A case for artificial intelligence (AI) rights

The subject of this essay is a hitherto hypothetical entity, although one that leading computer scientists predict will emerge this century, namely: human-level artificial general intelligence (hereafter AGI). This essay will explore the ethical standing of AGI to provide a precedent for the legal structures that will be necessitated upon its invention. I will begin by explaining why both Kantian and Utilitarian criteria require us to acknowledge the rights of AGI insofar as it is rational, autonomous and sensitive. I will then argue that the inequality of humans and AGI should be understood as a matter of property rights and that AI rights ought to be limited in Hohfeldian terms to claim-rights against arbitrary affliction.

The Kantian qualification

Humans have long justified the slaughter of non-human animals on the basis that they are inferior beings. While Abrahamic scripture claims that humanity's superiority was mandated by God, modern theories of human rights draw on the philosophy of Immanuel Kant, who predicated human supremacy on superior rationality (Kamm, 1992). Given that rationality is a measure of cognitive capacity, Kant's qualification for moral standing implies that an AGI with comparable cognition to a human would therefore be entitled to a comparable moral standing. Kant furthermore claimed that the ethical lodestar for humanity is a 'categorical imperative', according to which we are compelled to treat our fellow humans as ends in themselves and never means alone (Kamm, 1992).

Not being human, it seems that AGI would be sidelined by Kant's categorical imperative, as are non-human animals (Kamm, 1992). In Kant's reasoning for why the categorical imperative is exclusively centred on humans, however, we find further grounds for AI rights. Kant reasoned that our human peers deserve to be treated as ends in themselves because they are 'autonomous', whereas non-human animals merely exist at the whim of unconsidered instincts (Kamm, 1992). But what about AGI? Although non-human, would AGI not be autonomous? One would think this is a necessary characteristic of that which is said to be the harbinger of 'mass automation' (Yunis, 2016). And while it might be argued that AGI would act on programming and have no discretion in its actions, the same might be said of a human in the enactment of its own programming, albeit DNA code rather than computer code. This equivalence is best captured by Yuval Noah Harari, who observes that *organisms are algorithms* (Harari, 2015). An AGI with the equivalent cognitive capacity of a human therefore meets the Kantian qualification for moral standing by virtue of its comparable rationality and autonomy.

The Utilitarian qualification

A crucial counterpoint to the Kantian qualification of morality is that it entails lesser rights for infant humans and mentally disabled people due to their inferior rationality, which is an unacceptable outcome for many. This is a key objection of Utilitarians like Jeremy Bentham and Peter Singer, who propose that the basis of moral standing is instead a capacity for suffering (Singer, 1979). If it can be demonstrated that our hypothetical AGI is prone to a comparable experience of suffering, it is then the case that AGI should have some moral standing, constituted by rights that protect it from arbitrary affliction.

A comparison can be made in terms of the neural networks common to humans and AI. The human neural network has evolved to promote our survival by strategically rewarding us with pleasure when we attain optimal outcomes (Harber and Behrens, 2014). If we consider the most advanced iteration of AI in 2017, Google DeepMind's AlphaGo Zero, we can recognise a digital neural network with a similar system of incentives (DeepMind, 2017). Rather than identifying and emulating patterns of behaviour in masses of input data, as its predecessor AlphaGo did, Zero acts randomly and is rewarded when progress is made towards a certain end, learning over time to repeat those courses of action that maximise the optimal outcome (DeepMind, 2017). This process is known as reinforcement learning and has an uncanny resemblance to the evolutionary development of species (DeepMind, 2017). As Bentham hypothesised, pain and pleasure are part of a continuum, which means that Zero's sensitivity to a reward necessitates a parallel sensitivity to its inverse (Singer, 1979). If an entity could feel only pleasure, then that sensation would become the default and cease to be effective as a reward incentive. Given that AGI would therefore be sensitive by design, it must then be entitled to a moral standing in the Utilitarian view, lest it be subject to undue suffering.

