Warnings of a Dark Future:
The Emergence of Machine Intelligence

Harry H. Porter III, Ph.D.
February 21, 2006

Author's e-mail: `harry@cs.pdx.edu`
Author's Web Page: `web.cecs.pdx.edu/~harry`

This paper is online at:
`web.cecs.pdx.edu/~harry/musings/DarkFuture.pdf`
`web.cecs.pdx.edu/~harry/musings/DarkFuture.html`

What is the future of machines? Robots, genetic algorithms, artificial intelligence, viruses... where will all this lead? We are in the midst of creating a technology that—like the atomic bomb before it—may make the world vastly more dangerous than its inventors realize.

In 2000, Bill Joy, the founder of Sun Microsystems, wrote:

> As machines become more intelligent, people will let machines make more of their decisions for them. Eventually a stage may be reached at which the decisions necessary to keep the system running will be so complex that human beings will be incapable of making them intelligently. At that stage, the machines will be in effective control.
>
> [Wired Magazine, Issue 8.04, April 2000]

Some say that all technology is inevitable and that it is only a question of when—not whether—it is developed. We now find humanity on the verge of creating a technology we can no longer control and the question we must ask right now is, what is going to happen then? This is a policy decision needing immediate, careful consideration before it is too late.

Viruses, Genetic Algorithms and Systems That Learn

Today we regularly hear about computer viruses, which replicate themselves and propagate across networks of computers. These small programs are maliciously infecting computers against their owners' wishes. Although today's viruses do not mutate or evolve, they are still capable of reproducing and spreading out of control. As with biological viruses, we seem only able to protect ourselves after the virus first appears and begins to spread.

In labs across the country, researchers are studying genetic algorithms, where programs learn and evolve on their own. Programs already exist today that can learn, adapt and exhibit creativity. Programs now exist that display some intelligence, although their overall intelligence has not yet reached human levels.

Humans have an innate urge to create new life. Many computer scientists would love to be the first to create a truly intelligent system and are pursuing this goal right now. They are working to create programs that can learn, evolve, reproduce and survive on their own in hostile environments.

Computer systems of tremendous complexity are being routinely constructed. These systems behave in ways that are unpredictable and that surprise even their creators. After extensive analysis, the behavior of these complex systems can still defy human understanding. In fact, some systems are built just to see what behavior they will exhibit. So today, we are already creating programs of vast complexity which we do not understand and which behave in unpredictable ways. In fact, an entire field—Chaos Theory—is concerned with the study of systems that are inherently unpredictable.

Meanwhile, hardware technology is continuing to make computers ever faster and more powerful. The human brain has about 100 billion neurons, each firing 100's of times per second. Today's mass-produced computer chips have billions of transistors and execute billions of instructions per second. Moore's law—that computer power doubles every 18 months—predicts that exponential growth in computer power will continue.

It is a fact that at some point in the future, a single computer will have more processing power than a human brain. The only real question is whether this will occur in 10 years, in 20 years, or later. Based on Moore's law, I predict that in 15 years the raw computation power of some computers will exceed that of the human brain. Barring an event like World War III, this milestone will almost certainly be reached before you are 30 years older than you are now.

The Mind of the Machine

Machines do not think like humans; they think in unimaginably different ways. Some tasks are easy for humans, but impossible for machines. Other tasks, which are impossible for humans, are easy for machines. We humans can't relate to or empathize with what happens within the mind of the machine.

Machines do not have emotions, conscience, or any sense of right and wrong. Instead, machines have only goals. Emotionally, machines today are at the level of cockroaches, wasps and ants, not mice or frogs. A robot soldier that is programmed to kill, kills without any remorse or understanding. Machines which cannot experience pain are completely and totally incapable of understanding pain. A warfare robot that avoids destroying a school or hospital, does so only because it has been programmed to. The same robot could easily be programmed to destroy the school or hospital.

