autonomy and freedom, and fostering self-responsibility" (Hagendorff 2020, 112-113). In addition, this approach can deal positively with the concern raised by Morley et al (2019) that, "in a digital context, ethical principles are not simply either applied or not, but regularly re-applied or applied differently, or better,

or ignored as algorithmic systems are developed, deployed, configured … tested, revised and re-tuned…" (ibid., 18), as it allows for AI ethics as a dynamic adaptive ethical system within which it is active cultivation of techno-moral virtues, rational deliberation among all AI actors and mutual respect for concrete multi-disciplinary research that guide ethical decisions.

In conclusion, let us consider what the implications for the concept of trustworthy AI are, should we meet the quest for actionable AI in the terms described above. First, trustworthiness becomes a socio-technical concept, focused as much on the safety and robustness of AI technologies as it is on respect for every individual human AI actor. In this context, given the active role of AI actors in the AI ethics project, and their shared responsibility to action-alise AI ethics, trust becomes a benchmark for the social acceptance of AI technologies. Thus, there will be good reason to trust that AI technology brings benefits while adequate measures are taken to mitigate risks, as the trust at issue is not only in technology but trust in the actions of AI actors actively involved in contributing to the dynamic model of AI ethics.[27]

But, secondly, it becomes clear that trustworthy AI itself should be an adaptive concept given the fast pace at which AI technologies advance and the adaptive nature of AI ethics argued for here. To capture the adaptive nature of the concept of trustworthy AI, I suggest introducing a concept of *AI ethics capital* (AIEC) as outcome of the participative model of AI ethics, girded by state of the art multi-disciplinary research, argued for in this paper. This notion of AIEC is related to the notion of national AI capital (NAIC), suggested by Momčilović (2020). The concept of NAIC links to the Organisation for Economic Cooperation and Development's notion of human capital (https://www.oecd.org/insights/humancapital-thevalueofpeople.htm) as the "knowledge, skills, competencies and characteristics of individuals that facilitate the creation of personal, social and economic wellbeing" (https://medium.com /@acomomcilovic/introducing-concept-national-ai-capital-a233832796c1).

National AI capital is a "country's capacity to apply and develop, and cope with the challenges of various artificial intelligence systems, in order to increase the country's social and economic well-being and competitiveness" (ibid.). I suggest that a subset of NAIC is the AI ethics capital (AIEC) of a country as the state of the art multi-disciplinary knowledge, skills, and competencies of individual AI actors, which drive individual AI actors' ethical habits and inform a country's AI ethics guidelines; which as such, in their turn, facilitate the creation of personal, social and economic wellbeing as a result of the potential of harmonious and ethical co-existence of humans with technology thus created. Measuring AIEC may seem difficult from a quantitative perspective, but it can, for now, be correlated with the level of adherence to AI ethics guidelines on a global index of AI ethics, alluded to as one incentive for AI ethics adherence in the UNESCO First Draft of the Recommendation on the Ethics of AI (UNESCO 2020).

The above points to the urgent need for future research in AI ethics as well as exploring cooperation among all AI actors at all stages of the AI technology lifecycle in

[27] See the first version of the UNESCO First Draft of the Recommendation on the Ethics of AI (2020).

the name of actionable AI ethics to find workable counters against potential new harms coming from AI technologies. Again here, more engagement with Vallor's (2016) ideas on "a future worth wanting" will be enlightening. For now, it is worthwhile to note, given that portraying AI ethics as a static concept seems almost a category mistake in the context of the fast pace of AI advancement, a notion of an equally non-static core ingredient of trustworthy AI, namely AIEC, generated both by AI technological and ethical advancement (due to a participative adaptive model of AI ethics), seems like an essential element of a successful approach to engage with the wide scope of constantly changing challenges AI actors and AI ethicists are confronted with on a daily basis.

