ARTIFICIAL INTELLIGENCE

John Favaro

Viareggio, 7 November 1992

I have an article in Italian for you, that appeared last month in *La Repubblica*. In this article they claim that the four greatest problems that humanity is confronting today are:

1. understanding the structure of matter
2. the origins of life
3. the origins of the universe
4. artificial intelligence

There are a lot of very exciting topics in computer science, but I don't think that there is any topic that has captured the imagination of men more than the attempt to create thinking machines.

I also think there is no topic that managed to arouse the passions of men more than the question of artificial intelligence. There are those who think that the attempt to recreate intelligence in machines is the finest, most noble task that man can undertake. And there are those who think that it is at best presumptuous, at worst blasphemy.

Ever since artificial intelligence became a recognized field of study, there have been two opposing camps: Those who think artificial intelligence is possible and who work in the field, and those don't. It is really nothing less than a religious war.

It has become a favorite pastime for people on both sides to write terrible articles about each other and throw insults back and forth. For example, the most famous opponent of artificial intelligence today, a philosopher named Hubert Dreyfus, once wrote an article which he abrasively titled "Alchemy and Artificial Intelligence", and then expanded it into a book which he titled even more provocatively *What Computers Can't Do: The Limits of Artificial Intelligence*. And look at the picture on the front of the book: it is the computer's plug being pulled from the wall.

As you can imagine, a book like this has constituted nothing less than a declaration of war, and so the most famous champion on the other side, a computer scientist named Marvin Minsky, shot back with a review of the book, which he entitled: "The Artificial Intelligence of Hubert Dreyfus." Another article was written called "Artificial Intelligence Meets Natural Stupidity." And the war was on.

I'm going to try to make some sense out of this war today and sort out some of the issues for you, so that you will at least know what all the fuss is about.

And along the way I'd like to tell you about a couple of the early episodes in the history of artificial intelligence, because sometimes things went wrong in unexpected ways.

COMPUTER CHESS

I'm going to begin by taking a look at an area where some of the earliest attempts at artificial intelligence took place. These were man's attempts to get a machine to play chess. Now, for centuries, chess has been considered to be the supreme celebration of the human intellect. But oddly enough, for nearly as long, man has been fascinated by the idea of getting a machine to play chess. A couple of

John Favaro
Via Gamerra, 21
56123 Pisa - Italy

Tel +39 050 55 60 74
Fax +39 050 6143 1145
john@favaro.net

www.johnfavaro.com

hundred years ago there was actually a man touring around Europe with a beautiful robot-like machine with arms that hung over a chess board, that would play an excellent game of chess against all comers. However, it was discovered after some time that there was a man hidden inside of the robot who just happened to play a very good game of chess.

After that, there weren't many attempts to build a chess-playing machine until the first computers came along. Then it was revived again. In fact, programs were written for all kinds of games. Early on, a program was written to play checkers. The checkers-playing program learned from its opponents, and before long it could beat all human beings. But chess was a different story. No computer programs were ever written that could even come close to beating the great chess masters, and computer chess became the ultimate symbol of the battle between the believers and nonbelievers in the artificial intelligence war.

And this gives us a good illustration of the most fundamental question of all: "What is intelligence?"

Artificial intelligence has sometimes been defined as the use of computers to solve problems that previously could only be solved by applying human intelligence. Now, the problem is that this definition has a sliding meaning, as the computer scientist David Parnas has noted. In the middle ages, it was thought that arithmetic required intelligence. Now we realize that it is simply a mechanical process, and we've built pocket calculators to do it for us.

Chess is a bit like that. It used to be thought of as the supreme test of man's intelligence. But things are changing rapidly now, and computer chess playing programs are getting better and better.

Three years ago, a chess match made headlines all over the world. I read about it first in the *London Times*, but it also made the *New York Times*. The best computer chess program of the day was matched against the world champion Garry Kasparov.

The name of the program is Deep Thought. It's a takeoff, of course, on "Deep Throat", the nickname of the porno film star and of the mole who brought down the Nixon administration (presumably two different people). It was written by researchers at Carnegie-Mellon University. Deep Thought had beaten, for the first time ever, one of the grand masters.

