3rd December 2016   Why (or rather, when) suffering in AI is incoherent.

I've been arguing for some months now [http://joanna-bryson.blogspot.com/2016/09/what-makes-person-five-reasons-not-to.html] in public talks that AI cannot be a legal person because suffering in well-designed AI is incoherent. This is not actually my own argument, but rather is due to S. M. Solaiman from their brilliant recent article Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy [http://link.springer.com/article/10.1007/s10506-016-9192-3] .

The great thing about Solaiman's article is that they make it clear why corporations are legal persons but AI and chimpanzees aren't. Basically, the notion of legal person has been developed to synchronise with our system of justice. Justice among other things requires means of redress and coercion. A legal person most know and be able to claim their rights–they must be able to assert themselves as members of a society. This is why non-human animals (and some incapacitated humans) are not legal persons. I'm happy with definitions of "know" and "assert" that would mean that intelligent artefacts could do this, and indeed organisations of humans can meet these criteria. However, legal

[http://gunshowcomic.com/513]

From Gunshow by KC Green

persons must also like any person care about the kinds of sanctions that justice can use against them, so that justice can bind society together. Justice is part of the way a society composes itself of individuals, it is part of the glue that turns a set of individuals into a group. To date this has not involved direct "joystick" manipulation, but rather sanctions that individuals find aversive, such as loss of time and/or social status. Non-human animals can definitely suffer, as I said, the only thing that makes them not legal persons is their incapacity to understand and argue their rights. But can AI suffer? Solaiman concludes "not yet", but I would like to go further.

Pain, suffering, and concern for social status are things essential to a social species, and as such they are integral to our intelligence. I've read it argued that one of the characteristics of a sociopath is missing this part of what it is to be human, but I'm not an expert on clinical psychology so don't know whether this is ever truly possible even in these extreme cases. But safe, well architected (designed) AI tends to be modular. I and my students build systems of emotions for AI [http://www.cs.bath.ac.uk/~jjb/web/ai.html#emot] , and I'm happy to have some variable that represents an emotion on which other behaviour depends. I don't have an issue with saying a robot is excited or depressed if its action expression has been increased or inhibited, respectively.

But I have an issue with saying that a well-designed robot is suffering, because suffering is defined to be something sufficiently aversive that you would avoid it. But anything we insert into a well designed AI we (or conceivably it) could extract and isolate. This isn't the nature of suffering.

I am not saying I cannot conceive of AI that could suffer. This may not be likely or even possible, but if we did actually construct AI by scanning in a human or animal (really, it would have to be the whole thing, not just the brain – brain cells reference muscle cells and such like) then no doubt it could suffer. But that would not be well-designed AI; rather, that would be a sort of clone [http://joanna-bryson.blogspot.com/2015/10/clones-should-not-be-slaves.html] . Owning human clones strikes me as deeply unethical – exactly as unethical as owning humans. Building synthetic clones of animals also more generally strikes me as deeply inefficient; we would be better off working with the animals we have rather than constructing clones of them from materials that aren't as well suited to purpose as their original biology.

So I find it extremely unlikely we will ever have AI suffering, but even if we do, what I recommend (in keeping with the British EPSRC Principles of Robotics [http://joanna-bryson.blogspot.com/2016/03/the-meaning-of-epsrc-principles-of.html]) is that it should never be a legal product for manufacture and purchase. And therefore any AI which would ever be bought or sold should not be considered a legal person.

Note that there are also other reasons not to make AI a legal person, most importantly that taxing robots allows the corporations who actually decide to use them rather than humans off the hook [http://joanna-bryson.blogspot.com/2016/06/robots-are-owned-owners-are-taxed.html] for that decision, displaces those corporations' liability, and also afford the opportunity to hack up robots to minimise the amount of tax that will be paid. Robots and AI are not human; they do not come in discrete pre-defined units. They are artefacts, and as such, our own, authored responsibility. [http://joanna-bryson.blogspot.com/2015/03/robots-are-more-like-novels-than.html]

To get back to corporations, Solaiman says these can suffer to the extent that the humans in them suffer, and/or to the extent that losing their assets is equivalent to human suffering. The latter argument is weaker I think, but it has a lot of historic legal precedent. Nevertheless, the fact that corporations don't really suffer may well be why we have a number of problems with corporations being treated as legal persons. Confounding these problems by declaring AI or robots to be legal persons would almost certainly not be wise.

Posted 3rd December 2016 by Joanna Bryson

Labels: AI, ethics, policy

View comments

2 comments

Add a comment as Laurent Philippe

Top comments

Joanna Bryson via Google+   1 month ago - Shared publicly

An essential part of legal personhood turns out to be suffering. I review the key arguments in the article in which I learned that, then discuss why this means it makes no sense to talk about AI as a legal person.

1 · Reply

Graphcore   1 month ago - Shared publicly

#AI cannot be a legal person as suffering in AI is incoherent. Thought provoking article by @j2bryson
http://hubs.ly/H05ryhZ0

1 · Reply