References

1. Abdul, A., Vermeulen, J., Wang, D., et al.: Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. PROCEEDINGS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS - CHI '18, 1–18 (2018). https://doi.org/10.1145/3173574.3174156
2. Adams, F., Aizawa, K.: The Bounds of Cognition. 2nd Edn. Blackwell, Oxford (2010).
3. AI Ethics Global Inventory, last accessed 2020/9/20 https://inventory.algorithmwatch.org/
4. Allen, C., Varner, G., Zinser, J.: Prolegomena to any Future Artificial Moral Agent. Journal of Experimental & Theoretical Artificial Intelligence, 12(3), 251–261 (2000). https://doi.org/10.1080/09528130050111428
5. Alshammari, M., Simpson, A.: Towards a Principled Approach for Engineering Privacy by Design. In Schweighofer E., Leitold H., Mitrakas A., Rannenberg, K. (eds.) Privacy Technologies and Policy, Vol. 10518, pp. 161–177 (2017). https://doi.org/10.1007/978-3-31967280-9_9
6. Anabo, I. F., Elexpuru-Albizuri, I., & Villardón-Gallego, L.: Revisiting the Belmont Report's Ethical Principles in Internet-mediated Research: Perspectives from Disciplinary Associations in the Social Sciences. Ethics and Information Technology, 21(2), 137–149 (2019). https://doi.org/10.1007/s10676-018-9495-z
7. Ananny, M.: Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. Science, Technology, & Human Values, 41(1), 93–117 (2016).
8. Anderson, M. & Anderson, S. L.: Machine Ethics: Creating an Ethical Intelligent Agent. AI Magazine, 28(4), 15-26 (2007).
9. Asaro, P.M.: What Should We Want from a Robot Ethic? International Review of Information Ethics, 6(12), 9-16 (2006).
10. Asaro, P. M.: A Body to Kick, but Still No Soul to Damn: Legal Perspectives. In: Lin, P., Abney, K., Bekey G.A. (eds.) Robot Ethics: The Ethical and Social Implications of Robotics, pp.169-186. MIT Press, Cambridge (2012).
11. Audi, R.: Intrinsic Value and Reasons for Action. Southern Journal of Philosophy, 41, 30-56 (2003).
12. Barocas S., Selbst A.D.: Big Data's Disparate Impact. California Law Review, 104: 671-732 (2016).
13. Bekey, A. G.: Current Trends in Robotics: Technology and Ethics. In: Lin, P., Abney, K., Bekey, G.A. (eds.) Robot Ethics: The Ethical and Social Implications of Robotics, pp. 17-34. MIT Press, Cambridge MA (2012).

14. Benedikter, R., Siepmann, K., & Reymann, A.: 'Head-Transplanting' and 'Mind-Uploading': Philosophical Implications and Potential Social Consequences of Two Medico-Scientific Utopias. Review of Contemporary Philosophy, 16, 38–82 (2017).
15. Boden, M. A., Bryson, J. J., Caldwell, D., et al.: Principles of Robotics: Regulating Robots in the Real World. Connection Science, 29(2), 124–129 (2017).
16. Bostrom, N. & Yudkwosky, E.: The Ethics of Artificial Intelligence'. In: Frankish, K., Ramsey, W.M. (eds.) The Cambridge Handbook of Artificial Intelligence, pp. 316-334. Cambridge University Press, Cambridge (2014).
17. Brundage, M.: Limitations and Risks of Machine Ethics. Journal of Experimental & Theoretical Artificial Intelligence, 26(3), 355-372 (2014).
18. Campolo, A., et al.: AI Now 2017 Report (2017). https://assets.ctfassets.net/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be6 4c3ce0c78/_AI_Now_Institute_2017_Report_.pdf.
19. Chalmers, D.J.: Facing up to the Problem of Consciousness. Journal of Consciousness Studies, 2, 200-19 (1995).
20. Chalmers, D.J.: The Singularity: A Philosophical Analysis. Journal of Consciousness Studies, 17 (9–10), 7–65 (2010).
21. Clark, A.: Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence. Oxford University Press, Oxford (2003).
22. Clark, A.: Intrinsic Content, Active Memory and the Extended Mind. Analysis, 65(1), 1–11 (2005).
23. Clark, A. The Frozen Cyborg: A Reply to Selinger and Engström. Phenomenology of the Cognitive Sciences, 7, 343-346 (2008). https://doi.org/10.1007/s11097-008-9105-3
24. Clark, A. & Chalmers, D.: The Extended Mind. Analysis, 58, 7–19 (1998). https://doi:10.1093/analys/58.1.7
25. Comninos, A., Konzett, M.: Fabrics. Emerging AI Readiness. Martin Konzett KG, Zell am Moos, Austria (2018).
26. Corabi, J. & Schneider, S.: The Metaphysics of Mind Uploading. Journal of Consciousness Studies, 19(7–8), 26–44 (2012).
27. Couldry, N & Hepp, A.: The Mediated Construction of Reality. Polity Press, Cambridge UK (2017).
28. Crawford, K., et al.: The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term (2016). https://artificialintelligencenow.com
29. Crawford, K. & Calo, R.: There is a Blind Spot in AI Research. Nature, 538(7625), 311-313 (2016).
30. Danaher, J.: The Philosophical Case for Robot Friendship. Journal of Posthuman Studies, 3(1), 5-24 (2019). doi:10.5325/jpoststud.3.1.0005
31. Diakopoulos, N.: Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. Digital Journalism, 3(3), 398–415 (2015). https://doi.org/10.1080/21670811.2014.976411
32. Eliasmith, C.: How to Build a Brain: A Neural Architecture for Biological Cognition. Oxford University Press, Oxford (2013).
33. Floridi, L.(ed.): The Online Manifesto: Being Human in a Hyper Connected Era. Springer Open, Heidelberg (2015).
34. Floridi, L.: Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences, 374(2083), 1–13 (2016).
35. Floridi, L.: Establishing the Rules for Building Trustworthy AI. Nature Machine Intelligence, 1, 261-262 (2019a). https://doi.org/10.1038/s42256-019-0055-y