Kasparov beat Deep Thought easily. He said: "I can't visualize living with the knowledge that a computer is stronger than the human mind. I had to challenge Deep Thought for this match, to protect the human race."

But the fact that Deep Thought was so good actually moved the *New York Times* to write an editorial on the subject, which is amazing -- it is something that the *New York Times* never writes about. Here are some excerpts:

Against Deep Thought

"Chess playing is a notable accomplishment of the human mind. But computer programs have steadily improved and are now outwitting even the best human players. Is that a victory or defeat for human intelligence and creativity?

"...These chess programs are invading the special human domain of creativity ... If neurophysiologists finally comprehend the human brain, computers will doubtless be able to mimic the brain's behavior and, at the touch of a button, spew forth output equivalent to its finest creations. A mechanical Mozart? A programmable Newton? A Milton with vision restored, artificially? How could the human spirit not be discouraged?"

I don't happen to agree with the editorial.

Deep Thought can look at 720 thousand moves a second now. Now just about everybody accepts that within 5 or ten years a computer will be built that can beat any human being. The Carnegie-Mellon researchers think that within 5 years they will be able to build a computer that can scan a billion moves a second, and that should be able to do the job. And so what will it mean if we *do* have a computer that can beat any human being at chess? Will this computer be intelligent? I don't think so. It will just be one more thing that we used to associate with purely human intelligence, where in reality it was something mechanical. And I don't personally think it subtracts anything from our human dignity to have a chess program that can beat us.

So that leaves us again with the question, How do we define intelligence, especially machine intelligence? How would we even recognize it if we saw it?

Now I'm going to explain the most famous answer that has ever been given to that question; which was given by one of the most famous men in computer science forty years ago, when computers were in their infancy. The man himself was interesting enough that I'd like to start by telling you a little about him first.

THE TURING TEST

Alan Turing was born in London in 1912. This man was so important to the development of modern computer science that I'd like to say a few words about him before continuing. He was gifted in mathematics and went to Cambridge University, where in his mid-twenties he developed some of the most important theoretical concepts in all of computer science, before computers were ever even built!

He was a fanatic long-distance runner, and also a chess fanatic. He combined these hobbies with his invention of "round-the-house" chess. After your move, you run around the house once. If you get back before your opponent has moved, then you get another move.

When World War II came along, Turing became involved in the famous Ultra project. This was a top-secret attempt to decipher Germany's battlefield codes. The British had obtained one of the Germans' so-called Enigma machines, which were used to encode the cryptographic messages. It was Turing who led the group of cryptologists that designed a special computer to break the German codes, along with a group of cryptologists in Poland. The Ultra project was credited years later with being an important factor in the outcome of World War II, and in fact there was a play about this project on Broadway recently called "Breaking the Code".

But Turing was never revered as a hero. In the early 1950s he was arrested for being a homosexual and forced to undergo a year of so-called "therapy", which included being injected with female hormones. In 1954, with his wartime heroics still classified information, he committed suicide by eating an apple dipped in potassium cyanide. He was forty-one.

His legend has lived on in computer science, and now computer science's most prestigious award, the "Nobel Prize" of computer science, is called the Turing Award.

In 1950, Turing wrote the most famous article ever written in the field of artificial intelligence, which was entitled "Computing Machinery and Intelligence."

Turing's article begins with the sentence: "I propose to consider the question 'Can machines think?'" By the way, someone has once said this is like asking 'Can submarines swim?'. Turing proposed what he called the "imitation game". This has come to be known as the "Turing Test".

The idea is to put someone in a room with a computer terminal. He acts as the interrogator. On the other end of the terminal *is either* a human being or a computer program, but he doesn't know which one. The task of the interrogator is to try to figure out, simply by asking questions, whether he is communicating with a human being or a machine. The task of the computer program is to try to fool him, by imitating a human being as well as possible.