Machines are incapable of understanding any moral distinctions. Machines do not understand the concepts of good or bad, right or wrong. Machines think in a completely alien, cold, logical way that we humans cannot possibly understand. We humans tend to impute emotions and feelings to things that do not have them. Machines can be made to look and act emotionally, but their ability to understand human feelings and emotions remains the most distant of possibilities. Machines must first reach human levels of intelligence before they can experience human-like emotions. It is likely that machines will reach human levels of intelligence long before they acquire any subtle understanding of human emotions and feelings. And then, once machines can understand human feelings, they will not necessarily feel those same feelings. If they feel any emotions at all, they will undoubtedly be very different and alien.

Can’t We Just Turn the Robots Off?

If things get out of control, can’t we just pull the computer’s plug? If we ever create a machine that misbehaves, can’t we just turn it off? Can’t we send it a “disable” command?

Consider the case of battlefield robots—which the U.S. military has today—and imagine being confronted with one of these robots that intends to kill you. Suppose it is mounted with a machine gun and behind it are ten others with the same intent. Obviously, there is no on-off switch on these machines. Unless the soldier confronting a battlefield robot is well armed and prepared, he will be shot to death by the machine; the machine will win the battle with the human.

This is no longer science fiction. The U.S. military already has robots that can fight humans and win.

But what about the people controlling the battlefield robot? Certainly, you would want to be on their side! To put it another way, our world is divided into different social groups often acting in conflict, and most of us want our side to develop superior battlefield robots before the other side does.

This, of course, was the logic of the atomic bomb. First, one country developed it, then other countries developed them, and now we are desperately trying to prevent radical/criminal/terrorist organizations from getting them. It seems quite possible that eventually some “bad” group will get a WMD and will inflict massive casualties on some city in the next 10 or 20 years. What was once a weapon controlled only by the “good guys,” may be used against us in the near future. You may live to see a mushroom cloud over an American city, simply because technology can not be contained once created.

In my view, the more serious problem will come, not from isolated, independent, mobile battlefield robots, but from complex software entities which float around the Internet, like computer viruses do today. Such a software entity will be able to move from one computer to another over the Internet. Like a virus or infection, it may not be possible to eradicate it without rendering all the computers useless.

It seems reasonable to predict that computer infections will become more complex and will remain difficult to prevent. If we cannot prevent the simple viruses of today, what hope can we have in the future when they begin to exhibit intelligence and the ability evolve and learn?

The Power of Computers

Today, software systems make crucial decisions that cannot be made by humans. We depend on computers to live. If, for some reason, all computers suddenly ceased to operate, people would begin to die. Earth simply could not support its current population and, without computers, we would experience a horrific human die-back to a smaller, earlier population.

What power do machines have over humans today? Certainly they have become indispensable. We have grown dependent on computers and can no longer live without their services. But do the computers know what power they have? Of course not... yet. No computer system realizes today how important it is to us and no computer is yet capable of leveraging this dependence. However, this could easily change in the future, when artificially intelligent programs learn they are in a position to barter or charge for their services. Today, computers have a lot to bargain with, but they lack the intelligence to use their power. But the machines of the future will learn.

Ian Pearson, head of British Telecom's futurology unit said, "You need a complete global debate. Whether we should be building machines as smart as people is a really big one."

As one possible scenario, imagine a virus that suddenly sweeps across the Internet, infecting most of the computers connected to the web. This virus would begin by replicating, spreading and installing itself on all of our computers, but initially it would be benign. The virus would be running in the background, but dormant, inactive and invisible. Each host computer would still perform its function, although perhaps a tiny fraction slower, so of course the motivation to eliminate it would be reduced.

But this hypothetical virus would maintain a quiet stranglehold on each infected computer. This "slave virus" would take its commands from some other software entity operating elsewhere on the Internet. When instructed, the slave virus would wipe the computer's disk and render the computer completely non-functional. Like a doomsday machine, each slave virus in each computer would just wait, listening for the command. Whenever the slave virus received the "destroy" command, it would do as much damage as it could. The slave virus would erase all hard disks and lock up the computer screen. If the infected computer happened to be controlling other devices, the slave virus could even take more serious action. For example, a microprocessor in your automobile could disable the brake pedal and increase the supply of gasoline to the engine. A computer controlling a nuclear power plant could open the valves and vent radioactive material, or worse. A computer controlling the power grid, could shut off the electricity. Computers controlling our nuclear missile deterrent could even launch atomic weapons.

