Who's to blame when artificial intelligence systems go wrong?

August 17 2015, by Gary Lea

Robots in chains but are they really to blame when AI does something wrong? Credit: maxuser

There has been much discussion of late of the ethics of artificial intelligence (AI), especially regarding robot weapons development and a related but more general discussion about AI as an existential threat to humanity.

If Skynet of the Terminator movies is going to exterminate us, then it seems pretty tame – if not pointless – to start discussing regulation and liability. But, as legal philosopher John Donaher has pointed out, if these areas are promptly and thoughtfully addressed, that could help to reduce existential risk over the longer term.

In relation to AI, regulation and liability are two sides of the same safety/public welfare coin. Regulation is about ensuring that AI systems are as safe as possible; liability is about establishing who we can blame – or, more accurately, get legal redress from – when something goes wrong.

The finger of blame

Taking liability first, let's consider tort (civil wrong) liability. Imagine the following near-future scenario. A driverless tractor is instructed to drill seed in Farmer A's field but actually does so in Farmer B's field.

Let's assume that Farmer A gave proper instructions. Let's also assume that there was nothing extra that Farmer A should have done, such as placing radio beacons at field boundaries. Now suppose Farmer B wants to sue for negligence (for ease and speed, we'll ignore nuisance and trespass).

Is Farmer A liable? Probably not. Is the tractor manufacturer liable? Possibly, but there would be complex arguments around duty and standard of care, such as what are the relevant industry standards, and are the manufacturer's specifications appropriate in light of those standards? There would also be issues over whether the unwanted planting represented damage to property or pure economic loss.

So far, we have implicitly assumed the tractor manufacturer developed the system software. But what if a third party developed the AI system? What if there was code from more than one developer?

Over time, the further that AI systems move away from classical algorithms and coding, the more they will display behaviours that were not just unforeseen by their creators but were wholly unforeseeable. This is significant because foreseeability is a key ingredient for liability in negligence.

To understand the foreseeability issue better, let's take a scenario where, perhaps only a decade or two after the planting incident above, an advanced, fully autonomous AI-driven robot accidentally injures or kills a human and there have been no substantial changes to the law. In this scenario, the lack of foreseeability could result in nobody at all being liable in negligence.

Blame the AI robot

Why not deem the robot itself liable? After all, there has already been some discussion about AI personhood and possible criminal liability of AI systems.

But would that approach actually make a difference here? As an old friend said to me recently:

Will AI systems really be like Isaac Asimov's Bicentennial Man – obedient to the law, with a moral conscience and a hefty bank balance?

Leaving aside whether AI systems can be sued, AI manufacturers and developers will probably have to be put back into the frame. This might involve replacing negligence with strict liability – liability applied without any need to prove fault or negligence.

Strict liability already exists for defective product claims in many places. Alternatively there could be a no fault liability scheme with a claims pool contributed to by the AI industry.

Rules and regulations

On the regulatory side, development of rigorous safety standards and establishing safety certification processes will be absolutely essential. But designing and operating a suitable framework of institutions and processes will be tricky.

AI expert input will be needed in establishing any framework because of the complexity of the area and the general lack of understanding outside the AI R&D community. This also means that advisory committees to legislatures and governments should be established as soon as possible.

Acknowledging that there are potentially massive benefits to AI, there will be an ongoing balancing act to create, update and enforce standards and processes that maximise public welfare and safety without stifling innovation or creating unnecessary compliance burdens.

Any framework developed will also have to be flexible enough to take account of both local considerations (the extent of own production versus import of AI technology in each country) and global considerations (possible mutual recognition of safety standards and certification between countries, the need to comply with any future international treaties or conventions etc).

So as we travel down the AI R&D path, we really need to start shaping the rules surrounding AI, perhaps before it's too late.

We've already started discussions around driverless cars – see here and here – but there's so much more to deal with when it comes to AI.

What do we do next? Over to you.

This story is published courtesy of The Conversation *(under Creative*

Source: The Conversation

Citation: Who's to blame when artificial intelligence systems go wrong? (2015, August 17) retrieved 3 May 2024 from https://phys.org/news/2015-08-blame-artificial-intelligence-wrong.html