36. Floridi, L.: Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philosophy & Technology, 32, 185-193 (2019b). https://doi.org/10.1007/s13347-019-00354-x
37. Floridi, L., Cowls, J., et al.: AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines 28 (4), 689–707 (2018).
38. Floridi, L., Cowls, J.: A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1) (2019). https://doi.org/10.1162/99608f92.8cd550d1
39. Floridi, L., Taddeo, M.: What is Data Ethics? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083) (2016). https://doi.org/10.1098/rsta.2016.0360
40. Green, B.P.: Ethical Reflections on Artificial Intelligence. Scientia et Fides, 6(2), 9 (2018). https://doi.org/10.12775/SetF.2018.015
41. Greenhill, K.M, Oppenheim, B.: Rumor has It: The Adoption of Unverified Information in Conflict Zones. International Studies Quarterly, 61(3), 660–676 (2017). https://doi.org/10.1093/isq/sqx015
42. Gunkel, D. J.: The Machine Question: Critical Perspectives on AI, Robots, and Ethics. MIT Press, Cambridge MA (2012).
43. Hagendorff, T.: The Ethics of AI Ethics: An Evaluation of Guidelines. Minds & Machines 30, 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
44. Hansell, G.R.: H+/-: Transhumanism and Its Critics. Xlibris Corporation (2011).
45. Human Capital: The Value of People. https://www.oecd.org/insights/humancapital-thevalueofpeople.htm accessed 2020/9/21.
46. Innes, M., Dobreva, D., Innes, H.: Disinformation and Digital Influencing after Terrorism: Spoofing, Truthing and Social Proofing. Contemporary Social Science (2019). DOI: 10.1080/21582041.2019.1569714
47. Introducing Concept: National AI Capital. https://medium.com/@acomomcilovic/introducing-concept-national-ai-capital-a233832796c1 last accessed on 2020/9/21
48. Jobin, A., Ienca, M., Vayena, E.: The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 1, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
49. Kroll, J. A.: The Fallacy of Inscrutability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133) (2018). https://doi.org/10.1098/rsta.2018.0084
50. Latonero, M.: Governing Artificial Intelligence: Upholding Human Rights & Dignity'. *Data and Society*, USC (2018).
51. Leonelli, S.: Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems. Philosophical Transactions of the Royal Soceity A (2016). http://doi.org/10.1098/rsta.2016.0122
52. Lin, P., Abney, K. & Bekey, G.A.: Robot Ethics. The ethical and Social Implications of Robot Ethics. MIT Press, Cambridge, MA (2012).
53. McNamara, A., Smith, J., Murphy-Hill, E. Does ACM's Code of Ethics change Ethical Decision Making in Software Development? In Leavens, G.T., Garcia, A., Păsăreanu, C.S. (eds.) PROCEEDINGS OF THE 2018 26TH ACM JOINT MEETING ON EUROPEAN SOFTWARE ENGINEERING CONFERENCE AND SYMPOSIUM ON THE FOUNDATIONS OF SOFTWARE ENGINEERING—ESEC/FSE 2018, pp. 1–7. ACM Press, New York (2018).
54. Miller, C. & Coldicott, R.: People, Power and Technology: The Tech Workers' View. Retrieved from Doteveryone website: https://doteveryone.org.uk/report/workersview/ (2019).

