

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

Alcibiades Malapi-Nelson¹

¹ Humber College

Potential competing interests: No potential competing interests to declare.

The paper aims at addressing foundational issues, and such an ambitious aim carries pros and cons. The positive side is the tackling of relevant and perhaps even pressing issues that need addressing. The negative side is that one needs to deal with the origin of certain notions. In this vein, several sentences seem to be underpinned by non-articulated assumptions. For instance: "It is indubitable that machines with artificial intelligence (AI) will be an essential component in humans' quest to become a spacefaring civilization". Does the author mean that enhanced humans (i.e., machines with artificial intelligence) will be an essential part of long-term space travel?

Or the author claims that it is 'beyond doubt' that non-human machines will develop human-like intelligence *and* that humans will not space travel without them? Or that space-travelling will be impossible without humans getting enhanced first?

AI is tangentially addressed by a 'review' of well-known authors' ideas on the topic. However, the near impossibility of reaching a consensus regarding the possibility of GOFAI is not the main issue for the ambiguity of the quote above. One can choose one take on AI and work with it – even if that choice is to be superseded in the future. There is something deeper at stake: The notion of a machine is not addressed, and without a rigorous treatment of such notion, it is unclear how one can go on to address the question of 'machine ethics'.

I largely agree with the review of Neil Rowe. However, I would recommend a more practical course of action:

Cut out the first 2/3 of the piece, and focus on the last third one, aiming at delivering on the promised ethical insights regarding a future space-faring civilization – about which the body of the text barely acknowledges. There is plenty to talk about regarding that last part, and the other 2/3 can constitute another paper dealing with certain aspects of AI. Indeed, dealing with AI in itself can be so conceptually problematic that may merit writing a book on its own – at least (Again, speaking of foundational issues).

Reduce the abstract to 1/3 of it, avoid repetitive sentences, and hone down the argument. We can have a survey of the literature on certain aspects of AI by ourselves. What we are interested in (at least I am) is to know your claim about the ethics of space exploration aided by AI - whatever 'AI' means (choose one view of it and work with it for the sake of your argument – not because it is the correct view of AI).

Instead of the 2/3 removed, include a part where you devote some time to the articulation of the notion of a machine. Not much has been written about the nature of a machine, and that is likely the source of much confusion in both science and philosophy. To repeat, we would be adding more mystification when we talk about machine ethics when we do not know what a machine is in the first place. Are we referring to enhanced human ethics? Non-human machines' moralities? 'Intelligent' (but non-human intelligent) machines (qua 'things') as moral agents (a la Latour)? Or just human ethics, since we are machines, and to some extent, artificial, anyway (As several classical cyberneticians, particularly Walter Pitts, would claim)?

I would suggest seeing the work of W. Ross Ashby and John von Neumann on the nature of a machine. Regarding space exploration, including the possibility of enhancing humans for it (especially relevant if humans are to be artificial intelligence agents), the work of J.B.S. Haldane and J.D. Bernal will be found relevant.