Machine ethics and "Artificial Morality": Can and should machines act morally?

Prof. Dr. med. Mathias Goyen
Industry Adviser
******@***goyen.de

Professor of Diagnostic Radiology
- Hamburg University, Germany
Chief Medical Officer - GE
Healthcare, Europe

LinkedIn Twitter

Machine ethics is a new field of research at the interface of computer science and philosophy that aims to develop moral machines. It's all about creating machines that can make moral decisions based on computer technology. This project is inspired by the latest developments in artificial intelligence. If artificial intelligence (AI) is to be used in the context of machine ethics, then one speaks analogously of "Artificial Morality" (AM).

While AI's goal is to model or simulate people's cognitive abilities, AM is about equipping artificial systems with the capacity for moral judgment and action.

The more complex and autonomous artificial systems become, the more likely they are able to self-regulate their behavior to a certain extent. This implies that AI-systems also get into situations that demand moral decisions. The seemingly simplest alternative to systems that can exercise this control themselves is to demand permanent monitoring and "online" control by a human user who then makes the morally relevant decisions. In many areas, however, this will hardly be possible - be it because of lack of personnel, because rapid decisions are needed, because the operational situations are too dangerous or because human intervention itself is a risk factor. Which kind of moral decisions a system has to make depends on the context of application.

One field of application for moral machines is the care of the elderly. Due to demographic change the proportion of people in need of care will increase sharply in the coming decades. Artificial systems are repeatedly brought into play as a means to counteract the nursing calamity. However, systems to be used in this context face moral choices, such as: how often and insistently does a care system remind people to eat and drink and to take medication? When should a care system inform the relatives or call the medical service if someone does not move for a while? Should the system monitor the user around the clock and what should be done with the data collected?

In all of these situations, an artificial system has to balance certain moral values: in the first case, between the self-determination of the user and certain health risks that arise if a person does not take his medication as prescribed. In the second case, between the self-determination of the user and the concern of the relatives, who might like to be informed immediately.

The third case is about the self-determination of the user, the care of the relatives as well as the privacy of the data collected.

A second example of the need for moral machines is autonomous driving. Even fully automated vehicles face moral choices. So it is, for example, to program them so that in unavoidable dangerous situations, the protection of human life takes precedence over property and animal damage. But also animals should be spared if possible. Of particular difficulty are the moral dilemmas that may be encountered in this area of application, such as the need to decide whether to sacrifice a small number of lives to save a larger number if unavoidable.

Last but not least are the military applications. The dream is that soldiers no longer have to risk their lives on the battlefield, but instead autonomous machines are sent into battle. These should be equipped with the law of war and context-specific rules of engagement, which limit their room for maneuver and ensure that they behave properly and morally. So they have to decide when an action is militarily necessary and appropriate and how combatants can be distinguished from civilians.

It could be argued, however, that it is not the care system, the autonomous car or the combat robot that make a moral choice in these cases, but

© For personal and private use only. Reproduction must be permitted by the copyright holder. Email to copyright@mindbyte.eu.

the programmers of those devices. But as the advances in artificial intelligence become greater, developers are less able to plan and predict what decisions a system will make in a specific situation. Even a chess program plays far better than its programmers, who can not predict every single move in the system. This is all the more true for such a complex system as AlphaGo Zero, which initially knows only the basic rules of the game Go and then finds the optimal decision-making strategies by playing through a large number of games against itself. In no time, this system managed to beat its predecessor AlphaGo, the first artificial system to defeat some of the world's top human Go players.

But even if one concedes that it makes sense in many applications, if machines could act morally, it is not clear whether they are capable of doing so.

Can artificial systems act morally? The development of increasingly intelligent and autonomous technologies inevitably leads them to confront morally problematic situations. Therefore, it is necessary to develop machines that have a degree of autonomous moral decision-making. It is unclear on what ethical basis artificial systems should decide. This also depends on the field of application and should be the subject of a social discourse, especially in those areas of application that require generally binding rules. There are plenty of research groups and initiatives both in academia and in the industry that start thinking about the relevance of ethics and safety in AI.

Originally published on LinkedIn.

Published on : Tue, 19 Jun 2018

© For personal and private use only. Reproduction must be permitted by the copyright holder. Email to copyright@mindbyte.eu.