55. Mittelstadt, B.: Principles Alone cannot Guarantee Ethical AI. Nature Machine Intelligence 1, 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
56. Momčilović, A.: NAIC Foundations. (2020). https://www.ssbm.ch/blog/naic-foundations-is-human-capital-the-only-thing-becoming-and-remaining-important-by-aco-momcilovic-emba/ last accessed 2020/9/21
57. Moor, J.H.: What is Computer Ethics? Metaphilosophy, 16(4), 266-275 (1985).
58. Moor, J.H.: The Nature, Importance, and Difficulty of Machine Ethics. IEEE, 21(4), 18-21 (2006).
59. Moore, G.E.: *Philosophical Papers*, Allen and Unwin, London (1959).
60. Morley, J., Floridi, L., Kinsey, L., et al.: From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics 26, 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5
61. Müller, V.C.: Ethics of Artificial Intelligence and Robotics. *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2020/entries/ethics-ai/>.
62. Pearlberg, D., Schroeder, T.: Reasons, Causes, and the Extended Mind Hypothesis. Erkenntnis, 81, 41-57 (2015). https://doi.org/10.1007/s10670-015-9727-0
63. Pekka, A.-P., Bauer, W., et al.: The European Commission's High-level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI. Working Document for Stakeholders' Consultation. Brussels, pp. 1–37 (2018).
64. Pigliucci, M.: Mind Uploading: A Philosophical Analysis. In: Blackford, R., Broderick, D. Intelligence Unbound: Future of Uploaded and Machine Minds. Wiley & Sons (2014). https://doi.org/10.1002/9781118736302.ch7
65. Raso, F., et al.: AI and Human Rights. Opportunities and Risks. *Berkman Klein Centre for Internet and Society*, Harvard (2018).
66. Royakkers, L., Timmer, J., Kool, L. & van Est, R.: Societal and Ethical Issues of Digitization. Ethics and Information Technology, 20(2), 127–142 (2018). https://doi.org/10.1007/s10676-018-9452-x
67. Royakkers, L. & van Est, R.: A Literature Review on New Robotics: Automation from Love to War. International Journal of Social Robotics, 7, 549-570 (2015).
68. Sandberg, A.: Feasibilty of Whole Brain Emulation. In: Müller, V. (eds.) Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemological and Rational Ethics, Vol. 5. Springer, Berlin (2013). https://doi.org/10.1007/978-3-642-31674-6_19
69. Sandberg, A., Bostrom, N.: Whole Brain Emulation: A Roadmap, Technical Report #2008–3, Future of Humanity Institute, Oxford University (2008). [Online], www.fhi.ox.ac.uk/reports/2008-3.pdf
70. Schneider, S.: Mindscan: Transcending and Enhancing the Brain, in Schneider, S. (ed.) Science Fiction and Philosophy. Wiley- Blackwell Hoboken, NJ (2009)
71. Segun, S.T.: From Machine ethics to Computational Ethics. AI and Society (2020). https://doi.org/10.1007/s00146-020-01010-1
72. Sharkey, A.J., Sharkey, N.: Granny and the Robots: Ethical Issues in Robot Care for the Elderly. Ethics and Information Technology, 14(1), 27-40 (2010).
73. Siau, K., Wang, W.: Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. Journal of Database Management, 31(2), 74-87 (2020). DOI: 10.4018/JDM.2020040105
74. Spielkamp, M., Matzat, L., et al.: Algorithm Watch 2019: The AI Ethics Guidelines Global Inventory. (2019). Retrieved from https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/.

75. Steffensen, S.V.: Language, Languaging and the Extended Mind Hypothesis. Pragmatics and Cognition, 17(3), 677-697 (2009). https://doi.org/10.1075/pc.17.3.10ste
76. Taddeo, M. & Floridi, L.: How AI can be a Force for Good. Science, 361(6404), 751–752 (2018). https://doi.org/10.1126/science.aat5991
77. Taylor, P.: Normative Discourse. Prentice-Hall, New York (1961).
78. Turkle, S.: Alone Together: Why we expect more from Technology and less from Each Other. Basic Books, New York (2011).
79. Turkle, S.: The Second Self: Computers and the Human Spirit. Simon and Schuster, New York (1984).
80. UNESCO Ad Hoc Expert Group (AHEG) for the Preparation of a Draft text of a Recommendation the Ethics of Artificial Intelligence SHS/BIO/AHEG-AI/2020/4 REV.2 https://unesdoc.unesco.org/ark:/48223/pf0000374266.
81. Vallor, S.: Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press, Oxford (2016).
82. Vallor, S.: Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. Philosophy and Technology, 28,107-124 (2015).
83. Veale M. & Binns R.: Mitigating Discrimination without Collecting Sensitive Data. Big Data & Society, July-December 2017, 1-17 (2017).
84. Veruggio, G., Operto F.: Roboethics: Social and Ethical Implications of Robotics. In: Siciliano B., Khatib O. (eds.): Springer Handbook of Robotics. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-30301-5_65
85. Wachter, S., Mittelstadt, B., Floridi, L.: Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99 (2017). https://doi.org/10.1093/idpl/ipx005
86. Wallach, W., Allen, C.: Moral Machines: Teaching Robots Right from Wrong. Oxford University Press, Oxford (2009).
87. Walker, M.: Personal Identity and Uploading. Journal of Evolution and Technology, 22(1), 37-51 (2011).
88. Wang, W., Siau, K.: Ethical and Moral Issues with AI: A Case Study on Healthcare Robots. TWENTY-FOURTH AMERICAS CONFERENCE ON INFORMATION SYSTEMS, New Orleans, August (2018).
89. Wiley, K.: A Taxonomy and Metaphysics of Mind Uploading. Humanity+ Press and Alautun Press, Seattle (2014).
90. Winfield, A.: An Updated Round Up of Ethical Principles of Robotics and AI. Retrieved from http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-ofethical.html (2019).