

Current challenges in *Ethics of AI* or: old wine in new bottles

Contact information for questions and cooperation requests:

M. A. Leonie Seng, Ludwigsburg University of Education

Department: Philosophy and Media Ethics

Mail: leonie.seng@ph-ludwigsburg.de

First main thesis: Current discussions on *moral* development of *intelligent* machines/ robots/ algorithms/ programs in accordance to certain *ethical* criteria is old wine (in new bottles).

Why?

- Answering these questions requires first of all the systematic development of concrete concepts of *human beings* and their ideas of certain *moral values*. Floridi et al. 2018 distinguish in this context four fields and belonging questions:
 - autonomous self-realisation (*who we can become*),
 - human agency (*what we can do*),
 - individual and societal capabilities (*what we can achieve*) and
 - societal cohesion (*how we can interact with each other and the world*)
- The attempt to answer these and moreover *moral* questions (*what should we do*) has been part of most considerations since the very beginning of philosophical investigations.
- In this regard we deal with *old wine* when discussing current challenges.

Second main thesis: To answer the questions above, currently most approaches tend to a form of *general* Ethics of AI (Villani 2018; Floridi et al. 2018; Ethik-Kommission des BMVI 2017; Horvath 2015; Bostrom and Yudkowsky 2015; Ismajloska and Bakreski 2012 among others). Such approaches are not very promising.

Why?

- Talking *generally* about *Ethics of AI* necessarily leads to very general and hence superficial results, sometimes following the probably most prominent moral rules invented by Issac Asimov which he calls *Three Laws of Robotics*, first published in Asimov 1940.
- General *rules* or *laws* such as physical integrity of humans are very impractical though when developing (so called) *intelligent* machines or programs, because they are too vague.
- As Asimov's science-fiction stories (among others, see also McCarthy 2004¹) show, laws/ goals/ rules can come into conflict; moreover the tendency is to develop *learning* systems which are able to react spontaneously in certain situations.

¹While it has to be mentioned that the image of women drawn in this story unfortunately is surprisingly old school...

- PROBLEMS:

1. When trying to formulate *moral rules* more concretely, statements cannot be made but very specifically for certain domains.
2. Discussing utilitarian machine ethics contradicts current technological developments (see above).
3. Neither can discussing insoluble dilemma situations lead to a fruitful aim (cf. among many others Birnbacher and Birnbacher 2016) – qua concept! (To deepen this point cf. Seng 2017.)

- SOLUTIONS:

1. Not one general answer or definition can be considered as unique standard for all machines and *AI* programs in a pluralist world; it is rather necessary to develop concrete standards for robot development which comply with human moral ideas for different special areas in different countries and cultures. This results already from the fact that the development of military drones is of course much more delicate as the development of robot vacuum cleaners.
2. It is also absolutely necessary to install ethical supervisory bodies in every department where forms of artificial intelligence, androids or robots are developed, hence not only to name the word *ethics* but do what it stands for.
3. This is even more important as some countries already hold supervisory bodies which are called by name *ethics commission*, but hardly any ethicists take part in (cf. the Ethik-Kommission des BMVI 2017).

Some further remarks:

- Most approaches in *Ethics of AI* offer normative solutions so far (see examples above).
- For specifying normative rules it would be helpful to make declarative statements before to concrete general questions such as the classical Kantian, anthropological one *what is man?* or *how do we want to live?*.
- If not, normative rules are made up out of thin air instead of basing them on a systematic declarative ethical concept.
- OVERALL CONCLUSION: The challenge of an adequate approach to useful *Ethics of AI* therefore consists in drinking the appropriate old wine (means: defining systematically what to do on the base of philosophical ideas), depending on the particular question at hand (means: making a difference between sex dolls and vacuum cleaners) – instead of getting drunk by the alleged threats coming from AI (cf. Seng 2018b), may they be ever so present in the stories we are surrounded by (cf. Seng 2018a; Xanke and Bärenz 2012).

Bibliography

- ASIMOV, Isaac (1940): “Robbie”. In: *Super Science Stories*. New York: Popular Publications.
- BIRNBACHER, Dieter and BIRNBACHER, Wolfgang (2016): “Automatisiertes Fahren. Dieter und Wolfgang Birnbacher über ethische Fragen an der Schnittstelle von Technik und Gesellschaft”. In: *Information Philosophie* 4. Ed. by Peter MOSER.
- BOSTROM, Nick and YUDKOWSKY, Eliezer (2015): “The ethics of artificial intelligence”. In: *The Cambridge Handbook of Artificial Intelligence*. Ed. by Keith FRANKISH and William M. RAMSEY. Vol. 4. Cambridge: Cambridge University Press, pp. 316–334.
- ETHIK-KOMMISSION DES BMVI (2017): “Automatisiertes und vernetztes Fahren. Bericht: Juni 2017”. In: *Bundesministeriums für Verkehr und digitale Infrastruktur (BMVI)*, pp. 1–36. URL: <https://bit.ly/2IBhZZ8>. Aufgerufen am 19. Juni 2018.
- FLORIDI, Luciano et al. (2018): *AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Forthcoming in *Minds and Machines*, December 2018. URL: https://www.academia.edu/37691951/AI4People_-_An_Ethical_Framework_for_a_Good_AI_Society_Opportunities_Risks_Principles_and_Recommendations.
- HORVATH, Josip (2015): *The Ethics of Artificial Intelligence*. URL: <https://www.fsb.unizg.hr/brodogradnja/UZIR-2016-Essay-Josip-Horvat.pdf>.
- ISMAJLOSKA, Mersiha and BAKRESKI, Jane (2012): “The Ethics of Artificial Intelligence”. In: *ICT Innovations 2012 Web Proceedings*, pp. 539–542.
- MCCARTHY, John (2004): *The Robot and the Baby*. URL: <http://jmc.stanford.edu/articles/robotandbaby.html>.
- SENG, Leonie (2017): “Autonomes Fahren – Eine Frage der Ethik? Oder: Kant fährt Dein Auto gegen die Wand”. In: *Scilogs – Wissenschaftsblogs*. URL: <https://scilogs.spektrum.de/feuerwerk-der-neuronen/autonomes-fahren-eine-frage-der-ethik-oder-kant-faehrt-dein-auto-gegen-die-wand/>. Aufgerufen am 19. Juni 2018.
- (2018a): “Maschinenethik und Künstliche Intelligenz”. In: *Handbuch Maschinenethik*. Ed. by Oliver BENDEL. Wiesbaden: Springer VS, pp. 1–21.
- (2018b): “Mein Haus, mein Auto, mein Roboter? Eine (medien-) ethische Beurteilung der Angst vor Robotern und *künstlicher Intelligenz*”. In: *Brauchen Maschinen Ethik? Normative Grenzen autonomer Systeme*. Ed. by Matthias RATH, Friedrich KROTZ, and Matthias KARMASIN. Wiesbaden: Springer VS, pp. 57–72.
- VILLANI, Cédric (2018): *For a meaningful artificial intelligence. Towards a French and European Strategy*. URL: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.
- XANKE, Lisa and BÄRENZ, Elisabeth (2012): “Künstliche Intelligenz in Literatur und Film – Fiktion oder Realität?” In: *Journal of New Frontiers in Spatial Concepts* 4, pp. 36–43.