



## Is AI Near a Takeoff Point?

J. Storrs Hall\*

*Laporte, Pennsylvania, USA*

Ray Kurzweil has consistently predicted 2029 as the year to expect truly Turing-test capable machines. Kurzweil's estimates<sup>1</sup> are based on a broad assessment of the progress in computer hardware, software, and neurobiological science.

Kurzweil estimates that we need 10,000 teraops for a human-equivalent machine. Other estimates (e.g. Moravec<sup>2</sup>) range from a hundred to a thousand times less. The estimates actually are consistent, as Moravec's involve modeling cognitive functions at a higher level with ad hoc algorithms, whereas Kurzweil is assuming we'll have to simulate brain function at a more detailed level.

So, the best-estimate range for human-equivalent computing power is 10 to 10000 teraops. The Moore's Law curve for processing power available for \$1000 (in teraops) is:

**2000: 0.001    2010: 1    2020: 1000    2030: 1000000**

Thus, sophisticated algorithmic AI becomes viable in the 2010's, and the brute-force version in the 2020's, as Kurzweil predicts. (Progress in atomically precise nanotechnology is expected to keep Moore's Law on track throughout this period. Note that according to the usual definition, existing computer hardware with imprecise sub-100 nanometer feature sizes is *already* nanotechnology.)

However, a true AI would be considerably more valuable than \$1000. To a corporation, a good decision-maker would be worth at least a million dollars. At a million dollars, the Moore's law curve looks like this:

**2000: 1    2010: 1000    2020: 1000000**

In other words, based on processing power, sophisticated algorithmic AI is viable *now*. We only need to know how to program it.

Current brain scanning tools recently have become able to see the firing of a single neuron in real time. Brain scanning is on a track similar to Moore's law, in a number of critical figures

---

\* E-mail: jsh@mol-eng.com.

<sup>1</sup> Kurzweil, Ray (2005) *The Singularity Is Near: When Humans Transcend Biology* (Viking Adult).

<sup>2</sup> Moravec, Hans (1997) "When will computer hardware match the human brain?" (*Journal of Evolution and Technology*) <http://www.transhumanist.com/volume1/moravec.htm>.

of merit such as resolution and cost. Nanotechnology is a clear driver here, as more sophisticated analysis tools become available to observe brains in action at ever-higher resolution in real time.

Cognitive scientists have worked out diagrams of several of the brain's functional blocks, such as auditory and visual pathways, and built working computational models of them. There are a few hundred such blocks in the brain, but that's all.

In the meantime, purely synthetic computer-based artificial intelligence has been proceeding apace, beating Kasparov at chess, proving a thorny new mathematical theorem that had eluded any human mathematician, and driving off-road vehicles 100 miles successfully, in the past decade.

Existing AI software techniques can build programs that are experts at any well-defined field. The breakthroughs necessary for such a program to learn for itself easily could happen in the next decade. It's always difficult to predict breakthroughs, but it's quite as much a mistake *not* to predict them. One hundred years ago, between 1903 and 1907 approximately, the consensus of the scientific community was that powered heavier-than-air flight was impossible, *after the Wright brothers had flown*.

The key watershed in AI will be the development of a program that learns and extends itself. It's difficult to say just how near such a system is, based on current machine learning technology, or to judge whether neuro- and cognitive science will produce the sudden insight necessary inside the next decade. However, it would be foolish to rule out such a possibility: all the other pieces are essentially in place now. Thus, I see runaway AI as quite possibly the first of the "big" problems to hit, since it doesn't require full molecular manufacturing to come online first.

A few points: The most likely place for strong AI to appear first is in corporate management; most other applications that make an economic difference can use weak AI (many already do); corporations have the necessary resources and clearly could benefit from the most intelligent management (the next most probable point of development is the military).

Initial corporate development could be a problem, however, because such AIs are very likely to be programmed to be competitive first, and worry about minor details like ethics, the economy, and the environment later, if at all. (Indeed, it could be argued that the fiduciary responsibility laws would require them to be programmed that way!)

