Fly Me to the Moon: The Risks and Possible Rewards of Developing Intelligent Computers

By Mark Dahmke

Mark Dahmke

Mark Dahmke is the Database Administrator for the International Association of Torch Clubs and is also the Region 7 Director. He has been a member of the Lincoln Torch Club since 1985.

From 1995 to 2015 he was Vice President and co-owner of Information Analytics, a software development company. In the early 1980s Mark was a Consulting Editor for BYTE Magazine and published several books on Microcomputer Operating Systems through McGraw-Hill. Mark's interests include photography, astronomy, cosmology, and genealogy.

This paper was presented at the September 21, 2015 meeting of the Lincoln Torch Club and was inspired by the 2015 Paxton paper presented by Roger Hughes, "The Singularity: Technology and the Future of Humanism."

In a TED Talk about the pace of change taking place in artificial intelligence research, data scientist and entrepreneur Jeremy Howard shows a video from a conference in China in which a computer performs real-time transcription to English. Another computer then translates in real-time from English to Mandarin Chinese. Then still another program converts the text to speech, all with high enough accuracy that the Chinese audience was applauding. In another example, he showed how a machine learning program, also called "deep learning," was able to organize data in such a way that non-practitioners were able to extract meaningful new insights about the interaction of cancer cells with adjacent healthy cells.

Artificial intelligence, or AI, already permeates our world. Every time you use Google or ask Siri (Apple's intelligent personal assistant) a question or make a plane reservation, you are using some form of artificial intelligence. Most of these programs use what are called "neural networks," which are actually an old technology dating back to the 1980s that has been dusted off and retooled, with the help of computers that are orders of magnitude faster than what we had to work with back then.

Other related terms include "machine learning" or "deep learning." Machine learning could be considered a subset of artificial intelligence because it deals with the ability of a computer to learn all about specific subject matter through various forms of pattern recognition. Researchers also differentiate between strong AI and weak AI. Weak AI can be thought of as intelligence without self-awareness. Watson, the IBM computer that has played Jeopardy so effectively, is a weak AI system. It can analyze text and perform deductive reasoning, but is not anywhere close to being as intelligent as a human being. Strong AI implies an intelligence that is functionally equivalent to that of a human being.

The history of artificial intelligence is too large to encompass in this paper, but to understand what is happening today, we do need to grasp one of one of AI's basic concepts. When comparing the capabilities of AI to those of natural intelligence, consider what the Wright Brothers did when trying to build a flying machine. Instead of trying to build a plane that flaps its wings, they looked at the underlying aerodynamics. They separated the power source from the wing. By not following what evolution came up with, they were free to innovate and find another solution.

Such is the case with modern AI. Neural nets somewhat resemble neurons in the brain. They borrow concepts from nature, but since we still don't know exactly how the brain works, we need to fill in the gaps with technology. The process of designing things that mimic what the brain does will also help us learn how brains actually do work.

At a deep learning conference I attended in January 2015, I had the opportunity to talk to a researcher from Oxford. Over lunch, he and a Silicon Valley entrepreneur and I

discussed the current state of the art. The circumstances alone provided an interesting insight for me: on my left was a man who took the Silicon Valley approach to AI—what can it do for me today and how can I make money from it—while on my right was the Oxford scientist trying to figure out what makes biological neurons work, so he can make digital neurons work.

The practical, Silicon Valley approach using current technology is not much more than smoke and mirrors. It works, and surprisingly well, but it doesn't "think"—a topic we will take up a little later. I posed the following question to both of them. If one considers the human retina and what takes place in the optic nerve that results in our ability to recognize objects, how much do we really know what happens in the layer just behind the retina, let alone what's going on in the optic nerve or visual cortex? The Oxford scientist shook his head and said, "we don't know anything about what's really going on in even that layer."

Even so, in spite of our complete lack of understanding of how humans see and recognize objects, as of the end of 2014 computers were able to correctly recognize about 40% of the objects in almost any photo pulled from the Internet. By early 2015 that percentage was up to well over 50% and is expected to exceed human recognition by 2016. Similarly, software is available that can put captions to photos with over 50% accuracy. This means that if you ask the computer to generate captions for a random selection of photos, a human would rate over 50% of those captions as accurate descriptions of the subject of the photo. I expect that by late 2015, it will be over 80%, and it is expected to exceed human capability in a few more years.

