

Artificial intelligence researchers must learn ethics

August 29 2017, by James Harland



Drones deliver more than just food. Credit: www.routexl.com, CC BY-NC-SA

Scientists who build artificial intelligence and autonomous systems need a strong ethical understanding of the impact their work could have.

More than 100 technology pioneers recently published an [open letter to the United Nations](#) on the topic of lethal autonomous weapons, or "killer robots".

These people, including the entrepreneur Elon Musk and the founders of several robotics companies, are part of an effort that [began in 2015](#). The original letter called for an end to an arms race that it claimed could be the "third revolution in warfare, after gunpowder and nuclear arms".

The UN has a role to play, but responsibility for the future of these systems also needs to begin in the lab. The education system that trains our AI researchers needs to school them in ethics as well as coding.

Autonomy in AI

Autonomous systems can make decisions for themselves, with little to no input from humans. This greatly increases the usefulness of robots and similar devices.

For example, an autonomous delivery drone only requires the delivery address, and can then work out for itself the best route to take – overcoming any obstacles that it may encounter along the way, such as adverse weather or a flock of curious seagulls.

There has been a great deal of research into [autonomous systems](#), and delivery drones are currently being developed by companies such as [Amazon](#). Clearly, the same technology could easily be used to make deliveries that are significantly nastier than food or books.

Drones are also becoming smaller, cheaper and more robust, which means it will soon be feasible for flying armies of thousands of drones to be manufactured and deployed.

The potential for the deployment of weapons systems like this, largely decoupled from human control, prompted the letter urging the UN to "find a way to protect us all from these dangers".

Ethics and reasoning

Whatever your opinion of such weapons systems, the issue highlights the need for consideration of [ethical issues](#) in AI research.

As in most areas of science, acquiring the necessary depth to make contributions to the world's knowledge requires focusing on a specific topic. Often researchers are experts in relatively narrow areas, and may lack any formal training in ethics or moral reasoning.



A Google self-driving car. Credit: Roman Boed, CC BY-NC

It is precisely this kind of reasoning that is increasingly required. For example, driverless cars, which are [being tested in the US](#), will need to be able to make judgements about potentially dangerous situations.

For instance, how should it react if a cat unexpectedly crosses the road? Is it better to run over the cat, or to swerve sharply to avoid it, risking injury to the car's occupants?

Hopefully such cases will be rare, but the car will need to be designed with some specific principles in mind to guide its decision making. As Virginia Dignum put it when delivering her paper "[Responsible Autonomy](#)" at the recent International Joint Conference on Artificial Intelligence ([IJCAI](#)) in Melbourne: "The driverless car will have ethics; the question is whose?"

A similar theme was explored in the paper "[Automating the Doctrine of Double Effect](#)" by Naveen Sundar Govindarajulu and Selmer Bringsjord.

The [Doctrine of Double Effect](#) is a means of reasoning about moral issues, such as the right to self-defence under particular circumstances, and is credited to the 13th-century Catholic scholar [Thomas Aquinas](#).

The name Double Effect comes from obtaining a good effect (such as saving someone's life) as well as a bad effect (harming someone else in the process). This is a way to justify actions such as a drone shooting at a car that is running down pedestrians.

What does this mean for education?

The emergence of ethics as a topic for discussion in AI research suggests that we should also consider how we prepare students for a world in which autonomous systems are increasingly common.

The need for "[T-shaped](#)" people has been recently established. Companies are now looking for graduates not just with a specific area of technical depth (the vertical stroke of the T), but also with professional skills and personal qualities (the horizontal stroke). Combined, they are able to see problems from different perspectives and work effectively in multidisciplinary teams.

Most undergraduate courses in computer science and similar disciplines include a course on professional ethics and practice. These are usually focused on intellectual property, copyright, patents and privacy issues, which are certainly important.

However, it seems clear from the discussions at IJCAI that there is an emerging need for additional material on broader ethical issues.

Topics could include methods for determining the lesser of two evils, legal concepts such as criminal negligence, and the historical effect of technology on society.

The key point is to enable graduates to integrate ethical and societal perspectives into their work from the very beginning. It also seems appropriate to require research proposals to demonstrate how ethical considerations have been incorporated.

As AI becomes more widely and deeply embedded in everyday life, it is imperative that technologists understand the society in which they live and the effect their inventions may have on it.

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

Provided by The Conversation

Citation: Artificial intelligence researchers must learn ethics (2017, August 29) retrieved 2 May 2024 from <https://techxplore.com/news/2017-08-artificial-intelligence-ethics.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.