**Artificial Intelligence**

The ethical concerns about creating artificial intelligence are hot topics in these times of rapidly advancing technologies. As high speed, compact computing becomes exponentially more powerful (at least according to Moor’s law) the ability for people to create machines that model high level intelligence is becoming a reality. With this possibility comes the responsibility of evaluating the issues that may arise from these kinds of machines. Many popular science fiction novels and films explore consequences of this precise topic, ranging from the apocalyptic world of *The Matrix* to the comedic answer to the Ultimate Question of Life, the Universe, and Everything by Deep Thought in The Hitchhikers Guide to the Galaxy. As extreme or benign as these scenarios may seem, they stem from the possible realities that creating artificial intelligence can cause if the possible consequences are left unevaluated.

The first of many concerns deals with the use of Artificial Intelligence (AI). AI can be created to solve specific problems, such as winning a game of chess or solving a wider, more general range of problems. Depending on the application, AI can be used for either good or ill intensions. As a potentially powerful tool, the cost-benefit of this is a major concern. Another concern, which directly correlates with the consequential severity of AI, is about how powerful can and should an AI be. It is plausible to have AIs that are more intelligent than all humans. Allowing this to happen could be incredibly beneficial to humanity, but could also be very dangerous. A result of high-intelligence is sentience, an AI that could demand respect. This poses the question: what rights should AIs receive? The question may sound bizarre, but is a real issue when it comes to working with an AI who can think for itself. The emergence of new intellect requires a diligent exploration of the possible consequences of that technology.

AI can be used to solve a multitude of problems. As an intelligence, AI is meant to be able to solve specific problems that humans have not yet thought of. In other words, AI should be able to deal with new problems as they arise. The purpose of an AI is to do work for humans, as would be the case for any technology. The kind of work that AIs can uniquely do for humans is that which may require novelty. The definition of intelligence (by Merrian-Webster) is “the ability to acquire and apply knowledge and skills”. The two key components of this is acquiring and applying, which implies evaluation of a problem itself. This separates AI from a classical program, which can only apply knowledge that it is already given in a certain manner which has already been thought of by the programmer. An AI may then come up with an answer to a problem in a way that an onlooker may not be able to understand. An AI can give a “what” answer to a problem without being clear about the “how”. Therefore, an AI with high enough intelligence can be assigned to solve many different kinds of problems and come up with solutions that are not desirable for people, or solve problems specifically designed to be undesirable to humans. A crude example of this could be about finding the most efficient way to carry out a genocide. That may be extreme, but it is possible if used in the wrong hands. History has show that humans are not the most considerate people, so you can imagine what an AI without the human instinct of empathy can come up with. There are arguments about how to solve this precise issue, as those kinds of scenarios would imply that AIs are inherently dangerous (meaning they have the potential for destruction). One famous solution is that by Isaac Asimov, who formulated the three laws of robotics[[1]](#footnote-1). They are set in such a way that prioritizes the command and wellbeing of humans before that of an intelligent robot. These laws would need to be hardcoded into the AI, acting almost as instincts for the AI. There are also some scientists who believe that the creation of a high level AI would allow the AI to surpass these commands. If that is the case, some solutions involve full isolation of the AI or having kill switches implemented as failsafe’s. But this can also be dangerous. If an AI is powerful enough, it could convince someone to grant it further freedom by offering its superior intelligence as a service, effectively bribing someone to let it free. The main idea here is that powerful AIs can possibly have their own agendas, acting freely as they see fit. Even when thinking about AIs as bounded by machines, it is not hard to imagine such a program to spin out of control. These kinds of behaviors are thought to be a strong function of the sophistication of the AI. Therefore, the most direct method of regulating an AI is regulating its intelligence.

Designing an AI to be specifically intelligent, as opposed to generally intelligent, is a way to regulate an AI’s abilities. Many of the tasks humans would like AIs to complete are rather complex. The hope of creating an AI is that and AI can complete tasks that would generally take some human consideration but not necessarily humane consideration. A great example of this is using an AI to evaluate whether or not to invest in a certain market or product. Imagine a representative from a startup makes a pitch about why it would be smart to invest in their company. Typically, many statistics and objective data will be presented to make a case for how much an investment would be worth. Ultimately, the decision to invest is made by a human who either is or is not interested in the company. Why and how the investor decides to invest in a company is due to many reasons, some from weighing objective pros and cons, others from personal interest and instinct. It goes without saying that the investor is expecting their decision to make them money; no investor would like to waste their money on a bad investment. In reality, bad investments happen all the time, so there must be a reason as to why an investor would make an incorrect evaluation. This could come from miscalculation or just trusting too much in their gut – a “human” mistake. Either way, the human evaluation was flawed as the outcome was not as expected. An AI would hopefully be able to solve this sort of problem. To do this, though, the AI must have some sort of superior intelligence than the human in order to more accurately make investment decisions. The big question is what exactly is that sort of intelligence. Can an AI be built that can solve that problem and also not pose a threat to humans? Would an AI of that level of intelligence be inherently able to tackle unrelated problems? If it can, could an AI begin to learn to think in new ways? These questions are all asking about the ability of an AI to behave outside of its intended realm. The concern here is AI’s ability to learn and adapt, making it be able to increase its own intelligence over time. An AI must be useful, or else there is no reason to make one, so can a useful AI be made to solve complex human problems without creating one that has the ability to spiral out of control? Creating an AI to solve useful complex problems may require a level of general intellect that cannot be avoided.

Enough intellect in an AI could lead to many complications regarding the relationship between humans and AIs. If a superiorly intellectual AI is created in order to better solve problems than a human, it stands to reason that the AI is just as sentient as a human. An AI could have opinions, ones that differ with that of a human, and expect those opinions to be respected. An AI could be sentient enough to know that it has a superior intelligence to that of a human, expecting to be given rights just like a human. In an objective evaluation, an AI would consider itself more useful than a human. An AI could even think that humans are the main problem in furthering the progression of universe itself. This premise has been explored in movies such as *I, Robot* and *Eagle Eye*, where AIs believe the solution to solving humanity’s problems is to eliminate humanity. If humans are to live with machines of superior intelligence, steps must be taken to incorporate those machines into society in a way that leaves them satisfied. Unfortunately, it may believe human society is beneath an AI such as itself, causing it to want to take control and be free to exist as it sees fit with complete disregard to the human race. The modern world is inseparably intertwined with the technology that would house this kind of AI becoming instantly controlled by its will. A superior AI could very quickly become hostile if it so choses.

The level of intelligence that a useful AI may need to have can pose an inherent threat to humanity. Creating AI’s that are used for complex tasks may not be the correct way to approach those tasks. High-intelligence AIs have the possibility of becoming serious threats to humanity. I, for one, welcome our robot overlords.

Sources:

<http://www.auburn.edu/~vestmon/robotics.html>

<http://ieet.org/index.php/IEET/more/nanderson20150617>

<http://www.nickbostrom.com/ethics/artificial-intelligence.pdf>

<http://aitopics.org/topic/ethics-social-issues>

<http://www.aaai.org/ojs/index.php/aimagazine/article/view/540/476>

<http://www.wired.com/2015/01/ai-arrived-really-worries-worlds-brightest-minds/> (Thought it was in interesting article)

1. 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 orders given it by human being except where such orders would conflict with the First Law; 3) A robot must protect its own experience as long as such protection does not conflict with the Frist or Second Law [↑](#footnote-ref-1)