ABSTRACT debate contribution AI4EU workshop, Venice, 2-3 Sept. 2021

Evert F. Stamhuis

Erasmus University Rotterdam

Jean Monnet Centre of Excellence in Digital Governance

Another ingredient for good AI praxis

This debate contribution aims to explore the level of consensus on the steps required to arrive at ethically sound AI practices in the variety of contexts where AI systems are or will shortly be deployed. It intends to convey the message that we need further study of concrete contexts and emergent use cases to provide more solid ground for trustworthy AI practices. In that way, the principle based, rational-deductive approach will be supplemented by an empirical inductive approach that can inform concrete implementation policies and may even benefit the AI development process.

A level of consensus on AI ethical principles has clearly been achieved in the last few years. That was translated into a large package of assessment tools, with which users can assess or have assessed their deployment of AI in their practices. The EU proposal for an AI Act embraces these principles and translates several into concrete regulation. In the meantime, some concerns have been voiced within the AI Ethics community with regard to the turn from principles to practice. The AI industry's need for guidance is seen as a challenge (Vakkuria e.a. 2019; Morley e.a. 2020) and also the necessity of additional implementation of AI ethics for professions (Mittelstadt 2019).

Use cases emerge more quickly than good practices become available. We have ethical principles/values as architectural guidance, but not as concrete construction specifications. In the foreseeable future we will be able to add to that the (upcoming) certification systems that will help in providing good AI models. But still, we need a well-informed implementation, that we can only lay our hands on with a holistic approach. What actually will be the praxis is also determined by the ELSI aspects of the AI. That required factfinding research, not so much on machine behavior in concrete cases (Rahwan e.a. 2019), but more on the (pre)existing normative and empirical conditions in which the AI system, is meant to bring benefits.

Building on contextualization pleas that have already been published (Cowls e.a. 2019), we need to turn the attention of the AI ethics community to factfinding work, and therefore a focus on the bottom where the AI-system lands, the substrate in which it needs to form its roots and from which it must take its nutrition. Fortunately, we can benefit extensively from the cross-fertilization with advanced subfields of ethics that have already travelled the road from principles to empirically informed practices, such as bio/healthcare ethics and business ethics. These sister disciplines will hand us designs and methodologies to get a clear view in the actors in the case of AI. They can and need to be studied within the dynamics of the ecosystem in which they operate, which is their ethical atelier. ELSI is a common term from bioethics, adopted by the Dutch AI coalition for the next generation of funded research. Whether or not represented in this acronym, this factfinding considers socio/legal/ethical/techno/psychological dimensions, which, not surprisingly, all have some connection to the concept of trust and trustworthiness. As far as there are doubts in the AI Ethics community

regarding the relevancy of factfinding for normative ethics, we can learn from the stream of literature on the "is – ought" question in bioethics (Kon 2009, Frith 2012; Davies e.a. 2015; Spielthenner 2017). Empirical findings play an essential role in coming to arrangements in concrete cases for an ethically-sound AI deployment.

The presentation will conclude with sharing some provisional ideas for the design of the promoted factfinding endeavor. Adapting the model published by Susser and Grimaldi (Susser & Grimaldi 2021) and building on the discussions with researchers from Erasmus School of Law and Erasmus School of Health Policy and Management, the mapping of the legal landscape in this particular case and the study of pre-existing ethics are added to the traditional empirical methodologies. Those normative contextual factors will also affect and be affected by the integration of AI systems on the floor.

References

Davies R, J Ives, M Dunn (2015), A systematic review of empirical bioethics methodologies, 16 BMC Medical Ethics, nr. 15

Frith L (2012), Symbiotic Empirical Ethics: a practical methodology, Bioethics Vol 26 issue 4

Kon A A (2009) The role of empirical research in bioethics. 9 *American Journal of Bioethics*, 6–7, 59–65

Mittelstadt B (2019), Principles alone cannot guarantee ethical AI, Nature Machine Intelligence volume 1, 501–507

Morley J, L Floridi, L Kinsey, A Elhalal (2020), From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, arXiv:1905.06876

Rahwan L, M. Cebrian, N. Obradovich, J. Bongard, J. Bonnefon, C. Breazeal, J. W. Crandall, et al, 'Machine behaviour', *Nature,* 2019/568, p. 477 – 486

Spielthenner G (2017) The *Is-Ought* Problem in Practical Ethics, 24 HEC Forum, 277 – 292

Susser D & Vi Grimaldi (2021), Measuring Automated Influence: Between Empirical Evidence and Ethical Values. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021

Vakkuria V, K-K Kemella , J Kultanena , M Siponena , P Abrahamsson (2019), Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study, arXiv:1906.07946