

After 75 years, Isaac Asimov's Three Laws of Robotics need updating

17 March 2017, by Mark Robert Anderson



Credit: Shutterstock

When science fiction author Isaac Asimov devised his Three Laws of Robotics he was thinking about androids. He envisioned a world where these human-like robots would act like servants and would need a set of programming rules to prevent them from causing harm. But in the 75 years since the publication of the first story to feature his ethical guidelines, there have been significant technological advancements. We now have a very different conception of what robots can look like and how we will interact with them.

The highly-evolved field of robotics is producing a huge range of devices, from autonomous vacuum cleaners to military drones to entire factory production lines. At the same time, artificial intelligence and machine learning are increasingly behind much of the software that affects us on a daily basis, whether we're searching the internet or being allocated government services. These developments are rapidly leading to a time when robots of all kinds will become prevalent in almost all aspects of society, and human-robot interactions will rise significantly.

Asimov's laws are still mentioned as a template for guiding our development of robots. The [South Korean government](#) even proposed a [Robot Ethics](#)

[Charter](#) in 2007 reflecting the laws. But given how much robotics has changed and will continue to grow in the future, we need to ask how these rules could be updated for a 21st century version of [artificial intelligence](#).

The Three Laws

Asimov's suggested laws were devised to protect humans from interactions with robots. They are:

- A robot may not injure a human being or, through inaction, allow a human being to come to harm
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws

As mentioned, one of the obvious issues is that robots today appear to be far more varied than those in Asimov's stories, including some that are far more simple. So we need to consider whether we should have a threshold of complexity below which the rules might not be required. It is difficult to conceive a robotic vacuum cleaner having the capability of harming humans or even requiring an ability to obey orders. It is a robot with a single task that can be predetermined prior to it being switched on.

At the other end of the spectrum, however, are the robots designed for military combat environments. These devices are being designed for [spying, bomb disposal](#) or [load-carrying purposes](#). These would still appear to align with Asimov's laws, particularly as they are being created to reduce risk to human lives within highly dangerous environments.

But it is only a small step to assume that the ultimate military goal would be to create armed robots that could be deployed on the battlefield. In

this situation, the First Law – not harming humans – has been well researched in the field of AI and the development of rational behaviour has focused on limited, well defined areas. The role of the military is often to save the lives of soldiers and civilians but often by harming its enemies on the battlefield. So the laws might need to be considered from different perspectives or interpretations.

The laws' ambiguity has led [authors](#), including [Asimov](#), to explore how they could be misinterpreted or incorrectly applied. One issue is that they don't actually define what a robot is. As research pushes the boundaries of technology, there are emerging branches of robotics looking at more molecular devices.

For example, "robots" [made from DNA and proteins](#) could be used in surgery to correct gene disorders. In theory, these devices should really follow Asimov's laws. But for them to follow orders via DNA signals they would essentially have to become an integral part of the human they were working on. This integration would then make it difficult to determine whether the robot was independent enough to fall under the laws or operate outside of them. And on a practical level it would be impossible for it to determine whether any orders it received would cause harm to the human if carried out.

There's also the question of what counts as harming a human being. This could be an issue when considering the development of robot babies in Japan, for example. If a human were to adopt one of these robots it might arguably cause emotional or psychological harm. But this harm may not have come about from the direct actions of the robot or become apparent until many years after the human-robot interaction has ended. This problem could even apply to much simpler AI, such as the use of machine learning to create [music that elicits emotions](#).

Practical problems

The other big issue with the laws is that we would need a significant advancement in AI for robots to actually be able to follow them. The goal of AI research is sometimes described as developing machines that can [think and act rationally and like a human](#). So far, emulating human behaviour has not

With this in mind, a [robot](#) could only operate within a very limited sphere and any rational application of the laws would be highly restricted. Even that might not be possible with current technology, as a system that could reason and make decisions based on the laws would need considerable computational power.

Given all these issues, Asimov's laws offer little more than founding principles for someone wanting to create a robotic code today. We need to follow them with a [much more comprehensive set of laws](#). That said, without significant developments in AI, implementing such laws will remain an impossible task. And that's before we even consider the potential for hurt should humans start to fall in love with robots.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

APA citation: After 75 years, Isaac Asimov's Three Laws of Robotics need updating (2017, March 17)
retrieved 4 July 2022 from <https://techxplore.com/news/2017-03-years-isaac-asimov-laws-robotics.html>

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.