

Ethics of AI as practical ethics

Florian Richter

School of Humanities and Education

Instituto Tecnológico y de Estudios

Superiores de Monterrey

Monterrey, Mexico

richter@tec.mx

Abstract—Whereas we acknowledge in general certain values as crucial for the ethical debate in AI, like e.g., fairness, transparency, and accountability, these values often can stand in conflict with each other. E.g., more transparency can lead to less privacy. Introducing higher principles to balance the values faces two problematic issues: (1.) Principles also can stand in conflict with each other and defer the discussion into a purely theoretical realm. (2.) If a higher-level principle is introduced and this also stands in conflict with another principle, then a higher-higher-level principle is needed, and we will get into an infinite regress. Although ethics of AI is part of the so-called field of applied ethics and it seems therefore that it is about the application of principles and values and finding the right balance regarding certain ethical theories, e.g., Kantian ethics or utilitarianism ([1][2][3]), the traditional approaches in the field of applied ethics do not offer sufficient conceptual means to deal with practical problems. Thus, the problems that arise from the implementation of intelligent systems can also not be handled adequately, because the same issue as above arises: How can the values be balanced regarding the ethical theories? If higher-level principles are not a viable approach to resolve conflicts of values, the criteria under which they can be implemented should be taken into consideration. Therefore, it is proposed that specific criteria for the implementation of the values need to be made explicit, which will result at least in a clarification for the public debate about certain technological advancements in the field of AI. [4] Furthermore, it is proposed that to resolve these conflicts the implementation must be evaluated, whether it enables human intervention in the future instead of making further actions and interventions impossible.

Keywords—*philosophy of technology, ethics of AI, practical ethics*

REFERENCES

- [1] Peter Singer, *Practical Ethics*, Cambridge University Press, Kindle Edition.
- [2] Dieter Birnbacher, *Bioethik zwischen Natur und Interesse*, Frankfurt am Main, 2006.
- [3] Ludwig Siep, *Konkrete Ethik*, Frankfurt am Main, 2004.
- [4] AI Ethics Impact Group, “From Principles to Practice, An interdisciplinary framework to operationalise AI ethics,” www.ai-ethicsimpact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/ai eig---report---download-hb-data.pdf