Now this may seem like a pretty simple test to you, but stop and think for a while: how would *you* define intelligence? I think that you would have a hard time coming up with a better way of doing it. And even today, after nearly 40 more years of intensive research in artificial intelligence, the Turing Test is the most widely accepted definition of machine intelligence. "Intelligence is as intelligence does. If a machine appears to be thinking, then it probably is."

THE FIRST TURING TEST LAST YEAR

Last year, they actually tried the very first Turing Test, in the Computer Museum in Boston, Massachusetts. They took eight people as "judges," and sat them in front of computer terminals, and there were six candidate programs that tried to fool the people into thinking that they were human beings.

One of the programs actually did succeed in fooling the people. Its name was "Therapist III." This is the most recent version of one of the most famous programs ever written, and so I'd like to talk about it

now.

ELIZA

If you're thinking that it shouldn't be too difficult to decide whether a computer program is acting intelligently or not, I'd like to show you now that it's not quite all that easy.

In 1963, MIT computer scientist Joseph Weizenbaum wrote what is probably now the single most famous program in artificial intelligence. The name of the program is Eliza, as in Eliza Doolittle of *Pygmalion* and My *Fair Lady* fame. Eliza is a psychiatrist. It imitates the noncommittal style of questioning of a Rogerian psychoanalyst. You sit at the terminal and type in your problems, and Eliza will ask you questions and analyze you.

Here is Eliza's most famous conversation:

> Men are all alike. *In what way?*
>
> They're always bugging us about something or other. *Can you think of a specific example?*
>
> Well, my boyfriend made me come here. *Your boyfriend made you come here.*
>
> He says I'm depressed much of the time. *I'm sorry to hear you are depressed.*
>
> It's true. I'm unhappy. *Do you think coming here will help you not to be unhappy?*
>
> I need some help. That much seems certain. *What would it mean to you if you got some help?*
>
> Perhaps I could learn to get along with my mother. *Tell me more about your family.*
>
> My mother takes care of me. *Who else in your family takes care of you?*
>
> My father. *Your father.*
>
> You are like my father in some ways. *What resemblance do you see?*
>
> You are not very aggressive but I think you don't want me to notice that. *What makes you think I am not very aggressive?*
>
> You don't argue with me. *Why do you think I don't argue with you?*
>
> You are afraid of me. *Does it please you to think I am afraid of you?*
>
> My father is afraid of everybody. *What else comes to mind when you think of your father?*
>
> Bullies. *Does that have anything to do with the fact that your boyfriend made you come here?*

The conversation sounds impressive, especially the final exchange where Eliza brings up the boyfriend again, until one realizes that it's done largely with four simple tricks.

1. Eliza scans each input statement for a small list of words that trigger seemingly appropriate standard responses. The word "depressed" elicits a preprogrammed expression of condolences; the word "mother" sparks a comment about families. Sometimes this trick can be fairly subtle. When Eliza sees the word "like" in one sentence, it produces the response "What resemblance do you see?"

2. The patient's own comments are rearranged and turned into responses. At the beginning of the conversation, if the patient had said, "Well, my dog goes to the movies," Eliza would have responded, "Your dog goes to the movies. " For that matter, if the patient had said "My xyz had az37," Eliza would have mindlessly recycled the nonsense.

3. When an input statement contains no special words, or doesn't match one of the strategies Eliza uses to twist statements into replies, the program produces a noncommittal comment, like "Can you think of a specific example?" Each time it needs one of these nonreplies, it takes it from a list, repeating the list when the supply is exhausted. If the conversation had continued much longer the repetition would have become obvious, and Eliza would have seemed about as intelligent as a washing machine.

4. Early in the dialogue, Eliza stores one of the patient's comments. In this case, "My boyfriend made me come here" -- throwing it back many lines later, generating a feeling of surprise. The fact that, in this case, the comment seems particularly appropriate is really just a coincidence. If the patient had said "Eskimos" instead of "bullies", Eliza's reply would have been the same.