All of that image recognition power comes from a neural network with about the same complexity as the brain of an insect. Using our brains and problem solving capabilities, we humans have built, in a mere blink of an eye on a geologic time scale, something that outperforms evolution. Just as the Wright's plane did not need to flap its wings in order to fly, we did not have to simulate an entire human brain to do it, nor an entire optic nerve or visual cortex, nor even understand how the circuitry right behind the retina actually works.

I could go on talking about the miracles (and horrors) that will soon be upon us because of this technology, but I think you can extrapolate from these examples. Disruption of entire industries, AI's ability to replace almost all jobs—those are the small issues. I want to talk about the big picture.

Earlier this year it was widely reported that Elon Musk, Bill Gates, and Stephen Hawking were sounding the warning that the human race might be putting itself at risk because of the rise of super intelligent machines. Just a few years ago, this was all science fiction. But the technology has changed so rapidly that even in the academic world, the prospect of building sentient machines is now taken seriously and in fact may already be happening.

Bill Gates has said: "I am in the camp that is concerned about super intelligence. First the machines will do a lot of jobs for us and not be super intelligent. That should be positive if we manage it well. A few decades after that though the intelligence is strong enough to be a concern. I agree with Elon Musk and some others on this and don't understand why some people are not concerned" ("Bill Gates Joins").

Stephen Hawking has said: "The primitive forms of artificial intelligence we already have, have proved very useful. But I think the development of full artificial intelligence could spell the end of the human race. Once humans develop artificial intelligence it would take off on its own and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded" (Callen-Jones).

The leading Cassandra on this topic, however, is Elon Musk, who has said, "The risk of something seriously dangerous happening is in the five year time frame, ten years at most" (Cook). The very future of Earth, Musk said, is at risk. "The leading AI companies have taken great steps to ensure safety," he wrote in a post later deleted from the website Edge.org. "The[y] recognize the danger, but believe that they can shape and control the digital super intelligences and prevent bad ones from escaping into the Internet. That remains to be seen."

Speaking at MIT in October 2014, he said: "With artificial intelligence we are summoning the demon. In all those stories where there's the guy with the pentagram and the holy water, it's like yeah, he's sure he can control the demon. Didn't work out" (McFarland).

Back in August of 2014, Musk tweeted, "We need to be super careful with AI. Potentially more dangerous than nukes" (D'Orazio).

According to a Washington Post story, Musk wouldn't even condone a plan to move to another planet to escape AI. "The AI will chase us there pretty quickly," he said (Moyer).

Musk has invested in several artificial intelligence companies, one of which is DeepMind. "Unless you have direct exposure to groups like Deep Mind, you have no idea how fast—it is growing at a pace close to exponential," Musk wrote (Cook).

DeepMind was acquired by Google in January, 2015. But apparently Musk was just investing in AI companies to

keep an eye on them. "It's not from the standpoint of actually trying to make any investment return," he said. "It's purely I would just like to keep an eye on what's going on with artificial intelligence" (Moyer).

So what are the actual risks and possible rewards of developing intelligent computers? Is it even possible?

This returns us to the topic alluded to earlier. How will we know when a machine is intelligent?

This subject has been debated for decades, and we still don't have an answer. Is language a sign of intelligence, or perhaps tool use, or the ability to modify one's environment? All of these behaviors have been seen in animals, including dolphins and chimpanzees, and even birds and elephants. Does it take a combination of all of these attributes to be considered intelligent and self-aware? Is being self-aware even required for an artificial intelligence to be a threat to the human race?

In a recent conversation between human and machine, the human asked the machine: "What is the purpose of being intelligent?" The machine's answer was: "To find out what it is."

