We must be sure that robot AI will make the right decisions, at least as often as humans do

December 9 2014, by Sandor Veres

Artificial intelligences with the power of life or death over humans – what could possibly go wrong? Credit: Richard Greenhill and Hugo Elias, CC BY-SA

Your autonomous vacuum cleaner cleans your floors and there is no great harm if it occasionally bounces into things or picks up a button or a

scrap of paper with a phone number. But then again this latter case is irritating – it would be preferable if the machine was capable of noticing there was something written on it and alert you. A human cleaner would do that.

If your child has a toy robot, you are not worried much about its wheels, arms or eyes going wild occasionally during play. It can be just more fun for the kids. You know the toy has been designed not to have enough force to cause any harm.

But what about a factory robot designed for picking up cars pieces and fitting them into a car? Clearly you'd not want to be nearby when it goes berserk. You know it's been pre-programmed to do particular tasks and it may not welcome your proximity. This kind of robot is often caged or barred-off, even for operating personnel. But what about the case of some future [autonomous robot](#) with which you need to work in order to assemble something, or complete some other task? You may think that if it is powerful enough to be useful, it may also be powerful enough to do you an unexpected injury.

If you fly model aircraft then you may want to put a GPS-equipped computer on board and make it follow waypoints, perhaps to take a series of aerial photos. There may be two points of concern. First, the legality of flying your aircraft when occasionally out of your sight, as in the case of some trouble you would not notice that the automatic control needs to be overridden for safety. Second, whether its on-board software has been sufficiently well written for a safe emergency landing if required. Might it endanger the public, or cause damage to something else airborne or otherwise?

Your latest luxury car with its own intelligent sensor system for recognising the environment around it may be forced to choose between two poor options: to hit a car that suddenly appears in the street, or to

brake causing the car behind collide into you. As a passenger in an [autonomous car](#) travelling in a convoy of other [autonomous vehicles](#), you may wonder what the car might do if the convoy arrives at a junction or road works or if the vehicle in the convoy suffers a breakdown: can the [autonomous systems](#) be trusted to navigate itself through temporary barriers or sudden disruptions without harming the pedestrians or vehicles around it?

The right choices at the right time

These are questions that pose real challenges for those designing and programming our future semi-autonomous and autonomous robots. All possible dangerous situations need to be anticipated and accounted for, or resolved by the robots themselves. Robots also need to be able to safely recognise objects in their environment, perceive their functional relationship to them and make safe decisions about their next move, and when they are able to satisfy our requests.

For some applications, such as with [humanoid robots](#), it's not clear today where the responsibility lies: with the manufacturer, with the robot, or with its owner. In a case where damage or harm is caused, it may be that the user taught the robot the wrong thing, or requested something inappropriate of it.

There is still a legal framework to be introduced, something that at the moment is still entirely missing. If various software systems are used, how can we check that the robot's decisions are safe? Do we need a UK authority to certify autonomous robots? What will be the rules that robots need to keep to and how will it be verified that they are safe in all practical situations?

The EPSRC-supported research thgat we have [recently launched](#) at the universities of Sheffield, Liverpool and West of England in Bristol are

trying to establish answers and solutions to these questions that will make autonomous robots safer. The three-year project will examine how to formally verify and ultimately legally certify robots' decision-making processes. Laying down methods for creating this will in fact help define a legal framework (in consultation with lawyers) that will hopefully allow the UK robotics industry to flourish.

This story is published courtesy of The Conversation *(under Creative Commons-Attribution/No derivatives).*

Source: The Conversation