

# Abstract

Author: Maria Romanova Hynes

Degree: Master of Science in Interactive Digital Media

Title: Enlightened Intelligence: Towards an Ethical Framework for Strong AI

Supervisor: Dr. Mads Haahr

Year: 2016

Considering Artificial Intelligence one of humanity's most ambitious scientific endeavours of our time, this paper argues that the fundamental problems of ethics should be at the heart of AI research. After answering the question of whether contemporary scientists believe that it is possible in principle to create an artificial mind and conducting a historic overview of the past and present advancements in the field, a new definition of intelligence that has an ethical dimension to it and an ethical framework within which further developments on AI might be undertaken will be proposed. This paper argues that human intelligence is inappropriate as the model of intelligence for AI as humans are notorious for destroying their social and natural worlds by engaging in wars and environmentally devastating activities. Instead this paper presents a model of intelligence, called Enlightened Intelligence, inspired by the Buddhist concept of *enlightenment*, that considers technology a part of the common environment and implies that the only form of truly rational and intelligent behaviour for thinking machines would be to overcome their immediate interests and to reach beyond themselves and out to others out of the desire to protect the environment they are themselves a part of. It will be finally suggested that an inter-disciplinary, cross-religious, cross-cultural study is needed to develop a complete operational ethical model that would meet the needs of AI.

**Key-words:** artificial intelligence, machine intelligence, AI, ethics, Buddhism, enlightenment, intelligence, rational behaviour, strong AI, history of AI.