The gag is that Weizenbaum didn't write the program to show how wonderful Artificial Intelligence is, but in fact just the opposite. He wrote it as a joke. He wanted to show how easy it was to fool people into thinking they were dealing with something intelligent by using just a few simple tricks. But the problem was that it had the opposite effect. People were constantly fooled into believing they really were talking to a human being. Now this appalled Weizenbaum. He was appalled at how willingly people let themselves in for this sort of thing.

The last straw involved an episode with Weizenbaum's secretary. Now, she had been working for him the whole time he was developing the Eliza program, so she of all people should have known exactly what it was all about, that it was just a dumb program. Well, one day she began a conversation with Eliza with Weizenbaum looking on, and after a while she turned around with an embarrassed look on her face and asked him to please leave the room, because the conversation was getting too intimate!

By the way, the program that won that first Turing Test in Boston last year (Therapist III) was a slightly different version of Eliza: it pretended to be a paranoid schizophrenic. When the computer professor who wrote the program heard that the program had fooled many psychiatrists into thinking that it was a real person, he said a little sarcastically, "Well, it doesn't tell me much about my program - but it sure tells me a lot about psychiatrists!"

A LITTLE TURING TEST

Now I'd like to try a little Turing Test right here, but first I'll tell you about another program, this time one that produces English language sentences.

Douglas Hofstadter wrote a program that implements a so-called Augmented Transition Network. It generates syntactically correct English language sentences. Now, just having a sentence be syntactically correct doesn't mean that it has any real semantic meaning. The syntax is the easy part. Programs like this would reject even slightly ungrammatical sentences that a human would easily understand. And, even worse, they would accept grammatical sentences that were nonsense, such as "The gasoline can watched Mary explode" or linguist Noam Chomsky's frequently quoted example: "Colorless green ideas sleep furiously."

Hofstadter improved his program by adding a lot of rules to it, to try to generate sentences that made more sense. For example, he required that certain verbs (like "eat") be used only with animate objects. Although his program produced plenty of silly sentences, it produced some very convincing ones, too.

Now I want to try a little Turing Test with all of us. I have a set of sentences here. Some of them were written by humans, and some were produced by the program. I'll read them aloud, and we'll have a show of hands to see who thinks each sentence was written by the computer, and which by the human being.

(a) An attitude will often be held by the serfs of a strife-torn nation. (*computer*)

(b) Blurting may be considered as the reciprocal substitution of semiotic material (dubbing) for a semiotic dialogical product in a dynamic reflexion. *(person)*

(c) Think of that as a chain strength possibility of what, eventually, comes out as a product (epistemic conditions?) and the product is not a Frankfurt-ish packing-it-all-in. *(person)*

(d) Moreover, the Nobel Prizes will be achieved. By the same token, despite the consequence, the Nobel Prizes which will be achieved will sometimes be achieved by a woman. (*computer*)

Still, not everybody is in agreement with the Turing Test.

In 1984 John Searle, who is a philosopher at the University of California's Berkeley campus, was chosen to deliver the Reith lectures, a series of six thirty-minute talks presented each year by the BBC. The lectures, begun in 1948 by Bertrand Russell, are designed to introduce a general audience to some of the important intellectual and scientific issues of the day. For his presentation, Searle chose to speak

about the nature of the mind and what, if anything, it has in common with a digital computer.

Searle disagrees with the premise of the Turing Test. No matter how sophisticated a computer eventually becomes --even if it carries on a conversation convincing enough to pass the Turing test -- Searle would insist that it can never understand what it hears and says. Therefore it cannot be said to think.

"No one supposes that computer simulations of a five-alarm fire will burn the neighborhood down," he says. "No one thinks that a computer simulation of a rainstorm will leave us all drenched. Why on earth would anyone suppose that a computer simulation of understanding actually understood anything?"

TRANSLATION

As I think you can see by now, the issue really revolves around whether the computer can understand what it is dealing with.

One of the very first things that men tried to do with computers, way back in the early 1950's, was language translation. It's ironic now that they tried to do this, because we now realize how hard it is to do -- and it's too bad that one of the first things they tried with computers is actually one of the hardest.

