

For robotics and AI, great power comes with great responsibility

October 24 2017, by Allie Nicodemo



"AI and robotics have the potential to radically transform employment and displace a lot of workers," said John Basl, Assistant Professor of Philosophy. Basl's research focus on the ethical implications of new technologies. Credit: Matthew MODOONO/Northeastern University

Advances in artificial intelligence and robotics stand to make our lives

better. Dangerous jobs could be outsourced. Huge datasets could be analyzed instantly. Boring tasks could be automated.

But with any new technology comes risk. For example, the ability to organize and edit human genes—while promising for personalized medical treatments—raises questions about how comfortable we are exerting control over our DNA. Cellphones make navigation, communication, and countless other aspects of life more convenient. They also collect an unprecedented amount of personal information, forcing society to rethink the importance of privacy.

As robots join the workforce and intelligent algorithms are weaved into daily life, are we ready for what comes next?

John Basl, assistant professor of philosophy at Northeastern, studies the ethical implications of these emerging technologies. Here, he describes the moral questions facing society as robotics and [artificial intelligence](#) evolve, as well as the challenges still on the horizon, and those we must grapple with quickly.

What are the most important ethical concerns associated with robotics and artificial intelligence?

There are different ways in which [ethical concerns](#) can be important. They could be important in that they are pressing. Or, they could be important because, even if we don't need to deal with them right away, the cost of a mistake is ethically costly. As an example of this contrast, consider the difference between autonomous vehicles and artificial intelligence that is conscious and very much like us. Both of these raise ethical concerns, the first because we need to resolve it very soon, and the second, because even though we aren't close to having such an AI, its existence raises all sorts of ethical concerns that we don't know how to

resolve. People working in AI typically focus on the near term as the more pressing, but people like Elon Musk and others have serious worries about far-term ethical concerns.

Focusing on the near-term consequences you referenced, what are the ones that are most pressing?

One is related to economic justice. AI and robotics have the potential to radically transform employment and displace a lot of workers. As an example, consider that there have already been tests of autonomous semi-trucks. There are more than 3 million professional truck drivers in the U.S. This is just one industry where AI stands to displace workers. We are not yet, as a society, prepared for this economic impact, which is likely to have the greatest impact on those who are already socioeconomically disadvantaged.

Another important factor is accountability. Consider, again, autonomous cars. Who is to blame when a car gets in an accident? How do we assign responsibility? We need to find the right balance of values to maximize the benefits while minimizing the ethical costs. This will be an issue in any context where it is important to attribute responsibility, such as in the use of autonomous weapons or AI for medical diagnostics.

There is also the question of whether AI might result in socially unjust outcomes. Learning algorithms might be able to accurately make predictions by finding correlations in a huge dataset. But some of those correlations might be due to social biases or injustices, and so it might continue to propagate those or exacerbate them.

Also, dual-use problems arise for almost every technology. A dual-use problem is the problem of a technology designed for use in one context being adapted for another, potentially harmful, use. For example, a

technology developed for the military might find use in a civilian context and that might lead to bad outcomes.

Also, there has been some concern recently that AI and learning algorithms might have the effect of putting us in information bubbles. Facebook, in order to show me content I like, might use an algorithm that learns to filter out the political views I don't like or even news articles that I might not like. Doing so risks undermining the common ground we have to engage with one another on important issues. It can increase polarization.

You also referenced far-term ethical concerns. What are some examples?

AI safety concerns include how to ensure that AI doesn't become so intelligent that it poses an existential threat to humans. The worry is that AI might become so adept at problem solving or realizing its own ends that it results in the deaths of massive numbers of people. There are lots of scenarios that are considered, from conscious, super-intelligent AI that doesn't like us, to non-conscious AI that just runs amok but is intelligent enough that we can't stop it. This is far term because the probability of achieving such an AI might be very low, but if we do achieve it and can't control it, the cost is massive.

Another topic that I think is important, but also far-term, concerns our treatment of AI. If we are able to create conscious AI, we might very well mistreat it. Imagine if we simulated a dog's consciousness in a computer and then used it in experiments that emulate pain. Well, if the simulation is really conscious, we are causing real pain. In the case of animal and human subjects research, we have oversight to protect research subjects. We don't have anything like that yet for AI research. Some of my work concerns how to develop such oversight.

Provided by Northeastern University

Citation: For robotics and AI, great power comes with great responsibility (2017, October 24)
retrieved 24 April 2024 from

<https://phys.org/news/2017-10-robotics-ai-great-power-responsibility.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.