

CAN MORALITY BE COMPUTED? AN EXPLORATION OF WHETHER MACHINES CAN BE MORAL AGENTS

Gillian Eileen Cross

(Person Number: 8805370J)

A Research Report submitted to the Faculty of Humanities, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Arts, Applied Ethics for Professionals

31 May, 2015, Johannesburg

ABSTRACT

Technology is an integral part of our daily lives and continues to advance rapidly, impacting our physical and social worlds. We increasingly rely on advanced machines to act in more autonomous and sophisticated ways. What would happen if artificial forms of intelligence developed to the point where machines behaved more like us and we started treating them as people? Ethics should anticipate and account for such possibilities so that science does not move faster than our moral understanding.

My thesis states that when we are able to feel gratitude or resentment towards the actions of artificially intelligent machines we can be said to see them as morally responsible agents. I argue that standard ethics frames morality developmentally – only when we reach adulthood are we deemed able to enter into the type of relationships where we can hold one another morally responsible for our actions. I apply a more abstract notion of moral development to future versions of technology and couple this with a definition of morality as a relational or social construct. This allows me to argue that machines could develop to a point in the future where we react to them morally as we would to humans. Questions on whether we ought to react in this way are muted as relationally we quite simply would be unable to feel otherwise. Objections from definitions of moral agency based on innate qualities, specifically those associated with the concept of intelligence, are dispelled in favour of a relational definition.