

# Promise and Peril

by Ray Kurzweil

**Bill Joy wrote a controversial article in *Wired* advocating “relinquishment” of research on self-replicating technologies, such as nanobots. In this rebuttal, originally published in *Interactive Week*, Ray Kurzweil argues that these developments are inevitable and advocates ethical guidelines and responsible oversight.**

*Originally published October 23, 2000 at Interactive Week. Published on KurzweilAI.net April 9, 2001.*

## **A response to Bill Joy’s *Wired* article *Why the Future Doesn’t Need Us*.**

Bill Joy, cofounder of Sun Microsystems and principal developer of the Java programming language, has recently taken up a personal mission to warn us of the impending dangers from the emergence of self-replicating technologies in the fields of genetics, nanotechnology and robotics, which he aggregates under the label “GNR.”

Although his warnings are not entirely new, they have attracted considerable attention because of Joy’s credibility as one of our leading technologists. It reminds me of the attention that George Soros, the currency arbitrageur and arch capitalist, received when he made vaguely critical comments about the excesses of unrestrained capitalism.

According to Joy, the day is close at hand when it will be feasible to create genetically altered designer pathogens in college laboratories. Then, at a later date, we’ll have to contend with self-replicating entities created through nanotechnology, the field devoted to manipulating matter on the scale of individual atoms. Although nanoengineered “self-replicators” are at least one decade, and probably more than two decades, away, the specter that concerns Joy can be described as an unstoppable, non-biological cancer.

Finally, if we manage to survive these first two perils, we’ll encounter robots whose intelligence will rival and ultimately exceed our own. Such robots may make great assistants, but who’s to say that we can count on them to remain reliably friendly to mere humans?

Although I am often cast as the technology optimist who counters Joy’s pessimism, I do share his concerns regarding self-replicating technologies; indeed, I played a role in bringing these dangers to Bill’s attention. In many of the dialogues and forums in which I have participated on this subject, I end up defending Joy’s position with regard to the feasibility of these technologies and scenarios when they come under attack by commentators who I believe are being quite shortsighted in their skepticism. Even so, I do find fault with Joy’s prescription—halting the advance of technology and the pursuit of knowledge in broad fields such as nanotechnology.

Before addressing our differences, let me first discuss the salient issue of feasibility. Many long-range forecasts of technical feasibility dramatically underestimate the power of future technology for one simple reason: They are based on what I call the “intuitive linear” view of technological progress rather than the “historical exponential view.”

When people think of a future period, they intuitively assume that the current rate of progress will continue for the period being considered. In fact, the rate of technological progress is not constant, but since it is human nature to adapt to the changing pace, the intuitive view is that the pace will continue at the current rate. It is typical, therefore, that even sophisticated commentators, when considering the future, extrapolate the current pace of change over the next 10 years or 100 years to determine their expectations—the “intuitive linear” view.

But any serious examination of the history of technology reveals that technological change is at least exponential. There are a great many examples of this, including constantly accelerating developments in computation, communication, brain scanning, multiple aspects of biotechnology and miniaturization. One can examine these data in many different ways, on many different time scales and for a wide variety of phenomena. Whatever the approach, we find—at least—double exponential growth.

This phenomenon, which I call the “law of accelerating returns,” does not rely on a mere assumption of the continuation of Moore’s Law, which predicts, in effect, the quadrupling of computer power every 24 months. Rather, it is based on a rich model of diverse technological processes, a model I have been developing over the past couple of decades.

What it clearly shows is that technology, particularly the pace of technological change, has been advancing at least exponentially since the advent of technology. Thus, while people often overestimate what can be achieved in the short term because there is a tendency to leave out necessary details, we typically underestimate what can be achieved in the long term because exponential growth is ignored.

This observation also applies to rates of paradigm shifts, which are currently doubling approximately every decade. At that rate, the technological progress in the 21<sup>st</sup> century will be equivalent to changes that in the linear view would require on the order of 20,000 years.