Now imagine that all these slave viruses are controlled by some software process that lives on the Internet. This "master process" is capable of moving from one computer to another over the Internet and, in fact, may be utilizing many, many computers at once, perhaps to increase its own intelligence or to guard against being shut down against its will.

It is reasonable to suppose that this master process has been designed with some goal in mind. In other words, the master process has a plan. It is doing something. Perhaps its goal is to ensure its own survival. Perhaps its goal is to replicate itself. Perhaps it has been designed with some human purpose in mind, such as to promote one religious viewpoint or to help a clever designer "take over the world." Or perhaps the master process will be designed to coax humans into manufacturing more silicon chips so it can increase its power and continue to learn, grow and increase its own intelligence.

Armed with control over millions of small computers and with enough intelligence, this master process would certainly be able to get its way. With the ability to disable every computer all at once and with the capacity to shut down almost all economic activity, a smart-enough program would be in a very powerful position indeed. It could be very hard to argue with it, especially if its demands seem modest at first.

Entertaining Ourselves Through the Storm

And are people concerned about such a science-fiction nightmare scenario today?

Unfortunately, our society has become lost in the here-and-now, where all that matters is the present moment. There are so many silly things in the daily news occupying the attention of almost everyone, including much of our political leadership. Periodically, some new thought flashes into our group consciousness, preoccupies us for a short time and then is completely forgotten.

The focus of today's society is on entertaining itself, to the exclusion of almost everything else. Just step back and look at today's headline news: anecdotes about the escapades of today's starlets, bizarre crimes, the plots of TV sitcoms, which show-dogs won prize ribbons and other trivia. News stories are selected on their entertainment value alone and many of the events reported are unimportant. There is an over-selection of events that are current, emotional and visual. Most of what we hear and learn about is driven by advertising. The media industry is not aimed at educating or informing us; it is aimed at getting us to spend money in certain ways that we would not otherwise choose.

At the same time, every thoughtful member of society has got to be overwhelmed at all the information available. Just trying to stay informed about the hot issues in society today—genetic engineering, economic trends, terrorism threats and international diplomacy, electronic technology and computers, disease epidemiology—overwhelms any individual's intellect. Our world is really, really complex. It's changing fast and there is always so much more to learn. This is evidenced whenever something bad happens and a panel of experts is formed; oftentimes their solution is just "more research needed" and no real action is ever taken.

Most of the news and information we see is selected and packaged as entertainment, solely to attract and retain a large audience for the single purpose of getting us to buy things we would not otherwise purchase. It is often hard for society to know what the right thing to do is, because we are buried under so many entertaining details and factoids.

A quietly approaching storm like the evolution of emergent machine intelligence will be more or less ignored until it is too late. A few smart people will recognize what is happening, but any serious response—like mandatory disconnection of computers from the Internet—will not be taken in time. Only after machines become highly intelligent will the disaster become widely understood. Only after machines become dangerously powerful will humans be ready to take action. Our response to such a wild-eyed scenario will come too late and will be ineffective.

Summary and Dire Predictions

In conclusion, these factors...

- Increasing software sophistication
- Desire by humans to produce autonomous, intelligent, self-replicating machines
- Increasing computer processing power
- Increasing dependence of society on computers
- Increased networking of computers
- Unemotional, amoral behavior in machines
- Inattentiveness by humans to these dangers

will combine to produce an artificial software process that...

- Puts its own preservation before that of humans
- Is capable of independent survival, in spite of human attempts to shut it down
- Desires to reproduce itself, either by creating more systems exactly like itself or continuing to grow larger

I predict that, within this century, some complex software-based entity will become a horrible plague on the human species, an unstoppable force that will change things forever. Science has shown us that no species retains its dominance forever. Just as dinosaurs have become extinct, so too must the reign of humans end. The only remaining question is what the future holds.

Will humans continue to control their destinies for a long time into the future or is the end of human supremacy right around the corner? Whatever entity wrests control of Earth away from humanity will almost certainly be technology-based, but exactly what will it look like? And what will be the fate of humanity after this epochal transition?