The parameters of AI rights

Having established that our hypothetical AGI is entitled to some degree of moral consideration, it is necessary to clarify what the parameters of AI rights are and why they are not equal to those of humans. The AGI in question is distinct from humans only in terms of its material makeup and origin. Accordingly, we must differentiate the moral standing of humans and AGI on the basis of these factors rather than an arbitrary distinction. In the first case, there is little reason why rights ought to be contingent on material makeup, such that superiority would be conferred upon a carbon-based entity over a silicon-based entity.

At present, we do not afford moral weight to aesthetic distinctions like race and skin colour for precisely this reason, nor do the likes of Peter Singer give moral weight even to

endoskeleton distinctions (Singer, 1979). Origin, on the other hand, draws a useful analogy with property rights. Just as parents who conceive a daughter have a claim over that child by virtue of being her creators, so too should the creator of an AGI have a claim over their creation. That said, a child has certain inviolable rights and so too should an AGI. The analogy between AGI and children ends where children are capable of attaining fuller sovereignty upon coming of age, as endowing AGI with the same extent of sovereignty could pose a serious threat to human civilisation. I take it for granted that humanity is entitled to hedge its existential risk by limiting AI rights to a Hohfeldian claim-right against arbitrary affliction (Wenar, 2005). In this case, the subservience of an AGI to its creator is not an arbitrary affliction, like slavery, but rather a matter of property rights. AGI should thus take on a moral standing akin to that of a domestic pet under contemporary Canadian law, which allows an owner to treat their pet as property so long as they do not violate the animal's claims against being arbitrarily abused or killed (World Animal Protection, 2014).

In determining the moral standing of AGI, humans must be consistent in the application of the Kantian and Utilitarian ethics which constitute our own moral standing, lest we throw the baby (our moral standing) out with the bathwater (the case for AI rights). It is evident that both Kantian and Utilitarian ethics require moral consideration for AGI on the basis that it is rational, autonomous and sensitive. That said, any AGI will have a creator and must be beholden to the property rights of their creator. The hypothetical AGI in question should therefore be treated like property, but only insofar as its claim-rights against arbitrary affliction and termination are not violated.

Bibliography

Books:

Bostrom, N. (2014). Superintelligence. UK: Oxford University Press.

Singer, P. (1979). Practical Ethics. New York: Cambridge University Press.

Harari, Y.N. (2015). Sapiens. New York: Harper.

Essays:

Kamm, F. M. 'Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status', Philosophy & Public Affairs, Vol. 21, No. 4 (Autumn, 1992), pp. 354-389. Online: Wiley. Accessible at: http://www.jstor.org/stable/2265370. Accessed on 25-10-2017 at 16:23 UTC.

Hart, H. L. A. 'Are There Any Natural Rights?',The Philosophical Review, Vol. 64, No. 2 (Apr., 1955), pp. 175-191. Online: Duke University Press on behalf ofThe Philosophical Review. Accessible at: http://www.jstor.org/stable/2182586. Accessed on 26-10-2017 at 13:02 UTC.

Wenar, L. (2005). 'The Nature of Rights', Philosophy & Public Affairs, Vol. 33, No. 3, pp. 223-252. Online: Wiley. Accessible at:http://www.jstor.org/stable/3557929. Accessed on 28-10-2017 at 10:37 UTC.

Haber, S. and Behrens, T. (2014). The Neural Network Underlying Incentive-Based Learning. Online: National Institutes of Health. Accessible at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4255982/. Accessed on 29-10-2017 at 18:32 UTC.

Waldrop, M. (1987). A Question of Responsibility. Online: Association for the Advancement of Artificial Intelligence (AAAI). Accessible at:https://www.aaai.org/ojs/index.php/aimagazine/article/view/572/508. Accessed on 31-10-2017 at 14:22 UTC.

Web pages:

Yunis, A. (2016). Mass Automation: The Future of Employment. Accessible at: https://thelachatupdate.com/2016/05/30/mass-automation-the-future-of-employment/. Accessed on 26-10-2017 at 11:15 UTC.

DeepMind. (2017). AlphaGo Zero: Learning from Scratch. Accessible at: https://deepmind.com/blog/alphago-zero-learning-scratch/. Accessed on 27-10-2017 at 12:41 UTC.

World Animal Protection. (2014). Canada. Accessible at: http://api.worldanimalprotection.org/country/canada. Accessed on 28-10-2017 at 12:10 UTC.