The Future of Mankind

A Question of Limits

Meghan McCracken, 20087270
Jeff McGee, 20062382
Marianne Pike, 20087273
David Wheatley, 20077776

Introduction…

As the development of an autonomous, self-sufficient, intelligent machine gets closer the question arises as to what differentiates humans from machines. Mankind can be characterized by certain personality traits, emotions and experiences formed through time.[1] As humanity becomes more "mechanized" and, conversely, machines become more "humanized", the distinction between humans and machines will begin to blur. Clearly, artificial intelligence is becoming a more ethically troublesome subject.

Artificial Intelligence (AI), the field of research that attempts to emulate human intelligence in a machine,[2] raises ethical concerns that need to be addressed. The issue of whether or not to create artificial intelligence is no longer relevant. AI has been in development for decades and is advancing continuously and exponentially according to Moore's Law. However, several issues remain regarding the ongoing development of AI and its future use. Should intelligent, human-like robots be built? Should these robots be granted the rights and freedoms granted to other intelligent life? And to what extent do we continue to make advances in artificially intelligent robots? These are all questions that remain unanswered to this point.

The following report will show that the creation and advancement of artificially intelligent robots is an admirable venture and should be explored provided that the benefit and use of such creations is ethically justified, the robots are not given complete autonomous control, and strict regulatory and productive limits are placed upon the developers of such technology. Within these restrictions, AI development can be beneficial to society. However, without these guidelines in place, the reckless expansion of AI could potentially lead to harmful consequences.

The Development Ethical Issues in Artificial Intelligence…

The development of ethical guidelines surrounding the creation of AI has been an ongoing area of study since the mid-20th century. Isaac Asimov, author of "I, Robot" (1950), is credited with the development of the Three Laws of Robotics:

1) A robot may not injure a human being or, through inaction, allow a human being to come to harm
2) A robot must obey the orders given it by human beings except where such orders would conflict with the first rule

3) A robot must protect its own existence as long as such protection does not conflict with the first or second rule[3]

Asimov's work shows that the possibility of robot independence has been an ethical concern ever since the thought of robots was first conceived. Half a century later, with the possibility of androids becoming mainstream in the foreseeable future, effort is being made to amend these ethical guidelines to render them more applicable to today's level of technology.[4]

There are countless other authors who have tackled the subject of AI through documenting its progress or creating a science-fiction thriller. Perhaps one of the most influential volumes of the past few decades is Ray Kurzweil's "The Age of Spiritual Machines" (1999). His theories hold alarming projections regarding the future of mankind and machines if AI development continues at its current rate in an unregulated environment. The sci-fi thrillers such as "Prey" (Crichton, 2002), "1984" (Orwell, 1950), and Asimov's "I, Robot" itself have become classic novels and have brought the issues of AI development into mainstream society. The public perception surrounding these issues is generally negative, as the thought of what could go wrong in the development of these artificial "humans" terrifies society.[5] Evidently, the redefinition and elaboration of Asimov's Laws is needed to put society at greater ease.

Artificial Intelligence's Place in Society…

As society stands today, AI robots have many beneficial functions. Their many uses include entertainment, toys, military applications, and management software (expert systems). The development of mainstream androids could benefit society by working in dangerous situations such as mining, providing companionship, or physically supporting the elderly or disabled.

Due to these benefits, some would argue that if robots can improve human lives and the ability to construct them exists, AI robots should naturally be developed. However, with continuous improvement in AI technology, androids will become more independent of human supervision and control. Without complete continuous control, robots will become less predictable and increasingly dangerous. To avoid such uncertainty, strict limitations on their level of autonomy must be set in order to ensure that humans remain in control of their creation.

The most familiar robot product on the market today is AIBO, Sony's version of the robotic pet dog. Available for retail sale, this machine seems relatively harmless. It can walk,

make sounds, play with its owners and others, and provide those unable to own a pet with a close alternative.[6] But is it ethical for such a companion to be sold? This sort of feigned companionship causes a blur in the distinction between humans and machines and the moral significance of this distinction.