This exponential progress in computation and communication technologies is greatly empowering the individual. That’s good news in many ways, because those technologies are largely responsible for the pervasive trend toward democratization and the reshaping of power relations at all levels of society. But these technologies are also empowering and amplifying our destructive impulses. It’s not necessary to anticipate all the ultimate uses of a technology to see danger in, for example, every college biotechnology lab’s having the ability to create self-replicating biological pathogens.

Nevertheless, I do reject Joy’s call for relinquishing broad areas of technology—for example, nanotechnology. Technology has always been a double-edged sword. We don’t need to look any further than today’s technology to see this. Take biotechnology. We have already seen substantial benefits: more effective AIDS treatments, human insulin and many others. In the years ahead, we will see enormous gains in overcoming cancer and many other diseases, as well as in greatly extending human longevity, all presumably positive developments—although even these are controversial.

On the other hand, the means will soon exist in a routine biotechnology laboratory to create a pathogen that could be more destructive to humans or other living organisms than an atomic bomb.

If we imagine describing the dangers that exist today—enough nuclear explosive power to destroy all mammalian life, just for starters—to people who lived a couple of hundred years ago, they would think it mad to take such risks. On the other hand, how many people in the year 2000 would really want to go back to the short, disease-filled, poverty-stricken, disaster-prone lives that 99 percent of the human race struggled through a couple of centuries ago? We may romanticize the past, but until fairly recently, most of humanity lived extremely fragile lives, in which a single common misfortune could spell disaster. Substantial portions of our species still live this precarious existence, which is at least one reason to continue technological progress and the social and economic enhancements that accompany it.

People often go through three stages in examining the impact of future technology: awe and wonderment at its potential to overcome age-old problems, a sense of dread at a new set of grave dangers that accompany these new technologies, followed, finally and hopefully, by the realization that the only viable and responsible path is to set a careful course that can realize the promise while managing the peril.

Joy eloquently describes the plagues of centuries past and how new, self-replicating technologies, such as mutant bioengineered pathogens or “nanobots” (molecule-sized robots), run amok may bring back the

fading notion of pestilence. As I stated earlier, these are real dangers. It is also the case, which Joy acknowledges, that it has been technological advances, such as antibiotics and improved sanitation, that have freed us from the prevalence of such plagues.

Human suffering continues and demands our steadfast attention. Should we tell the millions of people afflicted with cancer and other devastating conditions that we are canceling the development of all bioengineered treatments because there is a risk that these same technologies might one day be used for malevolent purposes? That should be a rhetorical question. Yet, there is a movement to do exactly that. Most people, I believe, would agree that such broad-based relinquishment of research and development is not the answer.

In addition to the continued opportunity to alleviate human distress, another important motivation for continuing technological advancement is economic gain. The continued acceleration of many intertwined technologies are roads paved with gold. (I use the plural here because technology is clearly not a single path.) In a competitive environment, it is an economic imperative to go down these roads. Relinquishing technological advancement would be economic suicide for individuals, companies and nations.

### **The Relinquishment Issue**

Which brings us to the issue of relinquishment—the wholesale abandonment of certain fields of research—which is Joy’s most controversial recommendation and personal commitment. I do feel that relinquishment at the right level is part of a responsible and constructive response to genuine perils. The issue, however, is exactly this: At what level are we to relinquish technology?

Ted Kaczynski, the infamous Unabomber, would have us renounce all of it. This, in my view, is neither desirable nor feasible, and the futility of such a position is only underscored by the senselessness of Kaczynski’s deplorable tactics.

Another level would be to forgo certain fields, nanotechnology, for example, that might be regarded as too dangerous. But even these slightly less sweeping strokes of relinquishment are also untenable. Nanotechnology is simply the inevitable result of a persistent trend toward miniaturization that pervades all of technology. It is far from a single, centralized effort, but rather is being pursued by myriad projects with diverse goals.

One observer wrote:

A further reason why industrial society cannot be reformed . . . is that modern technology is a unified system in which all parts are dependent on one another. You can’t get rid of the ‘bad’ parts of technology and retain only the ‘good’ parts. Take modern medicine, for example. Progress in medical science depends on progress in chemistry, physics, biology, computer science and other fields. Advanced medical treatments require expensive, high-tech equipment that can be made available only by a technologically progressive, economically rich society. Clearly you can’t have much progress in medicine without the whole technological system and everything that goes with it.