A more subtle problem is that a learning system will necessarily be self-modifying. In other words, if we do begin by giving rules, boundaries, and so forth to a strong AI, there's a good chance it will find its way around them (note that people and corporations already have demonstrated capabilities of that kind with respect to legal and moral constraints).

In the long run, what self-modifying systems will come to resemble can be described by the logic of evolution. There is serious danger, but also room for optimism if care and foresight are taken.

The best example of a self-creating, self-modifying intelligent system is children. Evolutionary psychology has some disheartening things to tell us about children's moral development. The problem is that the genes, developed by evolution, can't know the moral climate an individual will have to live in, so the psyche has to be adaptive on the individual level to environments ranging from inner-city anarchy to Victorian small town rectitude.

How it works, in simple terms, is that kids start out lying, cheating, and stealing as much as

they can get away with. We call this behavior “childish” and view it as normal in the very young. They are forced into “higher” moral operating modes by demonstrations that they can’t get away with “immature” behavior, and by imitating (“imprinting on”) the moral behavior of parents and high-status peers.

In March 2000, computer scientist Bill Joy published an essay<sup>3</sup> in *Wired* magazine about the dangers of likely 21st-century technologies. His essay claims that these dangers are so great that they might spell the end of humanity: bio-engineered plagues might kill us all; super-intelligent robots might make us their pets; gray goo might destroy the ecosystem.

Joy’s article begins with a passage from the “Unabomber Manifesto,” the essay by Ted Kaczynski that was published under the threat of murder. Joy is surprised to find himself in agreement, at least in part. Kaczynski wrote:

First let us postulate that the computer scientists succeed in developing intelligent machines that can do all things better than human beings can do them. In that case, presumably all work will be done by vast, highly organized systems of machines and no human effort will be necessary. Either of two cases might occur. The machines might be permitted to make all of their own decisions without human oversight, or else human control over the machines might be retained.

But that either/or distinction is a false one (Kaczynski is a mathematician, and commits a serious fallacy applying pseudo-mathematical logic to the real world in this case).

To understand just how complicated the issue really is, let’s consider a huge, immensely powerful machine we’ve already built, and see if the terms being applied here work in its context. The machine is the U.S. government and legal system. It is a lot more like a giant computer system than people realize. Highly complex computer programs are not sequences of instructions; they are sets of rules. This is explicit in the case of “expert systems” and implicit in the case of distributed, object-oriented, interrupt-driven, networked software systems. More to the point, sets of rules are programs.

Therefore, the government is a giant computer program—with guns. The history of the twentieth century is a story of such giant programs going bad and turning on their creators (the Soviet Union) or their neighbors (Nazi Germany) in very much the same way that Kaczynski imagines computers doing.

Of course, you will say that the government isn’t *just* a program; it’s under human control, after all, and it’s composed of people. However, it is both the pride and the shame of the human race that we will do things as part of a group that we never would do on our own—think of Auschwitz. Yes, the government is composed of people, but the whole point of the rules is to make them do different things—or do things differently—than they would otherwise. Bureaucracies famously exhibit the same lack of common sense as do computer programs, and are just as famous for a lack of human empathy.

But, virtual cyborg though the government may be, isn’t it still under human control? In the case of the two horror stories cited above, the answer is: yes, under the control of Stalin and Hitler respectively. The U.S. government is much more decentralized in power; it was designed

---

<sup>3</sup> Joy, Bill (2000) “Why the future doesn’t need us.” (*Wired Magazine*, Issue 8.04) <http://www.wired.com/wired/archive/8.04/joy.html>.

that way. Individual politicians are very strongly tied to the wishes of the voters; listen to one talk and you'll see just how carefully they have to tread when they speak. The government is very strongly under the control of the voters, but no individual voter has any significant power. Is this "under human control"?

The fact is that life in the liberal western democracies is as good as it has ever been for anyone anywhere (for corresponding members of society, that is). What is more, I would argue vigorously that a major reason is that these governments *are not* in the control of individuals or small groups. In the 20th century, worldwide, governments killed upwards of 200 million humans. The vast majority of those deaths came at the hand of governments under the control of individuals or small groups. It did not seem to matter that the mechanisms doing the killing were organizations of humans; it was the nature of the overall system, and the fact that it was a centralized autocracy, that made the difference.

