

Could we forgive a machine? Study explores forgiveness in the context of robotics and AI

March 12 2020, by Ingrid Fadelli



Credit: Franck V. on Unsplash

As more robots make their way into society, it is important to consider the ethical and moral implications of having them complete tasks that

can have a significant impact on people's lives. If robots and machines are to become widely used in situations where they could seriously affect human lives, for instance by driving cars or giving elderly people the medication they need on a daily basis, developers should first consider the associated implications.

With this in mind, Michael Nagenborg, a researcher at the University of Twente in the Netherlands, has recently carried out a study investigating what could happen if a [robot](#) or machine makes a terrible mistake that has detrimental consequences for its users. His paper, published in Springer Link's *Technology, Anthropology, and Dimensions of Responsibility* journal, specifically explores the question of whether humans would be able to forgive a robot if it does something wrong, from a philosophical standpoint.

"The academic discussion on ethics is often focused on formulating [moral standards](#), while little attention is paid to what happens when people do not meet these standards," Nagenborg said. "Being able to forgive somebody, however, seems to be a crucial aspect of human existence and is closely linked with the way we hold each other responsible for our actions. I was wondering what role forgiveness could play when we consider robots as social entities."

In his recent study, Nagenborg explores a series of complex [ethical questions](#) related to blame and responsibility in robotics, such as 'how can human users deal with the fact that a robot made a crucial mistake' and 'can a robot be held fully responsible for its actions'? Rather than asking whether a robot can or should be punished for its mistakes, however, he focuses on whether humans would be able to forgive robots, even in cases where their actions resulted in irreversible and serious damage.

"Let's take a dramatic example: a self-driving car kills somebody you

love," Nagenborg explained. "Would you be willing and able to forgive the self-driving vehicle? I could imagine that I may forgive a human driver, but it seems to be more challenging to forgive a machine. I wanted to explore what this challenge reveals about our relationship to robots in the context of the academic debate about their morals status."

Many philosophy and ethics studies highlight forgiveness as a crucial aspect of interactions among human beings, as it can determine how they react to and learn to accept the mistakes of others. In his paper, Nagenborg assumes that forgiveness will also play a key role in a society where humans and robots coexist, particularly if the latter are assigned to tasks that can affect the lives of others.

"On the one hand, the difficulties to imagine what it could mean to forgive a robot, demonstrate that we should be very careful in holding a robot responsible for its actions," Nagenborg said. "On the other hand, if we are able to start forgiving machines in an honest and serious manner, we should be aware of the fundamental shift in how we perceive robots."

While Nagenborg's paper is merely based on speculations and philosophical ideas, it opens a window onto an important aspect of human-robot interactions that has rarely been explored before. His work could encourage others within the ethics community to consider the notion of forgiveness in robotics and AI development, particularly once robots start assisting humans in a wider variety of ways. Above all, Nagenborg also hopes to remind robotics researchers of the complexity of social interactions among humans and of how this complexity could one day also be reflected in human-robot interactions.

"Two of my current research projects are about the use of big data and AI in the [global south](#) for the purpose of 'slum mapping' and humanitarian responses," Nagenborg added. "One key issue here is the question of how we can explain and justify the decision taken on the

base on such technological systems to the people being mapped."

More information: Michael Nagenborg. *Can We Forgive a Robot?, Technology, Anthropology, and Dimensions of Responsibility* (2020).
[DOI: 10.1007/978-3-476-04896-7_11](https://doi.org/10.1007/978-3-476-04896-7_11)

© 2020 Science X Network

Citation: Could we forgive a machine? Study explores forgiveness in the context of robotics and AI (2020, March 12) retrieved 5 April 2024 from <https://techxplore.com/news/2020-03-machine-explores-context-robotics-ai.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
