

# What If AI Succeeds?

## The Rise of the Twenty-First Century Artefact

Hugo de Garis

---

*Within the time of a human generation, computer technology will be capable of producing computers with as many artificial neurons as there are neurons in the human brain. Within two human generations, intelligists (AI researchers) will have discovered how to use such massive computing capacity in brainlike ways. This situation raises the likelihood that twenty-first century global politics will be dominated by the question, Who or what is to be the dominant species on this planet? This article discusses rival political and technological scenarios about the rise of the artefact (artificial intellect, ultraintelligent machine) and launches a plea that a world conference be held on the so-called "artefact debate."*

0738-4602/\$3.50 © 1989 AAAI.

---

Many years ago, while reading my first book on molecular biology, I realized not only that living creatures, including human beings, are biochemical machines, but also that one day, humanity would sufficiently understand the principles of life to be able to reproduce life artificially (Langton 1989) and even create a creature more intelligent than we are. I made this discovery several years before I heard of the subject of AI, but when I did, I felt the same wave of curiosity and fascination as I felt earlier with molecular biology. The two subjects seemed to address similar questions—What is life? What is intelligence?

Today, I am a professional intelligist and just as fascinated with the idea of contributing toward creating an artificial intelligence, or *artefact*, as I was as a youth. At the time, the idea of building a machine smarter than its creator seemed like pure science fiction and at least a century or more away. Today, I believe that given current technological trends, humanity will have its first artefacts before the end of my lifetime, and if so, the consequences for humanity will be profound. It is difficult to find any social issue more important than the prospect of living in a world "peopled" by creations massively smarter than we are. It is an issue that ranks with those concerning the possibility of a nuclear holocaust or an ecological breakdown. In other words, it is concerned with the destiny of human beings as a species.

In fact, it is an even greater issue. The rise of the artefact in the twenty-first century, the issue that in my belief will dominate the global politics of the period, introduces something new into human affairs. For the first time, we are able to pose the question, What is humanity for? I do not mean to ask a religious question

but a real one concerned with the choice about whether humanity will or will not serve as the stepping-stone toward a higher form of evolution.

As is shown in this text, I believe that humanity will be sharply divided on the question about whether artefacts in an advanced form should be allowed to exist. The rival answers to this question are the major theme of this article.

### Technological Trends

Many intelligists believe that it is only a question of time before it will be technologically possible to build an artefact. Whether it will be ethical to do so is another question and, sadly enough, an issue that only a handful of intelligists have discussed over the years, either directly or indirectly (Turing 1950; Michie 1974; Evans 1979; McCorduck 1979; Jastrow 1981; Sinclair 1986; Drexler 1986; Kelly 1987; Hameroff 1987; Waltz 1988; Moravec 1988; de Garis 1989).

An increasing number of intelligists also think that AI is undergoing another paradigm change away from the current symbolic processing paradigm to that of massive parallelism, or parallel distributed processing (PDP), or simply connectionism (McClelland and Rumelhart 1986). This new paradigm is, in fact, an old one, reminiscent of the first AI paradigm of the 1950s and 1960s when self-organization had the upper hand. At the time, computer technology could not provide massively parallel machines to support parallel ideas, so the approach was not terribly successful. Intelligists' ideas adapted to the hardware that was available at the time, namely sequential, mono-processor, von Neumann machines.

Today, however, computer technology will soon be capable of providing massively parallel machines, and a

return to the original approach is warranted; this time, success should be much easier. In fact, as I soon show, the prospect of having billions of components in a single computer will place enormous pressure on the theorists to devise ways to use this hitherto undreamed of computing capacity in brainlike ways. This theorizing has already begun and is referred to as the *PDP, or connectionist, revolution*. However, to get a feel for the size of this coming hardware capacity revolution and the time scale over which it will occur, it is useful to consider an argument by Waltz (1988).

Waltz attempts to estimate how long it will be before computers have the same computing capacity as the human brain, meaning that these new machines will be able to process as many bits per second as the brain. To make this estimation, he needs an estimate of the processing capacity of the brain. He takes the following figures from the neurophysiologists.

There are approximately a hundred billion neurons in the brain, each of which is linked to roughly ten thousand others; hence, there are ten to the power 15 neuronal connections, or synapses as they are called. Each neuron fires roughly 10 times per second. Let us assume that the information content of each synapse (that is, the strength of the connection) is 4 bits. Thus, the total bit-processing rate of the brain is roughly 4 times 10 to the power 16 bits per second. Waltz compares this figure with a similar figure for the performance of the Connection Machine (Hillis 1985). The Connection Machine is massively parallel, with 65,536 separate processors that can function simultaneously; and each processor is able to communicate with any other through an inter-processor communications network.