Are Americans as a people so much more moral than Germans or Russians? Absolutely not. Those who will seek and attain power in a society, any society, are quite often ruthless and sometimes downright evil. The U.S. seems to have constructed a system that somehow can be more moral than the people who make it up. (Note that a well-constructed system being better than its components is also a feature of the standard model of the capitalist economy.)

This emergent morality is a crucial property to understand if we are soon to be ruled, as Joy and Kaczynski fear, by our own machines. If we think of the government as an AI system, we see that it is not under direct control of any human, yet it has millions of nerves of pain and pleasure that feed into it from humans. Thus in some sense it is under human control, in a very distributed and generalized way. However, it is not the way that Kaczynski meant in his manifesto, and his analysis seems to miss this possibility completely.

Let me repeat the point: It is possible to create (design may be too strong a word) a system that is controlled in a distributed way by billions of signals from people in its purview. Such a machine can be of a type capable of wholesale slaughter, torture, and genocide—but, if the system is properly controlled, people can live comfortable, interesting, prosperous, sheltered, and moderately free lives within it.

What about the individual, self-modifying, soon-to-be-superintelligent AI's? It shouldn't be necessary to tie each one into the "will of the people"; just keep them under the supervision of systems that are tied in. This is a key point: the nature (and particularly intelligence) of government will have to change in the coming era.

Having morals is what biologist Richard Dawkins calls an "evolutionarily stable strategy." In particular, if you are in an environment where you're being watched all the time, such as in a foraging tribal setting or a Victorian small town, you are better off *being* moral than just pretending, since the pretending is extra effort and involves a risk of getting caught. It seems crucial to set up such an environment for our future AI's.

Back to Bill Joy's *Wired* article: he next quotes from Hans Moravec's book *Robot: Mere Machine to Transcendent Mind*,<sup>4</sup> "Biological species almost never survive encounters with superior competitors." Moravec suggests that the marketplace is like an ecology where humans and robots will compete for the same niche, and he draws the inevitable conclusion.

---

<sup>4</sup> Moravec, Hans (2000) *Robot: Mere Machine to Transcendent Mind* (Oxford University Press, USA).

What Moravec is describing here is not true biological competition; he's just using that as a metaphor. He's talking about economic displacement. We humans are cast in the role of the makers of buggy whips. The robots will be better than we are at everything, and there won't be any jobs left for us poor incompetent humans. Of course, this sort of thing has happened before, and it continues to happen even as we speak. Moravec merely claims that this process will go all the way, displacing not just physical and rote workers, but everybody.

There are two separable questions here: Should humanity as a whole build machines that do all its work for it? And, if we do, how should the fruits of that productivity be distributed, if not by existing market mechanisms?

If we say yes to the first question, would the future be so bad? The robots, properly designed and administered, would be working to provide all that wealth for mankind, and we would get the benefit without having to work. Joy calls this "a textbook dystopia", but Moravec writes, "Contrary to the fears of some engaged in civilization's work ethic, our tribal past has prepared us well for lives as idle rich. In a good climate and location the hunter-gatherer's lot can be pleasant indeed. An afternoon's outing picking berries or catching fish — what we civilized types would recognize as a recreational weekend — provides life's needs for several days. The rest of the time can be spent with children, socializing, or simply resting."

In other words, Moravec believes that, in the medium run, handing our economy over to robots will reclaim the birthright of leisure we gave up in the Faustian bargain of agriculture.

As for the second question, about distribution, perhaps we should ask the ultra-intelligent AI's what to do.

### **About the author:**

J. Storrs Hall, PhD., is an independent scientist and author. His most recent book is *Nanofuture: What's Next for Nanotechnology* (Prometheus 2004). His current project is a book about AI and machine ethics. He was the founding Chief Scientist of Nanorex Inc. His research background includes AI, compilers, microprocessor design, massively parallel processor design, CAD software, and automated multi-level design (including software patent 6,367,052). His other inventions include swarm robotic systems, self-bootstrapping automated manufacturing systems, adiabatic logic, and agoric operating systems.