The main reason they were so interested in automatic translation can be summed up in two words: the military. The military has always had a special interest in computers, like any leading edge technology, and they have therefore been the single biggest source of research financing. Well, in the 50's, the Cold War was getting into high gear, and the military wanted to keep up with all of the written documents coming out of the Soviet Union, especially the scientific literature. But there was an acute shortage of Russian translators at the time, and they got the idea that computer translating programs would be a great source of cheap translating labor. They spent millions of dollars funding a program for Machine Translation in the 50's, and to put it bluntly, it was a total failure.

In fact, the best thing to come out of machine translation programs were the jokes. As you can imagine, the hardest things to translate are idiomatic phrases and slogans. They gave the program the following phrase to translate into Russian: "The spirit is willing, but the flesh is weak." The program came up with: "The vodka was great, but the meat was rotten." Another one that I like a lot was the saying "Out of sight, out of mind." This time the program came up with "Blind, insane." Another one for "Out of sight, out of mind" was "Invisible idiot".

And of course the problem is that you have to have the context of the sentence, you can't just understand it as it stands alone.

Take just the simple sentence "John had Mary for lunch." This is an ambiguous sentence, totally dependent on the context in which it appears. For example, it could mean that Mary was John's guest for lunch. But if Mary was, say, a fish, then it might mean that John ate Mary for lunch.

An early advocate of the project, the Israeli logician and philosopher Yehoshua Bar-Hillel, ruefully concluded that Machine Translation was probably impossible. In a famous paper he contended that a machine would never learn to understand both the sentence "The pen is in the box" and "The box is in the pen," realizing in the second case that "pen" must refer to a playpen, not a ball-point pen.

Bar-Hillel wrote that in order to make such distinctions, a machine would need not only a dictionary but an encyclopedia -- a body of knowledge about the world, and that just seemed inconceivable to him.

PHILOSOPHICAL ASPECTS

And this is the greatest problem that people have had with artificial intelligence. The whole thing seems to hinge on our philosophical vision of ourselves and our place in the universe.

I'd like to give you an overview of some of the philosophical arguments that have been advanced so far concerning artificial intelligence.

The arguments in favor of artificial intelligence are rather simple. They simply see no reason why it shouldn't be possible. My advisor at Yale, Alan Perlis, once told me "Why not? Nobody ever said we were the last word in evolution."

Furthermore, you should be careful to distinguish between *philosophical* arguments and purely *technical* arguments against the possibility of artificial intelligence. Leonardo da Vinci invented the helicopter five hundred years ago. The only reason it didn't exist then was because the technology did not exist. But there was no philosophical reason for the helicopter not to exist. Now, considering how much progress has already been made in computer technology in only forty years, you can imagine how much will be made in five hundred more years. So if technology were really the only obstacle to artificial intelligence, we could probably expect to succeed sooner or later, even though it might take another one or even two thousand years of technological advances.

But the philosophical objections are different. They claim that it will *never* be possible to achieve artificial intelligence, no matter how many technological advances are made. Now let's look at a couple of the philosophical objections.

KURT GOEDEL

Kurt Goedel was an Austrian mathematician who wrote a paper in 1931 that changed mathematics forever. The title of the paper was "On Formally Undecidable Propositions in *Principia Mathematica* and Related Systems I."

Now this sounds pretty bad, but it's actually pretty simple to understand what he was doing. It's the whole problem of *self-reflection* and has to do with when you try to turn logic around on itself. The most famous example is the so-called *liars paradox*. Like the statement: "This sentence is false." This is a genuine, real-live logical paradox. If it's true, then it's false, but then it's true again, but then … and so on.

The whole problem arises from trying to make the sentence say something about itself. Mathematicians had been trying to prove things about mathematics for a while by using mathematical ideas. And Goedel proved that it wouldn't work. That you can't get outside of the system and look back in on it. No matter how big the mathematical system is, there will always be something in it that can't be handled, something that will give you a paradox -- something that is neither true nor false, or even worse, both true and false at the same time. And that sentence is an example of the problem.

