

Review of: "More Human Than All Too Human: Challenges in Machine Ethics for Humanity Becoming a Spacefaring Civilization"

Krishna Deo Singh¹

¹ O.P. Jindal Global University

Potential competing interests: No potential competing interests to declare.

Disclaimer - I have reviewed the version the link to which I received earlier. If there are further changes in the paper, they may not reflect here. However, given the nature of feedback, I am hope full it would be useful.

Firstly, it must be said that the article is ambitious which is praiseworthy. It deals with an important set of issues, which must be addressed by researchers. However, some comments and suggestions follow:

"Stuart Russell and Peter Norvig (2009) define AI as machines that imitate the mental functions of humans (problem-solving and learning), and which can "[think] humanly" – please consider other definitions – the rational agent approach (in Russell and Norvig's standard textbook on AI), and also 'evaluative decision maker' approach given by Jacob Turner in his book 'Robot Rules'.

"Machines with AI have a normative function, but some argue that it can also be evaluated from the perspective of ethical norms" this statement requires clarification, as ethical norms are also normative in nature.

On machine ethics, please read the works of Wendell Wallach (along with several cowriters). In fact, Gunkel is far less relevant than Wallach as machine rights are a very marginal aspect of machine ethics.

"Strong AI is associated with artificial general intelligence, human-level intelligence, superintelligence and artificial consciousness (Braidotti, 2013). Ultra-intelligence is another term used for "a machine that can far surpass all the intellectual activities of any man, however clever"" – may be a good idea to keep consciousness distinct from super or ultra intelligence – they are not the same thing.

It's not clear what the discussion around SCOT and determinism adds to the paper.

"Whether machines can be considered fully ethical agents is a difficult question to answer, and until we develop the technology that allows us to create machines that have the capacity to become full ethical agents a definitive answer may continue to elude us" – even then, the problem of 'other minds' will hinder determination of AI as fully ethical agents.

One of the bigger weaknesses of the paper that needs to be addressed in almost complete lack of the discussion on the current forms of AI that actually exist. These AI, based on neural networks, exhibit surprising characteristics, which could very easily be a path to the kind of AI needed to make humans a spacefaring civilization, without the need for a 'mind'

(with ethics playing an important role of course). The analysis is focused on the theory of mind and cites Dreyfus - “Hubert Dreyfus (1979) highlights in his postphenomenological critique of AI that one of the main things that distinguish us from AI is our a conceptual and contextual experience.” But these lines are all relevant only to a world in which sub-symbolic neural networks are not functioning like they are today, which was very much the world of 1979. Even Dreyfus admits that his criticism is not applicable to neural networks.