

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

Zachary Goldberg¹

¹ Trilateral Research & Consulting

Potential competing interests: No potential competing interests to declare.

General review:

The author writes on a very interesting topic, namely: “adequately designed AI will determine the success or failure of humanity attempting to become a spacefaring civilization – and the adequacy of AI design for the purpose of long-distance space travel will greatly be influenced by our choice of conceptual and ethical foundations”. I encourage the author to continue to refine their approach and that these comments are helpful in that endeavor.

The author shows awareness of different philosophies and philosophers, yet the author does not delve carefully into these theories or the objections to them in the wider philosophical literature. For a blog or popular publication this approach may be acceptable, but not for academic philosophy. It depends on what the author’s aims for publication are.

This article summarizes many of the questions that belong to discussions surrounding the moral agency of machines. However, it is unclear why the author states that this article applies specifically to the space sector.

Specifics:

The sentence “Machines with AI have a normative function” is not explained. What does the author mean here by “normative” and which function or functions are referred to here?

The sentence “The notion of machine ethics is a controversial topic” needs clarification. Machine ethics understood as the ethics of developing and using machines, especially AI, is not controversial. Machine ethics understood as attributing moral agency to machines, discussing the nature of this agency, and how this agency does or does not elicit moral standing and moral responsibility, is under debate.

The references to Moor (2006) and Gunkel (2013) are fine, but there has been quite a large amount of philosophical discussion around this topic in recent years.

This sentence needs rewording: “the SCOT model is an example where theoretical assumptions inform models of mind, in turn informing AI, and if misguided inevitably leading to failure”

“technological determinism is based on the premise that cultural values and social structures are determined by

technology.” Is this the premise or the conclusion of this view?

“One ontic aspect – technology and economics – is prioritized.” The author has not yet introduced this vocabulary, and so it seems to come out of nowhere. What function does its inclusion serve towards achieving the aims of the article?

“A typical human that is not under the influence of psychoactive substances is normally considered a fully ethical agent.” There are other cases that might affect a human’s moral agency status or ascriptions of moral responsibility. There is a great deal of literature on this topic. To start, I suggest, *Responsibility from the Margins*, D. Shoemaker (Oxford).

“I agree with Moor and a graded ethical agent continuum for machines may be a better scale to use than a binary yes or no approach.” Again, the literature on the topic of agency is very rich. It is a bit of a strawman to claim that a binary approach is the main one adopted by professional philosophers.

“There is no conceptual clarity regarding who can be considered an ethical agent “. This statement is much too strong and, hence, incorrect. We have a great deal of conceptual clarity, although there are competing conceptualizations and interpretations on this topic.

“Scott (1990, p. 7) states that “[t]here appears to be more unanimity as regards the claim that in order for an individual to be a moral agent s/he must possess the relevant features of a person; or, in other words, that being a person is a necessary, if not sufficient, condition for being a moral agent.” Again, there is vast amount of philosophical literature on this topic that is not addressed. The claim made here is too simplistic.

The discussion on conceptual clarity is interesting but doesn’t seem related to the topic at hand.