Goedel's proof devastated the mathematical world. Back in 1931, people had thought that anything was possible with mathematics, and here was proof that said No, there will always be limitations no matter how good your system gets.

By the way, Goedel was very young when he wrote this paper, and his paper made him an immediate superstar in the mathematics world. Suddenly everybody wanted him. He was given a lifetime professorship at Princeton University -- and you know what – it has been claimed by some mathematicians that he never produced another significant piece of work for the rest of his life.

Now what does all this have to do with artificial intelligence? Well, the essence of Goedel's proof is that you can't use mathematics to think about itself, you can only do it by going outside or above the system. And that's what we do as human beings. We stand above mathematics and prove things inside of it. But with artificial intelligence we are trying to recreate our own minds. We are trying to think about ourselves, but necessarily at the same level. But of course we can't step outside of ourselves. Given Goedel's proof, maybe it is the case that only a higher intelligence can create a lower one; that it is theoretically impossible to create something as intelligent as ourselves. Only an intelligence that is higher than us (maybe God?) could do that.

Having said all that, I should say that the relevance of Goedel's work to artificial intelligence is not accepted by all philosophers, nor all computer scientists.

HUBERT DREYFUS

The most famous of the philosophical objectors today is the man I mentioned before, Hubert Dreyfus. He and his brother Stuart Dreyfus are the scourge of artificial intelligence researchers today.

To say the least, they do not believe that artificial intelligence is possible. Hubert once compared Artificial Intelligence researchers who write more and more powerful programs in hopes of one day achieving artificial intelligence to men who climb higher and higher trees in the hope of one day reaching the moon.

Hubert Dreyfus's book *What Computers Can't Do -- the Limits of Artificial Intelligence* is a kind of bible of the loyal opposition to artificial intelligence. In the book, Dreyfus draws on a school of philosophy called existential phenomenalism to argue that thinking cannot be reduced to computation.

The basic argument is that the essence of our humanity and our own powers of reasoning arise from the fact that we have bodies -- that we are at home in the world, and that we are rooted in the situations around us. In other words, until a computer has a body like we do, it cannot possibly hope to attain our kind of intelligence.

Dreyfus says that the human abilities of intuition, insight and comprehension -- the ability to immediately grasp complex situations -- cannot be rationally analyzed.

He says it this way: "Great artists have always sensed the truth, stubbornly denied by both philosophers and technologists, that the basis of human intelligence cannot be isolated and explicitly understood. In the book *Moby Dick*, Herman Melville writes of the tattooed savage Queequeg that he had 'written out on his body a complete theory of the heavens and the earth, and a mystical treatise on the art of attaining truth; so that Queequeg in his own proper person was a riddle to unfold; a wondrous work in one volume; but whose mysteries not even himself could read.'"

Then Dreyfus quotes William Butler Yeats: "Man can embody the truth, but he cannot know it."

Some researchers agree with this idea that a computer could only become intelligent by going through some of the same processes that a human must go through when it grows up. They have proposed a new version of the Turing Test: a computer must be given the intelligence of a young child, and you must show that it is capable of learning, of moving, and talking. In other words, it must "grow up" like a human. They say that if you can prove that, then you have a real Turing Test that proves that the computer is artificially intelligent.

STAR WARS

The issues I've discussed so far are all very interesting but you may not think that they have much relevance to you or to the state of world affairs. But the fact is that artificial intelligence and the controversy about what it is and what it can do is becoming more and more important because of what people are trying to do with it. Artificial Intelligence has come out of the laboratories and into the real world, where companies are risking great sums of money to try to gain an economic advantage with this new technology.

And now the governments and military organizations of the world are betting our own defense on the proposition that artificial intelligence is viable. I'd like to give you an example of this by telling you about one of the most interesting controversies to arise in recent years within the computer science community.

The subject is the Strategic Defense Initiative, which is of course also known as "Star Wars." This is the network of killer satellites and beam weapons that is supposed to set up the ultimate shield against nuclear weapons.

