Review of: "Do Androids Dread an Electric Sting?"

Igor Aleksander[1]

1 Imperial College London

Here is brave attempt to initiate a protection/fairness attitude towards machines under the assertion "Conscious sentient AI seems to be all but a certainty in our future." This is followed by these machines being "entities with the potential to experience more pain and suffering than any other living entity on Earth". There is a gaping problem with such assertions: not because it may never happen, but because it is not clear how 'conscious sentient AI', actually differs between biological organism and designed artefact. Were one to ask the entity, a machine with a generative language facility, "what are you conscious of ?" among its answers should be that it is conscious of being *a designed machine*. I personally, would reply that I am conscious of being a biologically created human. My consciousness contains my experience of being human from birth to date and a machine has the experience of being a machine and a projection of this into the future. The two are different with the machine being conscious of its technical character, and the human of its biological character. Language generative AI would support a discussion between the two. This is where the protection of the one by the other may arise in the conversation, but it is the difference between the two that is central to this issue and not the similarity of machine to an animal, human or otherwise.

The protection of such machines then boils down to methods of protecting both the design and implementation through conventional anti-destruction of property rules. But how about asking the machine "Can you be hurt by hearing adverse attitudes about you by humans?" A sensibly conscious machine would answer: 'not normally, but yes if you want me to be'. The protection needs of being a machine are fundamentally different and have a different source (design and technical adaptation) from those of a human/animal (biological source and development). A machine would not feel pain unless it had been designed for this (for very good reasons, maybe). But, the paper under consideration suggests that the feeling of pain will arise through some necessary emergence in a learning system. That this is possible or exists as a natural property of learning systems, both artificial and biological, remains to be argued.