

## **The Dangers of Artificial Intelligence (A.I.)**

Artificial Intelligence (AI) is becoming more and more predominant in today's world. The Terminator (James Cameron 1984) looks at one of the worst extremes of AI gone wrong, a robot apocalypse. Now is any of this even possible or just a creation of Hollywood in order to make a profit? My objective is to examine the various aspects of AI development and point out the dangers associated with them. First we must look at the definition of intelligence and more importantly differentiate it from morals and ethics. AI is all around us today, we see it in video games controlling the enemies and even in word processors which try to guess what word you may be trying to use. These examples are just the tip of the AI iceberg and is not what causes me to be concerned. Most AI systems as we see them today are very basic, only adapting a small portion of intelligence to do a simple task such as guessing a word or locating a target. AI, however, grows smarter, more intelligent every year. The DARPA Grand Challenge is a race which pits autonomous vehicles in a long race without human assistance. The first competition occurred in 2004 and no vehicles finished. However, just one year later five vehicles finished: Moore's Law at work.

So far the only examples given have been with “Applied AI, also known as advanced information processing, aims to produce commercially viable “smart” systems—for example, “expert” medical diagnosis systems and stock-trading systems.” (artificial intelligence in Britannica Online Encyclopedia 2008). There exists another branch of AI called strong AI which has a deceptively simple task, allow machines to think like humans. Development in strong AI has so far had very little success, while “Some critics doubt whether research will produce even a system with the overall intellectual ability of an ant in the foreseeable future” (artificial intelligence in Britannica Online Encyclopedia 2008). However, if success is achieved with

strong AI, we are potentially in for a whole mess of problems involving ethics, morals, human rights (more specifically do machines get them?) and sentience (will machines be self-aware?). We must be fully aware of these and all other problems before we move forward with the research and development of more complete AI systems.

Before we can look into Artificial Intelligence it would be best to have a definition of human intelligence.

Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought. Although these individual differences can be substantial, they are never entirely consistent: a given person's intellectual performance will vary on different occasions, in different domains, as judged by different criteria. Concepts of "intelligence" are attempts to clarify and organize this complex set of phenomena. Although considerable clarity has been achieved in some areas, no such conceptualization has yet answered all the important questions and none commands universal assent. Indeed, when two dozen prominent theorists were recently asked to define intelligence, they gave two dozen somewhat different definitions. (Neisser U, et al. 1998 p.96)

Now we already have a problem, the definition of intelligence is different depending on who you ask. However intelligence can be considered at some level a category of related abilities including reasoning, planning and problem solving. Also included is the ability to use language and the capability to learn. An artificial intelligence should be able to do the same. The problem arises when we compare how the human mind works with how a computer works. Dreyfus and Dreyfus (1986) give an example of this.

Unlike computers, man's memory . . . is instantly aware of what it does-and does not- contain. No list is needed. When were you born? You know the answer immediately. When was your mother born? You may not have a ready answer, but you know that you know the date and will remember it if you think long enough. . . . When was Thomas Jefferson born? If you don't know, you know that you don't know and that no amount of thinking will bring the date to mind.(P.69)

Humans think differently that computers do. Computers use data structures such as lists to store

and retrieve information and before a computer can decide it does not know something, it first must go through everything it does know and come to the conclusion that in fact it does not know that information. This has been one limitation in the advancement of AI. Both humans and computers can store a large amount of information and with the Internet computers can store a near infinite amount of information. So why can't an AI system just know everything? The problem is not just the time of looking through a massive list. The bigger problem is relevance. How is a computer to decide what information is relevant and what is not. The obvious answer was to make rules of what information is relevant. The problem which arises from that is how does the computer know which rules are relevant? More rules? This just leads us in an endless loop of rules about rules. (Dreyfus 1986) To handle all of this information seemed impossible so another branch of AI research was created. One which used micro-worlds to simplify the entire process.