An absolutely essential component of the proposed Star Wars system is computer software technology. The central nervous system will be Battle Management Software to identify incoming rockets and plan a defense strategy.

And one of the essential technical strategies that has been proposed for the Battle Management Software of Star Wars has been the use of artificial intelligence.

In early 1985, an organization called the "SDI Panel on Computing in Support of Battle Management" was created, and the foremost computer scientists of the time were appointed to membership on the panel. Their task was to direct the research in computing that was to be carried out during the course of the Star Wars project.

Not long afterwards, one of the most respected of these computer scientists, Professor David Parnas – mentioned earlier in this article -- resigned from the committee and wrote a long and detailed letter explaining the reasons for his resignation. He wrote:

"I am resigning my membership in the panel. I cannot, in good conscience, accept further payment for

useless effort. My conclusions are not based on political or policy judgments. They are [purely technical].

Parnas wrote a paper called "ARTIFICIAL INTELLIGENCE AND THE STRATEGIC DEFENSE INITIATIVE" and I'd like to read you a few excerpts from this paper.

"One of the technologies being considered for use in SDI is Artificial Intelligence. I find the approaches taken in Artificial Intelligence to be dangerous and much of the work misleading. On occasion I have had to closely examine the claims of a worker in Artificial Intelligence. I have always been disappointed.

"Artificial Intelligence has the same relation to intelligence as artificial flowers have to flowers. From a distance they may appear alike, but when closely examined they are quite different. ... Artificial intelligence techniques do not yield systems that one can trust."

In conclusion he wrote:

"The President has asked us, as scientists, to provide the means of rendering nuclear weapons obsolete. I believe that it is our duty to reply that we have no technological magic that will accomplish that. SDI is not going to solve the problem; the President and the public should know that."

Pretty strong words. That letter unleashed a storm of controversy within the computer science community, partly because he also accused researchers of knowing very well that SDI wouldn't work, but of supporting it anyway as a means of getting funding for their own research.

EXPERT SYSTEMS

Now let's turn now to the effects that artificial intelligence is having on the most important war of all that is going on today: the economic war between Europe, the United States, and Japan. I'm sure you've all heard about the trade war that is going on right now between the United States and Europe - they have just introduced an import tariff on wine from Europe.

In 1984 a man named Edward Feigenbaum, a professor at Stanford University, wrote a book called *THE FIFTH GENERATION: Artificial Intelligence and Japan 's Computer Challenge to the World.* This book created a huge uproar in both the scientific and economic worlds when it appeared.

In the early 1980s the Japanese began a massive multi-year program bent upon producing a new -- a fifth -- generation of computers that will be better than anything built before, and the cornerstone of the great leap forward that they plan to take is artificial intelligence, or what they call knowledge processing. The real power of these computers will lie not in their processing speed but in their capacity to reason.

Feigenbaum wrote: "The world is entering a new period. The wealth of nations, which depended upon land, labor, and capital during its agricultural and industrial phases --depended upon natural resources, the accumulation of money, and even upon weaponry -- will come in the future to depend upon information, knowledge, and intelligence."

That is, knowledge, stored and manipulated by intelligent computers, would become "a salable commodity like food and oil," a new wealth of nations. In the book, Feigenbaum warned that the Japanese government and computer industry, working jointly through the Institute for New Generation Computer Technology, were engaging in a half-billion dollar effort to monopolize the knowledge market. The result of this decade-long plan, which was announced in 1982, would be networks of sophisticated computers programmed to contain facts about the world.

The book galvanized the Americans into action, and they founded a institute in Texas to compete directly with the Japanese. It also galvanized the Europeans into action, and they founded a European Research Institute in Artificial Intelligence (The European Computing Research Center ECRC), which is situated in Munich, in Arabella Park.

The Japanese Fifth Generation Project ended last year - it was a failure. Even the American Research institute is about to close down, partly because American research in artificial intelligence is funded the most by the Department of Defense. And under President Clinton, the budget of the Defense Department will shrink even more.