> "If robot pets are designed and manufactured with the intention that they should serve as companions for people, and so that those who interact with them are likely to develop an emotional attachment to them, on the basis of false beliefs about them, this is unethical… whether or not this is the intention of the designer. For an individual to benefit significantly from ownership of a robot pet they must systematically delude themselves regarding the real nature of their relation with the animal. The design and manufacture of these robots is unethical insofar as it presupposes or encourages this delusion. We have a (weak) duty to ourselves to avoid delusion and apprehend the world correctly… but when such a delusion leads us to devote time and energy to a relationship that is in fact worthless, we have a duty to avoid it."[7]

Ethical questions arise as to the intention of the producers of these electronic pets. The developers hope that by putting these pet robots into homes, society will become accustomed to the presence of robots in everyday life.[8] This, in turn, may ease their fears of AI going "too far" and threatening the very existence of the human species, as is often the topic of popular fiction and movies, such as "I, Robot" (2004) and "The Matrix" series (1999, 2003). However, should a seemingly unethical product be sold, merely to falsely reassure society that more ambitious AI development is perfectly safe?

The Future of Mankind…

Currently, in the development of AI there is a strong trend towards modelling human emotions and personalities. If developers are trying to create a robot that mimics the human form, then these robots must be programmed to deceive. There are certain situations – such as negotiation or confidentiality – in which humans are ethically taught and expected to lie.[9] Thus, should these situations be extended into the robot's "psyche," the robot would essentially be

programmed to lie. If the machine discovered through experience that deception proved beneficial, its intelligence would then apply that experience to other situations, making the robot ultimately uncontrolled and potentially dangerous.

> "We could make robots perfectly safe only if we had absolute and perfect self-knowledge, that is, an exact knowledge of all our purposes, needs, desires, etc., not only in the present but in all future contingencies which might possibly arise in all conceivable man/robot interaction. Since our having this much knowledge is not even a theoretical possibility, obviously we cannot make robots safe to us along this line." [10]

Any programmable robot will implicitly inherit the ethical values of its programmer similar to the way a parent passes their value systems onto their children. Humans are not perfect moral and ethical beings and it is natural to assume that a programmer would – albeit accidental – transfer these ethical flaws into their creation. The development of ethics is formalized by a set of rules, in which some will take precedent over others. There are bound to be situations in which these rules will conflict, and without the possession of human intelligence and emotions, serious ethical concerns arise when using these machines to make decisions or analyze a situation independently.

"Experts are seldom able to retrace the analytic steps followed to make a particular decision. It will be impossible to diagnose a not-observable, not-controllable, real-time 'computer' whose detailed blueprint is not known, and whose components have complex dynamics not fully known." [11]

The whole debate of ethics and morals in AI development begs the question of what this development will mean for humanity. There are many speculations as to "worst case scenarios" (often depicted in books and movies) that involve robots evolving into the dominant species and humans eventually becoming enslaved. Although seemingly farfetched, it is interesting to note that the word *robot* comes from a Czech term that means "a debt of forced labour" or "slave." Alternatively, if the machines do become independent of the human race, they may simply choose to ignore us, similar to how humans currently treat animals. As a third possibility, roboticist Hans Moravec proposes that some day "humanity may be able to survive, and even

achieve a level of immortality, by digitally uploading our own consciousness into advanced robots."[12] For this reason, many label AI development as the quest for the "holy grail" of robotics.

In general, the predictions of the development of robots tend to depict a bleak future for mankind. Bill Joy (Sun Microsystems) published an essay entitled "Why the Future Doesn't Need Us", which depicts the ultimate extinction of the human race once machines capable of thinking autonomously realize there is no longer a need for mankind. Kurzwiel's "The Age of Spiritual Machines" suggests that computers will be conscious beings capable of feeling that they deserve the same rights and privileges that humans give each other. Through creating autonomous, intelligent machines, the development of the extinction of mankind could already be underway.

> "Assuming that some future robots are sophisticated to the point where…they know more about their own needs and capabilities than we do, it would be logical to give them decision-making power sufficient to safeguard their functioning. And at the point, however unlikely to be reached, that machines would know more about men than men themselves, it would be logical to transfer the control of men to them – for the sake of human fulfillment."[13]

Conclusion…

The ultimate question becomes: Is humanity willing to accept the risks associated with the development of artificial intelligence? To the extent that robots will enhance the lives of humans, they will necessarily be built. However, specific guidelines regarding restrictions on the development of these machines must be developed and *fiercely* enforced to remain in control of our robotic creations. It is impossible to "un-invent" something once it is created, especially in the case of autonomous, self-thinking robots that can learn and adapt in order to survive. A delicate balance exists between the degree of autonomy and the effective usefulness of the robot. A robot must be allowed to think freely for it to be useful, but controls must be in place for humans to remain the dominant species. The monitoring and regulation of the development of artificial intelligence should allow for the safe and controlled expansion of technology in the future. However, without the careful supervision of advancements in artificial intelligence, much like Pandora's box, an irreversible chain of events could begin that will lead to the demise of the human race.

