When robots commit wrongdoing, people may incorrectly assign the blame

April 5 2019

Credit: CC0 Public Domain

Last year, a self-driven car struck and killed a pedestrian in Tempe, Arizona. The woman's family is now suing Arizona and the city of Tempe for negligence. But, in an article published on April 5 in the

journal *Trends in Cognitive Sciences*, cognitive and computer scientists ask at what point people will begin to hold self-driven vehicles or other robots responsible for their own actions—and whether blaming them for wrongdoing will be justified.

"We're on the verge of a technological and social revolution in which autonomous machines will replace humans in the workplace, on the roads, and in our homes," says Yochanan Bigman of the University of North Carolina, Chapel Hill. "When these robots inevitably do something to harm humans, how will people react? We need to figure this out now, while regulations and laws are still being formed."

The article explores how the human moral mind is likely to make sense of robot responsibility. The authors argue that the presence—or perceived presence—of certain key capacities could make people more likely to hold a machine morally responsible.

Those capacities include autonomy, the ability to act without human input. The appearance of a robot also matters, as the more humanlike a robot looks, the more likely people are to ascribe a human mind to it. Other factors that can lead people to perceive robots as having "minds of their own" include an awareness of the situations they find themselves in as well as the ability to act freely and with intention.

Such issues have important implications for people in their interactions with robots. They're also critical considerations for the people and companies who create and operate autonomous machines—and the authors argue that there could be cases where robots that take the blame for harm caused to humans could shield the people and companies who are ultimately responsible for programming and directing them.

As the technology continues to advance, there will be other intriguing questions to consider, including whether robots should have rights.

Already, the authors note, the American Society for the Prevention of Cruelty to Robots and a 2017 European Union report have argued for extending certain moral protections to machines. They explain that such debates often revolve around the impact machine rights would have on people, as expanding the moral circle to include machines might in some cases serve to protect people.

While robot morality might still sound like the stuff of science fiction, the authors say that's exactly why it's critical to ask such questions now.

"We suggest that now—while machines and our intuitions about them are still in flux—is the best time to systematically explore questions of robot morality," they write. "By understanding how human minds make sense of morality, and how we perceive the mind of machines, we can help society think more clearly about the impending rise of robots and help roboticists understand how their creations are likely to be received."

As the early experience in Tempe highlights, people are already sharing roads, skies, and hospitals with autonomous machines. Inevitably, more people will get hurt. How robots' capacity for moral responsibility is understood will have important implications for real-world public policy decisions. And those decisions will help to shape a future in which people may increasingly coexist with ever more sophisticated, decision-making machines.

More information: *Trends in Cognitive Sciences*, Bigman et al.: "Holding Robots Responsible: The Elements of Machine Morality" www.cell.com/trends/cognitive- … 1364-6613(19)30063-4 , DOI: 10.1016/j.tics.2019.02.008

Provided by Cell Press