If each processor is connected to 10,000 others and sends a 32-bit message in each message cycle, 170 such cycles can be executed per second. A Connection Machine costs about \$4 million; so, for \$20 million (the upper limit of what people are prepared to pay for a single supercomputer), five such machines can be bought. This results in a bit-processing rate of  $6.5 \cdot 10^4 \cdot 10^4 \cdot 32 \cdot 170 \cdot 5 = 2 \cdot 10^{13}$  bits

per second, which is a factor of about 2000 short of the human figure.

If a similar comparison is made of the respective memory capacities of the human brain and the Connection Machine, the brain is even further ahead. The human memory capacity can be estimated at  $10^{11}$  neurons  $\cdot 10^4$  synapses per neuron  $\cdot 4$  bits per synapse =  $4 \cdot 10^{15}$  bits of memory.

With 256K memory chips, the comparable figure for the Connection Machine is  $6.5 \cdot 10^4 \cdot 64K$  bits of memory per processor  $\cdot 5$  machines =  $2.2 \cdot 10^{10}$  bits, which is about 200,000 times less than the brain; so, the machine has a long way to go in terms of memory capacity.

How long will it take for the machine to overtake the brain in memory capacity? (It is assumed that the processing capacity of the brain will be overtaken by the machine well before the brain's memory capacity). It is assumed that the price of a transistor on a very large scale integrated (VLSI) chip will continue to fall at roughly the same rate as it has over the last 35 years, namely, by a factor of 10 every five years.

If this rate is extrapolated, then humanity will have a machine of human memory capacity by, roughly, the year 2010, that is, a single human generation from now.

Needless to point out, this development will not stop at 2010. It is likely to go on, and the price of a massively parallel machine will continue to fall.

## Sixth and Seventh Generations

The historical development of computer technology has traditionally been divided into generations. The first generation was based on the valve, the second on the transistor, the third on the integrated circuit, and the fourth on the large scale and very large scale integrated circuit. The fifth generation, a term coined by the Japanese, is somewhat less explicit, but represents massive parallelism and heavy dependence on knowledge-based systems. Sixth and seventh generations are even less well defined, but for the purposes of this article, they are defined as neuronal computing and molecular computing, respectively.

This section deals with those aspects of research in computer science and related fields which will play a role in the rise of the twenty-first century artefact. The aim is to show various trends that will have an impact on future machine generations in the next human generation or two.

The most significant recent change in AI has been the renewed willingness to use the brain as a model for intelligence building. Until recently, the ignorance of the neurophysiologists about how the brain functions, plus the impracticality of building massively parallel machines, dampened any attempt to construct "electronic brains"; however, these days seem to be numbered. There is a growing awareness that the time is ripe for intelligists to renew their attack on building brainlike machines.

Mead (1987), for example, is using VLSI techniques to construct electronic devices with not only the usual transistors but capacitors and amplifiers as well to mimic the behavior of the neuron in silicon. With millions of such devices implanted in superchips, a brainlike device becomes possible. The neurochip is born.

The prospect of neural computers raises an interesting question about whether it will be the neurophysiologists or the intelligists who make the breakthroughs in elucidating the mysteries of the brain. The neurophysiologists are severely handicapped in that they have great difficulty in testing their hypotheses by simulation. The only systems they have to investigate are real brains themselves, with all their fantastic complexity.

Intelligists, however, will be able to test their hypotheses directly on their neuronal computers ("neuters"). They will be able to design machines which store the firing histories of the artificial neurons and which analyze the significance of certain neuronal groups firing and so on. No limit exists to the flexibility of these machines.

As silicon compilation techniques are perfected, it will be possible to design neurochips cheaply and easily so that neuronal hypotheses can be implemented directly into neurochips

***I believe that given current technological trends, humanity will have its first artifacts before the end of my lifetime, and if so, the consequences for humanity will be profound.***

at minimal cost. Silicon compilers will be designing chip layouts as easily as ordinary compilers translate high-level code into machine language.

Another possibility is to imagine neuronal computers designed with such flexibility that their architecture can be easily specified by software (Minsky 1986). This ability would avoid the need to redesign neurochips every time a new neurohypothesis required testing.

The neurophysiologists will be quick to profit from the existence of neural computers to test their brain theories. A marriage of the two subjects is, thus, likely; so, intelligists will become biologists to an increasing extent, and the neurophysiologists will be getting heavily into AI.

