How to/Not to Engineer Morally Correct Machines

Selmer Bringsjord

Rensselaer Polytechnic Institute

Ethics as a systematic field has been in operation for at least two millennia; it centers around such questions as: What ought we to do? What are we prohibited from doing? What is heroic? Why ought we to do what we ought to do? And so on. Notice the occurrences of ‘we’ in these questions. As humanity approached and moved into the third millennium, something radically new arrived on the scene: namely, machine ethics. Here, the idea isn’t our moral status and condition; rather, the focus is on the design and engineering of machines (including robots, autonomous weapon systems, etc.) that themselves do what ought to be done, don’t do what is forbidden, and — when relevant circumstances arise — behave heroically. I explain in this talk (1) why engineering morally correct machines is something we must do, (2) that such engineering, if based on the much-hyped “machine learning” of today will fail and likely get most of us killed, but (3) that there is a way to engineer morally correct machines, and thereby prevent catastrophe. This way is explained.

Selmer Bringsjord specializes in the logico-mathematical and philosophical foundations of artificial intelligence (AI) and cognitive science (CogSci), and in collaboratively building AI systems on the basis of computational logic. Though he spends considerable engineering time in pursuit of ever-smarter computing machines, he claims that “armchair” reasoning time has enabled him to deduce that the human mind will forever be superior to such machines.

The Center for Ethics is funded in part by the ENDOWMENT FUND for the TEACHING of ETHICAL DECISION-MAKING

Thursday, November 1, 2018 | 4:10 pm | Neville Hall 003