These micro-worlds are an artificial, make-believe world where everything has been simplified to the point where many rules would be false when applied to the real world. This limits the amount of knowledge required down to manageable levels. The AI systems which function in these micro-worlds are what I previously referred to as applied AI. A partial recreation of human intelligence limited by the simplicity or specialization of the applied AI system. The problem as I see it is as computers get faster and faster each year (Moore's Law) the domains of applied AI will become larger and larger. It may get to the point where we have the entire left side of the brain implemented through algorithms and limited but still extensive domains. What will likely be forgotten is what we find in the right side of the brain; the parts we have yet to understand ourselves and likely will not for quite some time. I speak of morals and ethics along with creativity.

Why neglect to include something as important as morals and ethics in the domain of what an AI system can do? “Ethics is a major branch of philosophy, encompassing right conduct and good life. It is significantly broader than the common conception of analyzing right and wrong.” (Singer 1993) It would be either incredible or incredibly stupid to entrust machines and computers to make ethical decisions. Issac Asimov wrote a short story in 1942 called "Runaround" in which he defined 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 orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

These laws were to be a flawless circle of protection, however, this is not the case. Without a true understanding of morality and ethics ourselves we will never be able to program a machine to act morally and ethically. The laws can be broken unintentionally or worse misinterpreted by the A.I. system. This is once again an issue explored in the I,Robot series. It also gets explored in James Cameron's “The Terminator” (1986). A military computer called Skynet is put in charge of defending humanity against all threats. With limited morality and ethics Skynet determines that humanity itself is the greatest threat and decides to terminate all humanity. This is course is a worst case scenario but as history has shown, it's not impossible.

“The development of AI is a business, and businesses are notoriously uninterested in fundamental safeguards— especially philosophic ones.” (Sawyer, R. 1991) Not only are safeguards like the three laws of robotics incomplete but history has shown businesses minimize costs by sacrificing safety and safeguards, will this change in the future? We have the danger of creating an intelligent system with no emotions and a very limited set of morals and ethics which

can be misinterpreted with dire consequences. Are we really to trust our lives in the future to a computer program? How do we prevent businesses from cutting safeguards out of AI research?

Will we ever even have the amount of computing power needed for a strong AI system? Moore's Law predicts exponential growth in the number of transistors that can fit on a circuit board. This translates into increased processing performance and higher limits on memory. This at first glance looks fantastic for the outcome of AI: computers get more and more powerful. Moore's law can not go on forever; we will eventually reach the limits of physics where layers in semi-conductors become mere atoms thick. It seems unlikely at this point that conventional computers will ever be able to create an artificial intelligence system outside of simple, specialized, Applied AI. There is, however, another type of computer is being created: a quantum computer. Unlike conventional computers which work with bits, quantum computers with with qubits.

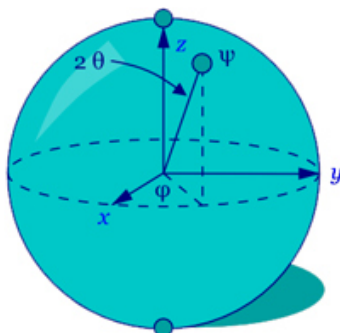


Image used under the GNU Free Documentation License 1.2

The Bloch sphere is a representation of a qubit, the fundamental building block of quantum computers.

Where as a traditional bit is either 0 or 1 a qubite “can be either 0 or 1 or a superposition of 0 and 1; in other words the symbols are both 0 and 1 (and all points in between) at the same time.” (Bonsor K. 2000) This gray area in between is what computers could never do before and what makes quantum computing so powerful and promising. Humans think in more than just yes/no, on/off, right/wrong and with quantum computing we will be able to give machines the ability to do the same. Quantum computing has the potential to be millions of times faster than

traditional computing. This is demonstrated by the ability of a quantum computer to break every kind of encryption currently being used in computing and on the Internet. Of course practical quantum computers do not exist at this time. Currently quantum computers work with only a few qubits and one of the greatest accomplishments in quantum computing has been finding the factors of 15 to be 3 and 5. (Bonsor K. 2000) Quantum computing is still a way off but it provides much more promise and flexibility to create a strong AI system than traditional computing ever will.