Another technology likely to have an impact is optical computing. Recent research on *bistability*, that is, the two-state behavior of certain nonlinear optic devices, allows computing to be entirely optical and, hence, able to overcome such problems as crosstalk, which plagues electronic computing. Optical computing would be much faster than electronic computing, so the interest in this new technology is significant and growing (Feitelson 1988).

A third technology that is not yet well developed is *molecular computing* (Drexler 1986; Hameroff 1987; Langton 1989), which aims to use genetic engineering techniques, among others, to create substances capable of computation but at molecular scales. Molecular computing is important because limits exist to the number of transistors one can cram onto a two-dimensional surface without running into quantum effects. However, these limits can be postponed to some extent by introducing a third dimension into chips, thus piling the number of layers until a solid

block is produced.

The great attraction of molecular computing is not only its (molecular) scale but the added advantages of biological adaptation, such as growth, self-repair, and learning. Recent and spectacular progress in superconductivity promises the possibility of superconducting proteins at room temperature, which would allow a huge quantity of such material to be packed together without worry of heat dissipation problems.

The Japanese Ministry of International Trade and Industry (MITI) is taking molecular computing seriously and, in 1984, promised \$36 million to such research. Unfortunately, the U.S. government has been much slower. The same story is true for the European community.

Recent American research has shown that genetically engineered polypeptides can be metallized, thus giving them the advantages of electric conductivity, so even if superconducting proteins are not found, biologically based computing technology can take advantage of electronic conduction speeds.

Molecular biology has made so much progress in the study of bacteria over the last decade that more and more biochemists are moving up to multicellular creatures and studying such molecular mechanisms as embryological development, including how neurons grow and connect with other neurons. As the principles of these processes are discovered, it will become possible to grow computer brains according to seventh generation technology.

In short, in AI circles, the brain is in again, and research money is starting to flow to support brain-oriented computing. The Japanese have launched two projects of this type. One is called simply the Sixth Generation Project

and the other the Human Frontiers Project. The National Science Foundation in the United States is now funding the Neuronal Computing Project, and the European Commission has launched its BRAIN project, so we should be seeing the first brainlike computing devices shortly.

Tomorrow's intelligists will probably be multidisciplinary experts in the fields of microelectronics (UltraLSI), molecular (nano)electronics, neurophysiology, embryology, optical computing, and so on. Today's symbolic computing on monoprocesor machines will probably be considered quaint.

### **As Machines Grow Smarter**

This section attempts to give a gut feel about what it might be like to live in a world where computers are rapidly increasing their intelligence and discusses the feelings this development might evoke in human beings.

In my view, the biggest impact that smart computers will have on ordinary people will occur when machines begin having conversations with them. This achievement is still some time away. I would say it will be another five years before the first commercial conversational systems are ready. These machines will be capable of recognizing and responding to the simple utterances of their owners. Over the years, however, the sophistication of these systems will increase, until one day people realize they are having relationships with their computers.

Such advanced systems will be capable of learning and will probably be the products of sixth-generation neural computers, using hardware which is based on brain modeling. They will speak well and understand

***The biggest impact that smart computers will have on ordinary people will occur when [they] begin having conversations with them.***

with breathtaking rapidity. Remember a computer thinks a million times faster than our neurons do.

I remember showing my girlfriend how a Lisp machine could calculate the factorial of 1000 in a matter of seconds and display the answer over several screens. She was aghast. She had never seen such analytic power before. I remember the same sense of wonder and even fear at seeing Macsyma, the mathematical assistant program, functioning for the first time before I knew how it was done and, hence, before the magic was taken away.

Another impact on the general public will come from household and commercial robots. These devices will be mobile, providing a range of different services to the public. While they are stupid and docile, we need not fear them; however, the steady increase in their intelligence year by year, as next year's model promises to be more emotionally aware than this year's, will sow doubts about where all this fabulous technology will end up. In 20 years time, it will be commonplace to say that the twenty-first century will be dominated by the machine if humanity so chooses—and maybe even if not.

The general public, the politicians, and, certainly, the intelligists will be discussing the fate of the artefact and the fate of humanity to a far greater extent than is the case today. In fact, I am rather annoyed by the current ostrichlike attitude of many intelligists with regard to the social implications of their work. I label such intelligists "the mice" because they have the horizons of mice.

It is only a question of time before enough people see the writing on the wall and start to seriously question just how far these artefacts should be allowed to develop. Today, this questioning is somewhat academic because we are still some time away from such realities, but it will be real and pressing in a generation or two and will constitute the dominant issue of the age.

One can expect that people will take sides and that considerable energy and passion will be devoted to pleading the various options, so it is now appropriate to discuss just what

the various options are.

## Options

Basically, I see two major options: We let the artefacts freely evolve, or we don't.