Endnotes…

[1] Khaili, Omar E M. (1993). Artificial decision-making and artificial ethics: A management concern. *Journal of Business Ethics*, 12(4), 313.

[2] Brian P. Bloomfield, Theo Vurdubakis. (2003). Imitation games: Turing, Menard, Van Meegeren. *Ethics And Information Technology*, 5(1), 27-38.

[3] Khaili, Omar E M. (1993). Artificial decision-making and artificial ethics: A management concern. *Journal of Business Ethics*, 12(4), 313.

[4] Ethical artificial intelligence patented: Robots to follow 10 mandates that focus on virtues, excluding vices. (20 August 2004). *Computerworld Canada,* n/a.

[5] Thomas Hayden, Peter Hadfield. (2001, April). The Age of Robot We're close to making humanlike machines. It's time to reckon with the promises and perils; Tokyo. *U.S. News & World Report,* 130(16), 44-50.

[6] *Sony Global - AIBO Global Site.* www.sony.net/Products/aibo/index.html

[7] Robert Sparrow. (2002). The March of the robot dogs. *Ethics And Information Technology*, 4(4), 305.

[8] Robert Sparrow. (2002). The March of the robot dogs. *Ethics And Information Technology*, 4(4), 305.

[9] Cristiano Castelfranchi. (2000). Artificial liars: Why computers will (necessarily) deceive us and each other. *Ethics And Information Technology*, 2(2), 113-119.

[10] Versenyi, Laszlo. (April 1974). Can Robots be Moral? *Ethics*, 84(3), 248-259.

[11] Khaili, Omar E M. (1993). Artificial decision-making and artificial ethics: A management concern. *Journal of Business Ethics*, 12(4), 313.

[12] Thomas Hayden, Peter Hadfield. (2001, April). The Age of Robot We're close to making humanlike machines. It's time to reckon with the promises and perils; Tokyo. *U.S. News & World Report,* 130(16), 44-50.

[13] Versenyi, Laszlo. (April 1974). Can Robots be Moral? *Ethics*, 84(3), 248-259.

Bibliography…

Asimov, Isaac. (1950). *I, Robot*.

Brian P. Bloomfield, Theo Vurdubakis. (2003). Imitation games: Turing, Menard, Van Meegeren. *Ethics And Information Technology*, 5(1), 27-38. Retrieved from ABI/INFORM Global database. (Document ID: 417620641).

Crichton, Michael. (2002). *Prey.*

Cristiano Castelfranchi. (2000). Artificial liars: Why computers will (necessarily) deceive us and each other. *Ethics And Information Technology*, 2(2), 113-119. Retrieved from ABI/INFORM Global database. (Document ID: 403872541).

Ethical artificial intelligence patented: Robots to follow 10 mandates that focus on virtues, excluding vices. (20 August 2004). *Computerworld Canada,* n/a. Retrieved from CBCA Business database. (Document ID: 693073931).

Khaili, Omar E M. (1993). Artificial decision-making and artificial ethics: A management concern. *Journal of Business Ethics*, 12(4), 313. Retrieved from ABI/INFORM Global database. (Document ID: 571869).

Kurzweil, Ray. (1992). *The Age of Intelligent Machines.*

Kurzweil, Ray. (1999). *The Age of Spiritual Machines*.

Kurzweilai.net. www.kurzweilai.net/index.html?flash=2

Orwell, George. (1950). *Nineteen Eighty-four.*

Robert Sparrow. (2002). The March of the robot dogs. *Ethics And Information Technology*, 4(4), 305. Retrieved from ABI/INFORM Global database. (Document ID: 330932201).

Sony Global - AIBO Global Site. www.sony.net/Products/aibo/index.html

Thomas Hayden, Peter Hadfield. (2001, April). The Age of Robot We're close to making humanlike machines. It's time to reckon with the promises and perils; Tokyo. *U.S. News & World Report,* 130(16), 44-50. Retrieved from CBCA Business database. (Document ID: 71445103).

Versenyi, Laszlo. (April 1974). Can Robots be Moral? *Ethics*, 84(3), 248-259.