New Technologies for the Good of Whom? Whose Values Win?

A position paper for the PDC 2020 Conference workshop ”Computing Professionals for Social Responsibility: The Past, Present and Future Values of Participatory Design”

Atte Kinnula (VTT Technical Research Centre of Finland), Marianne Kinnula (University of Oulu), Harko Verhagen (Stockholm University)[1]

In this position paper we take artificial intelligence (AI) in the spotlight, as a current case exemplifying the struggle between different forces driving technology development. We argue for an approach of *acceptable technologies* – technologies that their users consider as good for themselves and the society, and we see that PD and its value base are a perfect environment to further this thinking, as what we discuss here is nothing else but politics of design.

To start with, we want to ask: Are we going towards a cyberpunk-style dystopian future, where world is ruled by unscrupulous megacorporations with only profit in mind, where human rights are devalued, and new technologies such as AI stands supreme? A provocative question, but in many countries we have grown to take it granted that society is there to support us and to protect individuals, ensuring that merely being far more powerful does not mean you can trample over human rights. However, over the last decade we have seen how technology has amplified the power shift that globalization started, with corporations – whose paradigm is to make money for their owners – becoming increasingly powerful compared to the governments that primarily look to maintain the stability of the society and prosperity of the average citizens. We have already gone past the point where a company - even a relatively small one, such as Cambridge Analytica – can deploy AI technology to influence an entire nation’s future and is ready to do so for the highest bidder. With such a power available for those who are not required to weigh society’s benefit in their course of action, are we not, in fact, marching towards that very dystopia?

AI, as a technology, is cited to have potential to bring huge societal benefits (see, e.g., Russel et al. 2015). We should ask, however, who in practice will benefit of this new technology. There is a concern that AI and Big Data may in fact instigate and aggravate the Matthew effect (Merton 1968), where those who are already well-off benefit more, and those who have initial disadvantages fall into a negative feedback loop, pushing them further down and widening the gap between the two groups as a third level digital divide effect (Lutz 2019, Noble 2018). And while different parties have published or are working on ethical guidelines and defining means to systematically and ethically evaluate risks posed by AI (e.g. Ciupa and Abney 2018, see Jobin et al. 2019 for an overview and analysis of these guidelines), there does

[1] Authors in alphabetical order

Cite as: Kinnula A., Kinnula, M., Verhagen, H. (2020) New technologies for the good of whom? Whose values win? A position paper for the PDC 2020 Conference workshop ”Computing Professionals for Social Responsibility: The Past, Present and Future Values of Participatory Design”.

not seem to be much thinking on who should be the one who makes the call whether or not the risks inherent are acceptable. Is it the party creating the solution, or the audience being affected? Who sets the values in play? The danger we see here is that those with the power to deploy intelligent technology are not the same who are affected, and that those in power will put their own interests first or may overvalue the benefits they assume those subjected to the AI's decision-making get, and underestimate the drawbacks and risks to the target group.

Fortunately, there is an increased awareness for the social, legal, and ethical issues within the AI research community, which has resulted in the development of different takes on what is now broadly termed "responsible AI" (Dignum 2019). The main focus in this is on ethics, either *ethics in design* (awareness of AI developers), *ethics by design* (behaviour of the AI systems), or *ethics for design* (what designers are allowed or not allowed). However, given the possibility that AI artefacts learn and thus change behaviour over time, these three ethical aspects require continuous assessment of the AI behavior over the lifetime of the product, which means after the regular design phase has finished. Thus, we argue that *ethics after design* is currently lacking.

Two things are needed to reach that: 1) The fluctuations in artefact behaviour over time due to learning, as well as the possibility to use AI artefacts in a context different from its initially intended context (or context of training), cause the design process to be in some sense never ending. Therefore, *user empowerment during the use lifetime* is essential. 2) The guidelines and checklists for ethical or responsible AI remain look at it from the 'expert' point of view. What is needed is what we call *acceptable AI*, i.e., what does the empowered user accept as 'good AI,' rather than designing good AI isolated from the exact context of use.

Our vision is a future society where technology – AI as an example – is designed from the ground up to be inherently supportive of the society and acceptable in a societal sense, not just engineered to perform a task in a given action context (the tool viewpoint), as we argue that seeing AI only as a tool, it is easy to forget what possible wider effects it can have on the humans that are subjected to it and as a consequence society at large. We argue that this should be the viewpoint for defining the values and ethics the technology must incorporate, as it is the subject that feels the impact, both the positive and negative. Essential part of our vision is empowerment of citizens, to have the ability to steer how technology is developed and to have a choice of where and when they are subjected to it (see e.g. Dindler et al. 2020, Iivari & Kinnula 2018). To enable this, we need structures, but it also requires a lot from both technology development as well as from citizens who need to 1) have skills to steer the technology development, 2) feel empowered to use those skills, and 3) see where the world should go and what is potentially good and what is bad in technology, 4) be willing to be reflective and mindful over this, and over the general good of the society, 5) have agency in this. In essence, a utopia rather than dystopia.

References

Ciupa, M., & Abney, K. (2018). Conceptualizing AI risk. Proceedings of the 8th International Conference on Computer Science, Engineering and Applications (CCSEA 2018), pp. 73– 81 Melbourne, Australia, February 17-18, 2018.

Dignum, V. (2019). Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way, Springer.
Dindler, C., Smith, R. C., & Iversen, O. S. (2020). Computational Empowerment: Participatory Design in Education. Codesign: International Journal of Cocreation in Design and the Arts.
Iivari, N., & Kinnula, M. (2018). Empowering children through design and making: towards protagonist role adoption. In Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1 (pp. 1-12).
Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines, Nature Machine Intelligence, 1 (sept 2019), pp. 389-399.
Lutz, Ch. (2019). Digital inequalities in the age of artificial intelligence and big data. Human Behavior and Emerging Technologies 1(2).
Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. Science, 159(3810), 56-63.
Noble, S. U. (2018). Algorithms of Oppression – How Search Engines Reinforce Racism. New York University Press.
Russel, S., Dewey, D., Tegmark, M. (2015). Research Priorities for Robust and Beneficial Artificial Intelligence, AI Magazine, 36(4), pp. 105-114.