If we let them freely evolve, we take a risk because these machines might choose to modify themselves in random ways, similar to the chance mutations of biological evolution. Limits exist to the level of control one can place in machines. One can build in metalevel strategies to control the strategies, one can build meta-metalevels to control the metalevels, but ultimately at the top level, certain strategies simply have to be built in. To change these top-level strategies and choose between good changes and bad changes, the only resource left is survival. Our artefacts might choose to become subject to the same Darwinian forces as biological creatures and for the same reasons.

However, because ethical attitudes are in the limit merely a particular configuration of molecules, we could never be sure that the artefacts would treat human beings with the same level of respect as we would like. After all, when we kill mosquitoes or even cows, we think little of it because we believe mosquitoes and cows are such inferior creatures that we feel justified in exercising the power of life or death over them. We could not rule out a similar attitude on the part of the artefacts toward human beings.

However, a lot of people will start seeing humanity as a stepping-stone toward a higher form of evolution and will claim it is humanity's destiny to help the artefacts get off the planet and into their true environment—namely, the cosmos—perhaps in search of other hyperintelligences.

Some human beings might want to modify their own bodies and brains to become artefacts themselves. This is a third possibility. There might be others.

What is almost certain is that a great debate on the artefact issue will dominate the climate of global politics in the twenty-first century. It is quite likely that preliminary versions

of this great debate will occur among academic circles this century. It is the task of intellectuals to look into the future and anticipate major issues. The intelligists have a moral responsibility to do so, given that it is we who are creating this immense problem.

One of the final ironies of AI is that its long-term goal, which is explicit in the label of the subject itself, is to create an artificial intelligence (eight syllables) or an artelect (three syllables), but to date, too few intelligists are talking publicly about the consequences to humanity of AI succeeding, hence the title of this article.

### Scenario 1

The following scenario is my own contribution to the artelect debate. I might not necessarily believe this scenario will prove realistic, but I find it plausible and interesting.

I see humanity being split into two ideological camps, which I label, respectively, the Terras, (as they might colloquially come to be known) and the Cosmists.

#### The Terras

The Terras are the *terrestrialists*, that is, those people who believe that human beings must remain the dominant species on earth. All ethical systems to the Terras presuppose that human beings are the end and not the means by which actions are judged. The Terras will fear a possible takeover by the artelects or those human beings willing to be modified to become artelects themselves.

When artelect technology becomes capable of making genuinely intelligent machines, the artelect debate will reach its climax, and passions will be high. At this point, it is time to introduce the second ideological camp.

#### The Cosmists

The Cosmists have the opposite belief. The Cosmists will want to give the artelects the chance to develop themselves, escape their provincial terrestrial origins and venture into the cosmos, understand nature's mysteries, and perhaps search for other life forms in the universe.

At this point, possessing a good sci-

ence fiction background is an advantage because nothing else will help. The nature of the subject we are talking about demands it.

The Cosmists will invent a new religion and will defend it with passion because they will feel they are responsible for the next stage in the great upward movement toward . . .

## *We could never be sure that the artelects would treat human beings with the same level of respect as we would like.*

toward what?

Our scientific knowledge tells us that it is virtually certain advanced forms of life exist out there somewhere. With our puny human brains and our frail human bodies, we are not equipped to venture forth from the cradle we call earth, but a suitably adapted artelect could.

The dominant source of global political conflict in the twenty-first century will be between these two groups.

Global communications in 20 to 40 years will be such that every person will be able to communicate easily with everyone else, at least in the rich countries. English will have become the world language and by then nearly everybody will speak it. Robots will have become so productive that material wealth will no longer be an issue. Thus, the source of bitter ideological conflict in the nineteenth and twentieth centuries, namely, between capitalism and communism, will fade away. Who cares who owns capital when there is a surfeit of material goods?

I see the Cosmists forming their own ideological, geographic nation-state (analogous to the way the Zionists formed the state of Israel) as a reaction to the social pressure against them from the Terran majority in most, if not all, nations. The Terras

will be frightened that the experiments of the Cosmists will not only destroy the Cosmists but the Terras as well. The Terras will not permit the Cosmists to allow the artelects to evolve to an advanced state. In the extreme case, the Terras will be prepared to exterminate the Cosmists for the sake of the survival of the Terras.

The Cosmists, however, fully aware of the fears of the Terras, will pursue an age-old policy—mutual deterrence—so that the twenty-first century, politically speaking, will be similar to the twentieth century, only the weapon systems will be all the more horrific and artelectual.

