4th International Conference on Advanced Research Methods and Analytics (CARMA2022)
Universitat Politècnica de València, València, 2022
DOI: http://dx.doi.org/10.4995/CARMA2022.2022.15013

Ethical Behavior and Legal Regulations in Artificial Intelligence

Thomas Hauer
Department of Social Sciences, Technical University of Ostrava (VSB), Czech Republic.

Abstract

The European Union is also working intensively on the ethical development and use of artificial intelligence. The European Parliament, which commented on the issue of intelligent autonomous robots in January 2017, expressed the need to supplement the existing legal framework with ethical principles and an "effective ethical framework for the development, manufacture, use and modification of robots". This ethical framework should be based on the principles of expediency, harmlessness, autonomy and justice. The study analyzes the interconnection of ethical and legal rules in the field of AI and shows possible directions of development

Keywords: *machine ethics; values; ethical behavior; legal regulations; AI autonomous machines and platforms; moral philosophy.*

This work is licensed under a Creative Commons License CC BY-NC-ND 4.0
Editorial Universitat Politècnica de València

1. Introduction: Moral machine

The current state of affairs regarding the topic of Machine Ethics and the ethical autonomy of AI algorithms can be summarized as follows. Machines and platforms equipped with advanced AI and machine learning algorithms may differ in what their purpose is, even coming up with surprising solutions, plans and designs, but only in serving the goals that we set for them. The algorithm whose programmed goal is to "prepare a good dinner" may decide to serve steak, lasagne, or even a tasty new dish it creates by itself, but cannot decide to assassinate its owner, take his car and go to Iceland to rescue penguins because it became a vegetarian. Similarly, algorithms that are part of weapons systems, drones and unmanned aircraft, which choose their own targets without human intervention, fulfil their mission, adapt and respond to unforeseen circumstances with minimal human oversight, but cannot change or cancel their own mission if they had any moral reservations. It seems most rational to consider the ethical issues that may arise from these technologies before the technology is widely disseminated and deployed in practice (Allen et al., 2005). Moral machines capable of autonomous ethical reasoning and decision-making without any human oversight will necessarily emerge in the future. However, recent approaches to machine ethics have shown that researchers and programmers need to seek advice from philosophers and ethics to avoid novices' mistakes and to understand deep-rooted methodological problems in ethics better. If they fail to address these problems properly, their efforts to build adequate moral machines will be severely hampered. How to decide what steps are morally right is one of the most difficult questions in our lives. Understanding the ethical pitfalls and challenges associated with these decisions is essential to building intelligent, moral machines.

The first area of AI ethics, research deals with creating and applying ethical rules and standards. This area formulates recommendations that should respect fundamental rights, applicable regulations, and guiding principles and values, ensuring the ethical purpose of AI while ensuring their technical robustness and reliability. Ethical requirements and rules should be included in the various steps of the AI creation process, from research, through data collection, the initial design phase, testing the system to its deployment and practical application. Thus, this area of AI ethics mainly addresses questions about how developers should behave to minimize the ethical damage that may occur in AI, whether due to poor (unethical) design, inappropriate use, or misuse. This branch is commonly referred to as robotic ethics and has already led to the formulation of many declarations (Montreal Declaration for a Responsible Development of Artificial Intelligence), to postulating the main ethical principles and rules (Boddington, 2017; Boden 2016), formulating standards for producers and developers (International Organization for Standardization 2016), and designing best practices for developing and manufacturing platforms with AI (IEEE Standards Association 2017).