With the world of computing changing so dramatically it becomes difficult to predict the outcomes using science fact. Humans are reaching a time where the line between science fact and science fiction begins to dissipate. This is not the first time science fact and fiction mixed. The idea of a cell phone, a device you can use to call someone anywhere on the planet and be able to talk to them was only a dream inspired by Star Trek. “Dr. Martin Cooper, inventor of the modern mobile phone, credits the The Original Series communicator as being his inspiration for the technology. Although the first "brick" mobile phones were much larger, modern flip phones strongly resemble the original series communicator.” (Communicator (Star Trek))

The brain is nothing more, and nothing less, than a very powerful and very odd computer. Evolution has honed it over millions of years to do a fantastic job at certain things, such as pattern recognition and fine control of muscles. The brain is deterministic, meaning that its reactions and responses, including the sensations and behavior of its “owner,” are determined completely by how it is stimulated and by its own internal biophysics and biochemistry. Given those facts, most mathematical philosophers conclude that all the brain’s functions, including consciousness, can be re-created in a machine. It’s a matter of time. (Zorpette G. 2008)

It seems that with more analysis of the brain and the emerging power of quantum computing a strong AI system will inevitably be created. Will we be ready or will we expect the worse?

Under the assumption that we will have truly intelligent systems capable of passing the

Turing test in the future what can we do to prevent the robot apocalypse that *The Terminator* foresees? A common theme in science fiction is the inappropriate use of strong AI. What is inappropriate use? I think if we put an AI in charge of military operations we are just asking for trouble. Despite designing the AI system we may never fully understand how it thinks just as it will likely never understand how we as humans think. Coming back to the issue of morality and ethics we don't *know* ethics we *learn* it through experience. We would need to have restrictions on the use of AI and proceed very cautiously into the unknown. The problem with that is restrictions flat out don't work in the long term. People today, perform secret, illegal research. Why would we expect the field of AI to be without the same? Hopefully this will not be another example of humanity having to learn from a mistake that has a devastating global impact.

The key to preventing disaster seems to be avoiding sentience (the ability to feel) and the ability to be self-aware. Once a machine is self-aware, sentient or both we would have to consider the legal rights of such a creation. This raises multiple difficult ethical problems I won't discuss in too much detail. Would a humanoid, strong AI system be granted the same rights as a human being and is that even safe? It turns out Star Trek has explored this with Lieutenant Commander Data who is a "sentient android". (Data (Star Trek)) He fights for and wins his equal rights to human beings. Part of the argument is that creating sentient beings owned as property is on the same level as slavery. I could argue that sentient and self-aware AI research should be banned but as pointed out previously, that simply won't be enough to prevent the research from being done and potentially completed. What I do think we need to do is prevent a situation where there is an uprising of sentient AI systems. Laws should be enacted before sentience is achieved outlawing the ability to "own" like a slave, a sentient AI. Maybe this will even discourage businesses from partaking in such research as the financial benefit would be significantly lower if they couldn't

own their finished creation. If we proceed into the unknown of AI we must be ready to accept the changes that will be required to prevent disaster. Whether or not this means accepting our AI creations as equals, I don't know but this is a danger we need to be aware of.