However, a way out of this dilemma might be found. With twenty-first century technology, mass migration of a people might be possible, so the Cosmists might be rocketed to some outer planet to do what they want with themselves.

Meanwhile, the Terras will arm themselves to the hilt and destroy any foreign body approaching the earth, being all too conscious of their greatest weakness, namely, their human intellects.

### Scenario 2

The second scenario is probably more popular in science fiction. It is simply that the artelects will take over. Events might evolve too quickly for human beings to remain in control. If the artelects do take over, it will be difficult to predict what the outcome will be. Perhaps they will treat us as pets and ignore us, as we ignore most creatures, but we could never be sure this case would be true. Perhaps, the

artilects would quickly decide that their destiny was in space. However, the earth is a warm, cosy place in a bleak, black universe. Who knows? Perhaps they will decide that human beings are pests and exterminate us before we decide to exterminate them.

## World Conference

I would like to close this article by pleading for a world conference on this critical topic. I would like to see a group of top people from various fields bring their prestige and intellectual weight to this most important subject. Because the question of the rise of the twenty-first century artilect concerns everyone, a wide range of disciplines should be represented at such a conference. All perspectives should be given an airing.

Of course, because the aim of the conference is to bring the artilect debate to public consciousness, the media should be present in force. It is likely that the theme of the conference will ensure a massive interest on the part of the media. It is difficult to think of a stronger drawing card.

Why is it important to hold such a conference? My hope is that humanity can avoid what happened with the nuclear physicists in the 1930s when they began to realize a nuclear chain reaction might be possible, with a subsequent release of enormous energy. We now live in a world threatened by a nuclear holocaust in which everyone would die. If the nuclear physicists at the time had thought hard about the consequences of their work, perhaps they might not have continued their research, feeling that its consequences were too horrible.

## References

de Garis, H. 1989. The 21st Century Artilect, Moral Dilemmas Concerning the Ultra Intelligent Machine. To appear in *The International Philosophical Review* (Revue Internationale de Philosophie).  
Drexler, K. 1986. *Engines of Creation*. New York: Doubleday.  
Evans, C. 1979. *The Mighty Micro*. London: Coronet Books.  
Feitelson, D. 1988. *Optical Computing: A Survey for Computer Scientists*. Cambridge, Mass.: MIT Press.

Hameroff, S. 1987. Ultimate Computing. In *Biomolecular Consciousness and Nano Technology*. New York: North Holland.

Hillis, W. D. 1985. *The Connection Machine*. Cambridge, Mass.: MIT Press.

Jastrow, D. 1981. *The Enchanted Loom*. New York: Simon & Schuster.

Kelly, J. 1987. Intelligent Machines: What Chance? In *Advances in Artificial Intelligence*, eds. J. Hallam and C. Mellish, 17-32. New York: Wiley.

Langton, C. G., ed. 1989. *Artificial Life: The Synthesis and Simulation of Living Systems*. New York: Addison-Wesley.

McClelland, J. L., and Rumelhart, D. E., eds. 1986. *Parallel Distributed Processing, vols. 1 and 2*. Cambridge, Mass.: MIT Press.

McCorduck, P. 1979. Forging the Gods. In *Machines Who Think*, pp. 329-357. San Francisco: Freeman.

Mead, C. 1987. *Analog VLSI and Neural Systems*. Reading, Mass.: Addison-Wesley.

Michie, D. 1974. *On Machine Intelligence*. Edinburgh, Scotland: Edinburgh University Press.

Minsky, M. 1986. *The Society of Mind*. New York: Simon & Schuster.

Moravec, H. 1988. *Mind Children—The Future of Robot and Human Intelligence*. Cambridge, Mass.: Harvard University Press.

Sinclair, C. 1986. The Why and When of AI. Fax of Speech. Presented at the European AI Conference, 21-25 July, Brighton, U.K.

Turing, A. 1950. Computing Machines and Intelligence. Reprinted in *The Mind's I*. Hofstadter, D. R., and Dennett, D. C. eds., 53-67. New York: Bantam Books.

Waltz, D. 1988. The Prospects for Building Truly Intelligent Machines. In *The Artificial Intelligence Debate*. Cambridge, Mass.: MIT Press.

Former Australian Hugo de Garis is head of the Centre d'Analyses des Donnees et Processus Stochastiques (CADEPS) Artificial Intelligence Research Unit at the Universite Libre de Bruxelles, Ave. F.D. Roosevelt 50, C.P.194/7, B-1050, Brussels, Belgium. He is currently working with Prof. Michalski on machine learning at George Mason University, Virginia until October, 1989. His principal research interests are symbolic machine learning, data analysis, genetic algorithms, and neural networks.