2. Machine ethics and ethical autonomy of AI

The most important and rapidly developing area in terms of AI ethics is the layer of ethics of autonomous intelligent systems and AI platforms evolving over time through self-learning from Machine Ethics data. The second branch of AI ethics research deals with how robots and AI platforms can behave ethically autonomously (Allen et al., 2005). This area of AI ethics research is referred to as Machine Ethics. The main aims and assumptions of this branch of machine ethics have been formulated by the authors and participants of the AAAI Fall 2005 symposium[1]. On this basis, authors W. Wallach and C. Allen have developed the term - artificial moral agents (AMAs), which is now used in this area of AI ethics. AMAs research can be considered a scientific endeavour to answer the question of whether, in principle, it is possible to model moral behaviour – whether ethical rules are convertible to algorithms autonomously (Anderson & Anderson, 2010). In many areas, it is impractical to wait for human decision-making because the amount of data, the speed of response, and waiting for human intervention make the decision impractical. In recent years, the interdisciplinary field of machine ethics – how to use machine learning to create algorithms with ethical rules to become either implicit or explicit moral factors – has become extremely important due to current and expected technological developments in computer science, artificial intelligence (AI) and Robotics (Gunkel, 2014; Lin et al., 2012). On the basis of great technological advances in AI, the emergence of fully autonomous, human-like, intelligent robots capable of ethical reasoning and decision-making seems inevitable. “Robots with moral decision making will become a technological necessity (Wallach, 2007). “Artificial Moral Agents are necessary and, in a weak sense, inevitable!”

I use the term “AMAs” to refer only to algorithms, platforms, and robots that are explicit ethical agents (those who have an explicit set of normative principles that they can use in decision making). AMAs come into play only when the task is fully automated (Moor 2006). Regarding autonomous technology, the transfer of moral roles to machines is not a matter of specific choice. Conversely, such delegation depends on the general characteristics of the automated task. If we choose to automate a task that does not, in any way, require the exercise of moral authority when performed by humans, there is no need for moral machines (AMAs). And vice versa, if a task requires some form of moral authority when it is performed by humans, then delegating the same task to autonomous machines necessarily means transferring a moral role. However, if autonomous machines are deployed to perform tasks that, when performed by human beings, show a moral aspect, then we either decide not to address that side, or we need to find a way to implement some form of morally relevant data processing and action selection into the machines themselves.

[1] https://www.aaai.org/Library/Symposia/Fall/fs05-06.php

"As systems get sophisticated even more, and their ability to function autonomously in different contexts and environments expands, it will become more important for them to have 'ethical subroutines' of their own" (Allen et al. 2006, p. 14).

This, of course, would not mean that the algorithm would "become" human. Instead, it would provide us with tools for cooperation that would not offend our moral sensitivity and satisfy our moral expectations. The general objective of AMAs research is complex. Researchers want to create machines with autonomous ethical decision making. Thus, it seems that the moral issues arising from the autonomous functioning of the machine need to be specifically addressed (Allen et al. 2005). This is what Machine Ethics is trying to do: to build machines that work not only efficiently, not only safely, but also in a morally satisfactory way – that is, in a way that would ideally prevent moral harm and agree to confirm moral goodness. In the long run, AI will be ubiquitous, and it should, because in many areas, it can do better work than humans. Not only will their intellectual prowess exceed ours, but their moral judgment may be better.

3. Human Well-being with Autonomous and Intelligent Systems

We seem to be in an intermediate period before the mass diffusion of a new and fundamental technology, which is advanced AI algorithms (Anderson & Anderson, 2007; Allen et al., 2006; Boden, 2016; Moor, 2006). As a strategic technology, AI is now rapidly developed and used around the world. However, it also brings with it new risks for the future of jobs and raises major legal and ethical questions (Lin et al., 2012). AI technologies should be developed, deployed and used with an ethical purpose and based on respect for fundamental rights, taking into account societal values and ethical principles of beneficence, non-maleficence, human autonomy, justice and explainability (Moor, 2006; Wallach, 2007). It is a prerequisite for ensuring the credibility of AI. In order to address the ethical risks and make the most of the opportunities that AI brings, the European Commission has published a European strategy on the Ethics of AI. It puts humans at the centre of AI development and defines so-called Human Centric Artificial Intelligence (HCAI).