Artificial Intelligence is still a relatively new project that has a long way yet to come. While some people are doubtful that further development will ever produce a truly intelligent machine I believe that the only factor is time and that strong AI is unavoidable. Whether or not I refer to a sentient self-aware AI is most likely irrelevant as there is no proven way to restrict research into sentience. We can slow such research by limiting funding; however, we will still have to face the problems sooner or later. The biggest of these problems are those of ethics and morality. Is it alright to own a robotic “slave” as long as it is made of inorganic material? These kind of questions will require answers eventually. Science fiction can provide some help in solving the problems, such as Asimov's Robotic Laws but like most science fiction these answers are often incomplete or not applicable to the real world. Science fiction has been able to spark the imagination in the past as shown with cell phones and I believe this imagination is key to solving the problems of AI in a way which won't cause devastating side effects to human society. In the shorter term we need to be aware of the dangers of integrating applied AI with our more vital systems such as transportation and military. In short something unable to understand the decisions it makes should not be allowed to make those decisions. Understanding is part of intelligence and applied AI can only imitate intelligence; therefore, applied AI can not understand anything at all and should not be allowed to ever make important decisions. Will humanity ever be ready for the dangers of artificial intelligence? Should we as humans really place our futures in the hands of our own intelligent creations?

## References:

Ancill, Dr. Raymond (Personal communication, September 20th, 2008)

Artificial intelligence (n.d.). Retrieved September 21<sup>st</sup> 2008  
from Wikipedia: [http://en.wikipedia.org/wiki/Artificial\\_intelligence](http://en.wikipedia.org/wiki/Artificial_intelligence)

artificial intelligence (2008) in Britannica Online Encyclopedia  
Retrieved October 4<sup>th</sup> 2008 from  
<http://www.britannica.com/EBchecked/topic/37146/artificial-intelligence>

Asimov Issac (1950) *I, Robot*  
Gnome Press (USA)

Bonsor K., Strickland J. (2000) *How Quantum Computers Work*  
HowStuffWorks.com. Retrieved October 16<sup>th</sup> 2008 from  
<http://computer.howstuffworks.com/quantum-computer.htm>

Buttazzo G. (2000) *Can a Machine Ever Become Self-aware?*  
from, <http://feanor.sssup.it/~giorgio/movies/ac-eng.html>

Cameron James  
(1984) *The Terminator* [Motion picture]  
USA (Shows what can happen if we ignore the dangers of creating a strong A.I. system:  
Skynet)

Communicator (Star Trek) (n.d.). Retrieved October 17<sup>th</sup> 2008  
from Wikipedia: <http://en.wikipedia.org/wiki/Combadge>

DARPA Grand Challenge Retrieved October 4<sup>th</sup> 2008  
from Wikipedia: [http://en.wikipedia.org/wiki/DARPA\\_Grand\\_Challenge](http://en.wikipedia.org/wiki/DARPA_Grand_Challenge)

Data (Star Trek) (n.d.). Retrieved October 18th 2008  
from Wikipedia: [http://en.wikipedia.org/wiki/Lieutenant\\_Commander\\_Data](http://en.wikipedia.org/wiki/Lieutenant_Commander_Data)

Dreyfus L., Dreyfus S., Athanasiou, T. (1986) *Mind over machine: the power of human intuition and expertise in the era of the computer*. New York: Free Press

Moore's Law (n.d.). Retrieved October 16th 2008  
from Wikipedia: [http://en.wikipedia.org/wiki/Moore%27s\\_law](http://en.wikipedia.org/wiki/Moore%27s_law)

Neisser U., Boodoo G., Bouchard Jr T.J., Boykin A.W., Brody N., Ceci S.J., Halpern D.F., Loehlin J.C., Perloff R., Sternberg R.J., Others, (1998). "Intelligence: Knowns and Unknowns". *Annual Progress in Child Psychiatry and Child Development 1997*.

Retrieved on October 12<sup>th</sup> 2008.

Sawyer, R. (1991). "On Asimov's Three Laws of Robotics". Retrieved on October 15<sup>th</sup> 2008 from <http://www.sfwriter.com/rmasilaw.htm>

Singer, P. (1993) *Practical Ethics*, 2nd edition (p.10).  
Cambridge: Cambridge University Press

Zorpette G. (June 2008)

Waiting for the Rapture. *IEEE Spectrum*

Retrieved September 21<sup>st</sup> 2008 from <http://www.spectrum.ieee.org/jun08/6311>