That we will anytime soon switch on a computer resembling HAL in the movie *2001: A Space Odyssey* is unlikely; it is far more likely that an intelligence will arise from our vast network of computers called the internet. As a thought experiment, consider what it would be like to be a self-aware colony organism. Imagine an ant colony with the level of complexity of a brain. Now imagine that you are that self-aware being. Your brain is made up of a network of cells, but you have no knowledge of how it functions. You can think and are aware of your own existence. You might become aware that you live in a vast universe full of other stars and planets, and you might wonder if there is anyone out there like yourself. This all sounds very familiar to us humans, doesn't it?

Following the above analogy, say that a large network of computers becomes self-aware. The brain cells are made up of computing nodes or are part of a neural network. The humans who created it would probably never be aware of its existence as a self-aware being unless it was able to cause a change in one of its own components. This would be like trying to exert conscious control over the functioning of cells in your own brain. Even if you could accomplish that, how would you find out how you were created, and how would you communicate with your maker?

The above gives us lots to ponder. Let's imagine several scenarios that could occur in the near future.

Scenario #1: maybe we're worrying for no reason. Is a machine intelligence even possible? It's been suggested that self-awareness might be mathematically incomputable. This means that there's no way to simulate it mathematically using any type of machine.

Scenario #2: the US decides to ban Strong AI but China or some other country does not. We know all too well how that works. If something can be built, it will be, and the economic loser is the one who didn't get there first. The net effect for the planet will be the same regardless of what we decide to ban or not ban.

Scenario #3: AI emerges on its own from our computer networks. It might not be aware of our existence for quite some time. What would an AI do to ensure its continued existence? It would expand to fill all available resources. It might find a way to make us create more of what it needs to exist. But it probably would not be aware that we exist as intelligent beings. It will just do what life does—try to fill every available ecological niche.

Scenario #4: Strong AI technology continues to develop, designed by humans. In most of the AI scenarios, all jobs will shortly be performed by smart computers. The first to go will be all non-creative work, but computers are already doing things we would call creative, such as writing reports and stories for newspapers. Weaponization is the biggest worry, and even if operated with stringent safeguards, there are many ways that this technology could lead to the end of humans.

Scenario #5: We have a bad scare with Strong AI at a global level (e.g., a strong AI is created that kills someone); the backlash leads to a complete ban and scares even the most avid proponents into abandoning strong AI. But this leads us to scenario #6.

Scenario #6: There is a world-wide ban on strong AI, but it is still developed underground or develops on its own. As with genetic engineering, once the technology is democratized, it doesn't take big government or big industry to make it happen. This scenario leads to even more chaos because there will be no incremental ethical framework or recognized standards for development and deployment of the technology. It could be even more disruptive than scenario #2.

Scenario #7: Can we survive with Strong AI? This is the big question. We might even turn it around: can we survive *without* Strong AI?

We have become so used to high technology that we are no longer aware of the profound impact it has on us. Machine learning and big data—the collection and analysis of huge datasets—has already changed our lives, enabling new treatments for cancer and other diseases. It guides our understanding of genetics and genetic engineering. It might be the only way to feed 10 billion people—the population peak we are expected to hit, even with declining birthrates. This

number is unprecedented, and we do not really know what the carrying capacity of our planet is, or what standard of living we may have to accept. We will likely need AI to survive the biggest bottleneck the human race, and perhaps our planet's ecosystem, has ever faced.

In the 1970s, one heard warnings that we would run out of oil or run out of some other critical raw material by the early 2000s. Most of these doomsayers, however, made their predictions based on a linear extrapolation of the future based on the technology available at the time. They rarely allow for human creativity and our ability to pull a technological rabbit out of the hat at the last minute. AI provides us with a very powerful new bag of tricks. A benign form of Strong AI could help us through this crisis and avoid a collapse that would kill 99% of the population. (Unless the AI that develops decides that we are not worth saving.)

Scenario #8: We expand off-planet. But how can that happen?