4. Artificial Intelligence for Europe

At the European level, the European Commission's Communication Artificial Intelligence for Europe[2] and the Coordinated Plan on Artificial Intelligence "Made in Europe"[3] issued by the European Commission in December 2018 are the starting documents in the field of

[2] https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

[3] https://ec.europa.eu/digital-single-market/en/artificial-intelligence#Coordinated-EU-Plan-on-Artificial-Intelligence

AI. This Coordinated Plan sets out the European Union's strategic objectives and priorities in the field of artificial intelligence. It is the overarching European strategy for AI, which was developed in collaboration with the Member States and calls on the Member States at the national level to -implement the Coordinated Plan. The Member States are thus required to submit national AI strategies by the end of 2019 at the latest, including setting investment measures and implementation plans. In April 2018, the European Commission published a Communication on Artificial Intelligence for Europe, proposing a comprehensive and integrated European approach to AI. According to this document, the EU should respond to the current developments in AI and create a pan-European initiative focusing on three pillars:

- increasing technological and industrial capacity and deploying artificial intelligence across the economy,
- focus on socio-economic issues arising in the context of artificial intelligence (AI)
- providing an ethical and legal framework for AI technology.

The third pillar of EC Communication deals with legal and ethical issues related to AI. The European Commission has committed to developing ethical standards and guidelines for the use of AI. In this context, the High-Level Expert Group on Artificial Intelligence[4], which brings together AI experts, has been established to develop guidelines and recommendations on AI ethics. As part of the Communication, the EC also initiated the creation of the so-called European Artificial Intelligence Alliance, a broad discussion platform for various interest groups. The main strategic documents on AI ethical issues, which also provide a framework for this area, are:

1. Draft Ethics guidelines for trustworthy AI – published on 18 December 2018[5]
2. Communication: Building Trust in Human Centric Artificial Intelligence – published on 8 April 2019[6]

The third pillar, focusing on the ethical and legal context of AI development, is based on the above-mentioned strategic documents of the European Commission and formulates the main objective based on them. Credible human-centred AI has two components:

1. it should respect fundamental rights, applicable regulations, and the guiding principles and values shared in the EU, thereby ensuring the "ethical purpose" of AI
2. it should be technically robust and reliable because even without intentional malice, AI technologies can cause unintended harm or damage.

[4] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[5] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top

[6] https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence

5. Conclusion: Trustworthy AI

The ethical requirements for trustworthy AI should be incorporated into every step of the AI algorithm development process, from research, data collection, initial design phases, system testing, and deployment and use in practice. And how things really are? Do we really consider the benefits of AI versus the possible risks? Do we currently emphasize the ethical dimension of the development and implementation of new innovations in robotics and artificial intelligence?

Acknowledgements

Published with the support of the Technology Agency of the Czech Republic (TA CR), project number TL01000299.

References

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 28(4), pp. 15-26

Anderson, M., & Anderson, S. L. (2010). Robot be good: A call for ethical autonomous machines. Scientific American, 303(4), pp. 15–24

Allen, C., Smit, I., & Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. Ethics and Information Technology, 7(3), pp. 149–155. https://doi.org/10.1007/s10676-006-0004-4.

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? IEEE Intelligent Systems, 21(4), 12–17. https://doi.org/10.1109/MIS.2006.83.

Boddington, P. (2017). Towards a Code of Ethics for Artificial Intelligence (Artificial Intelligence: Foundations, Theory, and Algorithms), Springer; 1st ed

Boden, A. M. (2016). AI: Its Nature and Future, 1st Edition, Oxford University Press

Bryson, J. (2008). Robots should be slaves. In Y. Wilks (Ed.), Close Engagements with artificial companions: Key social, psychological, ethical and design issue (pp. 63–74). Amsterdam: John Benjamins Publishing.

Dignum, V. 2017. Responsible Artificial intelligence: Designing AI for human values, ITU Journal: ICT Discoveries, Special Issue No. 1, 25 Sept. 2017,

Gunkel, D. J. (2014). A vindication of the rights of machines. Philosophy & Technology, 27(1), pp. 113–132. https://link.springer.com/article/10.1007/s13347-013-0121-z

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4), pp. 18–21. https://doi.org/10.1109/MIS.2006.80

Lin, P., Abney, K., Bekey, G. (2012). Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, MA

Wallach, W. (2007). Implementing moral decision making faculties in computers and robots. AI & Society, 22(4), pp. 463–475. https://doi.org/10.1007/s0014 6-007-0093-6.