The observer I am quoting is Kaczynski. Although one might properly resist him as an authority, I believe he is correct on the deeply entangled nature of the benefits and risks of technology. Where Kaczynski and I clearly part company is in our overall assessment of the relative balance between the two. Joy and I have engaged in dialogues on this issue both publicly and privately, and we concur that technology will and should progress and that we need to be actively concerned with its dark side. If Bill and I disagree, it’s on the granularity of relinquishment that is both feasible and desirable.

Abandonment of broad areas of technology will only push these technologies underground where development would continue unimpeded by ethics or regulation. In such a situation, less stable, less responsible practitioners—for example, terrorists—would have a monopoly on deadly expertise.

I do think that relinquishment at the right level needs to be part of our ethical response to the dangers of 21<sup>st</sup> century technologies. One salient and constructive example of this is the proposed ethical guideline by the Foresight Institute, founded by nanotechnology pioneer Eric Drexler. This guideline would call on nanotechnologists to relinquish the development of physical entities that can self-replicate in a natural environment. Another example is a ban on self-replicating physical entities that contain their own codes for self-replication. In a design that nanotechnologist Ralph Merkle calls the “Broadcast Architecture,” such entities would have to obtain such codes from a centralized secure server, which would guard against undesirable replication.

The Broadcast Architecture is impossible in the biological world, which represents at least one way in which nanotechnology can be made safer than biotechnology. In other ways, nanotech is potentially more dangerous because nanobots can be physically stronger than protein-based entities and more intelligent. But it will eventually be possible to combine the two by having nanotechnology provide the codes within biological entities (replacing DNA), in which case we can use the much safer Broadcast Architecture.

As responsible technologists, our ethics should include such “fine-grained” relinquishment, among other professional ethical guidelines. Other protections will need to include oversight by regulatory bodies, the development of technology-specific “immune” responses, as well as computer-assisted surveillance by law enforcement organizations. Many people are not aware that our intelligence agencies already use advanced technologies such as automated word spotting to monitor a substantial flow of telephone conversations. As we go forward, balancing our cherished rights of privacy with our need to be protected from the malicious use of powerful 21<sup>st</sup> century technologies will be one of many profound challenges. This is the reason recent issues of an encryption “trap door,” in which law enforcement authorities would have access to otherwise secure information, and the FBI’s Carnivore e-mail snooping system have been so contentious.

As a test case, we can take a small measure of comfort from how we have dealt with one recent technological challenge. There exists today a new form of fully non-biological, self-replicating entity that didn’t exist just a few decades ago: the computer virus. When this form of destructive intruder first appeared, strong concerns were voiced that as such viruses became more sophisticated, software pathogens had the potential to destroy the computer network medium in which they live. Yet the “immune system” that has evolved in response to this challenge has been largely effective.

Although destructive, self-replicating software entities do cause damage from time to time, the injury is but a small fraction—much less than one-tenth of 1 percent—of the benefit we receive from the computers and communication links that harbor them.

One might counter that computer viruses lack the lethal potential of biological viruses or of destructive nanotechnology. Although true, this strengthens my observation. The fact that computer viruses are not usually deadly to humans only encourages more people to create and release them. It also means that our response to the danger is relatively relaxed. Conversely, when it comes to self-replicating entities that are potentially lethal on a large scale, our response on all levels will be vastly more intense.

Technology will remain a double-edged sword, and the story of the 21<sup>st</sup> century has not yet been written. So, while we must acknowledge and deal with the dangers, we must also recognize that technology represents vast power to be used for all humankind’s purposes. We have no choice but to work hard to apply these quickening technologies to advance our human values, despite what often appears to be a lack of consensus on what those values should be.

Originally published in *Interactive Week*, October 23, 2000. Copyright c2001 Ziff Davis Media Inc. All rights reserved. This text from KurzweilAI.net.