Still, there were successes in some areas: one is the area of the so-called expert systems; they are the first programs to reach the market successfully today. The purpose of an expert system is to capture the knowledge that a human expert in some field has.

The expert systems work best in fields where there are human experts with very specialized knowledge, like medicine, or law, or engineering. A human expert is interviewed by what they call a "knowledge engineer", who tries to squeeze the knowledge out of him and into the computer. These expert systems are easily the most successful artificial intelligence programs that have yet been built, and there are several on the market today.

A brochure for one of the commercial expert systems on the market says: " Buy our expert system. It won't get sick, resign, or take early retirement."

MYCIN

The most famous of all expert systems is the MYCIN program, which diagnoses meningitis and blood infections.

The MYCIN program has been acknowledged by human experts, to be able to make medical diagnoses -- in its area of specialty -- better than any humans.

And now will the lawyers in the audience please listen carefully. The MYCIN system -- in spite of being acknowledged as better than any human being -- *is still not used in one single place*. Why is that?

It is because of the matter of legal responsibility. Nobody has been willing to take on the legal responsibility for diagnoses made by a computer program. No doctor has been willing to place his reputation behind the program. Nor have lawyers been willing to back up the statements made by legal expert systems in a court of law.

This is a very real barrier to the use of expert systems in the real world. Now, in spite of any problems there might be with artificial intelligence, it doesn't seem right that a program like MYCIN, that can demonstrably do a better job than a human being in making life-saving diagnoses, should be banned from use. It is an issue that really has to be settled.

LEGAL ASPECTS

Let's stay on this topic for a while and think about what could happen if we do succeed in building more and more intelligent systems.

As Artificial Intelligence begins to move out of the laboratory and into our lives, the political and moral implications of thinking machines will become as important as the scientific questions. When we live in a world where artificial systems make decisions, who will take responsibility for bad judgments? Can a computer program be held legally and morally accountable?

This issue is essentially a matter of whether it is possible to blame a computer, to consider it guilty of something.

For something to be blamed, it must, in some sense, be considered a *person*. Over the years, societies have extended personhood to previously excluded groups such as blacks and women. Some people would like to see it granted to fetuses. In many ways corporations are considered persons in the eyes of the law. So why not intelligent machines as well?

An attorney named Marshal Willick pointed out that whether a computer is responsible for its actions is as much a legal as a philosophical question. Isn't that funny? In spite of all the philosophical questions surrounding artificial intelligence, the question of whether a program is intelligent might be decided by a judge and a jury, as the result of a lawsuit filed by someone who feels damaged by a negligent machine. This might happen sooner than we think. After all, for a machine to be held responsible in a court of law, it probably wouldn't have to be of superhuman intelligence. In fact, proving that a computer should be legally considered a person might simply be a matter of showing that it is not dead: When a person dies, he loses all of his rights. The law in this area tends to be very conservative, so that any person is considered "alive" basically if he can be shown to be "not dead".

You may recall that the concept of 'brain death' is now critical in medicine: When does the brain officially die? Since many computers today can exhibit far more 'intelligent' behavior than that of

comatose human beings (who do enjoy legal recognition), a legal minimum standard test of personality could probably be satisfied by a computer system in the proper circumstances.

What do we mean by this definition of "personality?" A computer with a personality? How could the law handle something like that? There are some ideas already now about how to do that: The emergence of the modern corporation provides the most subtle means by which computer systems might achieve legal recognition. Corporations have names, can buy and sell property, and can commit crimes, but they cannot be drafted, be married, or vote. They are persons, but they are owned, as Willick notes.

This is the concept of *partial personality*. This concept is applied in many ways in modern society. Young people, for example, slowly acquire rights and obligations as they grow older ... while rights are removed from the retarded and the insane ...

“The legal system is thus equipped with a variety of approaches with which to decide the extent and variety of rights that should be given to computers that are recognized as persons,” Willick says.

EPILOGUE

Is this the way it will be in the future? I don't think so. I personally think that Artificial Intelligence will be a long, long time in coming, if ever. But as you can see, many think otherwise. I hope today’s lecture will help you to form your own opinion.