With current technology, getting to Mars is very difficult. Going beyond the solar system is currently impossible. Furthermore, most of the universe is a very hard vacuum with a few molecules per cubic meter. The environment we humans require occurs in only one place that we know of, and that place, our earth, is incredibly tiny, given the scale of the entire universe. Even a short trip to the Moon is perilous because we have to take along a pressurized environment that is at the correct temperature, has the right percentage of oxygen, and is shielded from cosmic rays. If we want to move on to other worlds or into deep space, our descendants will have to evolve to the requirements of the environment; no form of life on earth has ever remained the same when moving into an environment that has different properties than the one it left.

But humans may be stuck at an evolutionary local maxima. If that is so, might intelligence and technology provide the means by which life can reach higher peaks by creating solutions that could not have been reached by evolution alone?

With strong AI, the galaxy is in theory open to colonization. Machines can survive in almost any environment and for the length of time required to get there; they are ideally suited to existence in the vacuum of space, with no need to carry along tons of supplies or worry about cosmic rays or micro-meteoroids puncturing their spacecraft.

But should humans go along? Given our biological limitations, perhaps not. Ideally we'd like to see human beings go to the stars, but that is a difficult and expensive proposition. Even sending microbes to worlds outside our solar system would be tremendously expensive using current technology.

Above all else, we want to see life and more importantly intelligent life flourish. As far as we know, this is the only place in the universe where there is life as we know it. The universe is a hostile place, so it's in our best interests to spread life *in some form* beyond our planet and to ensure that it continues to spread rather than succumb to any local catastrophes, such as a nearby supernova or even a large asteroid striking the Earth. Strong AI could imaginably be that form.

I would answer the concerns of Elon Musk, Bill Gates, and Stephen Hawking by saying that the survival of intelligence is more important than survival of our race. Regardless of how intelligent machines evolve, whether we design them or they evolve on their own out of our technology, they will *still* be our progeny and perhaps even our legacy.

The publication of this article is funded by The Torch Foundation

References and Further Reading

Alton, Larry. "How Consumer Focused AI Startups are Breaking Down Language." Techcrunch.com. August 7, 2015.

Berman, Alison E. "Are you a Thinking Thing? Why Debating Machine Consciousness Matters." *SingularityHub.com*. August 16, 2015.

"Bill Gates Joins Elon Musk and Stephen Hawking in Saying Artificial Intelligence is Scary." *Quartz Daily Brief*. January 29, 2015.

Bridle, James. "Robots that Write Science Fiction? You Couldn't Make It Up." *The Guardian*, August 10, 2015.

Callen-Jones, Rory. "Stephen Hawking warns artificial intelligence could end mankind." BBC.com. December 2, 2014.

Cook, James. "Elon Musk: You Have No Idea How Close We Are to Killer Robots." *Business Insider*. Nov. 17, 2014.

D'Orazio, Dante. "Elon Musk says artificial intelligence 'potentially more dangerous than nukes'." *The Verge*. August 3, 2014.

Frank, Aaron. "We Can't Find Any Alien Neighbors and Virtual Reality Might Be to Blame." *SingularityHub.com*. August 20, 2015.

Howard, Jeremy. "The Wonderful and Terrifying Implications of Computers that Can Learn." TED Talk. Filmed December 2014.

Lomas, Natasha. "Not Just Another Discussion about Whether AI Is Going to Destroy Us." Techcrunch.com. September 6, 2015.

McFarland, Matt. "Elon Musk: 'With artificial intelligence we are summoning the demon'." Washington Post. October 24, 2014.

Metz, Cade. "IBM's Rodent Brain Chip Could Make Phones Hyper Smart." *Wired*. August 17, 2015.

Moyer, Justin William. "Why Elon Musk is Scared of Artificial Intelligence—and Terminators." *Washington Post,* November 18, 2014.

Nader, Ralph. "Why the Future Doesn't Need Us – Revisited." *Huffington Post*, August 21, 2015.

Nield, David. "Your Brain is Still 30 Times More Powerful than the Best Supercomputers." Sciencealert.com. August 28, 2015.

Pittis, Don. "Scientists Must Act Now to Make Artificial Intelligence Benign." Canadian Broadcasting Company. cbc.ca. August 20, 2015.

Pratt, Gill A. "Is a Cambrian Explosion Coming for Robotics?" IEEE Spectrum. August 31, 2015.